

UNIVERSIDADE ESTADUAL DO CENTRO-OESTE, UNICENTRO-PR

**MODELOS DE PREDIÇÃO GENÔMICA ADITIVO E
ADITIVO-DOMINANTE, PARA VOLUME E
QUALIDADE DA MADEIRA EM *Eucalyptus benthamii*
Maiden et Cabbage**

DISSERTAÇÃO DE MESTRADO

JOÃO GABRIEL ZANON PALUDETO

IRATI – PR

2020

JOÃO GABRIEL ZANON PALUDETO

**MODELOS DE PREDIÇÃO GENÔMICA ADITIVO E ADITIVO-DOMINANTE,
PARA VOLUME E QUALIDADE DA MADEIRA EM *Eucalyptus benthamii* Maiden et
Cambage**

Dissertação apresentada ao Programa de Pós-Graduação de Ciências Florestais da Universidade Estadual do Centro-Oeste, para obtenção do título de Mestre em Ciências Florestais.

Orientador: Prof. Dr. Evandro Vagner Tambarussi

Coorientadora: Dra. Regiane Estopa

Coorientador: Dr. Caio César Salgado

Coorientadora: Prof. Dra. Fabiana Schmidt Bandeira Peres

IRATI - PR

2020

Catálogo na Publicação
Rede de Bibliotecas da Unicentro

P184m Paludeto, João Gabriel Zanon
Modelos de predição genômica aditivo e aditivo-dominante, para volume e qualidade da madeira em *Eucalyptus benthamii* Maiden et Cambage / João Gabriel Zanon Paludeto. -- Irati, 2020.
ix, 74 f. : il. ; 28 cm

Dissertação (mestrado) - Universidade Estadual do Centro-Oeste, Programa de Pós-Graduação em Ciências Florestais, 2020.

Orientador: Evandro Vagner Tambarussi

Coorientadora: Regiane Estopa

Coorientador: Caio César Salgado

Coorientadora: Fabiana Schmidt Bandeira Peres

Banca examinadora: Roberto Fritsche Neto, Bruno Marco de Lima

Bibliografia

1. Ciências Florestais. 2. Capacidade preditiva. 3. Efeito de dominância. 4. Viés de predição. 5. GBLUP *singlestep*. I. Título. II. Programa de Pós-Graduação em Ciências Florestais.

CDD 634.9



TERMO DE APROVAÇÃO

Defesa Nº 145

João Gabriel Zanon Paludeto

“Modelos de predição genômica aditivo e aditivo-dominante para volume e qualidade da madeira em *Eucalyptus benthamii* Maiden et Cabbage”

Dissertação aprovada em 05/03/2020, como requisito parcial para obtenção do grau de Mestre, no Programa de Pós-Graduação em Ciências Florestais, área de concentração em Manejo Sustentável de Recursos Florestais, da Universidade Estadual do Centro-Oeste, pela seguinte Banca Examinadora:


Dr. Roberto Fritsche Neto


Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo
Primeiro Examinador


Dr. Bruno Marco de Lima

Bracell Ltda
Segundo Examinador


Dra. Regiane Abjaud Estopa

Klabin
Terceira Examinadora


Dr. Carlo César Salgado

Klabin
Quarto Examinador


Dr. Evandro Vagner Tambarussi

Universidade Estadual do Centro-Oeste
Orientador e Presidente da Banca Examinadora

Irati - PR
2020

Home Page: <http://www.unicentro.br>

Campus Santa Cruz: Rua Salvatore Renna – Padre Salvador, 875, Bairro Santa Cruz – Cx. Postal 3010 – Fone: (42) 3621-1000 – FAX: (42) 3621-1090
CEP 85.015-430 – GUARAPUAVA – PR

Campus Cedeteg: Alameda Élio Antonio Dalla Vecchia, 838, Bairro Vila Carli – Fone: (42) 3629-8100 – CEP 85.040-167 – GUARAPUAVA – PR

Campus de Irati: Rua Professora Maria Roza Zanon de Almeida, Bairro Engenheiro Gutierrez – Cx. Postal, 21 – Fone: (42) 3421-3000
CEP 84.500-000 – IRATI – PR

AGRADECIMENTOS

Primeiramente aos meus pais, Sonia e João, e à toda minha família pelo apoio, incentivo e sacrifício durante toda a minha jornada acadêmica.

A Laís Mazur, por ser meu pilar. Por estar ao meu lado em todos os momentos difíceis, pelo carinho, compreensão e acima de tudo, sempre me incentivar a seguir em frente.

Ao meu orientador, Prof. Dr. Evandro V. Tambarussi (Tamba) por toda a orientação, incentivo, amizade, paciência, por confiar em minha capacidade e por abrir meus olhos para esta área incrível e fascinante.

A minha coorientadora Dra. Regiane A. Estopa, pela orientação, apoio, e principalmente por enxergar potencial em minha pessoa a ponto de firmar esta grande parceria.

Ao meu coorientador Caio César Salgado, pelo incentivo, parceria e auxílio.

À Klabin, aos demais pesquisadores e funcionários, que de alguma forma contribuíram para meu crescimento profissional e para a concretização desta etapa.

Ao Prof. Dr. Dário Grattapaglia por todos os *insights* durante a pré-defesa, os quais contribuíram substancialmente no aperfeiçoamento deste trabalho.

Ao Prof. Dr. Fabio Marroni, por auxiliar um simples estudante que entrou em contato.

Aos amigos e parceiros de laboratório: Mateus Chagas, Matheus Perek, Renan, Rafael, Leonardo, Weverton e João Francisco. Pelos momentos de descontração, amizade e apoio.

Aos amigos que sempre estiveram ao meu lado e, muitas vezes sem saber, contribuíram imensamente na minha trajetória: Alesson Bitar, Alexandre Antoszczyszen (Thwui), André Perussolo (Deco), Caio Augusto Gomes, João Estefano, Marcos Taffarel (Kikss), Rafael Mazur e Rodrigo Cararo (Panzer).

A todos os professores que contribuíram em minha formação, desde a base até agora.

Aos funcionários da Unicentro, a Adriane do PPGF e a todo o Departamento de Engenharia Florestal, por toda a atenção durante o mestrado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Código de Financiamento 001, pelo financiamento do presente estudo.

“I guess I can go anywhere I want.

If only I knew where to go.”

Layne Staley

RESUMO

O *Eucalyptus benthamii* é frequentemente apontado como espécie potencial para uso em áreas com ocorrência de geadas severas. Assim como outras espécies do gênero *Eucalyptus*, é considerada de rápido crescimento. Todavia, um dos maiores desafios do melhoramento florestal tradicional deste gênero, é o tempo para completar um ciclo seletivo. A seleção genômica tem mostrado grande potencial para acelerar significativamente programas de melhoramento, sendo capaz de auxiliar a seleção de indivíduos geneticamente superiores de maneira precoce e confiável. Desta forma, esta pesquisa teve como objetivo construir modelos preditivos de caráter aditivo (GA), aditivo-dominante (GAD), híbrido *single-step* (H) e baseado em *pedigree* esperada (At e Ag) para lignina (LIG), extrativos (EXT), carboidratos (CBO) e densidade básica da madeira (DBM) aos quatro anos de idade e volume individual aos seis anos de idade (VOL6). No estudo de associação genômica (GWAS), encontrou-se um baixo número de associações, as quais explicaram de 0,9% até 6,5% da variância genética dos caracteres. Herdabilidades no sentido restrito (\hat{h}_a^2), variaram de 0,15 (EXT – GAD) até 0,70 (DBM – At). Os modelos genômicos apresentaram desempenho superior em relação aos modelos de *pedigree*. As capacidades preditivas (r_{gy}) variaram de 0,12 (VOL6 – Ag) até 0,44 (DBM – H/GA/GAD). A inclusão do efeito de dominância (modelo GAD) influenciou positivamente na capacidade preditiva e viés de predição para VOL6. O modelo híbrido *single-step* (H) demonstrou ser capaz de aplicar a seleção genômica em maior escala, incluindo indivíduos não genotipados, com capacidade preditiva equivalente ou superior aos demais modelos testados.

Palavras-chave: Capacidade preditiva, efeito de dominância, viés de predição, GBLUP *single-step*.

ABSTRACT

Eucalyptus benthamii is often indicated as a potential species for use in areas with severe frost. Although several species of the *Eucalyptus* genus, including *E. benthamii*, are considered fast-growing, one of the major challenges in traditional forest breeding of the genus is the time required to complete a selective cycle. Genomic selection has shown great potential to significantly accelerate breeding programs, enabling a trustworthy early selection of genetically superior individuals. Thus, this study aimed to construct additive (GA), additive-dominant (GAD), *single-step* hybrid (H) and *pedigree*-based (At and Ag) predictive models for lignin (LIG), extractives (EXT), carbohydrates (CBO) and basic wood density (DBM) at four years of age and individual volume at six years of age (VOL6). The genomic association study (GWAS) presented a low number of associations, which explained from 0.9% to 6.5% of the genetic variation of traits. Genomic narrow-sense heritability (\hat{h}_a^2) ranged from 0.15 (EXT - GAD) to 0.70 (DBM - At). Genomic models performed better than *pedigree* models. The predictive abilities (r_{gy}) ranged from 0.12 (VOL6 - Ag) to 0.44 (DBM - H/GA/GAD). The inclusion of the dominance effect (GAD model) positively influenced predictive ability and prediction bias for VOL6. The *single-step* hybrid model (H) has been shown to be applicable for genomic selection on a larger scale, including non-genotyped individuals, with a predictive ability equivalent or superior to the other evaluated models.

Key Words: Predictive ability, dominance effect, prediction bias, *single-step* GBLUP.

LISTA DE FIGURAS

- Figura 1** - Esquema de melhoramento tradicional de *Eucalyptus* spp., e com o uso de seleção genômica. Adaptado de Resende et al. (2012) e Resende et al. (2017)..... 12
- Figura 2** - Distribuição geográfica do *Eucalyptus benthamii* em seu local de origem. Fonte: BUTCHER et al. (2005). 14
- Figura 3** - Populações utilizadas na prática da Seleção Genômica (GS)..... 23
- Figura 4** – Parentesco genômico e de pedigree. Heatmap e barplot das frequências de parentesco contidas na matriz de parentesco esperado (A) e parentesco genômico realizado (B) para um teste de progênies de polinização aberta de *E. benthamii*, Otacílio Costa - SC. 38
- Figura 5** - Decaimento do desequilíbrio de ligação (r^2) para um teste de progênies de polinização aberta de *E. benthamii*, Otacílio Costa - SC. 39
- Figura 6** – Manhattan e Quantil-quantil (Q-Q) plots para (A) DBM, (B) EXT, (C) LIG, (D) CBO de um teste de polinização aberta de *E. benthamii* (Otacílio Costa – SC) aos quatro anos de idade e (E) VOL6 aos seis. A linha azul representa a correção de Bonferroni ($\alpha = 0,05$), a linha vermelha indica a correção FDR (5%) e a linha verde pontilhada indica o limite de significância ad hoc ($p\text{-valor} \leq 8 \times 10^{-5}$)..... 44
- Figura 7** - Capacidade preditiva (r_{gy}) dos modelos avaliados para os caracteres densidade básica (DBM), extrativos (EXT), lignina (LIG), carboidratos (CBO) aos quatro anos de idade e volume aos seis anos de idade (VOL6) em um teste de progênies de polinização aberta de *E. benthamii*, Otacílio Costa – SC. 53

LISTA DE TABELAS

Tabela 1 - Dados gerais do experimento avaliado.....	26
Tabela 2 - Normas utilizadas na análise química da madeira.....	28
Tabela 3 - Comparação entre o parentesco esperado obtido via pedigree e o parentesco genômico realizado par a par entre indivíduos para um teste de progênies de polinização aberta de <i>E. benthamii</i> , Otacílio Costa - SC.....	37
Tabela 4 – Associações significativas entre os caracteres fenotípicos densidade básica da madeira (DBM), extrativos (EXT), lignina (LIG), carboidratos (CBO) aos quatro anos de idade e volume aos seis anos de idade (VOL6) para um teste de progênies de polinização aberta de <i>E. benthamii</i> e marcadores SNP, Otacílio Costa - SC.	42
Tabela 5 - Parâmetros genéticos estimados por meio dos modelos avaliados para os caracteres densidade básica (DBM), extrativos (EXT), lignina (LIG), carboidratos (CBO) aos quatro anos de idade e volume aos seis anos de idade (VOL6) para um teste de progênies de polinização aberta de <i>E. benthamii</i> , Otacílio Costa – SC.	48

SUMÁRIO

1. INTRODUÇÃO	10
2. REVISÃO BIBLIOGRÁFICA	13
2.1. O gênero <i>Eucalyptus</i> - <i>E. benthamii</i> Maiden et Cambage	13
2.2 Seleção Genômica – <i>Background</i>	15
2.3 Estudo de associação genômica (GWAS)	18
2.4 GBLUP-D e método <i>single-step</i>	20
2.5 Predição de valores, acurácia dos modelos e tamanho efetivo	22
3. OBJETIVOS	25
3.1. Objetivos Específicos	25
4. MATERIAL E MÉTODOS	26
4.1 Material experimental e fenotipagem	26
4.2 Genotipagem e controle de qualidade	28
4.3 Estudo de associação genômica (GWAS)	29
4.4 Computação das matrizes de efeito aditivo, dominante e híbrida	30
4.5 Parentesco, tamanho efetivo e desequilíbrio de ligação	32
4.5.1 Parentesco esperado e realizado	32
4.5.2 Tamanho efetivo populacional	32
4.5.2 Análise da extensão do desequilíbrio de ligação (LD).....	32
4.6 Modelagem e performance dos modelos	33
4.6.1 Ajuste dos modelos e parâmetros genéticos.....	33
4.6.2 Performance dos modelos	35
5. RESULTADOS E DISCUSSÃO	37
5.1 Parentesco esperado e realizado	37
5.2 Desequilíbrio de Ligação e tamanho efetivo	39
5.3 Estudo de associação genômica (GWAS)	41
5.4 Parâmetros e Capacidade Preditiva	46
5.5 O efeito de dominância o modelo <i>single-step</i> na seleção genômica	56
6. CONCLUSÃO	58
7. REFERÊNCIAS BIBLIOGRÁFICAS	59

1. INTRODUÇÃO

O Brasil é referência mundial na silvicultura do *Eucalyptus*. Em 2018, a área total de árvores plantadas no país totalizou 7,83 milhões de hectares, onde 5,7 milhões de hectares são de *Eucalyptus*. Também no ano de 2018, o Brasil foi líder no *ranking* global de produtividade florestal de *Eucalyptus*, alcançando uma média de 36,0 m³/ha ao ano (IBÁ, 2019).

Em plantações florestais no sul do Brasil, a geada é o fator limitante mais importante a ser observado. Para estes casos, o *Eucalyptus benthamii* Maiden et Cambage é a principal espécie recomendada (SILVA et al., 2012). Para o melhoramento, seja do *E. benthamii* ou de qualquer outra espécie, o melhorista dispõe de diferentes métodos, seja mais simples como a seleção massal, ou um método mais elaborado, como as seleções baseadas no valor genético individual. São tais métodos que tornam o uso da seleção recorrente uma metodologia eficiente de acumulação de alelos favoráveis e recombinação de genótipos (BORÉM e MIRANDA, 2009). Em resumo, o sucesso do melhoramento genético depende do correto uso de todas as ferramentas disponíveis, visando a criação e seleção de materiais superiores (RESENDE et al., 2008; VELLO e NAZATO, 2017).

Dentre as ferramentas disponíveis ao melhorista, uma que possui bastante destaque nos programas de melhoramento da atualidade é o BLUP (*Best Linear Unbiased Prediction*). Resumidamente, o BLUP é um procedimento utilizado na predição dos valores genotípicos a partir de dados fenotípicos (DAETWYLER, 2009). Apesar de ser um método consolidado, o BLUP realiza inferências sobre efeitos genotípicos a partir de informações fenotípicas, o que, de acordo com alguns autores, pode acarretar em uma baixa acurácia do método, resultando em informações imprecisas (RESENDE et al., 2008; CROSSA et al., 2010). A fim de aumentar a eficiência desse procedimento e consequentemente melhorar a acurácia, Lande e Thompson (1990), propuseram a seleção assistida por marcadores (*Marker Assisted Selection* - MAS). Esta metodologia propõe o uso de dados fenotípicos e moleculares (simultaneamente), e, basea-se na existência de ligações gênicas entre os locos de características quantitativas (*Quantitative Trait Loci* – QTL), e os marcadores moleculares.

Os caracteres de interesse agrônômico são, na maioria, poligênicos, ou seja, caracteres quantitativos que são controladas por vários genes, onde cada um deles exerce um pequeno efeito no caráter. Este fato é, na realidade, problemático para a MAS, já que dificulta a detecção de um grande número de QTLs capazes de explicarem efetivamente a variação genética do

caráter (BEAVIS, 1998; FERREIRA e GRATTAPAGLIA, 1998; BERNARDO, 2008; BORÉM e CAIXETA, 2009).

Apesar das limitações da MAS acerca da detecção de QTLs, a genética molecular apresenta grande potencial, pois possibilita a utilização direta de informações genéticas na prática da seleção, o que permite alta eficiência seletiva e obtenção de ganhos em menor tempo e custo quando comparada à metodologia padrão de análise baseada em dados fenotípicos (RESENDE et al., 2008). Sabendo das limitações de uso da seleção assistida por marcadores, e visando aproveitar o potencial da genética molecular, Meuwissen et al. (2001) propuseram a seleção genômica (*Genomic Selection* - GS).

Existem grandes expectativas quanto à GS aplicada ao melhoramento genético de espécies perenes, especialmente do *Eucalyptus*. Esta abordagem apresenta inúmeras vantagens em relação à metodologia tradicional de análise e seleção, isto se dá principalmente devido ao massivo ganho de tempo. Já que, devido à natureza poligênica e complexa dos caracteres de interesse agrônômico, estes, geralmente, só podem ser acessados ao fim de um ciclo seletivo, criando uma demanda de acelerar os programas de melhoramento de *Eucalyptus*.

O tempo para completar um ciclo seletivo, é uma dificuldade clássica do melhoramento de essências arbóreas, sendo retratado por vários autores (KAGEYAMA e VENCOSKY, 1983; PEREIRA et al., 1997; GONÇALVES et al., 1998; PALUDETO et al., 2017). Quando consolidado, o uso da GS resulta em uma redução significativa dos ciclos de seleção, permitindo “pular” etapas de um ciclo de melhoramento, como o teste de progênie, o teste clonal, ou até mesmo ambos, sendo capaz de reduzir, em relação ao modelo tradicional, o ciclo de melhoramento de 18-10 anos para até 9-5 anos (RESENDE et al., 2012; GRATTAPAGLIA, 2014). (Figura 1).



Figura 1 - Esquema de melhoramento tradicional de *Eucalyptus* spp., e com o uso de seleção genômica. Adaptado de Resende et al. (2012) e Resende et al. (2017).

Além do ganho em tempo, a GS tem capacidade de fornecer maior acurácia na predição de fenótipos e/ou valores genéticos em relação ao método padrão baseado no uso do parentesco esperado entre indivíduos (*pedigree*). A GS permite, devido à genotipagem inerente ao método, a estimação e inclusão de efeitos além do aditivo *per se*, como o efeito de dominância, o qual é extremamente complexo de ser estimado apenas com o uso de *pedigree* (VITEZICA et al., 2013). A inclusão do efeito de dominância é recomendada por vários autores, visto que, se determinado caráter é influenciado por este efeito, sua inclusão na predição tende a melhorar a capacidade preditiva, bem como as estimativas atreladas ao modelo (DENIS e BOUVET, 2011; SU et al., 2012; MUÑOZ et al., 2014; TAN et al., 2017; DE ALMEIDA FILHO et al., 2019). É importante ressaltar, que até o presente momento, este é o primeiro estudo a avaliar o efeito de dominância associado ao *E. benthamii*, bem como à uma espécie pura.

Apesar de todo o apelo da GS, a genotipagem caracteriza-se como uma operação de alto custo, principalmente no melhoramento de essências florestais, já que os testes de progênes geralmente possuem muitos indivíduos (PÉREZ-RODRÍGUEZ et al., 2017; RATCLIFFE, 2019). Com o objetivo de reduzir os custos da genotipagem, uma técnica inicialmente proposta para o melhoramento animal vem sendo sugerida e amplamente estudada, o *single-step* (LEGARRA et al., 2009; AGUILAR et al., 2010). Este método consiste no uso simultâneo de indivíduos genotipados e não genotipados em uma mesma avaliação, de modo a combinar a

matriz de parentesco obtida a partir da genotipagem, com a matriz de *pedigree*. Esta combinação aumenta o tamanho amostral de inferência do modelo, possibilitando a aplicação de uma abordagem genômica em indivíduos não genotipados, ou seja, necessita apenas da genotipagem de parte do experimento. Ou seja, o *single-step* consiste em um método capaz de aplicar a GS em maior escala, com menor custo.

2. REVISÃO BIBLIOGRÁFICA

2.1. O gênero *Eucalyptus* - *E. benthamii* Maiden et Cabbage

O *Eucalyptus* é pertencente à família Myrtaceae e subfamília Leptospermidae (ANDRADE, 1961). É um gênero natural da Austrália, Indonésia e ilhas próximas, como Flores, Alor e Wetar, com mais de 700 espécies descritas (LADIGES et al., 2003). Além disso, apresenta grande plasticidade e ampla dispersão mundial, e por isso é capaz de se desenvolver em diversas situações edafoclimáticas.

Atualmente, o *Eucalyptus* é extensivamente plantado por vários motivos, dentre eles sua elevada produtividade, rápido crescimento, produção de madeira de importância comercial para diversos usos e grande capacidade adaptativa (MORA e GARCIA, 2000). Entretanto, existem certas limitações quanto a sua expansão na região sul do Brasil, pois é uma região caracterizada por temperaturas mais baixas e geadas frequentes (NISGOSKI et al., 1998).

O *E. benthamii* Maiden et Cabbage, comumente conhecido como “Camden White Gum” é uma espécie subtropical de distribuição natural restrita entre a latitude 33°S e 34°S na região costeira do estado australiano de Nova Gales do Sul, entre 50 e 800 m de altitude. Suas principais exigências climáticas são: temperatura média anual entre 13 °C e 17 °C e precipitação anual entre 900 e 1.400 mm. A espécie é encontrada em áreas restritas localizadas ao sudoeste de Sydney, Austrália (Figura 2) (HALL e BROOKER, 1973; PRYOR, 1981). Sua ocorrência, de acordo com a classificação de Köppen e Geiger (1928) é predominante em locais com tipologia climática do tipo Cfb, um clima com temperatura média no mês mais frio abaixo de 18°C e no mês mais quente abaixo de 22°C, chuvas uniformemente distribuídas e sem estação seca definida; e em menor proporção no tipo Cfa, um clima que tende a concentrar as chuvas nos meses de verão, entretanto sem estação seca definida, cujo a temperatura média no mês mais frio é inferior a 18°C, e no mês mais quente, acima de 22°C (FLORES et al., 2016).

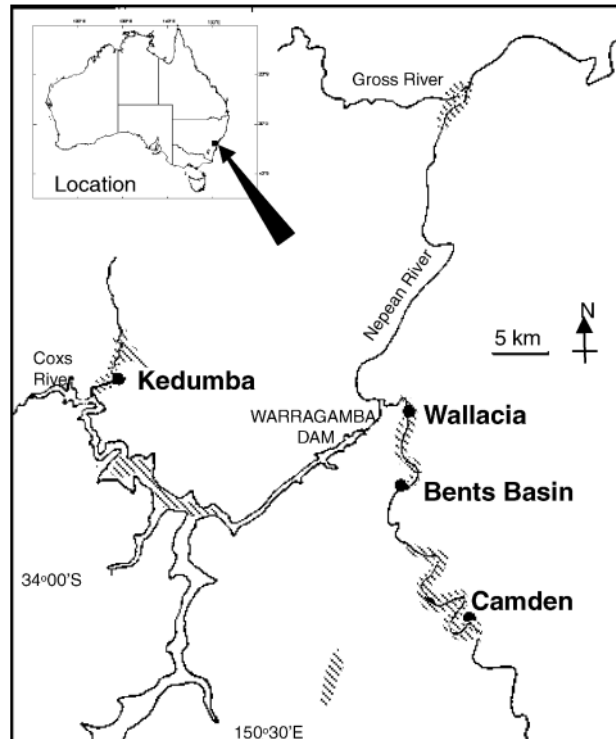


Figura 2 - Distribuição geográfica do *Eucalyptus benthamii* em seu local de origem. Fonte: BUTCHER et al. (2005).

O *E. benthamii* é uma espécie de baixa regeneração natural, a qual teve sua base genética reduzida drasticamente devido a ações direta ou indiretamente antrópicas, tais como introdução de outras espécies, aumento de atividades agrícolas, inundações, aumento de áreas urbanas e incêndios (HALL e BROOKER, 1973; NSW, 2000; KJAER et al., 2004; BUTCHER et al., 2005). Por isso, esta espécie hoje é considerada vulnerável a extinção (NSW, 2000; KJAER et al., 2004).

O principal interesse na espécie, se dá pela sua alta tolerância à geada, sendo capaz de se desenvolver em regiões com temperatura média anual de até 14,5 °C, e suportar temperaturas mínimas absolutas de -6 °C até -10°C (SWAIN, 1999; LIN et al., 2003; PALUDZYSZYN FILHO et al., 2006). Em seus primeiros plantios experimentais, a espécie já foi apontada como potencial para uso em áreas com ocorrência de geadas severas, sendo cada vez mais procurada por produtores brasileiros (HIGA et al., 2012). Em áreas subtropicais da Argentina, China e Uruguai, o *E. benthamii* tem mostrado bom desenvolvimento e resistência á geadas (YONQI et al., 1994).

O programa de melhoramento de *E. benthamii* ainda é recente, porém visando aproveitar seu elevado potencial de tolerância à geada, programas de melhoramento em regiões

subtropicais, como o sul do Brasil, sudeste dos E.U.A, Uruguai, Argentina, Chile e China, utilizam tanto o *E. benthamii* como espécie pura ou o combinada com outras espécies, como o *E. dunni*, com o objetivo de agregar tolerância à geada aos híbridos (BUTCHER et al., 2005; RESENDE e ASSIS, 2008; BRONDANI et al., 2011; YU e GALAGHER, 2015).

2.2 Seleção Genômica – *Background*

Os marcadores moleculares foram desenvolvidos na década de 1980 e são sem dúvida um grande marco na biotecnologia (NADEEM et al., 2017). Alguns anos após, a partir de 1990, os avanços na genética molecular motivaram amplamente o desenvolvimento de ferramentas genômicas, pois trouxeram esperança de que o uso de informações em nível de DNA levaria a maiores ganhos genéticos que o uso de dados fenotípicos (MEUWISSEN e GODDARD, 1996; MEUWISSEN et al., 2016). A partir daí, despertou-se grande interesse na detecção e seleção de locos de caracteres quantitativos, os chamados QTLs (*Quantitative Trait Loci*) (COLLARD et al., 2005), pois, a identificação destes locos, permite o mapeamento exato das regiões genômicas que afetam o caráter de interesse em estudo (GRATTAPAGLIA et al., 2009), possibilitando uma seleção de genótipos assertiva e direcionada. É importante ressaltar, que os marcadores precisam estar em ligação com o loco alvo (MACKILL e NI, 2008). Esta ligação ocorre entre alelos de maneira não aleatória em locos separados, idealmente, do mesmo cromossomo, e é conhecida como desequilíbrio de ligação (*Linkage disequilibrium – LD*) (WILLIAMS e CUMMINGS, 1997; MACKAY e POWELL, 2007). Isto tudo deu origem à uma técnica conhecida como seleção assistida por marcadores (*Marker-assisted selection – MAS*).

Os primeiros registros de ideias precursoras da MAS datam de Sax (1923), o qual observou em sua pesquisa, a existência de um tipo de associação entre a cor (caráter monogênico) e o peso (caráter poligênico) das sementes de feijão (*Phaseolus vulgaris* L.). O autor concluiu que o único gene que controla a cor da semente, deveria possuir algum tipo de ligação com um ou mais genes que controlam o peso. O conceito da seleção assistida por marcadores (MAS) foi sugerido primeiramente por Soller e Backman (1983) e Smith e Simpson (1986), a partir da ideia de que “*é improvável que os polimorfismos identificados pelas novas técnicas de detecção são os próprios QTLs, entretanto, muitos deles podem estar em algum tipo de ligação com os QTLs, e por isso, devem tornar possível a prática de seleção indireta.*” Ainda, vale destacar o trabalho de Lande e Thompson (1990), os quais, visando o uso da MAS

na seleção para caracteres quantitativos em toda uma população, propuseram o uso de um grande número de marcadores em desequilíbrio de ligação com os QTLs.

Entretanto, alguns pesquisadores se mostraram menos otimistas acerca do potencial de uso da MAS para caracteres poligênicos. O fato de os caracteres agronômicos mais importantes serem quantitativos, torna a MAS menos efetiva, já que este tipo de caráter é influenciado por uma grande quantidade de pequenos efeitos, os quais são muito pequenos para serem eficientemente capturados (NAKAYA e ISOBE, 2012). Devido a isso, encontram-se um pequeno número de QTLs de grandes efeitos que, além de não serem capazes de explicar toda a variabilidade genética de forma confiável, podem resultar de superestimativas devido ao pequeno tamanho populacional utilizado em tais análises (GRATTAPAGLIA e KIRST, 2008; GRATTAPAGLIA e RESENDE, 2011). Isso caracteriza um grande problema para a MAS, já que para essa técnica ser útil, cada gene/QTL que participa no controle do caráter de interesse deve estar identificado e mapeado, pois o sucesso desta técnica depende tanto de alta precisão na detecção dos QTLs, quanto em uma elevada proporção da variação genética explicada por eles (FERREIRA e GRATTAPAGLIA, 1998; BERNARDO, 2008; SCHUSTER, 2011).

Apesar do proposto por Lande e Thompson (1990), as densidades de marcas utilizadas na MAS eram muito baixas a ponto de poder encontrar marcadores em desequilíbrio de ligação com QTLs influentes em toda uma população (ROOS, 2011). Já que o número de marcas era um fator restritivo, seus efeitos deveriam ser estimados com poucas famílias segregantes, o que diminuía significativamente o tamanho amostral. Ao fazer isso, o mapeamento somente detecta um pequeno número de QTLs de grandes efeitos, o que causa uma grande superestimativa dos efeitos das marcas devido ao chamado *Beavis effect* (BEAVIS, 1994; 1998). Esta superestimativa ocorre devido ao fato da análise não ser capaz de detectar QTLs de pequenos e médios efeitos, os quais sabem-se ser os principais influentes em caracteres quantitativos (BERNARDO e CHARCOSSET, 2006; BERNARDO, 2008). Apesar da existência de uma grande quantidade de estudos, bem como expectativas e esperanças envolvendo a MAS e a detecção de QTLs, aparentemente o uso desta técnica mostrou-se algo muito mais ilusório do que se pensava inicialmente (GRATTAPAGLIA, 2014). Como uma alternativa, visando o melhoramento de caracteres de herança complexa, um método capaz de contabilizar os pequenos efeitos genéticos foi sugerido, a seleção genômica (XU et al., 2020).

O termo “seleção genômica” (*Genomic selection* – GS) foi introduzido por Haley e Visscher (1998), mas foram Meuwissen et al. (2001) que, cientes dos problemas da MAS, propuseram o conceito e a metodologia da técnica. O método consiste no uso de um denso

painel de marcas (dezenas ou centenas de milhares) visando cobrir todo genoma, de forma que todos os QTLs, de grandes, médios e pequenos efeitos, estejam em desequilíbrio de ligação com pelo menos uma parte dos marcadores (GRATTAPAGLIA, 2014). Devida a alta densidade de marcas, a GS não necessita de testes de significância para os marcadores, focando exclusivamente em eficiência de seleção (MEUWISSEN et al., 2011; GODDARD e HAYES, 2009; GRATTAPAGLIA, 2014).

É importante ressaltar que a viabilidade econômica da GS ocorreu devido à redução progressiva do custo de marcadores SNP (*Single Nucleotide Polymorphism*). A redução nos custos da genotipagem possibilitou o acesso à uma grande quantidade de SNPs distribuídos no genoma, o que conseqüentemente aumenta a probabilidade de encontrar marcadores em desequilíbrio de ligação com os polimorfismos que causam variações fenotípicas (MEUWISSEN et al., 2011). Estes marcadores, quando utilizados para mapear todo o genoma, permitem estimativas do parentesco entre pares de indivíduos, já que são capazes de capturar os componentes de segregação mendeliana, o que conseqüentemente torna as estimativas de parâmetros mais acuradas em relação às estimativas obtidas com o parentesco esperado (HAYES et al., 2009; LEE et al., 2010).

Grande parte da redução de custo e tempo das genotipagens se deu devido ao desenvolvimento de *chips* de genotipagem para as diferentes espécies de interesse (RASHEED et al., 2017). O primeiro *chip* de SNPs de alta densidade foi construído em 1996 para uso em seres humanos (WANG et al., 1998). Desde então, *chips* de alta densidade já foram gerados e validados para uma grande variedade de animais e espécies agrícolas (CAETANO, 2009; RASHEED et al., 2017). Para as espécies florestais, a variedade de espécies contempladas com *chips* é menor, isso se dá porque os estudos com GS no melhoramento de plantas ainda podem ser considerados recentes (NAKAYA e ISOBE, 2012). Dentre os *chips* destinados às essências florestais, vale destacar os desenvolvidos por Howe et al. (2013) para *Pseudotsuga menziesii* (Mirb.) Franco; Chancerel et al. (2013) para *Pinus pinaster* Ait; Gerald et al. (2013) para *Populus trichocarpa* Torr. & A. Gray ex. Hook.; Pavy et al. (2013) para *Picea glauca* (Moench) Voss; Silva-Junior et al. (2015) para *Eucalyptus* spp.

O principal objetivo em plantas com o estudo da GS é uma redução significativa nos ciclos de seleção, de modo a aumentar o ganho por unidade de tempo e reduzir os custos do programa de melhoramento como um todo (HEFFNER et al., 2010; GRATTAPAGLIA et al., 2018). Para as espécies florestais, a GS tem o potencial de reduzir o ciclo pela metade, já que a seleção pode ser realizada de maneira precoce (a nível de muda), permitindo até mesmo pular

etapas como o teste de progênie, o teste clonal, ou até mesmo ambos em um programa de melhoramento (GRATTAPAGLIA e RESENDE, 2011; RESENDE et al., 2012).

2.3 Estudo de associação genômica (GWAS)

Estudos de associação genômica (GWAS), tem como principal objetivo, identificar polimorfismos de nucleotídeo único (SNPs), que são significativamente influentes e ligados à variação fenotípica de um dado caráter, buscando identificar possíveis regiões do genoma de maior efeito sobre o fenótipo em estudo (INGVARSSON e STREET, 2011; YANG et al., 2013; HIRAOKA et al., 2018). A técnica, que já possui amplo uso na medicina, vem sendo amplamente estudada no melhoramento de animais (FAN et al., 2015; HENRIQUE et al., 2016; HIGGINS et al., 2018; YIN e KÖNIG, 2019) espécies agrícolas (MORRIS et al., 2013; SPINDEL et al., 2015; XIAO et al., 2017; QUERO et al., 2018), e florestais (ALLWRIGHT et al., 2016; LU et al., 2017; RESENDE et al., 2016; MÜLLER et al., 2017; 2018; HIRAOKA et al., 2018; CHHETRI et al., 2019).

O sucesso da GWAS, depende do número de indivíduos avaliados e da disponibilidade de um painel de marcas de alta densidade que seja capaz de captar o desequilíbrio de ligação (LD) entre SNPs e caracteres fenotípicos (MYLES et al., 2009; ZEGEYE et al., 2014). O LD, por sua vez, descreve uma combinação de alelos que ocorrem dentro de uma população de forma não aleatória, ou seja, a presença de um determinado alelo em um loco é condicionada a presença de outro alelo em loco diferente (WILLIAMS e CUMMINGS, 1997; MACKAY e POWELL, 2007).

A GWAS tem o potencial de aumentar a credibilidade da predição genômica, a partir da inclusão de SNPs, em forte associação com os caracteres fenotípicos, nos *chips* utilizados em predições genômicas de rotina, assim como auxiliar no entendimento da herança de caracteres complexos. (LOPES et al., 2015; CERICOLA et al., 2017). O uso da GWAS, possibilitou a identificação de locos para o estudo de doenças importantes em humanos, como a esquizofrenia, diabetes e até mesmo caracteres de herança complexa, como a altura (POCIOT, 2017; YANG et al., 2018; XUE et al., 2018; PARDIÑAS et al., 2019). Em relação a espécies florestais, os estudos se intensificaram com o advento dos *chips* de genotipagem de alta densidade, entretanto, poucas associações foram encontradas, o que já era inicialmente esperado, devido à complexidade dos caracteres de interesse, os quais possuem natureza poligênica, assim como o

baixo número de indivíduos avaliados (CAPPA et al., 2013; PORTH et al., 2013; RESENDE et al., 2016; ALLWRIGHT et al., 2016; LU et al. 2017; MÜLLER et. al., 2018).

Em situações de múltiplas comparações, ou seja, múltiplos testes de significância, que é o caso da GWAS, é de extrema importância o uso de métodos de controle de falsos positivos, também conhecidos como erros de tipo I (MILLER, 1981; ROEDER e WASSERMAN, 2009; JOO et al., 2016). Dentre os métodos de controle, os dois mais utilizados são a correção de Bonferroni e o FDR (*False Discovery Rate*) (KALER e PURCELL., 2019).

O método de Bonferroni, consiste em um limite de significância de p-valores pré-estabelecido de $0,05/N$, onde N é o número de marcadores e 0,05 a taxa de erro experimental (BALDING, 2006). Entretanto, este método é muito conservativo, e tende a inflar erros do tipo II (falsos negativos), principalmente em casos onde as comparações não são totalmente independentes, que é o caso da GWAS, onde certamente há algum nível de correlação entre os SNPs em desequilíbrio de ligação (PERNEGER, 1998; BALDING, 2006). O FDR (BENJAMINI e HOCHBERG, 1995). O FDR, por sua vez, considera a distribuição dos p-valores como uma mistura entre p-valores que aceitam a hipótese de nulidade (p-valores distribuídos de maneira uniforme), e a rejeitam (p-valores que tendem a zero), desta forma, definindo quais comparações são realmente significativas. Trata-se um método que também é considerado conservativo para a GWAS, porém em menor intensidade que Bonferroni (WANG et al., 2016a; MÜLLER et al., 2018; KALER e PURCELL, 2019).

A presença de estrutura de população, também é um problema relacionados à detecção falsos positivos em estudos de associação (VILHJALMSSON e NORDBORG, 2013; HUANG e HAN, 2014). Na verdade, são ditos como os principais problemas causadores de erros tipo I (YU et al., 2006). Estrutura de população é a presença de diferentes frequências alélicas entre subpopulações, subgrupos ou famílias devido à ancestralidade, já o parentesco representa as relações familiares entre pares de indivíduos (MÜLLER et al., 2017). Muitas ferramentas estatísticas foram propostas e utilizadas para controlar estes efeitos (ZHANG et al., 2008). Dentre os vários métodos, atualmente, o uso de modelagem mista é o mais amplamente utilizado para controlar tanto falsos positivos, quanto falsos negativos (KANG et al., 2010; YIN e KÖNIG, 2019).

2.4 GBLUP-D e método *single-step*

O ABLUP, ou *pedigree*-BLUP é a metodologia padrão na predição de valores genéticos a partir de um parentesco esperado entre indivíduos (CROSSA et al., 2010). O método foi primeiramente sugerido por Fisher (1918) e consiste no uso do *pedigree* para o cálculo da matriz de parentesco (\mathbf{A}) para a predição de valores genéticos baseados no modelo aditivo infinitesimal. O método GBLUP, por sua vez, é derivado do ABLUP, porém diferente da matriz \mathbf{A} , faz uso de uma matriz estimada a partir de genótipos e marcadores genéticos. Esta matriz é denominada \mathbf{G} e trata-se da proporção real do genoma que dois indivíduos compartilham, diferente da matriz \mathbf{A} , que é uma proporção esperada (GODDARD et al., 2011). O uso da matriz \mathbf{G} permite, não apenas retificar a baixa acurácia da matriz de *pedigree*, mas também capturar a variação dentro de famílias, causada pela segregação Mendeliana aleatória (GRATTAPAGLIA et al., 2018).

Combinar as informações de marcadores genéticos em uma matriz de parentesco foi primeiramente sugerido por Nejati-Javaremi et al. (1997), mas o método apenas foi utilizado no contexto da seleção genômica vários anos depois (VAN RADEN, 2008). Atualmente, várias formas de computação da matriz \mathbf{G} são conhecidas e amplamente utilizadas (HAYES e GODDARD, 2008; VAN RADEN, 2008; ASTLE e BALDING, 2009; YANG et al., 2010; GODDARD et al., 2011).

Na predição genômica, desvios da linearidade são extremamente comuns devido às complexas interações gênicas existentes em caracteres quantitativos (CUEVAS et al., 2016). Tais desvios podem ser melhor contabilizados por abordagens semiparamétricas, como a Regressão Kernel de Reprodução do Espaço de Hilbert (*Reproducing Kernel Hilbert Spaces – RKHS*) (KIMELDORF e WAHBA, 1971).

O método RKHS necessita de uma matriz de Kernel (\mathbf{K}) baseada em qualquer tipo de similaridade entre indivíduos, esta matriz tem o objetivo de restringir o ajuste do modelo definindo o espaço provável de soluções de uma função contido dentro de um amplo espaço de funções, i.e., espaço de Hilbert (AKHIEZER e GLAZMAN, 1963; DE LOS CAMPOS et al., 2009; WANG et al., 2014). O Kernel tem como função na RKHS, definir implicitamente uma coleção de funções reais dentro de um espaço por ele mesmo estabelecido (RESENDE et al., 2014).

A matriz \mathbf{G} de parentesco genômico aditivo se encaixa nos requisitos de uma matriz de Kernel, já que computa as similaridades genômicas entre indivíduos genotipados, seu uso como

matriz \mathbf{K} , torna o método GBLUP um caso especial de regressão RKHS, proporcionando uma abordagem bayesiana semiparamétrica da metodologia, porém equivalente (RESENDE et al., 2014; DE LOS CAMPOS et al., 2009; MOROTA e GIANOLA, 2014; JIANG e REIF, 2015, CUEVAS et al., 2016). Apesar da metodologia ser semiparamétrica, a escolha de um Kernel paramétrico, como por exemplo a matriz \mathbf{G} de parentesco aditivo, proporciona à variância associada ao Kernel \mathbf{G} , a interpretação de variância genética aditiva (σ_a^2) (DE LOS CAMPOS et al., 2010).

Atualmente, há um elevado número de estudos que mostram que os métodos baseados em Kernel são superiores ou equivalentes a métodos paramétricos (PÉREZ-RODRÍGUEZ et al., 2012; MOROTA e GIANOLA, 2014; JACQUIN et al., 2016).

Sabe-se que a dominância é o fenômeno em que o valor genotípico do heterozigoto, desvia da média do valor genotípico dos dois homozigotos (FALCONER e MACKAY, 1996). E apesar do interesse em incluir este efeito na predição de valores genéticos, estimar estes efeitos de maneira acurada apenas com uso de *pedigree*, é algo desafiador e complexo (VITEZICA et al., 2013). Entretanto, com os recentes avanços na genotipagem de alta densidade, a inclusão deste efeito tornou-se passível de ser realizada (LEGARRA et al., 2009). Atualmente, a inclusão do efeito de dominância em modelos de predição genômica é recomendada por vários autores (SU et al., 2012; VITEZICA et al., 2013; MUÑOZ et al., 2014; XIANG et al., 2016; TAN et al., 2017), já que na presença de um efeito significativo de dominância sobre o caráter, a inclusão deste efeito tende a melhorar as estimativas e a capacidade preditiva em relação aos modelos genômicos contendo apenas o efeito aditivo (DENIS e BOUVET, 2011; WELLMAN e BENNEWITZ, 2012; MUÑOZ et al., 2014; NISHIO e SATOH, 2014; DUENK et al., 2017; DE ALMEIDA FILHO et al., 2019).

A inclusão do efeito de dominância em modelos RKHS ocorre a partir de uma combinação de Kernels, as quais podem aprimorar predições genômicas (VITEZICA et al., 2013). A matriz de parentesco genômico de dominância (\mathbf{D}) (SU et al., 2012) pode ser computada como Kernel paramétrico junto à \mathbf{G} em modelos RKHS, e, desta forma, estimar tanto os efeitos aditivos quanto os efeitos de dominância (MOROTA et al., 2013; MOROTA e GIANOLA, 2014; SANTOS et al., 2017).

Muitos estudos confirmam acerca das superioridade de modelos preditivos que fazem uso do parentesco genômico realizado (\mathbf{G}) em detrimento do parentesco esperado (\mathbf{A}) (BEALIEU et al., 2014; MUÑOZ et al., 2014; RESENDE et al., 2014; CAPPA et al., 2017;

GRATTAPAGLIA, 2018; TAN et al., 2018). Entretanto, existem vários relatos acerca do potencial incremento no desempenho de modelos de seleção genômica ao utilizar uma combinação destas duas matrizes, superando a capacidade preditiva obtida, tanto por modelos de parentesco esperado (ABLUP), quanto parentesco genômico realizado (GBLUP) (FORNI et al., 2011; CHRISTENSEN, 2012; BEALIEU et al., 2014; CAPPÀ et al., 2017; 2019; RATCLIFFE et al., 2017; RATCLIFFE, 2019; VELAZCO et al., 2019; ZHAO et al., 2019).

Este método foi proposto por pesquisadores do melhoramento animal (LEGARRA et al., 2009, MISZTAL et al., 2009; AGUILAR et al., 2010; CHRISTENSEN e LUND, 2010) para casos onde se tinham disponíveis as *pedigrees* de todos os indivíduos fenotipados, porém apenas parte destes foram genotipados. A metodologia foi denominada ssGBLUP (*single-step* GBLUP), mas também é conhecida como HBLUP, e consiste em considerar indivíduos genotipados (presentes na matriz **G** e **A**) e não genotipados (presentes apenas na matriz (**A**) na mesma avaliação, por meio do uso de uma matriz híbrida (**H**) entre as matrizes de parentesco supracitadas, visando combinar informações fenotípicas, *pedigree* e genômicas em uma só avaliação (PÉREZ-RODRÍGUEZ et al., 2017).

A informação adicional gerada ao combinar as matrizes **G** e **A**, age como uma “ponte de parentesco” que conecta progênes e genitores, gerando informações mais acuradas e aumentando a probabilidade de um assertivo ordenamento de indivíduos geneticamente superiores (CAPPÀ et al., 2017; 2019).

O HBLUP, além de melhorar a predição de valores genéticos, permite realizar inferências sobre a totalidade dos indivíduos de um teste genético, tendo apenas genotipado parte destes indivíduos, e por isso, é bastante indicado para o melhoramento florestal, visto que, normalmente, os testes de progênie possuem muitos indivíduos, o que torna a genotipagem do teste todo um processo de elevado custo financeiro (PÉREZ-RODRÍGUEZ et al., 2017; RATCLIFFE, 2019).

2.5 Predição de valores, acurácia dos modelos e tamanho efetivo

Para a prática da seleção genômica (GS), são necessários três conjuntos de dados, ou três populações (MEUWISSEN et al., 2001; GODDARD e HAYES, 2007; RESENDE et al., 2008): A população de descoberta, ou treinamento (PD); população de validação (PV); e população de seleção (PS) (Figura 3).

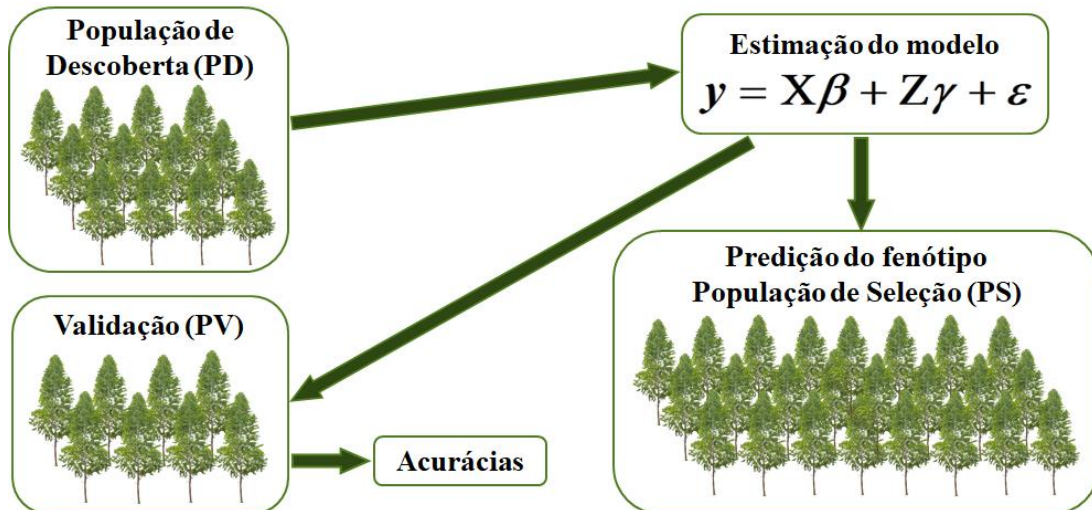


Figura 3 - Populações utilizadas na prática da Seleção Genômica (GS).

A população de descoberta é fenotipada para os caracteres de interesse e genotipada com um grande número de marcadores. Este conjunto de dados é utilizado para a obtenção de equações que sejam capazes de prever os valores genéticos genômicos (VGG) e/ou os fenótipos dos caracteres de interesse. A população de validação consiste em um conjunto de dados menor que a população de descoberta que também são fenotipados e genotipados. Os dados são então submetidos ao modelo para obtenção dos valores preditos. A população de seleção não precisa ser fenotipada, apenas genotipada. O conjunto de dados consiste apenas em marcas SNPs avaliadas nos indivíduos da população em questão. Os valores genéticos genômicos e os fenótipos futuros, são estimados a partir das equações derivadas da população PD, mas a acurácia de seleção é àquela estimada na população PV (RESENDE et al., 2008).

Independente do ambiente em que os indivíduos se encontram, seus VGGs podem ser preditos, desde que possuam equações de predição com alta acurácia para tais ambientes (ESTOPA, 2017). A acurácia obtida a partir da própria correlação entre o valor genético verdadeiro de um indivíduo e aquele estimado a partir de genotipagem e/ou fenotipagem (RESENDE et al., 2011; GRATTAPAGLIA, 2014). Esta estimativa é dita como um parâmetro chave, pois é altamente relacionada aos ganhos na seleção e útil em comparações entre modelos de GS (XU, 2013). A acurácia dos modelos de predição genômica é influenciada pelo método estatístico empregado, o tamanho da população de descoberta, número de marcas utilizadas e o parentesco entre a população de descoberta e a população de validação (TAN et al., 2017). A GS é capaz de captar com sucesso a variação genética e por isso é capaz de atingir acurácias

seletivas para os caracteres de crescimento e qualidade madeiras tão boas ou até melhores do que aquelas obtidas via seleção fenotípica convencional (RESENDE et al., 2012).

Vários autores reportam que uma maior população de descoberta, tende a melhorar a acurácia na predição dos VGGs (NAKAYA e ISOBE, 2012). Em geral, quanto maior a diversidade genética da espécie em estudo, maior deve ser o tamanho da população para a obtenção de VGGs com alta acurácia (MUJIBI et al., 2011).

O conceito de tamanho efetivo populacional (N_e) foi introduzido por Wright (1931; 1933) e até os dias atuais é tido como um parâmetro chave de extrema importância tanto na genética de populações quanto na genética quantitativa, pois pode ser interpretado como uma medida de representatividade genética dentro de uma população, que permite a distinção entre o número amostral e a representativa genética total da amostra (FALCONER e MACKAY, 1996; VENCOVSKY, 1987; WANG, 1997; VENCOVSKY e CROSSA, 1999; FRANKHAM et al., 2010; WANG et al., 2016b).

Para a GS, o N_e tem grande importância e influência na acurácia e capacidade preditiva de modelos de seleção genômica (HAYES et al., 2009; LEE et al., 2017; GRATTAPAGLIA e RESENDE, 2011; BEAULIEU et al., 2014). Quanto maior o N_e de uma dada população ou amostra, maior o número de marcadores necessários para cobrir o genoma e capturar todo o desequilíbrio de ligação (LD) entre marcadores e QTLs, de modo a atingir uma acurácia preditiva satisfatória para a prática da seleção genômica (GRATTAPAGLIA e RESENDE, 2011). Em outras palavras, a densidade de marcas necessária para a prática da GS depende do nível de LD entre marcas e QTLs, o qual, por sua vez, é dependente do N_e (GRATTAPAGLIA, 2014).

3. OBJETIVOS

Avaliar a capacidade preditiva para volume e qualidade da madeira em modelos de *pedigree*, aditivo, aditivo-dominante e híbrido *single-step* em um teste de progênies de polinização aberta de *E. benthamii*

3.1. Objetivos Específicos

Os objetivos específicos do presente estudo são: 1) verificar as diferenças entre a matriz de parentesco esperado e a matriz de parentesco genômico realizado; 2) identificar regiões de maior efeito nos fenótipos em estudo; 3) conferir acerca do viés de predição (*bias*) dos diferentes modelos avaliados ; 4) comparar o efeito da inclusão da matriz de dominância (**D**) na modelagem e 5) verificar a eficiência do método *single-step* em considerar indivíduos genotipados e não genotipados em uma mesma análise.

4. MATERIAL E MÉTODOS

4.1 Material experimental e fenotipagem

A população em estudo é a mesma avaliada por Estopa (2017). Consiste em um teste de progênies de polinização aberta de *E. benthamii* localizado em Otacílio Costa, Santa Catarina, plantado em outubro de 2010, sendo 81 progênies com 20 blocos e uma planta por parcela (*Single Tree Plot* – STP) (Tabela 1). O material foi selecionado em dois testes de progênies implantados em duas localidades (procedências) diferentes do município de Telêmaco Borba - PR. Entretanto, ambos os testes possuem material de origem de Kedumba Valley – NSW, Austrália (Tabela 1).

Tabela 1 - Dados gerais do experimento avaliado.

Dados do experimento	
Espécie	<i>E. benthamii</i>
Famílias de polinização aberta	81
Blocos	20
Tipo de parcela	<i>Single tree plot</i> – STP
Indivíduos genotipados	780
Indivíduos não genotipados (NIR)	367
Indivíduos genotipados após controle de qualidade	671
Famílias após o controle de qualidade	77
Número de SNPs	14.638
Idade da fenotipagem*	4 anos; 6 anos
Origem do material	Kedumba Valley – NSW (AU)
Local de implantação	Otacílio Costa – PR
Número de caracteres avaliados	5

*: 4 anos para os caracteres de qualidade da madeira; 6 anos para volume individual.

Os indivíduos foram fenotipados aos três anos de idade para DAP (Diâmetro a Altura do Peito, cm) e altura total (m), sendo então possível estimar o volume de madeira (Vol em m³). Foram selecionados os melhores 780 indivíduos para o volume de madeira previamente estimado por meio do método de estimação REML (*Restricted Maximum Likelihood*),

associado à metodologia BLUP (*Best Linear Unbiased Prediction*) a partir do seguinte modelo misto:

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

Em que y é o vetor de observação dos n dados, b é o efeito fixo associado ao bloco, u é o efeito genético aditivo aleatório associado ao indivíduo, e é o vetor associado ao erro residual aleatório, \mathbf{x} e \mathbf{z} são as matrizes de incidência para os respectivos efeitos.

A população previamente selecionada (780 indivíduos) foi fenotipada para caracteres de qualidade da madeira aos quatro anos de idade. Duas baguetas foram coletadas com um trado de incremento à 1,3 m, uma delas foi utilizada para a determinação da densidade da madeira, a outra para a caracterização química.

A densidade básica da madeira (DBM) foi determinada por meio do método da balança hidrostática, de acordo com as normas da ABNT (Associação Brasileira de Normas Técnicas), por meio da relação direta entre o peso absolutamente seco (P_s) e o volume saturado (V_u) da amostra:

$$DBM = P_s / V_u \quad (2)$$

As baguetas destinadas para a caracterização química foram transformadas em serragem e encaminhadas para as análises no laboratório de Celulose e Papel da Universidade Federal de Viçosa (UFV). Os caracteres tecnológicos associadas à produção de polpa celulósica que foram constatadas são: teor de lignina (LIG), extrativos (EXT) e carboidratos (CBO- soma dos polímeros que compõem a celulose) por meio das normas apresentadas na Tabela 2.

Tabela 2 - Normas utilizadas na análise química da madeira.

Caracteres analisados	Metodologia
Extrativos – solubilidade em acetona	TAPPI T280 PM-99 (2000)
Lignina solúvel	GOLDSCHMID (1971)
Lignina insolúvel	GOMIDE & DEMUNER (1986)
Carboidratos	WALLIS et al. (1996)

Fonte: Estopa (2017).

Utilizando os 780 indivíduos selecionados, Estopa (2017) construiu modelos de predição via espectroscopia NIR para os caracteres de qualidade da madeira, realizando a predição de 367 indivíduos, os quais não foram genotipados (Tabela 1).

No ano de 2016, ocorreu uma nova fenotipagem do DAP e altura da população em estudo, proporcionando a estimativa do volume aos seis anos de idade (VOL6).

4.2 Genotipagem e controle de qualidade

Foram coletadas amostras de folhas para a extração de DNA dos 780 indivíduos da população de descoberta. A extração do DNA e genotipagem foram realizadas pela empresa Heréditas (Brasília, DF) utilizando o método CTAB (GRATTAPAGLIA e SEDEROFF, 1994). As amostras dos 780 indivíduos foram genotipadas com 60.728 SNPs pela Geneseek (Lincoln, NE), por meio do EUChip60K, um *chip* de alta densidade (SILVA-JUNIOR et al., 2015). O controle de qualidade foi realizado em três etapas por meio do software PLINK (PURCELL et al., 2007):

a) *Call rate*: As marcas passaram por um filtro de SNPs com 95 % de *call rate*, o qual busca eliminar valores perdidos na genotipagem, ou seja, eliminar marcas que não estejam presentes em ao menos 95% dos indivíduos. Os indivíduos também foram filtrados (*call rate* > 90%), eliminando aqueles que não possuem ao menos 90% das marcas. Ao fim desta etapa, restaram 671 indivíduos com 33.933 SNPs cada (Tabela 1).

b) *Minor Allele Frequency* (MAF): Esta etapa buscou eliminar alelos fixos (MAF > 0),

restando 15.293 SNPs.

c) Equilíbrio de Hardy-Weinberg (EHW): Foram eliminados 655 SNPs que mostraram fortes desvios do equilíbrio de Hardy-Weinberg (p-valor $< 1 \times 10^{-6}$), restando 14.638 SNPs (Tabela 1).

Após o controle de qualidade, os dados perdidos foram imputados com a média de cada loco.

4.3 Estudo de associação genômica (GWAS)

Com o objetivo de verificar a existência de associação dos marcadores SNP com os caracteres, foi realizado um estudo de associação genômica (GWAS) utilizando o método EMMAX (KANG et al., 2010) por meio do pacote *cpgen* (HEUER, 2018). O método EMMAX consiste no uso de modelagem de efeitos mistos aliada à matriz de parentesco genômico na criação de uma matriz de variâncias e covariâncias entre fenótipos que represente e modele a estrutura populacional e o parentesco (KANG et al., 2010). A raiz quadrada inversa desta matriz é utilizada na computação de p-valores via método dos mínimos quadrados generalizados (GLS) a partir do seguinte modelo misto:

$$y = \mathbf{X}\beta + \mathbf{z}_i\alpha_i + u + e \quad (3)$$

Onde y é a medida fenotípica do caráter em avaliação; β é o vetor de efeitos fixos; \mathbf{z}_i é o vetor de incidência de SNPs (0, 1, ou 2 para a i -ésima marca); α_i é o efeito de substituição alélica para o i -ésimo marcador e u é o efeito aleatório do modelo com variância $\text{Var}(u) = \sigma_g^2 \mathbf{G}$, onde \mathbf{G} é a matriz $n \times n$ de parentesco genômico aditivo (VanRaden, 2008) entre indivíduos e e é a matriz $n \times n$ de efeitos residuais com variância $\text{Var}(e) = \sigma_e^2 \mathbf{I}$. A matriz de variância e covariância entre fenótipos é representada como $\text{Var}(y) = \sigma_g^2 \mathbf{ZGZ}' + \sigma_e^2 \mathbf{I}$.

O efeito de substituição alélica relatado no modelo acima (α) foi obtido, utilizando a seguinte expressão:

$$\alpha = \hat{a} + (q - p)\hat{d} \quad (4)$$

Onde p e q são as frequências alélicas; \hat{a} e \hat{d} são estimativas dos efeitos aditivo e dominante, respectivamente, obtidos por meio de regressão linear.

As marcas consideradas significativas no estudo de associação genômica, tiveram suas herdabilidades (h_{GWAS}^2) estimadas por:

$$h_{GWAS}^2 = \frac{2pq\alpha^2}{\sigma_F^2} \quad (5)$$

Onde σ_F^2 é a variação fenotípica do respectivo caráter em estudo; p , q e α são definidos como na equação 4.

De posse dos p-valores e dos *scores* ($-\log_{10}p$) de cada marca, os gráficos *Manhattan plots* e *QQ-plots* foram feitos com o uso do pacote *GWASTools* (GOGARTEN et al., 2012). Afim de constatar a significância de possíveis associações, a correção de Bonferroni ($\alpha = 0,05$) foi aplicada para controlar possíveis erros do tipo I, assim como o controle de falsos positivos (FDR) de Benjamini e Hochberg (1995) a 5%. Além das correções tradicionais, foi utilizado um limite de significância *ad hoc* (p-valor $\leq 8 \times 10^{-5}$), baseado em Müller et al. (2018).

4.4 Computação das matrizes de efeito aditivo, dominante e híbrida

A matriz de parentesco baseada em informação de *pedigree* (\mathbf{A}), foi computada utilizando o parentesco genético esperado entre os indivíduos das progênes.

Com informações dos marcadores SNPs, a matriz de parentesco genômico aditivo (\mathbf{G}) foi obtida de acordo com VanRaden (2008):

$$\mathbf{G} = \frac{\mathbf{ZZ}'}{2 \sum_{i=1}^p p_i(1-p_i)} \quad (6)$$

Seja \mathbf{M} a matriz de incidência que especifica quais alelos cada indivíduo herdou, com dimensões $n \times m$, onde n é o número de indivíduos e m o número de locos. Dado que a frequência do segundo alelo em um loco i seja p_i , e que a matriz \mathbf{P} contenha as frequências alélicas expressas como uma diferença de 0,5 e multiplicadas por 2, de modo que cada coluna i de \mathbf{P} seja $2(p_i - 0,5)$, subtrair \mathbf{P} de \mathbf{M} , seria igual \mathbf{Z} . Esta correção proporciona com que as médias dos efeitos alélicos sejam zero, e além disso, dá mais créditos a alelos raros do que alelos comuns no cálculo de \mathbf{G} . A divisão por $2 \sum p_i(1-p_i)$ escala a matriz de parentesco para ser análoga à matriz de parentesco dos coeficientes de Wright (matriz \mathbf{A}).

A matriz de parentesco genômico de dominância (\mathbf{D}) foi obtida de acordo com Vitezica et al. (2013):

$$\mathbf{D} = \frac{\mathbf{SS}'}{\sum_{i=1}^p \{2p_i(1-p_i)\}^2} \quad (7)$$

Onde \mathbf{S} é a matriz codificada como $-2(1-p_i)^2$ para o homozigoto alternativo, $2p_i(1-p_i)$ para heterozigotos e $-2p_i^2$ para o homozigoto de referência; p_i é o mesmo utilizado na matriz \mathbf{G} .

A matriz híbrida (\mathbf{H}) que inclui *pedigree* e parentesco genômico foi obtida utilizando a abordagem de Legarra et al. (2009) e Aguilar et al. (2010):

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}_{nn} + \mathbf{A}'_{gn} \mathbf{A}_{gg}^{-1} (\mathbf{G}_a - \mathbf{A}_{gg}) \mathbf{A}_{gg}^{-1} \mathbf{A}_{gn} & \mathbf{A}_{gn} \mathbf{A}_{gg}^{-1} \mathbf{G}_a \\ \mathbf{G}_a \mathbf{A}_{gg}^{-1} \mathbf{A}_{gn} & \mathbf{G}_a \end{bmatrix} \quad (8)$$

As matrizes \mathbf{A}_{gg} , \mathbf{A}_{nn} e \mathbf{A}_{gn} são submatrizes de \mathbf{A} , e contêm o parentesco entre indivíduos genotipados (\mathbf{A}_{gg}), não genotipados (\mathbf{A}_{nn}) e entre indivíduos genotipados e não genotipados (\mathbf{A}_{gn}) (CHRISTENSEN et al., 2012). \mathbf{G}_a é a matriz de parentesco genômico (\mathbf{G}) ajustada a partir da seguinte expressão:

$$\mathbf{G}_a = \beta \mathbf{G} + \alpha \quad (9)$$

Onde β e α são obtidos ao resolver o seguinte sistema de equações:

$$\text{Avg}\{\text{diag}(\mathbf{G})\} \beta + \alpha = \text{Avg}\{\text{diag}(\mathbf{A}_{gg})\}, \quad (10)$$

$$\text{Avg}(\mathbf{G}) \beta + \alpha = \text{Avg}(\mathbf{A}_{gg}) \quad (11)$$

Este ajuste tem como objetivo reescalonar \mathbf{G} a fim de obter a matriz \mathbf{G}_a , de modo que a média de sua diagonal se iguale à média da diagonal de \mathbf{A}_{gg} , bem como a média de todos os elementos de \mathbf{G}_a se iguale a média de todos os elementos de \mathbf{A}_{gg} .

4.5 Parentesco, tamanho efetivo e desequilíbrio de ligação

4.5.1 Parentesco esperado e realizado

Com o objetivo de comparar o parentesco esperado (*pedigree*) da população de polinização aberta com os diferentes níveis de parentesco genômico realizado, foram construídos *heatmaps* das matrizes \mathbf{A}_{gg} (matriz de parentesco esperado contendo apenas os indivíduos genotipados) e \mathbf{G} utilizando o pacote *gplots* (WARNES et al., 2019) e *barplots* das frequências de parentesco contidas em ambas as matrizes por meio do pacote *ggplot2* (WICKHAM, 2016).

4.5.2 Tamanho efetivo populacional

O tamanho efetivo populacional (N_e) foi estimado com base no método do desequilíbrio de ligação (LDN_e) de Waples e Do (2008) implementado no software NeEstimator v2.1 (DO et al., 2014). Foi utilizado o modelo de segregação aleatória, assim como um $MAF \geq 0,02$ na exclusão de alelos raros. O LDN_e foi então, obtido por meio do seguinte estimador:

$$LDN_e = \frac{2}{3(\hat{r}^2 - 1/S)} \quad (12)$$

Onde \hat{r}^2 é a média dos valores de correlação entre diferentes locos e é obtido no estimador não viesado do desequilíbrio de ligação de Burrows, calculado para cada par de alelos em cada par de locos (WEIR, 1979; WAPLES, 2006), e S é o tamanho amostral.

Buscando inferir acerca da confiabilidade da estimativa, foram estimados intervalos de confiança a 95% de probabilidade por meio do método *Jackknife* (WAPLES e DO, 2008).

4.5.2 Análise da extensão do desequilíbrio de ligação (LD)

Estimativas par a par de LD foram obtidas por meio da estimativa clássica da correlação (r^2) entre pares de SNP separados por uma distância máxima de 100 kilobases (kb), por meio do software PLINK (PURCELL et al., 2007).

O decaimento de r^2 com a distância entre os SNPs (*LD decay*), foi ajustado utilizando a expressão de Hill e Weir (1988), por meio de um script de *R* adaptado de Marroni et al. (2011):

$$E(r^2) = \left[\frac{10 + C}{(2 + C)(11 + C)} \right] \left[1 + \frac{(3 + C)(12 + 12C + C^2)}{n(2 + C)(11 + C)} \right] \quad (13)$$

Onde n é o tamanho amostral; C é o parâmetro a ser estimado, o qual representa o produto do parâmetro de recombinação populacional ($\rho = 4N_e r$; N_e é o tamanho efetivo populacional e r a fração de recombinação entre os pares de SNPs).

4.6 Modelagem e performance dos modelos

4.6.1 Ajuste dos modelos e parâmetros genéticos

A matriz \mathbf{G} , responsável por ajustar os modelos GBLUP aditivo e aditivo-dominante, contém apenas indivíduos genotipados. Já a matriz \mathbf{H} , por sua vez, contempla a totalidade do teste, ou seja, indivíduos genotipados e não genotipados, no caso, tanto os indivíduos genotipados após controle de qualidade (671), quanto os indivíduos não genotipados (367) (Tabela 1), gerando diferentes tamanhos amostrais entre estes modelos. Devido a isso, para comparação, o modelo A (ABLUP) foi ajustado com dois *datasets*: um contendo apenas os indivíduos genotipados (modelo Ag) e outro contemplando a totalidade do teste (modelo At).

O ajuste de ambos os modelos A foi realizado com o pacote *rrBLUP* (ENDELMAN, 2011), por meio do seguinte modelo linear misto:

$$y = \mathbf{X}\beta + \mathbf{Z}_a u + e \quad (14)$$

Onde y é a medida fenotípica do caráter em avaliação; β é o vetor de efeito fixo associado ao bloco; u é o vetor aleatório de efeitos aditivos; e é o efeito residual aleatório; \mathbf{X} e \mathbf{Z}_a são as matrizes de incidência para os respectivos efeitos. Para este modelo, assumiu-se que $a \sim N(0, \mathbf{A}\sigma_a^2)$.

Os demais modelos foram ajustados como casos particulares de Modelo de Regressão Bayesiana de Kernel de Reprodução do Espaço de Hilbert (*Reproducing Kernel Hilbert Spaces* – RKHS), por meio do pacote estatístico BGLR (PEREZ e DE LOS CAMPOS, 2014).

Para o ajuste do modelo GBLUP que considera apenas o efeito aditivo (GA), a matriz \mathbf{G} de parentesco genômico foi fornecida como matriz de *Kernel* para o ajuste do seguinte modelo misto:

$$y = \mathbf{X}\beta + \mathbf{Z}_a a + e \quad (15)$$

Onde y é a medida fenotípica do caráter em avaliação; β é o vetor de efeito fixo associado ao bloco; a é o vetor aleatório de efeitos aditivos; e é o efeito residual aleatório; \mathbf{X} e \mathbf{Z}_a são as matrizes de incidência para os respectivos efeitos. Para a análise, as seguintes distribuições *a priori* foram assumidas:

$$p(\mu) \propto \text{constante}; a | K_1, \sigma_a^2 \sim N(0, K_1 \sigma_a^2)$$

Para o ajuste do modelo que considera tanto o efeito aditivo quanto dominante (GAD), as matrizes \mathbf{G} de parentesco genômico aditivo e \mathbf{D} de dominância foram fornecidas como matrizes de *kernel* para o ajuste do seguinte modelo misto:

$$y = \mathbf{X}\beta + \mathbf{Z}_a a + \mathbf{Z}_d d + e \quad (16)$$

Onde y é a medida fenotípica do caráter em avaliação, b é o vetor de efeito fixo associado ao bloco; a é o vetor aleatório de efeitos aditivos; d é o vetor aleatório de desvios de dominância; e é o efeito residual aleatório; \mathbf{X} , \mathbf{Z}_a e \mathbf{Z}_d são as matrizes de incidência para os respectivos efeitos. Para a análise, as seguintes distribuições *a priori* foram assumidas:

$$p(\mu) \propto \text{constante}; a | K_1, \sigma_a^2 \sim N(0, K_1 \sigma_a^2) \text{ e } d | K_2, \sigma_d^2 \sim N(0, K_2 \sigma_d^2)$$

Nas pressuposições dos modelos GA e GAD acima, K_1 e K_2 são as matrizes de *Kernel* \mathbf{G} e \mathbf{D} respectivamente.

Para o ajuste do modelo híbrido *single-step* HBLUP (H), a matriz \mathbf{G} foi substituída pela matriz \mathbf{H} no modelo 12. As mesmas *prioris* do modelo GA foram assumidas.

Para a parametrização dos componentes de variância associados à cada modelo (σ_a^2 , σ_d^2 e σ_e^2), foram assumidas as distribuições padrão do pacote BGLR (PÉREZ e DE LOS CAMPOS, 2014): $\sigma_i^2 \sim \chi^{-2}(\sigma_i^2 | df_i, S_i)$ com df_i graus de liberdade e parâmetro de escala $S_i > 0$.

Os modelos para os diferentes caracteres avaliados foram ajustados por meio do algoritmo amostrador de Gibbs (*Gibbs sampling*) (GERMAN e GERMAN, 1984; CASELLA e GEORGE, 1992). Após análises prévias nas cadeias de Markov com o auxílio do pacote *coda* (PLUMMER et al., 2006), foram consideradas 60.000 iterações, sendo as 2.000 primeiras excluídas como *burn-in* para garantir estacionariedade, e um intervalo (*thinning*) de 100 para garantir a independência e eliminar autocorrelação entre os passos da cadeia.

As herdabilidades, tanto fenotípicas (modelo de *pedigree*) quanto genômicas foram estimadas no sentido amplo e restrito como:

$$h_G^2 = \frac{\sigma_a^2 + \sigma_d^2}{\sigma_a^2 + \sigma_d^2 + \sigma_e^2} \quad (17)$$

$$h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_d^2 + \sigma_e^2} \quad (18)$$

Onde h_G^2 é a herdabilidade no sentido amplo (somente estimada para o modelo GAD); h_a^2 é a herdabilidade no sentido restrito; σ_a^2 é a variância genética aditiva associada à matriz de parentesco, σ_d^2 é a variância genética associada à dominância (somente contabilizada nos modelos GAD) e σ_e^2 é a variância residual.

Buscando entender se a variância de dominância, quando considerada na modelagem, exerce algum tipo de interação na estimativa da variância aditiva, foi calculada uma razão entre a variância aditiva estimada considerando o efeito de dominância (modelo GAD - $\sigma_{a(GAD)}^2$) e a variância aditiva estimada sem considerar o efeito de dominância (modelo GA - $\sigma_{a(GA)}^2$):

$$\sigma_{a(GAD)}^2 / \sigma_{a(GA)}^2 .$$

Com o objetivo de verificar acerca da magnitude da variância associada ao efeito de dominância em relação à variância aditiva estimada com o modelo GAD, foi calculada uma razão entre as variâncias: σ_d^2 / σ_a^2 .

4.6.2 Performance dos modelos

A performance dos modelos ajustados foi comparada por dois critérios:

- a) Viés da predição (*bias*): Estimado a partir do coeficiente de regressão obtido entre os valores preditos pelo modelo e os valores reais, onde 1 é o valor ideal (i.e., ausência de viés).
- b) Capacidade preditiva (r_{gy}): Foi acessada por meio de validação cruzada do tipo 10-*fold*. Os dados foram divididos em dois *subsets*, um com 90% dos indivíduos (população de descoberta - PD) e outro com 10% dos indivíduos restantes (população de validação - PV). Os indivíduos da PV tiveram seus fenótipos preditos a partir do ajuste do modelo utilizando a PD. Este processo foi repetido 10 vezes, de maneira que todos os indivíduos tiveram seus fenótipos preditos. Ao final do processo, os fenótipos preditos foram correlacionados com os respectivos valores reais.

5. RESULTADOS E DISCUSSÃO

5.1 Parentesco esperado e realizado

Dentre os parentescos capturados pela matriz A_{gg} de *pedigree* (Figura 4 – A), 98,5% dos indivíduos foram considerados não relacionados geneticamente e 1,5%, como meios irmãos. Tal comportamento era esperado, afinal, não há como capturar parentescos maiores que meios irmãos em testes de progênie de polinização aberta, que é o caso do presente estudo.

A estratégia de progênie de polinização aberta é mais simples, ou seja, demanda menor custo de tempo, dinheiro e mão-de-obra do que estratégias baseadas em polinização controlada (GAMAL EL-DIEN et al., 2016; KLÁPŠTĚ et al., 2018). Entretanto este tipo de estratégia não é capaz de captar completamente o parentesco existente em uma população, gerando problemas retratados na literatura como “parentesco escondido” (*hidden relatedness*). O “parentesco escondido”, por sua vez, pode afetar a acurácia de estimativa dos parâmetros genéticos, valores genotípicos, capacidade preditiva e inflar a variância aditiva (SQUILLACE, 1974; NAMKOONG et al., 1998; KLÁPŠTĚ et al., 2018; TAMBARUSSI et al., 2018).

Tabela 3 - Comparação entre o parentesco esperado obtido via *pedigree* e o parentesco genômico realizado par a par entre indivíduos para um teste de progênie de polinização aberta de *E. benthamii*, Otacílio Costa - SC.

<i>Range</i>	Não aparentados (442758) ^a		Meios-irmãos (6812)	
	<i>Pedigree</i>	Genômico	<i>Pedigree</i>	Genômico
Mínimo	0	-0,137	0.25	-0,106
Média	0	-0,005	0.25	0,249
Máximo	0	1,117	0.25	1,01

^a: O número em parênteses indica o número de parentescos par a par.

O uso de marcadores moleculares pode captar o “parentesco escondido”, assim como potenciais erros na matriz de *pedigree* (KLÁPŠTĚ et al., 2018). Isso pode ser observado na matriz **G** de parentesco realizado (Figura 4 – B; Tabela 3), que, por sua vez, apresenta uma gama maior de parentescos, revelando a presença de uma estrutura mais complexa do que àquela captada com o uso de *pedigree*.

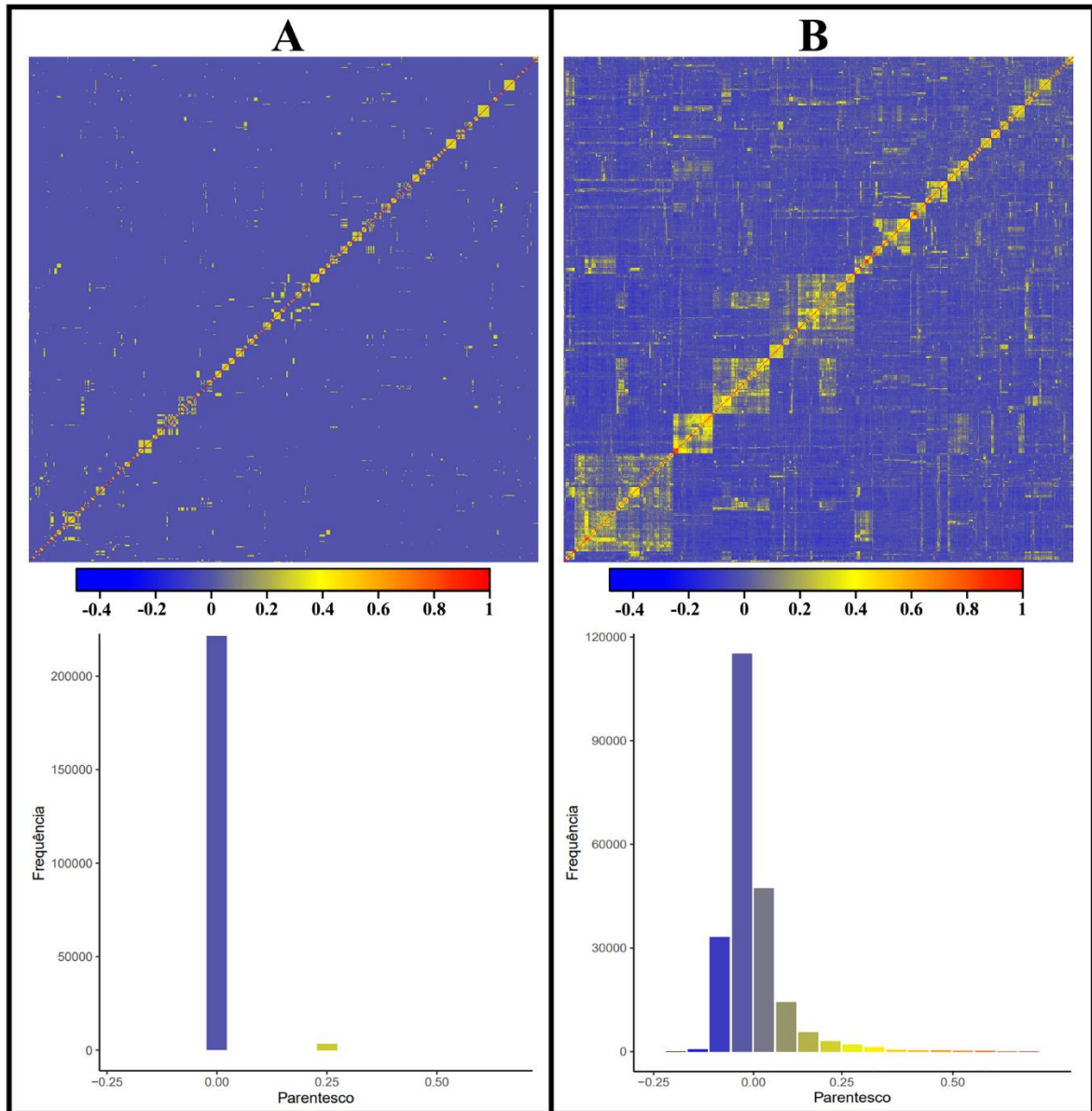


Figura 4 – Parentesco genômico e de *pedigree*. *Heatmap* e *barplot* das frequências de parentesco contidas na matriz de parentesco esperado (A) e parentesco genômico realizado (B) para um teste de progênies de polinização aberta de *E. benthamii*, Otacílio Costa - SC.

Apesar do parentesco médio captado com o uso dos marcadores ser condizente com o parentesco esperado via *pedigree* (Tabela 3), percebe-se que a matriz **G** foi capaz de corrigir parentescos entre pares de indivíduos que a matriz A_{gg} erroneamente considera como meios-irmãos ou não aparentados, revelando a existência de parentescos mais acentuados na população, como a presença de irmãos completos ou irmãos de autofecundação.

5.2 Desequilíbrio de Ligação e tamanho efetivo

A extensão do LD varia entre populações dentro de espécies, assim como entre o genoma de espécies (HYTEN et al., 2007). Em plantas alógamas, o LD tende a decair em distâncias curtas (BALLESTA et al., 2018). O mesmo usualmente acontece para espécies de *Eucalyptus*, apesar do seu sistema misto de reprodução. Na literatura, existem vários relatos do desequilíbrio de ligação decaindo rapidamente (após 500-1.500 bp) para: populações naturais e controladas de *E. globulus* (THAVAMANIKUMAR et al., 2011; CAPPA et al., 2013; DURÁN et al., 2017); *E. grandis* (GRATTAPAGLIA e KIRST, 2008; SILVA-JUNIOR e GRATTAPAGLIA, 2015); *E. urophylla* (DENIS et al., 2013); *E. nitens* (THUMMA et al., 2005); *E. benthamii* (MÜLLER et al., 2017). O mesmo ocorre no presente estudo, onde o LD cai abaixo de 0.2 em 9,95 Kb (Figura 5).

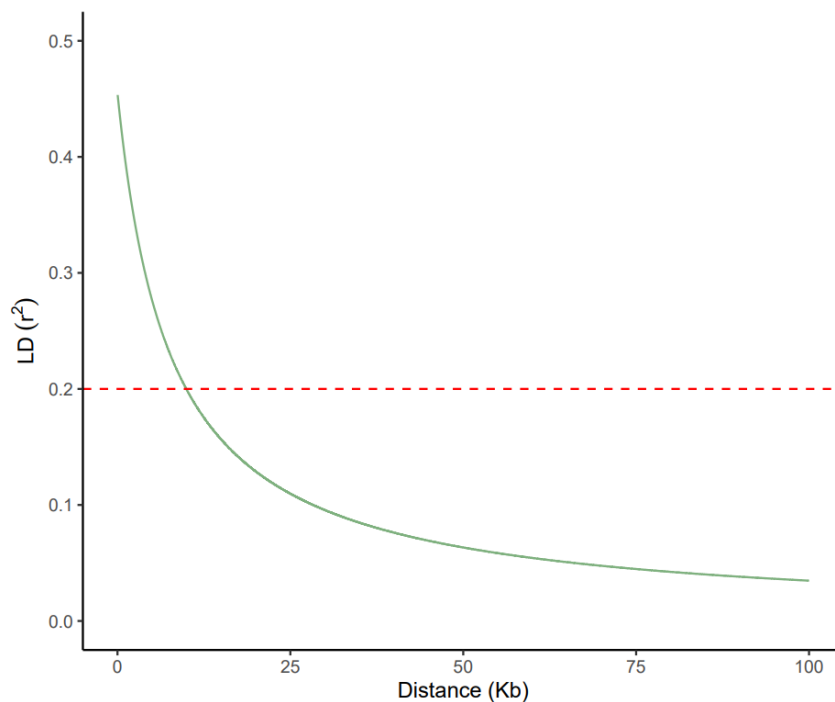


Figura 5 - Decaimento do desequilíbrio de ligação (r^2) para um teste de progênes de polinização aberta de *E. benthamii*, Otacílio Costa - SC.

Em seu estudo com *E. benthamii*, Müller et al. (2017) relatou um valor de decaimento de LD abaixo de 0,2 na distância de 7,7 Kb, bastante semelhante ao do presente estudo (9,95 Kb). De acordo com Müller et al. (2017), as populações de melhoramento de *E. benthamii* provêm de sementes coletadas das poucas populações naturais remanescentes, as quais sofreram um histórico de *bottlenecks*, e, portanto, apresentam diversidade genética limitada

(BUTCHER et al., 2005), por isso, apesar de serem populações de melhoramento, as populações de *E. benthamii* ainda assemelham-se muito a populações naturais, inclusive apresentando decaimento de LD em distâncias semelhantes ao encontrado para populações naturais de *E. grandis* (4-6 Kb) (SILVA-JUNIOR e GRATTAPAGLIA, 2015).

O tamanho efetivo populacional obtido com base no método do desequilíbrio de ligação (WAPLES e DO, 2008), foi de $LDN_e = 75,7$ com $IC_{95\%} = [71,6; 80,0]$. Frequências alélicas muito próximas de zero podem afetar a estimativa do tamanho efetivo (N_e) (WAPLES, 2006), por isso, adotou-se o um $MAF \geq 0,02$, o qual proporciona um bom balanço entre maximizar a precisão da estimativa e diminuir o viés (WAPLES e DO, 2010).

Sabe-se que N_e corresponde ao número de indivíduos em uma população idealizada que é capaz de gerar progênes com a mesma dispersão de frequência alélica sob deriva genética, ou a mesma quantidade de endogamia que a população real em consideração (FISHER, 1930; WRIGHT, 1931). O N_e pode ser interpretado como uma medida de representatividade genética dentro de uma amostra de indivíduos, e é um parâmetro importante tanto na genética de populações quanto na genética quantitativa, (VENCOVSKY, 1987; WANG, 1997; VENCOVSKY e CROSSA, 1999), o qual também permite a distinção entre o número amostral e a representatividade genética da amostra (VENCOVSKY e CROSSA, 2003).

Para diminuir a probabilidade de oscilações genéticas associadas à deriva genética e a endogamia, a manutenção do tamanho efetivo é de extrema importância (FRANKHAM, 1995; LYNCH, 1996; REED, 2005). Valores de N_e de aproximadamente 25 para populações melhoradas e 50 para populações pouco melhoradas, são recomendados pela literatura (PEREIRA e VENCOVSKY, 1988; RESENDE, 2015). No melhoramento de essências florestais, populações com N_e entre 20 e 50 são capazes de suportar seleção e proporcionar ganhos genéticos por várias gerações (NAMKOONG et al., 1988). De acordo com Resende (2002), um N_e de aproximadamente 40 a 50 são valores típicos em populações de melhoramento elite de plantas alógamas. Apesar da espécie em estudo, possuir na realidade um sistema reprodutivo misto ($t_m = 0,62$) (BUTCHER et al., 2005), vale lembrar que este tamanho efetivo (75,7) foi estimado com apenas a porção genotipada da população (Tabela 1), ou seja, provavelmente o N_e da população total é ainda maior.

Quanto às implicações do tamanho efetivo (N_e) na seleção genômica, basicamente quanto maior o N_e de uma dada população, mais marcas são necessárias para capturar todo o desequilíbrio de ligação (LD) entre marcas e QTLs ao longo do genoma, a fim de se atingir valores satisfatórios de capacidade preditiva (GRATTAPAGLIA e RESENDE, 2011; SPINDEL et al., 2015; LEE et al., 2017). De acordo com Grattapaglia e Resende (2011), para um N_e acima de 60, a fim de capturar toda a extensão do LD, é necessária uma densidade de SNPs de cerca de 20 marcas/cM a fim de atingir uma capacidade preditiva satisfatória, algo facilmente atingido utilizando um *chip* de alta densidade, como o EUChip60K (SILVA-JUNIOR et al., 2015). Tamanhos efetivos entre 10 e 100, no caso de populações de melhoramento de *Eucalyptus* (assim como no presente estudo), são satisfatórios para a prática da seleção genômica (GRATTAPAGLIA et al., 2009; GRATTAPAGLIA e RESENDE, 2011).

5.3 Estudo de associação genômica (GWAS)

A presença de estrutura de população nos dados deve ser corrigida por algum método estatístico, já que tanto a presença de estrutura de população, quanto o *hidden relatedness* entre os indivíduos, tendem a atrapalhar os resultados de um GWAS, reduzindo o poder da técnica e aumentando a detecção de falsos positivos (PRICE et al., 2010; HUANG e HAN, 2014). Enquanto a inclusão dos primeiros componentes principais como covariáveis na modelagem, quando estes explicam grande parte da variância, captura efetivamente a estrutura de população, modelar o parentesco par a par entre indivíduos captura tanto a estrutura de população, quanto o parentesco (KANG et al., 2010; PRICE et al., 2010; KENNY et al., 2011). O uso do método EMMAX (KANG et al., 2010) realiza este tipo de correção, a qual, por meio do modelo 4, utilizou a matriz de parentesco genômico realizado (\mathbf{G}) para fazê-lo.

O sucesso da correção e do ajuste do modelo pode ser constatado visualmente nos Q-Q plots (Figura 6), onde percebe-se que os p-valores observados foram consistentes com os p-valores esperados ao longo da diagonal do gráfico. A análise resultou na detecção de associação significativa de pelo menos um SNP para cada caráter em estudo, considerando, além de Bonferroni e FDR, o limite *ad hoc* ($p\text{-valor} \leq 8 \times 10^{-5}$) (Figura 6; Tabela 4).

Tabela 4 – Associações significativas entre os caracteres fenotípicos densidade básica da madeira (DBM), extrativos (EXT), lignina (LIG), carboidratos (CBO) aos quatro anos de idade e volume aos seis anos de idade (VOL6) para um teste de progênies de polinização aberta de *E. benthamii* e marcadores SNP, Otacílio Costa - SC.

Caráter	SNP ^a	MAF	β	h^2_{GWAS}	p-valor
DBM	EuBR10s10423880	0,05	-30,19	0,065	6,87E-06 ^b
	EuBR07s52222610	0,05	-29,46	0,063	9,48E-06 ^b
EXT	EuBR03s22568438	0,01	0,86	0,019	7,30E-05 ^c
LIG	EuBR10s17804495	0,03	3,99	0,018	5,99E-05 ^c
CBO	EuBR10s17804495	0,03	-3,97	0,018	6,63E-05 ^c
VOL6	EuBR03s28554582	0,20	0,09	0,009	1,26E-06 ^b

MAF: Frequência do alelo menos comum (*Minor allele frequency*); β : Valor do coeficiente de regressão associado ao marcador; ^aSNP EuBR10s10423880 lê-se: SNP no cromossomo 10, na posição 10.423.880 pares de base. ^bMarcas consideradas significativas pela correção de Bonferroni com $\alpha = 0,05$. ^cSNPs considerados significativos pelo p-valor *ad hoc* (p-valor $\leq 8 \times 10^{-5}$) de Müller et al. (2018).

Para espécies florestais, é comum uma baixa taxa de detecção de marcas significativamente associadas com os fenótipos em estudo. Isto vale tanto para caracteres de crescimento, quanto para os relacionados à qualidade da madeira em *Populus* (PORTH et al., 2013; MCKOWN et al., 2014; ALLWRIGHT et al., 2016; DU et al., 2016; FAHRENKROG et al., 2016; CHHETRI et al., 2019), *Pinus* (BARTHOLOMÉ et al. 2016; LU et al. 2017) e *Eucalyptus* (CAPPA et al. 2013; RESENDE et al., 2016; TORRES-DINI, 2017; MÜLLER et al., 2017; 2018). A baixa detecção de marcas associadas em espécies florestais, pode estar relacionada com o baixo número de indivíduos utilizados nestas análises, as quais frequentemente utilizam populações inferiores a algumas centenas de indivíduos (VISSCHER et al., 2017).

Este fato provavelmente influenciou os resultados do presente estudo, que avaliou 671 indivíduos. Vale ressaltar, que os caracteres de interesse em programas de melhoramento florestal, são, em sua maioria, de herança complexa, ou seja, poligênicos, o que significa que são controlados por um grande número de QTLs de pequenos efeitos, algo que certamente dificulta a detecção de marcas associadas, já que são necessárias técnicas extremamente

poderosas para detectar um número de marcas capazes de explicar uma porção considerável da variância genética (VISSCHER et al., 2017; MÜLLER et al., 2018).

As correções tradicionais de FDR (*False Discovery Rate*) e Bonferroni (especialmente Bonferroni) são bastante conservadoras para estudos de associação genômica contendo milhares de marcas, de maneira que locos importantes podem não passar por seus critérios altamente restritivos (PERNEGER, 1998; BALDING, 2006; WANG et al., 2016a; MÜLLER et al., 2018). Sabendo disso, e com base em um estudo de Resende et al. (2016) que estabeleceram valores de limites de significância para associação entre marcas e caracteres em *Eucalyptus* spp., Müller et al. (2018) definiram um valor de limite de significância *ad hoc* menos restritivo ($p\text{-valor} \leq 8 \times 10^{-5}$), o qual também foi utilizado como limite *ad hoc* no presente estudo. Apesar da correção de Bonferroni ser mais conservadora e restritiva que a correção por FDR, isto não influenciou no presente estudo, já que as associações significativas para FDR, também foram significativas para Bonferroni, i.e., DBM e VOL6 (Tabela 4).

O uso do limite *ad hoc*, por sua vez, permitiu a detecção de associações significativas com os caracteres EXT, LIG e CBO. Ainda assim, tal afirmação deve ser tratada com cautela, já que se baseia em um relaxamento não usual dos limites de significância frequentemente utilizados em estudos de associação.

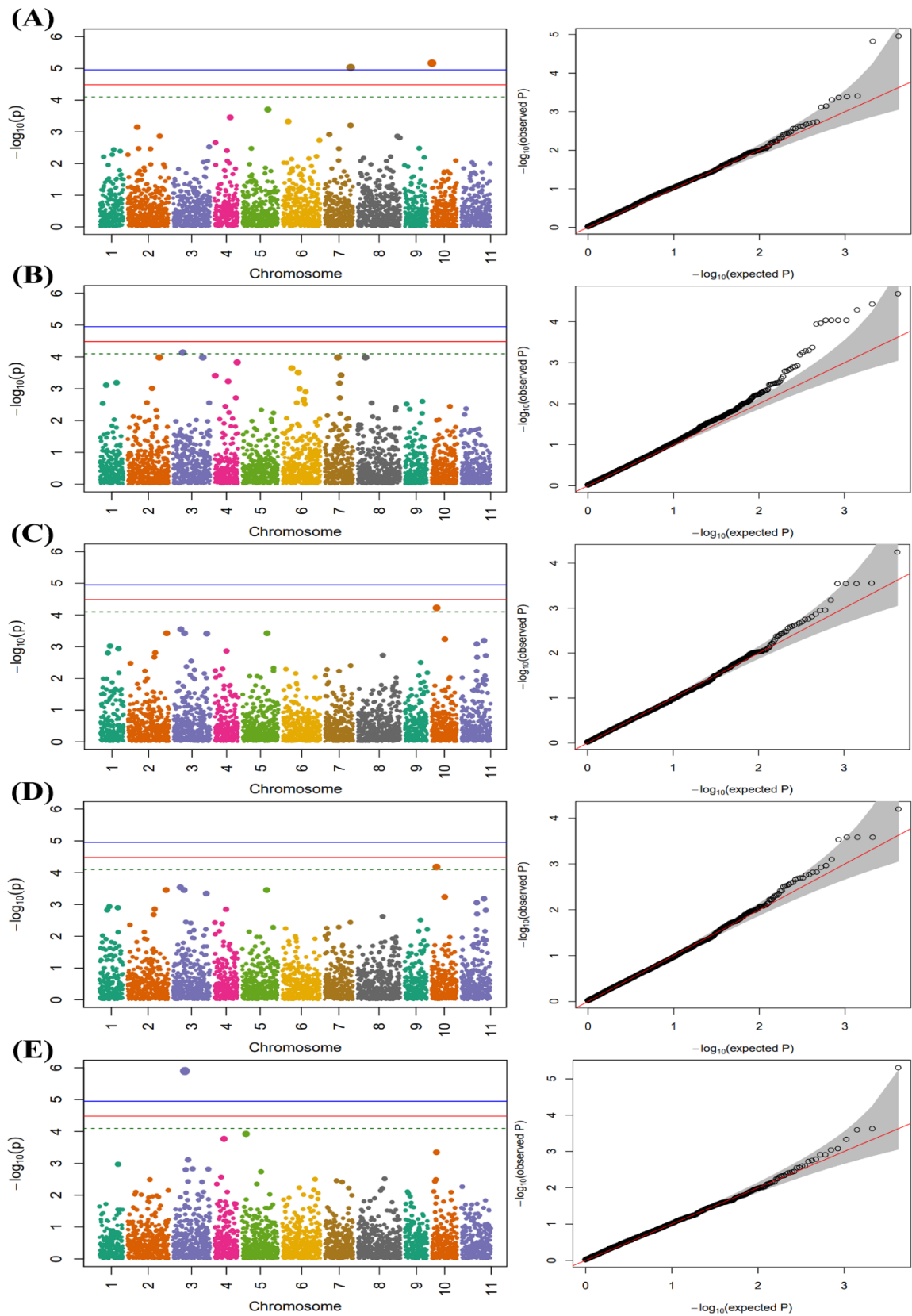


Figura 6 – Manhattan e Quantil-quantil (Q-Q) plots para (A) DBM, (B) EXT, (C) LIG, (D) CBO de um teste de polinização aberta de *E. benthamii* (Otacílio Costa – SC) aos quatro anos de idade e (E) VOL6 aos seis. A linha azul representa a correção de Bonferroni ($\alpha = 0,05$), a linha vermelha indica a correção FDR (5%) e a linha verde pontilhada indica o limite de significância ad hoc ($p\text{-valor} \leq 8 \times 10^{-5}$).

Os primeiros estudos visando a detecção de marcas (utilizando marcadores RAPD) significativamente associadas com caracteres fenotípicos em *Eucalyptus* datam da década de 90 (GRATTAPAGLIA et al., 1995; 1996). Porém o primeiro trabalho com GWAS em *Eucalyptus*, foi o de Cappa et al. (2013), os autores, utilizando marcadores DArT, detectaram poucas associações (16 para caracteres de crescimento e dois para relação S:G em lignina). Os trabalhos com marcadores SNP no gênero *Eucalyptus* apenas surgiram após o desenvolvimento do EUChip60K de Silva-Junior et al. (2015), o qual também foi utilizado no presente estudo.

Fazendo uso do mesmo painel de marcas, Müller et al. (2017) encontraram cinco associações significativas com volume de madeira para *E. pellita* e nenhuma para *E. benthamii*. Diferente do presente estudo, o qual foi capaz de detectar um SNP associado com o caráter volume (EuBR03s28554582) para *E. benthamii* (Tabela 4).

Para *Eucalyptus grandis* x *Eucalyptus urophylla*, Resende et al. (2016), considerando o mesmo limite de significância *ad hoc* utilizado no presente estudo, encontraram três SNPs associados significativamente com caracteres de crescimento (DAP e altura), quatro associados com DBM, além de outros seis associados com ferrugem causada por *Puccinia psidii* e caracteres relacionados à produção de polpa celulósica. Trabalhando com quatro populações de melhoramento não relacionadas de *Eucalyptus grandis* x *Eucalyptus urophylla*, Müller et al. (2018) dispôs de um maior tamanho amostral (3.373 indivíduos) e encontraram um maior número de marcas associadas com os caracteres de crescimento DAP (de 3 a 11 SNPs) e altura (de 3 a 78 SNPs).

Para os caracteres LIG e CBO, os quais consideram o limite *ad hoc* para declarar significância, o SNP significativamente associado foi o mesmo (EuBR10s17804495). Tal resultado não é inesperado, já que estes dois caracteres são inversamente proporcionais ($r_{xy} = -0,99$).

Em estudos de mapeamento de QTLs em *Eucalyptus* spp., foram mapeados QTLs para diâmetro a altura do peito, um caráter de crescimento sabidamente correlacionado com volume, e teor de extrativos (EXT) no cromossomo 3, o mesmo cromossomo em que se encontraram SNPs significativos nesta pesquisa (THUMMA et al., 2009; FREEMAN et al., 2013). Na literatura existem QTLs mapeados para densidade básica da madeira, relação S/G (Siringil/Guaiacil) da lignina, teor de celulose e teor de lignina no cromossomo 10 (THUMMA et al., 2009; GION et al., 2011; FREEMAN et al., 2013). No estudo de Cappa et al. (2013), com marcadores DArT, os autores encontraram um marcador associado com relação S/G no

cromossomo 10, a cerca de 1Mbp de distância do gene *ferrulate 5-hydroxylase* (F5H), relacionado com a produção de lignina. Lima (2014), utilizando marcadores SNPs, encontrou um *cluster* de efeitos para o mesmo caráter (relação S/G) também no cromossomo 10. No presente estudo, foram encontrados dois SNPs associados com caracteres de qualidade da madeira, um com DBM e um com LIG e CBO. As associações encontradas neste estudo, bem como as da literatura, sugerem a existência de um *hotspot* no cromossomo 10, relacionado caracteres de qualidade da madeira, especialmente à produção de lignina e celulose.

Dentre as associações encontradas no presente estudo para os caracteres de qualidade da madeira, todas são com SNPs de baixa frequência ($MAF \leq 0,05$). É necessário cautela para afirmar acerca da significância destas marcas, visto que é necessário um grande número de amostras de uma marca para se afirmar significância, além de serem marcas menos informativas (HONG e PARK, 2012; FOURNIER et al., 2019). Diferente do SNP em associação com VOL6, o qual possui maior frequência e pode ser considerado comum ($MAF = 0,20$).

Esta baixa magnitude na realidade é esperada, e apenas corrobora quanto a natureza complexa e poligênica destes caracteres, os quais são controlados possivelmente por centenas ou milhares de pequenos efeitos distribuídos por todo o genoma da espécie. Esta inabilidade da GWAS em capturar toda a variação genética associada a um dado caráter gera uma lacuna entre a verdadeira herdabilidade e àquela capturada pela técnica, a qual é referida na literatura como “herdabilidade perdida” (*missing heritability*). A *missing heritability* pode ser interpretada como: “a proporção de variância genética que não é explicada pelos QTLs que o estudo de associação foi capaz de detectar” (MAHER, 2008; MANOLIO et al., 2009). Até os dias atuais, o paradoxo da *missing heritability* ainda é amplamente discutido na literatura (VISSCHER et al., 2014; YANG et al., 2017; YOUNG, 2019).

Vale ressaltar, que nenhum dos SNPs em associação encontrados neste estudo foram previamente relatados por outros estudos que fizeram uso do EUChip60K.

5.4 Parâmetros e Capacidade Preditiva

As herdabilidades no sentido restrito (\hat{h}_a^2) estimadas por meio de modelos genômicos (H, GA e GAD) são de magnitude inferior às herdabilidades no sentido restrito estimadas com o uso da matriz de *pedigree* (modelos Ag e At) para todos os caracteres avaliados, com exceção de VOL6, onde o modelo Ag apresentou o menor valor deste parâmetro. (Tabela 5).

Devido aos diferentes tamanhos amostrais entre o tamanho amostral do *dataset* que contempla os indivíduos genotipados (Ag, GA e GAD) e o *dataset* de todos os indivíduos (At e H), as comparações das herdabilidades entre os métodos genômicos e fenotípicos, serão realizadas apenas para modelos ajustados com mesmo tamanho amostral.

As maiores diferenças entre as herdabilidades estimadas por meio de métodos genômicos e de *pedigree* para o *dataset* com apenas indivíduos genotipados, se dão para os caracteres EXT e LIG, ambos para o modelo aditivo-dominante (GAD). A estimativa obtida com o modelo GAD representa 53,4% da herdabilidade obtida com o modelo Ag (0,15/0,28) para EXT, e 55% para LIG. As menores diferenças, por sua vez, ocorrem para os caracteres DBM (modelo GA) e VOL6 (modelo GAD), onde as estimativas genômicas representam, respectivamente 90,9% e 116,7% da estimativa com modelo Ag.

Para o *dataset* contendo todos os indivíduos, as maiores diferenças entre a estimativa genômica *single step* (H) e a baseada em *pedigree* (At), se dão para LIG e CBO, representando 54,9% e 55,7% da herdabilidade fenotípica, respectivamente. Já as menores diferenças ocorreram para VOL6, onde as estimativas são praticamente iguais (94,6%) e DBM (76,1%).

Resultados similares para herdabilidades genômicas em contraste com a obtida via *pedigree* já foram encontrados por outros autores estudando espécies florestais como a *Picea* (BEAULIEU et al., 2014; GAMAL EL-DIEN et al., 2015), híbridos entre *E. grandis* x *E. urophylla* (RESENDE et al., 2017; LIMA et al., 2019) e até mesmo a espécie do presente estudo, o *E. benthamii* (MÜLLER et al., 2018). Entretanto, isto não aparenta ser uma regra. Há estudos com diferentes espécies de *Eucalyptus*, onde estas duas estimativas tem magnitude relativamente semelhantes, com alguns desvios para mais ou para menos, tanto para caracteres de crescimento, quanto qualidade da madeira (LIMA, 2014; SUONTAMA et al., 2019). Há também, para *Eucalyptus* spp., estudos onde a magnitude da herdabilidade genômica, supera a fenotípica (TAN et al., 2017; 2018), assim como VOL6, no presente estudo, para o *dataset* contemplando apenas indivíduos genotipados.

Tabela 5 - Parâmetros genéticos estimados por meio dos modelos avaliados para os caracteres densidade básica (DBM), extrativos (EXT), lignina (LIG), carboidratos (CBO) aos quatro anos de idade e volume aos seis anos de idade (VOL6) para um teste de progênies de polinização aberta de *E. benthamii*, Otacílio Costa – SC.

Caráter	Modelo	bias	$\hat{h}_a^2 \pm \text{S.E}$	$\hat{h}_G^2 \pm \text{S.E}$	$\sigma_{a(\text{GAD})}^2 / \sigma_{a(\text{GA})}^2$	σ_d^2 / σ_a^2
DBM	At	1,03	0,67 ± 0,08	-	-	-
	H	1,01	0,51 ± 0,05	-	-	-
	Ag	1,05	0,55 ± 0,06	-	-	-
	GA	0,99	0,50 ± 0,06	-	-	-
	GAD	0,99	0,43 ± 0,06	0,63 ± 0,07	0,86	0,46
EXT	At	0,99	0,70 ± 0,04	-	-	-
	H	1,02	0,42 ± 0,06	-	-	-
	Ag	1,26	0,28 ± 0,04	-	-	-
	GA	0,82	0,22 ± 0,04	-	-	-
	GAD	0,83	0,15 ± 0,04	0,35 ± 0,07	0,68	1,28
LIG	At	1,03	0,51 ± 0,09	-	-	-
	H	1,03	0,28 ± 0,04	-	-	-
	Ag	1,04	0,40 ± 0,05	-	-	-
	GA	1,02	0,28 ± 0,05	-	-	-
	GAD	0,98	0,22 ± 0,05	0,40 ± 0,06	0,81	0,80
CBO	At	1,04	0,61 ± 0,09	-	-	-
	H	1,04	0,34 ± 0,05	-	-	-
	Ag	1,02	0,37 ± 0,05	-	-	-
	GA	0,99	0,28 ± 0,05	-	-	-
	GAD	1,01	0,21 ± 0,04	0,40 ± 0,06	0,78	0,87
VOL6	At	1,02	0,37 ± 0,08	-	-	-
	H	1,15	0,35 ± 0,05	-	-	-
	Ag	1,18	0,18 ± 0,04	-	-	-
	GA	0,87	0,30 ± 0,06	-	-	-
	GAD	1,01	0,21 ± 0,05	0,51 ± 0,07	0,69	1,38

At: modelo de *pedigree* para o *dataset* de todos os indivíduos; Ag: modelo de *pedigree* para o *dataset* contendo apenas indivíduos genotipados; H: modelo híbrido *single-step*; GA: modelo aditivo; GAD: modelo aditivo-dominante; *bias*: viés de predição do modelo; \hat{h}_a^2 : herdabilidade individual no sentido restrito; \hat{h}_G^2 : herdabilidade no sentido amplo; $\sigma_{a(\text{GAD})}^2 / \sigma_{a(\text{GA})}^2$: relação entre a variância aditiva estimada com modelo aditivo-dominante e modelo aditivo; σ_d^2 / σ_a^2 : relação entre variância de dominância e variância aditiva.

Um dos motivos para a magnitude superior da herdabilidade estimada por meio de ambos os modelos A (At e Ag) em relação às herdabilidades genômicas neste estudo, pode ser o próprio uso da matriz **A**, pois sabe-se que a *pedigree* não é a abordagem ideal na estimativa de parâmetros, visto que trata-se de um parentesco esperado, o qual negligencia “parentescos escondidos” na população (Figura 4), de modo a superestimar a variância genética aditiva existente, o que tende a ser ainda mais acentuado em populações de polinização aberta, como a do presente estudo (SQUILLACE, 1974; NAMKOONG et al., 1998; GODDARD et al., 2011; KLÁPŠTĚ et al., 2018; RATCLIFFE et al., 2017; TAMBARUSSI et al., 2018). Negligenciar os “parentescos escondidos” tem ainda mais impacto em espécies com sistema reprodutivo misto, como é o caso do *E. benthamii* ($t_m = 0.62$) (BUTCHER et al., 2005; TAMBARUSSI et al. 2018).

Alguns dos autores que obtiveram resultados divergentes entre as herdabilidades via *pedigree* e genômica em espécies florestais (BEAULIEU et al., 2014; GAMAL EL-DIEN et al., 2015), argumentam que as herdabilidades genômicas representam melhor o verdadeiro parentesco entre os indivíduos e que o uso do parentesco genômico realizado (**G**) fornece estimativas mais assertivas do verdadeiro valor genético (GODDARD e HAYES, 2007; RESENDE et al., 2014).

Apesar das diferenças entre estimativas genômicas e fenotípicas, a magnitude das herdabilidades genômicas no sentido restrito (\hat{h}_a^2) obtidas para VOL6 (variando de 0,21 até 0,35 para os modelos GAD e H respectivamente) está de acordo com o único trabalho de seleção genômica em *E. benthamii*, o de Müller et al. (2017), onde este parâmetro variou de 0,10 até 0,28 nos vários modelos genômicos por eles testados para este caráter. Outros estudos com o gênero *Eucalyptus*, estimaram herdabilidades genômicas no sentido restrito para DBM de 0,30 (TAN et al., 2018), 0,42 até 0,52 (SUONTAMA et al., 2019), 0,57 (LIMA et al., 2019) e 0,67 (RESENDE et al., 2017), baseado nestes estudos, pode-se afirmar que o valor encontrado para este caráter nos modelos estudados (0,43 – GAD, até 0,51 – H) está de acordo com a literatura. Para LIG, Lima et al. (2019) obtiveram valores que variaram de 0,58 até 0,68. Já de Moraes et al. (2018), encontraram estimativas de \hat{h}_a^2 que variaram de 0,23 até 0,24, valores condizentes com os encontrados para este caráter no presente estudo (0,22 – GAD até 0,28 – H/GA). Para EXT e CBO, não foram encontrados estudos para comparação, entretanto, sabe-se que CBO e LIG são inversamente proporcionais ($r_{xy} = -1$), portanto, espera-se que o comportamento do caráter CBO, seja semelhante ao de LIG (Tabela 5).

As herdabilidades no sentido restrito obtidas com modelo GAD, ou seja, incluindo o efeito de dominância, são inferiores às aquelas estimadas com modelo GA. Quando o modelo GA é utilizado, o efeito de dominância não é modelado e, por isso, se torna parte do resíduo (DUENK et al., 2017). Isso faz com que o modelo GA proporcione estimativas menos acuradas da variância aditiva, pois ele assume que os resíduos são identicamente distribuídos, entretanto, sabe-se que isto não ocorre, já que os desvios de dominância variam de genótipo para genótipo (OTT e LONGNECKER, 2010). Na presença de dominância, os modelos GAD tendem a estimar melhor a variância aditiva (DUENK et al., 2017). Além disso, existem diversos relatos na literatura de uma não independência entre os efeitos aditivos e dominantes, os quais erroneamente são tratados como ortogonais com frequência (XIANG et al., 2018; TAN et al., 2018).

A não ortogonalidade entre os efeitos aditivos e não aditivos, foi relatada em espécies florestais por Tan et al. (2018), os autores, comparando modelos aditivos com modelos aditivo-dominantes para caracteres de crescimento e qualidade da madeira em híbridos de *Eucalyptus*, constataram que, ao incluir o efeito de dominância, este corrigia parte da variância aditiva, reduzindo sua magnitude e conseqüentemente diminuindo a estimativa genômica da herdabilidade no sentido restrito. A partir da relação $\sigma_{a(\text{GAD})}^2 / \sigma_{a(\text{GA})}^2$, pode-se observar que o mesmo ocorre no presente estudo, onde, quanto mais próximo da unidade, mais ortogonais são os efeitos para aquele dado caráter. Os valores da relação $\sigma_{a(\text{GAD})}^2 / \sigma_{a(\text{GA})}^2$ mostraram que a independência entre os efeitos variou de 0,68 (EXT) até 0,86 (DBM), revelando que o modelo GAD obteve uma estimativa de variância aditiva referente a 68% e 86% da variância aditiva estimada pelo modelo GA para EXT e DBM, respectivamente. Isto ocorre pois apenas em condições ideais do modelo epistático descrito Cockerham (1954), onde assume-se independência de frequências alélicas entre locos (equilíbrio de ligação), há verdadeira ortogonalidade entre os efeitos aditivos e não aditivos. Porém como estas condições não são facilmente atingidas em populações de melhoramento, estes efeitos podem se confundir (MUÑOZ et al., 2014). Este fenômeno também foi relatado em outros estudos (SU et al., 2012; MUÑOZ et al., 2014; BOUVET et al., 2016; VITEZICA et al., 2017).

Para *Eucalyptus*, o efeito de dominância costuma variar muito entre caracteres, espécies e híbridos do gênero (TAN et al., 2018). Entretanto, é comum uma alta influência deste efeito em caracteres de crescimento. O valor da relação σ_d^2 / σ_a^2 encontrado no presente estudo para VOL6 (1,38) está de acordo com outros estudos envolvendo caracteres de crescimento no

gênero (BISON et al., 2006; VOLKER et al., 2008; RESENDE et al., 2017; CHEN et al., 2018; TAN et al., 2018; LIMA et al., 2019; MORA et al., 2019). Há casos onde a variância de dominância é tão superior a aditiva, que a relação σ_d^2 / σ_a^2 atinge valores muito superiores a unidade, como em Tan et al. (2018) que obtiveram 3,2 para circunferência a altura do peito; Mora et al. (2019), de 4,16 para volume individual e Lima et al. (2019), de 8,25 para altura.

Para DBM, existem diferentes resultados na literatura. Resende et al. (2017) e Lima et al. (2019), não encontraram influência do efeito de dominância em seus estudos ($\sigma_d^2 / \sigma_a^2 \cong 0$), corroborando a afirmação de Potts e Dungey (2004) e Bison et al. (2006) acerca da herança para este caráter ser predominantemente aditiva entre os híbridos de *Eucalyptus*. Todavia, no presente estudo, é encontrada uma relação σ_d^2 / σ_a^2 de 0,46. Este valor é compatível com o estudo de Chen et al. (2018), o qual encontraram valores que variavam de 0,54 até a unidade para híbridos de *E. urophylla* x *E. tereticornis*. Tan et al. (2018), também se depararam com uma relação acima de zero para DBM (0,32).

Para lignina (LIG) e carboidratos (CBO), existem poucos estudos acerca do efeito de dominância em espécies florestais até então. Na realidade, para CBO não foi encontrado nenhum estudo, porém, para teor de celulose e hemicelulose, os quais compõem o caráter CBO, existem alguns. Dentre eles, os estudos de Chen et al. (2018) e Lima et al. (2019) podem ser mencionados, os quais encontraram baixos valores de relação σ_d^2 / σ_a^2 , variando de 0 até 0,06 para celulose. Já para hemicelulose, Lima et al. (2019) encontraram valores de magnitude semelhante (0 até 0,08), entretanto, Chen et al. (2018), constataram uma relação σ_d^2 / σ_a^2 variando de 0 até 3,5. O valor encontrado para CBO no presente estudo (0,87), revela certa influência de σ_d^2 no caráter, a qual, provavelmente pode ser atribuída à hemicelulose, porém são necessários mais estudos para se afirmar isto com clareza. Para LIG, no presente estudo, o valor de σ_d^2 / σ_a^2 encontrado (0,80) foi muito próximo ao de CBO, algo já esperado, visto que estes dois caracteres são inversamente proporcionais. Na literatura, foram encontrados valores variando de 0,19 até 1,15 por Chen et al. (2018) e de 0 até 0,19 por Lima et al. (2019). Segundo De Assis (2000), o teor de lignina e hemicelulose são caracteres herdados de maneira predominantemente aditiva. Os resultados do presente estudo, e dos estudos citados não corroboram totalmente com tal afirmação.

O caráter teor de extrativos (EXT) apresentou a segunda maior relação σ_d^2 / σ_a^2 entre todos os caracteres avaliados (1,28). Entretanto, apenas um estudo com *Picea abies* envolvendo

o efeito de dominância e este caráter foi relatado na literatura, o qual não obteve efeito de dominância associado à sua expressão (HANNRUP et al., 2004). O presente estudo observou um grande efeito de dominância para os caracteres avaliados.

Apesar da magnitude do efeito de dominância associado aos caracteres, no geral, estar de acordo com outras estimativas na literatura, não se pode descartar a hipótese de uma superestimativa da variância de dominância, uma vez que isto já foi relatado estar associado ao método RKHS e ao modelo GAD (AZEVEDO et al., 2015; SANTOS et al., 2017; DE ALMEIDA FILHO et al., 2019). De acordo com Morota et al. (2014), a superestimativa da variância de dominância pode ser atribuída à falta de ortogonalidade entre as matrizes **G** e **D**, ou seja, entre os efeitos aditivo e dominante. Na realidade, esta hipótese faz sentido, visto que aparentemente há uma relação entre a não ortogonalidade dos efeitos ($\sigma_{a(\text{GAD})}^2 / \sigma_{a(\text{GA})}^2$) e a magnitude da variância de dominância sobre a aditiva (σ_d^2 / σ_a^2), ou seja, quanto menor a magnitude da razão $\sigma_{a(\text{GAD})}^2 / \sigma_{a(\text{GA})}^2$, maior a razão σ_d^2 / σ_a^2 (Tabela 5). Ainda assim, a inclusão do efeito de dominância é de grande importância, principalmente em espécies que podem ser propagadas vegetativamente, que é o caso do *E. benthamii*.

O viés de predição (*bias*) (Tabela 5), é capaz de mostrar se os modelos avaliados subestimam (valores acima de 1) ou superestimam (valores abaixo de 1) os valores preditos. Para o caráter densidade básica da madeira (DBM), lignina (LIG) e carboidratos (CBO), nenhum dos modelos testados apresentou grandes desvios da unidade, significando que para estes caracteres, todos os modelos mostraram-se livres de viés de predição. Para EXT, percebe-se uma superestimativa dos modelos genômicos que consideram apenas os indivíduos genotipados (0,82 – GA; 0,83 – GAD), bem como uma subestimativa do modelo de *pedigree* (1,26 – Ag). Entretanto, para *dataset* que contém o teste todo, não há viés nos modelos de *pedigree* (At) e híbrido *single-step* (H), o que significa que a inclusão de informações fenotípicas, assim como o aumento do tamanho amostral, mostrou-se benéfica na predição de valores para este caráter. Para VOL6, um caráter de crescimento sabidamente influenciado pelo efeito de dominância, a inclusão da matriz **D** corrigiu a superestimativa dos valores preditos pelo modelo GA (0,87), bem como a subestimativa do modelo Ag (1,18), resultando em uma estimativa não viesada (1,01 – GAD). Ainda para VOL6, em relação ao modelo H, percebe-se que este apresentou uma leve subestimativa dos valores preditos, mostrando haver um certo viés de predição associado (1,15).

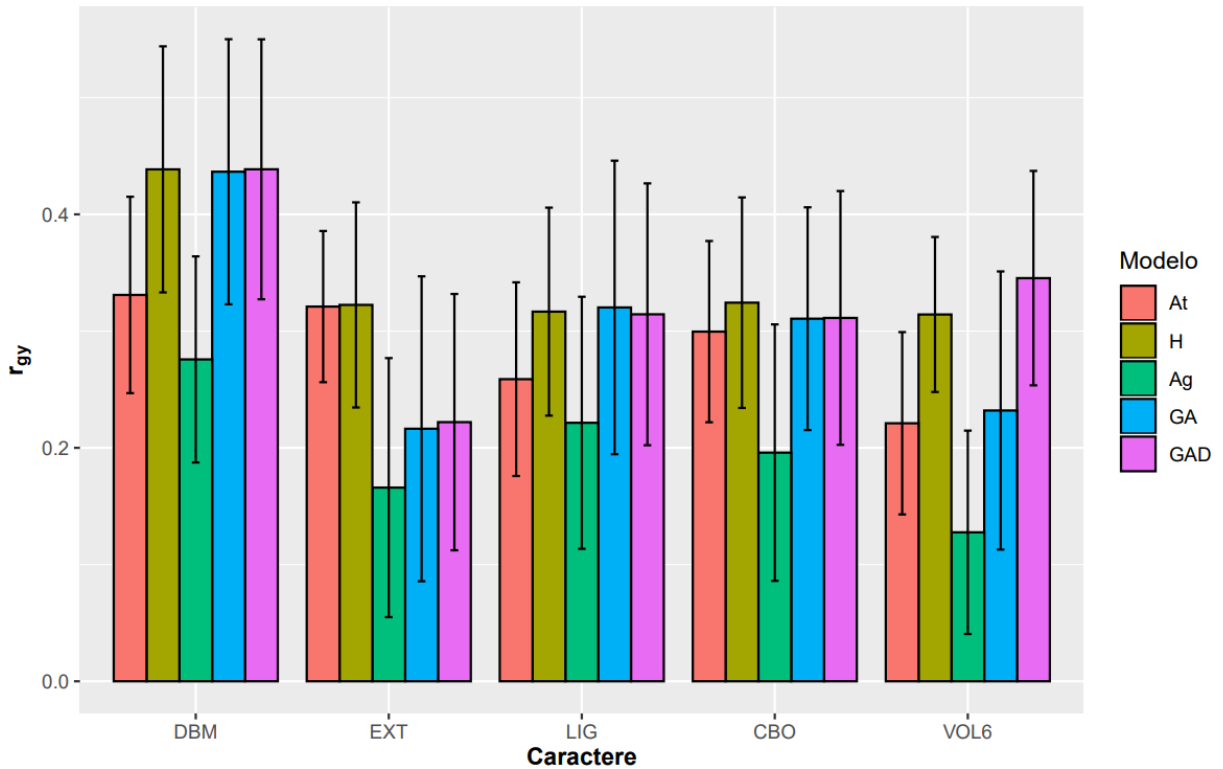


Figura 7 - Capacidade preditiva (r_{gy}) dos modelos avaliados para os caracteres densidade básica (DBM), extrativos (EXT), lignina (LIG), carboidratos (CBO) aos quatro anos de idade e volume aos seis anos de idade (VOL6) em um teste de progênies de polinização aberta de *E. benthamii*, Otacílio Costa – SC.

As capacidades preditivas (Figura 7) variaram de 0,12 (VOL6 – Modelo Ag) até 0,44 (DBM – Modelos H, GA e GAD). Na realidade, para DBM, no gênero *Eucalyptus* são comuns maiores valores de capacidade preditiva em relação aos demais caracteres usualmente avaliados: 0,54 até 0,64 (RESENDE et al., 2017), 0,50 (TAN et al., 2018), 0,46 a 0,52 (SUONTAMA et al., 2019).

Para *Eucalyptus*, existem casos onde a capacidade preditiva do modelo baseado em *pedigree* (A) se iguala ou até mesmo supera a capacidade preditiva de modelos genômicos (MÜLLER et al., 2017; CAPPÀ et al., 2019), bem como casos onde modelos genômicos desempenham melhor que o modelo A (TAN et al., 2017; MÜLLER et al., 2017; KAINER et al., 2018).

Para todos os caracteres avaliados, os modelos genômicos (GA e GAD) mostraram superioridade em relação a capacidade preditiva do modelo Ag. O mesmo ocorre entre o modelo híbrido *single-step* (H) e o modelo At, com exceção de EXT, onde ambos os modelos apresentaram igual desempenho. Tais resultados são de extrema importância, visto que o

principal objetivo da seleção genômica é proporcionar uma seleção mais assertiva, confiável e em menor tempo do que os métodos usuais baseados apenas em *pedigree*.

O método HBLUP, ou *single-step* GBLUP, nomeado no presente estudo como modelo H, une o benefício de uma estimativa com maior tamanho amostral, aliado à robustez proporcionada pela genotipagem. Apesar disso, apresentou performance equivalente aos outros métodos genômicos para os caracteres DBM, LIG e CBO. Entretanto performances superiores para EXT e VOL6 foram alcançadas. Os resultados do presente estudo estão de acordo com a literatura, onde este método tem-se mostrado como superior ou pelo menos comparável aos modelos baseados em *pedigree* (A) e genômicos aditivos (GA) em relação a capacidade preditiva (BEAULIEU et al., 2014; LEGARRA et al., 2014; ASHRAF et al., 2016; CAPPA et al., 2017; 2018; PÉREZ-RODRÍGUEZ et al., 2017; RATCLIFFE et al., 2017; MORAIS JÚNIOR et al., 2018; ZHAO, 2019; IMAI et al., 2019).

Para o caráter extrativos (EXT), o tamanho amostral aparentou ter grande influência na estimativa da capacidade preditiva, visto que há grande diferença entre as estimativas para os dois *datasets* avaliados, e empate entre os modelos contendo todos os indivíduos. Isto provavelmente ocorre, devido a existência de maior parentesco entre a população de treinamento e a população de validação, já que os modelos At e H utilizaram a totalidade de indivíduos no teste, não apenas a porção genotipada, resultando em uma população de treinamento de maior tamanho, o que conseqüentemente gera uma predição mais confiável, com menor viés de predição (Tabela 5) e maior capacidade preditiva (Figura 7) (DAETWYLER et al., 2010; GRATTAPAGLIA e RESENDE, 2011; TAN et al., 2017).

Dentre as abordagens genômicas, a superioridade do modelo H em relação à modelos genômicos tradicionais (GA) se torna mais aparente para caracteres de crescimento (CAPPA et al., 2019). Isto é observado no presente estudo, visto que a capacidade preditiva com modelo H foi 35,5% superior àquela obtida com modelo GA para VOL6. Para este caráter, o modelo H apenas se mostrou inferior ao modelo GAD, tanto em capacidade preditiva (Figura 7), quanto em *bias* (Tabela 5), o que provavelmente ocorreu devido a, como dito anteriormente, VOL6 ser um caráter de crescimento, o qual é sabidamente influenciado por uma porção considerável do efeito de dominância.

Ainda que o modelo H não tenha sido o melhor para VOL6, sua capacidade preditiva (0,31) foi próxima ao modelo GAD (0,34), assim como superior em relação aos demais modelos testados (0,12 – Ag; 0,22 – At; 0,23 – GA). Para os demais caracteres, apesar do modelo H não se destacar como o melhor, mostrou-se tão bom quanto os melhores modelos avaliados (Figura

7). Algo que ainda é positivo, visto que o modelo H permite uma seleção da totalidade de indivíduos no teste, não apenas da porção genotipada, com uma capacidade preditiva equivalente ou até superior a alguns modelos genômicos.

Quando o efeito de dominância constitui uma porção considerável da variância fenotípica, a inclusão deste efeito na modelagem, tende a aumentar a acurácia na predição dos valores (NISHIO e SATOH, 2014; DE ALMEIDA FILHO et al., 2016; DUENK et al., 2017), porém mesmo que uma pequena variância associada ao efeito de dominância tenha sido observada para DBM (0,46), LIG (0,80), CBO (0,87) (Tabela 5), o modelo GAD demonstrou capacidade preditiva semelhante aos outros modelos genômicos testados (H e GA) (Figura 7). O mesmo ocorreu nos estudos de Resende et al. (2017) para DBM e Tan et al. (2018) para DBM e produção de polpa, onde, apesar de encontrado um efeito de dominância associado este não influenciou na capacidade preditiva.

Para *Eucalyptus* spp., Lima (2014) e Cappa et al. (2019), trabalhando com o mesmo conjunto de dados, obtiveram capacidades preditivas superiores à do presente estudo para LIG, as quais variaram de 0,59 até 0,63. Entretanto, para *Pinus taeda*, Resende et al. (2012), obtiveram capacidades preditivas para LIG que variaram de 0,11 a 0,17 dependendo da metodologia utilizada. Para CBO, como relatado anteriormente, não existem estudos de seleção genômica. Porém, para celulose e hemicelulose (os caracteres que compõem CBO), foram estimadas capacidades preditivas que variaram de 0,39 até 0,55 em *Eucalyptus* spp. (LIMA, 2014; CAPPÀ et al., 2019).

No caso de VOL6, percebe-se uma influência positiva na capacidade preditiva ao se modelar a matriz **D**. Segundo a literatura, ao se considerar o efeito de dominância na modelagem de caracteres que possuem este efeito, a capacidade preditiva tende a aumentar (NISHIO e SATOH, 2014; DE ALMEIDA FILHO et al., 2016; DUENK et al., 2017). Apesar de todos os caracteres do presente estudo apresentarem efeito de dominância, EXT e VOL6 apresentaram maior influência deste efeito, superando até mesmo o efeito aditivo ($\sigma_d^2 / \sigma_a^2 > 1$) (Tabela 5). O que resultou, para VOL6, em uma maior capacidade preditiva para o modelo GAD (0,34) em relação aos demais modelos testados. Entretanto, para EXT, apesar da magnitude da variância de dominância superar a aditiva, o modelo GAD não apresentou performance superior. Vale ressaltar, que para EXT não foram encontrados estudos na literatura para comparação das capacidades preditivas.

Em estudos com híbridos interespecíficos entre *E. grandis* e *E. urophylla*, alguns autores

também se depararam com um aumento da capacidade preditiva em caracteres de crescimento, ao considerar na modelagem o efeito de dominância: Tan et al. (2018) observaram um acréscimo de cerca de 10% para circunferência a altura do peito e altura quando comparados modelos GA e GAD. Para IMA (volume de madeira em m³/ano/ha), Resende et al. (2017), observaram um aumento de 5 a 14%. No presente estudo, o acréscimo foi de 32,4% para VOL6. Além da notável influência da matriz **D**, este grande aumento da capacidade preditiva, também pode estar relacionado ao método de estimativa utilizado, já que há relatos na literatura de que métodos não paramétricos, ou semi-paramétricos, como o RKHS (utilizado no presente estudo), tendem fornecer maiores capacidades preditivas em caracteres de crescimento, especialmente na presença de efeitos não aditivos (TAN et al., 2017; MOMEN et al., 2018; DE ALMEIDA FILHO et al., 2019).

As capacidades preditivas obtidas para VOL6 no presente estudo, superam a única estimativa existente para volume em *E. benthamii* (0,14) na literatura (MÜLLER et al., 2017), e estão dentro do *range* das estimativas para este caráter em outras espécies do gênero, como o *E. pellita* (0,42) (MÜLLER et al., 2017), *E. globulus* (0,04 até 0,42) (BALLESTA et al., 2018), híbridos de *E. urophylla* x *E. grandis* (0,27 até 0,30) (TAN et al., 2017), e *Eucalyptus* spp. (0,42) (LIMA, 2014).

5.5 O efeito de dominância o modelo *single-step* na seleção genômica

Diversos autores afirmam que, na presença de efeito significativo de dominância sobre um caráter, a inclusão deste efeito na predição, melhora as estimativas e a capacidade preditiva de modelos genômicos (DENIS e BOUVET, 2011; WELLMAN e BENNEWITZ, 2012; MUÑOZ et al., 2014; NISHIO e SATOH, 2014; DUENK et al., 2017; DE ALMEIDA FILHO et al., 2019). Algo visível para o caráter volume aos seis anos de idade (VOL6), devido ao alto efeito de dominância associado, superando até mesmo o efeito aditivo ($\sigma_d^2 / \sigma_a^2 > 1$). Apesar do caráter teor de extrativos (EXT) também apresentar alto efeito de dominância associado, este não aparentou influenciar na capacidade preditiva (Figura 7). Entretanto, vale ressaltar, que este é o primeiro estudo com este caráter, portanto, são necessários mais estudos a fim de compreender de maneira satisfatória o seu comportamento.

Em resumo, em caracteres de crescimento, o efeito de dominância em *Eucalyptus* spp. é bastante pronunciado e relatado por diversos autores na literatura (BISON et al., 2006; VOLKER et al., 2008; RESENDE et al., 2017; CHEN et al., 2018; TAN et al., 2018; LIMA et

al., 2019; MORA et al., 2019), entretanto, para caracteres de qualidade da madeira, existem resultados mistos (RESENDE et al., 2017; CHEN et al., 2018; TAN et al., 2019; LIMA et al., 2019).

O método *single-step* (HBLUP), tem demonstrado em diversos estudos, uma excelente alternativa aos métodos usuais, tanto genômicos quanto fenotípicos de *pedigree* (FORNI et al., 2011; CHRISTENSEN, 2012; BEALIEU et al., 2014; CAPPA et al., 2017; 2019; RATCLIFFE et al., 2017; RATCLIFFE, 2019; VELAZCO et al., 2019; ZHAO et al., 2019). O uso da matriz híbrida (**H**) pode ser visto como uma reconstrução da *pedigree* total do teste a partir da genotipagem de uma porção de famílias relacionadas, permitindo a implementação da seleção genômica em maior escala, corrigindo possíveis erros de ordenamento, e, assim como a seleção genômica *per se*, corrigindo possíveis erros de parentesco da matriz de *pedigree*, com uma capacidade preditiva equivalente ou até superior às demais metodologias. Este método é especialmente útil no melhoramento florestal, visto que a genotipagem da totalidade de indivíduos que compõem um teste de progênie é bastante onerosa (BEAULIEU et al., 2014; RATCLIFFE et al., 2017).

O presente estudo busca ressaltar a necessidade de maiores estudos acerca da influência e estimativas do efeito de dominância na seleção genômica para caracteres de interesse em espécies florestais, visto que sua contabilização pode ser benéfica e contribuir significativamente para o sucesso da técnica. Assim como demonstrar que o uso da matriz híbrida *single-step*, mesmo que para alguns caracteres apresente desempenho equivalente às outras metodologias, é de grande valia, visto que permite, com custo significativamente menor, estimativas com a totalidade da população em avaliação.

6. CONCLUSÃO

- ✓ Foram encontradas poucas associações no estudo de associação genômica (GWAS), as quais explicam de 0,9% até 6,5% da variância genética dos caracteres em estudo.
- ✓ O efeito de dominância não teve grande influência na capacidade preditiva e viés de predição para caracteres de qualidade da madeira. Entretanto, para volume (VOL6), um caráter de crescimento, sua contabilização teve impacto positivo.
- ✓ O modelo híbrido *single-step* permite com sucesso a aplicação da seleção genômica em indivíduos genotipados e não genotipados com igual ou até mesmo superior capacidade preditiva em relação aos métodos usuais.
- ✓ Os modelos genômicos, no geral, apresentaram desempenho superior em relação ao modelo padrão (*pedigree*).
- ✓ A seleção genômica pode ser empregada com sucesso para um massivo ganho de tempo em relação à metodologia tradicional de análise e seleção.

7. REFERÊNCIAS BIBLIOGRÁFICAS

AGUILAR, I., MISZTAL, I.; JOHNSON, D.L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T.J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **J. Dairy Sci.** vol. 93. p. 743–752. 2010.

AKHIEZER, N.I.; GLAZMAN, I. M. **Theory of Linear Operators in Hilbert Space** (Teoriia lineiynykh operatorov v Gil'bertovom prostranstve). vol. 1. M. Nestell, trans. from Russian. Frederick Ungar, New York, NY. 400 p. 1963.

ALLWRIGHT, M.R.; PAYNE, A.; EMILIANI, G.; MILNER, S.; VIGER, M.; ROUSE, F.; KEURENTJES, J.J.B.; BÉRARD, A.; WILDHAGEN, H.; FAIVRE-RAMPANT, P. et al. Biomass traits and candidate genes for bioenergy revealed through association genetics in coppiced European *Populus nigra* (L.). **Biotechnol Biofuels.** vol. 9. n. 195. 2016.

ANDRADE, E. N. **O Eucalipto**. São Paulo: Cia. Paulista de Estradas de Ferro. 2ª Edição, 1961, 680p.

ASHRAF, B.; EDRISS, V.; AKDEMIR, D.; AUTRIQUE, E.; BONNET, D.; CROSSA, J. et al. Genomic prediction using phenotypes from pedigreed lines with no marker data. **Crop Sci.** vol. 56 p. 957–964. 2016.

ASTLE, W. AND BALDING, D. Population Structure and Cryptic Relatedness in Genetic Association Studies. **Statistical Science.** vol. 24. p. 451-471. 2009.

AZEVEDO, C.F.; RESENDE, M.D.V.; SILVA, F.F. Ridge, Lasso and Bayesian additive-dominance genomic models. **BMC Genet.** vol. 16. n. 105. 2015.

BALDING, D.J. A tutorial on statistical methods for population association studies. **Nat Rev Genet.** vol. 7. n. 10. p. 781-791. 2006.

BALLESTA, P.; SERRA, N.; GUERRA, F.P.; HASBÚN, R.; MORA, F. Genomic Prediction of Growth and Stem Quality Traits in *Eucalyptus globulus* Labill. at Its Southernmost Distribution Limit in Chile. **Forests.** vol. 9. n. 779. 2018.

BARTHOLOMÉ, J.; BINK, M.C.; VAN HEERWAARDEN, J.; CHANCEREL, E.; BOURY, C.; LESUR, I.; ISIK, F.; BOUFFIER, L.; PLOMION, C. Linkage and association mapping for two major traits used in the maritime pine breeding program: Height growth and stem straightness. **PLoS One.** vol. 11. p. 1–21. 2016.

BEAULIEU, J.; DOERKSEN, T.; CLÉMENT, S.; MACKAY, J.; BOUSQUET, J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. **Heredity.** vol. 113. p. 343–352. 2014.

BEAVIS, W.D. The power and deceit of QTL experiments: lessons from comparative QTL studies, in **Proceedings of the Forty-Ninth Annual Corn & Sorghum Industry Research Conference**, pp. 250–266. American Seed Trade Association, Washington, DC. 1994.

- BEAVIS, W.D. QTL analyses: power, precision, and accuracy, have missing genotypes at the marker, the variance. **Molecular Dissection of Complex Traits**, CRC Press, New York. pp. 146-162. 1998.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Royal Statistical Society**. vol. 57. p. 289–300. 1995.
- BERNARDO, R.; CHARCOSSET, A. Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. **Crop Science**. vol. 46, n. 2, p. 614–621, 2006.
- BERNARDO, R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. **Crop Science**, Madison, vol. 48, n. 5, p. 1649-1664, 2008.
- BISON, O.; RAMALHO, M.; REZENDE, G.; AGUIAR, A.; RESENDE, M.D.V. Comparison between open pollinated progenies and hybrids performance in *Eucalyptus grandis* and *Eucalyptus urophylla*. **Silvae Genet.** vol. 55. p. 192-196. 2006.
- BORÉM, A.; CAIXETA, E.T. **Marcadores moleculares**. 2 ed. Viçosa, MG: Editora UFV. 2009.
- BORÉM, A.; MIRANDA, G.V. **Melhoramento de Plantas**. 5 ed. Viçosa, MG. Editora UFV, 2009. 529 p.
- BOUVET, J.M.; MAKOUANZI, D.C.; VIGNERON, P. Modeling additive and non-additive effects in a hybrid population using genome-wide genotyping: prediction accuracy implications. **Heredity**. vol. 116. p. 146-157. 2016.
- BRONDANI, G.E.; DUTRA, L.F.; WENDLING, I.; GROSSI, F.; HANSEL, F.A.; ARAUJO, M.A. Micropropagation of an *Eucalyptus* hybrid (*Eucalyptus benthamii* x *Eucalyptus dunnii*). **Acta Sci Agron**. vol. 33. p. 655–663. 2011.
- BUTCHER, P.A.; SKINNER, A.K.; GARDINER, C.A. Increases inbreeding and inter-species gene flow in remnant populations of the rare *Eucalyptus benthamii*. **Conservation Genetics**. vol. 6. p. 213-226. 2005.
- CAETANO, A.R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia** v.38: 64-71. 2009.
- CAPPA, E.P.; EL-KASSABY, Y.A.; GARCIA, M.N.; ACUNA, C.; BORRALHO, N.M.G.; GRATTAPAGLIA, D.; MARCUCCI POLTRI, S.N. Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. **PLoS ONE**. vol. 8. p. 1–16. 2013.
- CAPPA, E.P.; EL-KASSABY, Y.A.; MUÑOZ, F.; GARCIA, M.N.; VILLALBA, P.V. et al. Improving accuracy of breeding values by incorporating genomic information in spatial-competition mixed models. **Mol. Breeding**. vol. 37. n. 125. 2017
- CAPPA, E.P.; LIMA, B.M.; SILVA-JUNIOR, O.B.; GARCIA, C.C.; MANSFIELD, S.D.; GRATTAPAGLIA, D. Improving genomic prediction of growth and wood traits in *Eucalyptus* using phenotypes from non-genotyped trees by single-step GBLUP. **Plant Science**. vol. 284. p. 9-15. 2019.

CASELLA, G.; GEORGE, E.I. Explaining the Gibbs Sampler. **The American Statistician**. vol. 46. n. 3. p. 167–174. 1992.

CERICOLA, F.; JAHOOOR, A.; ORABI, J.; ANDERSEN, J.R.; JANSSE, L.L.; JENSEN, J. Optimizing training population size and genotyping strategy for genomic prediction using association study results and pedigree information. A case of study in advanced wheat breeding lines. **PLoS One**. vol. 12. n. 1. 2017.

CHANCEREL, E.; LAMY, J.B.; LESUR, I.; NOIROT, C.; KLOPP, C.; EHRENMANN, F.; BOURY, C.; LE PROVOST, G.; LABEL, P.; LALANNE, C. et al. High-density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. **BMC Biology**. vol. 11. 2013.

CHEN, S.; WENG, Q.; LI, F.; LI, M.; ZHOU, C.; GAN, S. Genetic parameters for growth and wood chemical properties in *Eucalyptus urophylla* x *E. tereticornis* hybrids. **Annals of Forest Science**. vol. 75. n. 16. 2018.

CHHETRI, H.B.; MACAYA-SANZ, D.; KAINER, D.; BISWAL, A.K.; EVANS, L.M.; CHEN, J.; COLLINS, C.; HUNT, K.; MOHANTY, S.S.; ROSENSTIEL, T.; RYNO, D. et al. Multi-trait genome-wide association analysis of *Populus trichocarpa* identifies key polymorphisms controlling morphological and physiological traits. **New Phytologist**. vol. 223. p. 293-309. 2019.

CHRISTENSEN, O.F.; LUND, M.S. Genomic prediction when some animals are not genotyped. **Genet. Sel. Evol.** vol. 42. n. 2. 2010.

CHRISTENSEN, O.F. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. **Genetics Selection Evolution**. vol. 44. n. 37. 2012.

COCKERHAM, C.C. An extension of the concept of partitioning hereditary variance for analysis of covariance among relatives when epistasis is present. **Genetics**. vol. 39. p. 859-882. 1954.

COLLARD, B.C.Y.; JAHUFER, M.Z.Z.; BROUWER, J.B.; PANG, E.C.K. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. **Euphytica**. vol. 142. p. 169-196. 2005.

CROSSA, J.; CAMPOS, G. DE LOS.; PÉREZ, P.; GIANOLA, D.; BURGUEÑO, J.; ARAUS, J. L.; MAKUMBI, D.; SINGH, R. P.; DREISIGACKER, S.; YAN, J.; ARIEF, V.; BANZIGER, M.; BRAUN, H. J.; YAN, I. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**. vol. 186, n. 2, p. 713–724, 2010.

CUEVAS, J.; CROSSA, J.; SOBERANIS, V.; PÉREZ-ELIZALDE, S.; PÉREZ-RODRIGUEZ, P.; DE LOS CAMPOS, G.; MONTESINOS-LÓPEZ, O.A.; BURGUEÑO, J. Genomic Prediction of Genotype x Environment Interaction Kernel Regression Models. **The Plant Genome**. vol. 9. n. 3. p. 1-20. 2016.

DAETWYLER, H.D. **Genome-Wide Evaluation of Populations**. Ph. D. Thesis: Animal Breeding and Genomics Centre, Wageningen University, Wageningen, NL. 192p. 2009.

DAETWYLER, H.D.; PONG-WONG, R.; VILLANUEVA, B.; WOOLLIAMS, J. A. The impact of genetic architecture on genome-wide evaluation methods. **Genetics**. vol. 185. n. 3. p. 1021-1031, 2010.

DE ALMEIDA FILHO, J.E.; GUIMARÃES, J.F.R.; FONSECA E SILVA, F.; RESENDE, M.D.V.; MUÑOZ, P.; KIRST, M.; RESENDE JÚNIOR, M.F.R. The contribution of dominance to phenotype prediction in a pine breeding and simulated population. **Heredity**. vol. 117. p. 33-41. 2016.

DE ALMEIDA FILHO, J.E.; GUIMARÃES, J.F.R.; FONSECA E SILVA, F.; RESENDE, M.D.V.; MUÑOZ, P.; KIRST, M.; RESENDE JÚNIOR, M.F.R. Genomic Prediction of Additive and Non-additive Effects Using Genetic Markers and Pedigrees. **G3**. vol. 9. p. 2739-2747. 2019.

DE ASSIS, T.F. Production and use of Eucalyptus hybrids for industrial purposes. In: DUNGEY, H.S.; DIETERS M.J.; NIKLES, D.G. **Hybrid Breeding and Genetics of Forest Trees. Proceedings of QFRI/CRC-SPF Symposium**, 9–14th April 2000 Noosa, Queensland, Australia. Department of Primary Industries, Brisbane, pp. 63–74. 2000.

DE LOS CAMPOS, G.; GIANOLA, D.; ROSA, G.J. Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. **J. Anim. Sci.** vol. 87. p. 1883–1887. 2009.

DE LOS CAMPOS, G.; GIANOLA, D.; ROSA, G.J.; WEIGEL, K.A.; CROSSA, J. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel hilbert spaces methods. **Genet. Res.** vol. 92. p. 295–308. 2010.

DENIS, M.; BOUVET, J-M. Genomic selection in tree breeding: testing accuracy of prediction models including dominance effect. **BMC Proceedings**. vol. 5 n. 3. 2011.

DENIS, M.; FAVREAU, B.; UENO, S.; CAMUS-KULANDAIVELU, L.; CHAIX, G.; GION, J.M.; NOURRISIER-MOUNTOU, S. et al. Genetic variation of wood chemical traits and association with underlying genes in *Eucalyptus urophylla*. **Tree Genetics & Genomes**. vol 9. p. 927– 942. 2013.

DO, C.; WAPLES, R.S.; PEEL, D.; MACBETH, G.M.; TILLET, B.J.; OVENDEN, J.R. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. **Molecular Ecology Resources**. vol. 14, n. 1. p. 209-214. 2014.

DU, Q.; GONG, C.; WANG, Q.; ZHOU, D.; YANG, H.; PAN, W.; LI, B.; ZHANG, D. Genetic architecture of growth traits in *Populus* revealed by integrated quantitative trait locus (QTL) analysis and association. **New Phytologist**. vol. 209. p. 1067-1082. 2016.

DUENK, P.; CALUS, M.P.L; WIENTJES, Y.C.J.; BIJMA, P. Benefits of Dominance over Additive Models for the Estimation of Average Effects in the Presence of Dominance. **G3**. vol. 7. p. 3405-3414. 2017.

DURÁN, R.; ISIK, F.; ZAPATA-VALENZUELA, J.; BALOCCHI, C.; VALENZUELA, S. Genomic predictionns of breeding values in a cloned *Eucalyptus globulus* population in Chile. **Tree Genet Genomes**. vol. 13. n. 74. 2017.

ENDELMAN, J.B. Ridge regression and other kernels for genomic selection with R package rrBLUP. **Plant Genome**. vol. 4. n. 3. p. 250-255. 2011.

ESTOPA, R.A. Fenotipagem via NIR e Predição Genômica em *Eucalyptus benthamii*. 2017, 104p. Tese (Doutorado em Produção Vegetal) – Universidade Federal do Paraná – UFPR. Curitiba.

FAHRENKROG, A.M.; NEVES, L.G.; RESENDE JR.; M.F.R.; VAZQUEZ, A.I.; DE LOS CAMPOS, G.; DERVINIS, C.; SYKES, R.; DAVIS, M.; DAVENPORT, R.; BARBAZUK, W.B.; KIRST, M. Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. **New Phytologist**. vol. 213. p. 799-811. 2017.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to Quantitative Genetics**. Harlow: Longmans Green. 464 p. 1996.

FAN, H.; WU, Y.; ZHOU, X.; XIA, J.; ZHANG, W.; SONG, Y.; LIU, F.; CHEN, Y.; ZHANG, L.; GAO, X.; GAO, H.; LI, J. Pathway-based genome-wide association studies for two meat production traits in simmental cattle. **Sci Rep**. vol. 5. n. 18389. 2015.

FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3. ed. Brasília: EMBRAPA-CENARGEN, 1998. 220p.

FISHER, R. A. The correlation between relatives on the supposition of Mendelian inheritance. **Trans. R. Soc. Edinb**. vol. 52. p. 399–433. 1918.

FISHER, R.A. **The genetical theory of natural selection**. Oxford, UK: The Clarendon Press. 1930.

FLORES, T.B.; ALVARES, C.A.; SOUZA, V.C.; STAPE, J.L. **Eucalyptus no Brasil: zoneamento climático e guia para identificação**. Piracicaba: IPEF, 2016. 448 p.

FORNI, S.; AGUILAR, I.; MISZTAL, I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetics Selection Evolution**. vol. 43. n. 1. 2011.

FOURNIER, T.; SAADA, O.A.; HOU, J.; PETER, J.; CAUDAL, E.; SCHACHERER, J. Extensive impact of low-frequency variants on the phenotypic landscape at population-scale. **eLife**. 2019.

FRANKHAM, R. Inbreeding and extinction – a threshold effect. **Conservation Biology**, Oxford, vol. 9. p. 792-799. 1995.

FRANKHAM, R.; BALLOU, J.D.; BRISCOE, D.A. **Introduction to Conservation Genetics**. Cambridge University Press: Cambridge, UK. 2010. 618 p.

FREEMAN, J.S.; POTTS, B.M.; DOWNES, G.M.; PILBEAM, D.; THAVAMANIKUMAR, S.; VAILLANCOURT, R.E. Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus*. **New Phytologist**. vol. 198. n. 4. p. 1121-1134. 2013.

GAMAL EL-DIEN, O.; RATCLIFFE, B.; KLÁPŠTĚ, J.; CHEN, C.; PORTH, I.; EL-KASSABY, Y.A. Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. **BMC Genomics**. vol. 16. n. 370. 2015.

GAMAL EL-DIEN, O.; RATCLIFFE, B.; KLÁPŠTĚ, J.; PORTH, I.; CHEN, C.; EL-KASSABY, Y.A. Implementation of the Realized Genomic Relationship Matrix to Open-Pollinated White Spruce Family Testing for Disentangling Additive from Nonadditive Genetic Effects. **G3**. vol. 6. p. 743 – 753. 2016.

GERALDES, A.; DIFAZIO S.P.; SLAVOV, G.T.; RANJAN, P.; MUCHERO, W.; HANNEMANN, J; GUNTER, L.E.; WYMORE, A.M.; GRASSA, C.J.; FARZANEH N et al. A 34k SNP genotyping array for *Populus trichocarpa*: design, application to the study of natural populations and transferability to other populus species. **Molecular Ecology Resources**. vol. 13. p. 306–323. 2013.

GERMAN, S.; GERMAN, D. Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. **IEEE Trans Pattern Anal Machice Intelligence**. vol. 6. p. 721-741. 1984.

GION, J.M.; CAROUCHE, A.; DEWEER, S.; BEDON, F.; PICHAVANT, F.; CHARPENTIER, J.P.; BAILLERES, H. et al. Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: *Eucalyptus*. **BMC Genomics**. vol. 12. n. 301. p. 1-19. 2011.

GODDARD, M.E.; HAYES, B.J. Genomic selection. **Journal of Animal Breeding and Genetics**. vol. 124. p. 323-330, 2007.

GODDARD, M.E.; HAYES, B.J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. **Nat Rev Genet**. vol. 10. p. 381–391. 2009.

GODDARD, M.E.; HAYES, B.J.; MEUWISSEN, T.H.E. Using the genomic relationship matrix to predict the accuracy of genomic selection. **J Anim Breed Genet**. vol. 128. n. 6. p. 409–21. 2011.

GOGARTEN, S. M.; BHANGALE, T.; CONOMOS, M.P.; LAURIE, C. A.; MCHUGH, C. P.; PAINTER, I.; ZHENG, X.; CROSSLIN, D. R. et al. GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. **Bioinformatics**, vol. 28, n. 24, p. 3329-3331, 2012.

GOLDSCHMID, O. Ultraviolet spectra. In: Sarkanen, K.V. e Ludwig,C.H. Ed. **Lignins**. New York: Wiley-Interscience. p. 241-266. 1971.

GOMIDE, J.L. and DEMUNER, B.J. Determination of lignin content in woody material: modification Klason Method. **O Papel**. vol. 47. 1986.

GONÇALVES, P.S.; BORTOLETTO, N.; FONSECA, F. S., BATAGLIA, O. C.; ORTOLANI, A. A. Early selection for growth vigor in rubber tree genotypes in northwestern São Paulo State (Brazil). **Genetics and Molecular Biology**. vol. 21, n. 4, p. 1-15, 1998.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic Linkage Maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. **Genetics**. vol. 137. n. 4. p. 1121-1137. 1994.

GRATTAPAGLIA, D.; BERTOLUCCI, F.L.; SEDEROFF, R.R. Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudotestcross strategy and RAPD markers. **Theor Appl Genet.** vol. 90. p. 933–947. 1995.

GRATTAPAGLIA, D.; BERTOLUCCI, F.L.G.; PENCHEL, R.; SEDEROFF, R.R. Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. **Genetics.** vol. 144. p. 1205–1214. 1996.

GRATTAPAGLIA, D.; KIRST, M. Eucalyptus applied genomics: from gene sequences to breeding tools. **New Phytol.** vol. 179. p. 911–929. 2008.

GRATTAPAGLIA, D.; PLOMION, C.; KIRST, M.; SEDEROFF RR; Genomics of growth traits in forest trees. **Curr. Opin. Plant. Biol.** vol. 12. p. 148–56. 2009.

GRATTAPAGLIA, D.; RESENDE M.D.V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, vol. 7, n. 2, p. 241-255, 2011.

GRATTAPAGLIA, D. Breeding forest trees by genomic selection: current progress and the way forward. In: TUBEROSA, R.; GRANER, A.; FRISON, E. **Genomics of plant genetic resources.** Berlin: Springer, 2014. chap. 26, p. 652-678.

GRATTAPAGLIA, D.; SILVA-JUNIOR, O.B.; RESENDE, R.T.; CAPP, E.P.; MÜLLER, B.S.F.; TAN, B.; ISIK, F.; RATCLIFFE, B.; EL-KASSABY, Y.A. Quantitative Genetics and Genomics Converge to Accelerate Forest Tree Breeding. **Frontiers in Plant Science.** vol. 9. n.1693. 2018.

HALEY, C.S.; VISSCHER, P. M. Strategies to utilize marker-quantitative trait loci associations. **Journal of Dairy Science**, vol. 81 (Suppl 2), p. 85-97. 1998.

HALL, N.; BROOKER, I. **Camden White Gum: *Eucalyptus benthamii*** Maiden et Cambage. Canberra: Department of National Development Forestry and Timber Bureau, 1973. 4 p. (Forest Tree Series, 57).

HANNRUP, B.; CAHALAN, C.; CHANTRE, G.; GRABNER, M.; KARLSSON, B.; LE BAYON, I. LLOYD, G. et al. Genetic parameters of growth and wood quality traits in *Picea abies*. **Scand. J. For. Res.** vol. 19. p. 14-29. 2004.

HAYES, B. J.; GODDARD, M. E. Technical note: Prediction of breeding values using marker-derived relationship matrices. **Journal of Animal Science** vol. 86. p. 2089-2096. 2008.

HAYES, B. J.; VISSCHER, P. M.; GODDARD, M. E. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics Research**, New York, vol. 91. n. 1. p. 47-60, 2009.

HEFFNER, E.L.; LORENZ, A.J.; JANNINK, J-L.; SORRELLS, M.E. Plant breeding with genomic selection: gain per unit time and cost. **Crop Science.** vol. 50. p. 1681–1690. 2010.

HENRIQUE, M.; SANTANA, D.A.; ANTÔNIO, G.; JUNIOR, O.; SILVA, A.; CESAR, M.; FREUA, M.C.; GOMES, C.; LEME, P.R.; FUKUMASU, H. et al. Copy 138 number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the

feed conversion ratio in beef cattle. **Journal of Applied Genetics**. vol. 57. n. 4. p. 495-504. 2016.

HEUER, C. **cpgen: Parallel Genomic Evaluations**. R package version 0.2. 2018.

HIGA, R.C.V.; WREGGE, M.S.; HIGA, A.R.; SANTOS, G.A. Considerações sobre Exigências Climáticas de *Eucalyptus benthamii*. In: SILVA, L. D.; HIGA, A. R.; SANTOS, G. A. **Silvicultura e melhoramento genético de *Eucalyptus benthamii***. Curitiba: FUPEF, 2012, 150 p.

HIGGINS, M.G.; FITZSIMONS, C.; MCCLURE, M.C.; MCKENNA, C.; CONROY, S.; KENNY, D.A.; MCGEE, M.; WATERS, S.M.; MORRIS, D.W. GWAS and eQTL analysis identifies a SNP associated with both residual feed intake and GFRA2 expression in beef cattle. **Scientific Reports**. vol. 8. n. 14301. 2018.

HIRAOKA, Y.; FUKATSU, E.; MISHIMA, K.; HIRAO, T.; TESHIMA, K.M.; TAMURA, M.; TSUBOMURA M. et al. Potential of Genome-Wide Studies in Unrelated Plus Trees of a Coniferous Species, *Cryptomeria japonica* (Japanese Cedar). **Frontiers in Plant Science**. vol. 9. n. 1322. 2018.

HONG, E.P.; PARK, J.W. Sample size and statistical power calculation in genetic association studies. **Genomics & Informatics**. vol. 10. n. 2. p. 117-122. 2012.

HOWE, G.T.; YU, J.; KNAUS, B.; CRONN, R.; KOLPAK, S.; DOLAN, P.; LORENZ, W.W.; DEAN, J.F.D. A SNP resource for Douglas-fir: *de novo* transcriptome assembly and SNP detection and validation. **BMC Genomics**. vol. 14. p. 137. 2013.

HUANG, X.; HAN, B. Natural Variations and Genome-Wide Association Studies in Crop Plants. **Annual Reviews of Plant Biology**. vol. 65. p. 531–551. 2014.

HYTEN, D.L.; CHOI IY, SONG. Q.; SHOEMAKER, R.C.; NELSON, R.L. et al. Highly variable patterns of linkage disequilibrium in multiple soybean populations. **Genetics**. vol. 175. p. 1937–1944. 2007

IBÁ. Indústria Brasileira de Árvores. Relatório IBÁ 2019. Brasília, 79 p.

IMAI, A.; KUNIGA, T.; YOSHIOKA, T.; NONAKA, K.; MITANI, N.; FUKAMACHI, H.; HIEHATA, N.; YAMAMOTO, M.; HAYASHI, T. Single-step genomic prediction of fruit-quality traits using phenotypic records of non-genotyped relatives in citrus. **PLoS One**. vol 14. n. 8. 2019.

INGVARSSON, P.K.; STREET, N.R. Association genetics of complex traits in plants. **New Phytologist**. vol. 189. n. 4. p. 909-922. 2011.

JACQUIN, L.; CAO, T.; AHMADI, N. A Unified and Comprehensible View of Parametric and Kernel Methods for Genomic Prediction with Application to Rice. **Front. Genet**. vol. 7. 2016.

JIANG, Y.; REIF, J. C. Modeling epistasis in genomic selection. **Genetics**. vol. 201. p. 759-768. 2015

JOO, J.W.J.; HORMOZDIARI, F.; HAN, B.; ESKIN, E. Multiple testing correction in linear mixed models. **Genome Biology**. vol. 17. n. 62. 2016.

KAGEYAMA, P.Y.; VENCOVSKY, R. Variação genética em progênes de uma população de *Eucalyptus grandis* (Hill) Maiden. IPEF, Piracicaba, vol. 24. p. 9-26, 1983.

KALER, A.S.; PURCELL, L.C. Estimation of a significance threshold for genome-wide association studies. **BMC Genomics**. vol. 20. n. 618. 2019.

KANG, H.M.; SUL, J.H.; SERVICE, S.K.; ZAITLEN, N.A. KONG, S.Y.; FREIMER, N.B. SABATTI, C.; ESKIN, E. Variance component model to account for sample structure in genome-wide association studies. **Nat. Genet.** vol. 42. n. 4. p. 348-354. 2010.

KENNY, E.E.; KIM, M.; GUSEV, A.; LOWE, J.K.; SALIT, J. et al. Increased power of mixed models facilitates association mapping of 10 loci for metabolic traits in an isolated population. **Human Molecular Genetics**. vol. 20. p. 827–39. 2011.

KIMELDORF, G.; WAHBA, G. Some results on Tchebycheffian spline functions. **J. Math. Anal. Appl.** vol. 33. p. 82–95. 1971.

KÖPPEN, W.; GEIGER, R. **Klimate der Erde**. Gotha: Verlag Justus Perthes. 1928.

KJAER, E.; AMARAL, W.; YANCHUK, A.; GRAUDAL, L. Strategies for conservation of forest genetic resources. Conservation of *Eucalyptus benthamii*: an endangered eucalypt species from eastern Australia. In: FAO, FLD, IPGRI. **Forest genetic resources conservation and management: overview, concepts and some systematic approaches**. Rome, Italy: International Plant Genetic Resources Institute, v. 1, 2004. p. 5-24.

KLÁPŠTĚ, J.; SUONTAMA, M.; DUNGEY, H.S.; TELFER, E.J.; GRAHAM, N.J. LOW, C.B. STOVOLD, G.T. Effect of Hidden Relatedness on Single-Step Genetic Evaluation in an Advanced Open-Pollinated Breeding Program. **Journal of Heredity**. vol. 109. n. 7. p. 802-810. 2018.

LADIGES, P.Y.; UDOVICIC, F.; NELSON, G. Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. **Journal of Biogeography**. vol. 30. n. 7. p. 989-998. 2003.

LANDE, R.; THOMPSON, R. Efficiency of marker assisted selection in the improvement of quantitative traits. **Genetics**, vol. 124, p. 743-756, 1990.

LEE, S.H.; GODDARD, M.E.; VISSCHER, P.M.; VAN DER WERF, J.H.J. Using the realized relationship matrix to disentangle confounding factors for the estimation of genetic variance components of complex traits. **Genetics Selection Evolution**, London, vol. 42, n. 22, p. 1-14, 2010.

LEE, S.H.; CLARK, S.; VAN DER WERF, J.H.J. Estimation of genomic prediction accuracy from reference populations with varying degrees of relationship. **PLoS One**. vol 12. n. 12. 2017.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information, **J. Dairy Sci.** vol. 92. p. 4656-4663. 2009.

LEGARRA, A.; CHRISTENSEN, O.F.; AGUILAR, I.; MISZTAL, I. Single Step, a general approach for genomic selection. **Livest. Sci.** vol. 166. p. 54–65. 2014.

LIMA, B.M. **Bridging genomics and quantitative genetics of Eucalyptus: genome-wide prediction and genetic parameter estimation for growth and wood properties using high-density SNP data.** Doctoral Thesis. University of São Paulo: Piracicaba, Brazil. 92 p. 2014.

LIMA, B.M.; CAPPA, E.P.; SILVA-JUNIOR, O.B.; GARCIA, C.; MANSFIELD, S.D.; GRATTAPAGLIA, D. Quantitative genetic parameters for growth and wood properties in *Eucalyptus* “urograndis” hybrid using near-infrared phenotyping and genome-wide SNP-based relationships. **PLoS ONE.** vol. 14. n. 6. 2019.

LIN, M.; ARNOLD, R.; LI, B.; YANG, M. Selection of cold-tolerant eucalypts for Hunan Province. In: TURNBULL J.W. (ed) *Eucalypts in Asia: proceedings of a international conference held in Zhanjiang, Guangdong, people’s Republic of China.* ACIAR Proceedings, Zhanjiang, Guangdong, China, pp. 107–116. 2003.

LOPES, M.S.; DREISIGACKER, S.; PEÑA, R.J.; SUKUMARAN, S.; REYNOLDS, M.P. Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. **Theor Appl Genet.** vol. 128. p. 453-464. 2015.

LU, M.; KRUTOVSKY, K.V.; NELSON, C.D.; WEST, J.B.; REILLY, N.A.; LOOPSTRA, C.A. Association genetics of growth and adaptive traits in loblolly pine (*Pinus taeda* L.) using whole exome-discovered polymorphisms. **Tree Genet Genomes.** 2017.

LYNCH, M. A quantitative-genetic perspective on conservation issues. In: AVISE, J.; HAMRICK, J. (Eds.). **Conservation genetics: case histories from nature.** New York: Chapman and Hall, 1996. p. 471-501.

MACKAY, L.; POWELL, W. Methods for linkage disequilibrium mapping in crops. **Trends Plant Sci.** vol. 12. p. 57-63. 2007.

MACKILL, D. J.; NI, J. Molecular mapping and marker-assisted selection for major-gene traits in rice. **Rice Genetics.** vol. 4. p. 137–151. 2008.

MAHER, B. The case of the missing heritability. **Nature.** vol. 456. p. 18–21. 2008.

MANOLIO, T.A.; COLLINS, F.S.; COX, N.J.; GOLDSTEIN, D.B.; HINDORFF, L.A.; HUNTER, D.J.; MCCARTHY, M.I.; RAMOS, E.M. et al. Finding the missing heritability of complex diseases. **Nature.** vol. 461. p. 747–753. 2009.

MARRONI, F.; PINOSIO, S.; ZAINGA, G.; FOGOLARI, F.; FELICE, N.; CATTONARO, F.; MORGANTE, M. Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (CAD4) gene. **Tree Genet Genomes.** vol. 7. p. 1011–1023. 2011.

MCKOWN, A.D.; KLÁPSTE, J.; GUY, R.D.; GERALDES, A.; PORTH, I.; HANNEMANN, J.; FRIEDMANN, M.; MUCHERO, W.; TUSKAN, G.A.; EHLTING, J. et al. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. **New Phytologist.** vol. 203. p. 535–553. 2014.

MEUWISSEN, T.H.E.; GODDARD M.E. The use of marker haplotypes in animal breeding schemes. **Genetics Selection Evolution.** vol. 28. p. 161-76. 1996.

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

MEUWISSEN T.H.E.; LUAN T.; WOOLLIAMS J.A. The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. **Journal of Animal Breeding and Genetics**. vol. 128. n. 6. p. 429–439. 2011.

MEUWISSEN T.H.E.; HAYES B.; GODDARD M. Genomic selection: A paradigm shift in animal breeding. **Animal Frontiers**. vol. 6. p. 6–14. 2016.

MILLER, R.G. **Simultaneous Statistical Inference**. 2. ed. New York: Springer-Verlag New York. 1981. 299 p.

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **J. Dairy. Sci.** vol. 92. p. 4648–4655. 2009.

MOMEN, M; MEHRGARDI, A.A.; SHEIKHI, A.; KRANIS, A.; TUSSEL, L.; MOROTA, G.; ROSA, G.J.M.; GIANOLA, D. Predictive ability of genome-assisted statistical models under various forms of gene action. **Scientific Reports**. vol. 8. n. 12309. 2018.

MORA, A.L.; GARCIA, C.H. **Eucalypt cultivation in Brazil**. São Paulo: Sociedade Brasileira de Silvicultura, 2000. 112p.

MORA, F.; BALLESTA, P.; SERRA, N. Bayesian analysis of growth, stem straightness and branching quality in full-sib families of *Eucalyptus globulus*. **Bragantia**, Epub August 08. 2019.

MORAIS JÚNIOR, O.P.; DUARTE, J.B.; BRESEGHELLO, F.; COELHO, A.S.G.; MORAIS, O.P.; MAGALHÃES JÚNIOR, A.M. Single-step reaction norm models for genomic prediction in multi-environment recurrent selection trials. **Crop. Sci.** vol. 58. p. 592-607. 2018.

MOROTA, G.; KOYAMA, M.; ROSA, G.J.M.; WEIGEL, K.A.; GIANOLA, D. Predicting complex traits using a diffusion kernel on genetic markers with an application to dairy cattle and wheat data. **Genet. Sel. Evol.** vol. 45 n. 17. 2013.

MOROTA, G.; GIANOLA, D. Kernel-based whole-genome prediction of complex traits: A review. **Front. Genet.** vol. 5. p. 1–13. 2014.

MOROTA, G.; BODDHIREDDY, P.; VUKASINOVIC, N.; GIANOLA, D.; DENISE, S. Kernel-based variance component estimation and whole-genome prediction of precorrected phenotypes and progeny tests for dairy cow health traits. **Frontiers in Genetics**. vol.5. n. 56. 2014.

MUJIBI, F.D.N.; NKUMAH, J.D.; DURUNNA, O.N.; STOTHARD, P.; MAH, J.; WANG, Z.; BASARAB, J.; PLASTOW, G.; CREWS, D.H.; MOORE, S.S. Accuracy of genomic breeding values for residual feed intake in crossbred beef cattle. **Journal of Animal Science**. vol. 89. p. 3353– 3361. 2011.

MÜLLER, B.S.F.; NEVES, L.G; DE ALMEIDA FILHO, J.E.; RESENDE, M.F.R; MUÑOZ, P.R.; DOS SANTOS, P.E.T.; FILHO, E.P.; KIRST, M.; GRATTAPAGLIA, D. Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus*. **BMC Genomics**. vol. 18. n. 524. 2017.

MÜLLER, B.S.F.; FILHO, J.E.A.; LIMA, B.M.; GARCIA, C.C.; MISSIAGGIA, A.; AGUIAR, A.M.; TAKAHASHI, E.; KIRST, M. et al. Independent and Joint-GWAS for growth traits in *Eucalyptus* by assembling genome-wide data for 3373 individuals across four breeding populations. **New Phytologist**. 2018.

MUÑOZ, P.R.; RESENDE, M.F.; GEZAN, S.A.; RESENDE, M.D.V.; DE LOS CAMPOS, G. et al. Unraveling additive from nonadditive effects using genomic relationship matrices. **Genetics**. vol. 198. p. 1759–1768. 2014.

MYLES, S.; PEIFFER, J.; BROWN, P.; ERSOZ, E.; ZHANG, Z.; COSTICH, D.; BUCKLER, E. Association mapping: critical considerations shift from genotyping to experimental design. **Plant Cell**. vol. 21. n. 8. p. 2194–2202. 2009

NADEEM, M.A.; NAWAZ, M.A.; SHAHID, M.Q.; DOGAN, Y.; COMERTPAY, G. et al. DNA molecular marker in plant breeding: current status and recent advancements in genomic selection and genome editing. **Biotechnology & Biotechnological Equipment**. vol. 32. n. 2. p. 261-285. 2017.

NAMKOONG, G.; KANG, H.C.; BROUARD, J.S. **Tree breeding: principles and strategies**. New York: Springer-Verlag. 1988. 177p.

NAKAYA, A.; ISOBE, S.N. Will genomic selection be a practical method for plant breeding?. **Annals of Botany**. vol. 110. p. 1303-1316. 2012.

NEJATI-JAVAREMI, A.; SMITH, C.; GIBSON J.P. Effect of total allelic relationship on accuracy of evaluation and response to selection. **J Anim Sci**. vol. 75. p. 1738–1745. 1997.

NISGOSKI, S.; MUÑIZ, G.I.B.; KLOCK, U. Caracterização anatômica da madeira de *Eucalyptus benthamii* Maiden et Cambage. **Ciência Florestal**, Santa Maria, n.1, p.67-76, 1998.

NISHIO, M.; SATOH, M. Including dominance effects in the genomic BLUP method for genomic evaluation. **PloS One**. vol. 9. 2014.

NSW - NATIONAL PARKS & WILDLIFE SERVICE. **Threatened species information: *Eucalyptus benthamii* Maiden and Cambage**. Hurstville: National Parks & Wildlife Service, 2000. 3 p.

OTT, L.; LONGNECKER, M. **An Introduction to Statistical Methods and Data Analysis**. Brooks/Cole. 2010. 1273p.

PALUDETO, J.G.Z.; ESTOPA, R.A.; TAMBARUSSI, E.V. Eficiência da seleção precoce em clones de *Eucalyptus grandis* W. Hill ex Maiden x *Eucalyptus urophylla* S. T. Blake. **Rev. Inst. Flor**. vol. 29 n. 2. p. 169-179. 2017.

PALUDZYSZYN FILHO, E.; SANTOS, P. E. T.; FERREIRA, C. A. Eucaliptos indicados para plantio no Estado do Paraná. **Colombo: Embrapa Florestas - CNPF**, 2006. 45 p. (Documentos, 129).

PARDIÑAS, A.F.; HOLMANS, P.; POCKLINGTON, A.J.; ESCOTT-PRICE, V.; RIPKE, S.; CARRERA, N. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. **Nat Genet**. vol. 51. n. 1193. 2019.

PAVY, N.; GAGNON, F.; RIGAUULT, P.; BLAIS, S.; DESCHENES, A.; BOYLE, B.; PELGAS, B.; DESLAURIERS, M.; CLEMENT, S.; LAVIGNE, P. et al. c. **Molecular Ecology Resources**. vol. 13. p. 324–336. 2013.

PEREIRA, A.B.; MARQUES JÚNIOR, O.G.; RAMALHO, M. A. P.; ALTHOFT, P. Eficiência da seleção precoce em famílias de meios-irmãos de *Eucalyptus camaldulensis* Dehnh., avaliadas na região noroeste do estado de Minas Gerais. **Cerne**. vol. 3. n. 1, p. 67-81, 1997.

PEREIRA, M.B.; VENCOVSKY, R. Limites da seleção recorrente. I - Fatores que afetam o acréscimo das frequências alélicas. **Pesquisa Agropecuária Brasileira**. vol. 23, n. 7, p. 769-780, 1988.

PEREZ, P.; DE LOS CAMPOS, G. Genome-wide regression and prediction with the BGLR statistical package. **Genetics**. vol. 198. n. 2. p. 483-495. 2014.

PÉREZ-RODRÍGUEZ, P.; GIANOLA, D.; GONZÁLEZ-CAMACHO, J.M.; CROSSA, J.; MANÈS, Y.; DREISIGACKER, S. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. **G3**. vol. 2. p. 1595–1605. 2012.

PÉREZ-RODRÍGUEZ, P.; CROSSA, J.; RUTKOSKI, J.; POLAND, J.; SINGH, R.; LEGARRA, A.; AUTRIQUE, E.; DE LOS CAMPOS, G.; BURGUEÑO, J.; DREISIGACKER, S. Single-Step Genomic and Pedigree Genotype × Environment Interaction Models for Predicting Wheat Lines in International Environments. **Plant Genome**. vol. 10. 2017.

PERNEGER, T.V. What's wrong with Bonferroni adjustments. **BMJ**. vol. 316. n. 7139. p. 1236–1238. 1998.

PLUMMER, M.; BEST, N.; COWLES, K.; VINES, K. CODA: Convergence Diagnosis and Output Analysis for MCMC, **R News**. vol 6. p. 7-11. 2006.

POCIOT, F. Type 1 diabetes genome-wide association studies: not to be lost in translation. **Clin Transl Immunology**. vol. 6. n. 12. 2017.

PORTH, I.; KLÁPSTE, J.; SKYBA, O.; HANNEMANN, J.; MCKOWN, AD.; GUY, RD.; DIFAZIO, S.P.; MUCHERO, W.; RANJAN, P.; TUSKAN, G.A. et al. Genome-wide association mapping for wood characteristics in *Populus* identifies an array of candidate single nucleotide polymorphisms. **New Phytologist**. vol. 200. p. 710–726. 2013.

POTTS, B.M.; DUNGEY, H.S. Interspecific hybridization of *Eucalyptus*: key issues for breeders and geneticists. **New Forest**. vol. 27. p. 115-138. 2004.

PRICE, A.L.; ZAITLEN, N.A.; REICH, D.; PATERSON, N. New approaches to population stratification in genome-wide association studies. **Nature Reviews Genetics**. vol. 11. p. 459-463. 2010.

PRYOR, L.D. **Australian endangered species: Eucalyptus**. Canberra: Commonwealth of Australia. 1981. 139 p.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A.; BENDER, D.; MALLER, J.; SKLAR, P.; de BAKKER, P.I.; DALY, M. J.; SHAM, P.C. PLINK: a tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**. vol. 81, n. 3, p. 559-575. 2007.

QUERO, G.; GUTIÉRREZ, L.; MONTEVERDE, E.; BLANCO, P.; DE VIDA, F.P.; ROSAS, J.; FERNÁNDEZ, S.; GARAYCOCHEA, S. et al. Genome-Wide Association Study Using Historical Breeding Populations Discovers Genomic Regions Involved in High-Quality Rice. **The Plant Genome**. vol. 11. n. 3. 2018.

RASHEED, A.; HAO, Y.; XIA, X.; KHAN, A.; XU, Y.; VARSHNEY, R.K.; HE, Y. Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. **Molecular Plant**. vol. 10. p. 1047-1064. 2017.

RATCLIFFE, B.; GAMAL EL-DIEN, O.; CAPP, E.P.; PORTH, I.; KLAPSTE, J.; CHEN, C.; ELKASSABY, Y. Single-Step BLUP with Varying Genotyping Effort in Open-Pollinated *Picea glauca*. **G3**. vol. 7. n. 3. p. 935-942. 2017.

RATCLIFFE, B. **Exploring Genomic Selection in Conifers**. Doctoral Thesis. The University of British Columbia: Vancouver, 138 p. 2019.

REED, D.H. Relationship between population size and fitness. **Conservation Biology**, Oxford. vol. 19. p. 563-568. 2005.

RESENDE, M.D.V.; **Genética biométrica e estatística no melhoramento de plantas perenes**. Embrapa Informação Tecnológica, Brasília: 2002. 975 p.

RESENDE, M.D.V.; LOPES, P.S.; SILVA, R.L.; PIRES, I.E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, Colombo, n. 56, p. 63-77. 2008.

RESENDE, M.D.V.; ASSIS, T.F. Seleção recorrente recíproca entre populações sintéticas multi-espécies (SRR-PSME) de *Eucalyptus*. **Pesquisa Florestal Brasileira**. vol. 57. p. 57-60. 2008.

RESENDE, M.D.V.; SILVA, F.F.; LOPES, P.S. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial**. Viçosa: UFV, Departamento de Estatística. 2011. 180p.

RESENDE, M.D.V.; RESENDE JR., M.F.R.; SANSALONI, C.P.; PETROLI, C.D.; MISSIAGGIA, A.A.; AGUIAR, A.M.; ABAD, J.M. et al. Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**. vol. 194. 2012.

RESENDE, M.D.V.; SILVA, F. F. E.; AZEVEDO, C. F. **Estatística matemática, biométrica e computacional: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência**. Visconde do Rio Branco: Suprema, 2014. vol. 1, 881 p.

RESENDE, M.D.V. **Genética quantitativa e de populações**. Viçosa- MG: Suprema. 2015, 463 p.

RESENDE, R.T.; RESENDE, M.D.V.; SILVA, F.F.; AZEVEDO, C.F.; TAKAHASHI, E.K.; SILVA-JUNIOR, O.B.; GRATTAPAGLIA, D. Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in *Eucalyptus*. **New Phytologist**. vol. 213. p. 1287–1300. 2016.

RESENDE, R.T.; RESENDE, M.D.V.; SILVA, F.F.; AZEVEDO, C.F.; TAKAHASHI, E.K.; SILVA-JUNIOR, O.B.; GRATTAPAGLIA, D. Assessing the expected response to genomic selection of individual and families in *Eucalyptus* breeding with and additive-dominant model. **Heredity**. vol. 119. p. 245–255. 2017.

ROEDER, K.; WASSERMAN, L. Genome-wide significance levels and weighted hypothesis testing. **Statistical Science**. vol. 24. n. 4. p. 398-413. 2009.

ROOS, A.P.W. **Genomic selection in dairy cattle**. Ph.D. thesis, Wageningen University, the Netherlands. 183 p. 2011.

SANTOS, V.S.; MARTINS FILHO, S.; RESENDE, M.D.V.; FONSECA E SILVA, F. Proposta de BLUP genômico com efeitos aditivos e de dominância em ambiente R. **Rev. Bras. Biom.** vol. 35, n. 2, p. 361-375. 2017.

SAX, K. The association of size differences with seed-coat pattern, and pigmentation in *Phaseolus vulgaris*. **Genetics**, vol. 8, p. 552–560. 1923.

SCHUSTER, I. Marker-assisted selection for quantitative traits. **Crop Breeding and Applied Biotechnology**. vol. 11 p. 50-55, 2011.

SILVA, L.D.; HIGA, A.R.; SANTOS, G.A. **Silvicultura e melhoramento genético de *Eucalyptus benthamii***. Curitiba: FUPEF, 2012. 150 p.

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

SILVA-JUNIOR, O.B.; GRATTAPAGLIA, D. Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. **New Phytologist**. vol. 208, n. 3. p. 830-845. 2015.

SMITH, C.; SIMPSON, S.P. The use of genetic polymorphisms in livestock improvement. **Journal of Animal Breeding and Genetics**, vol. 103. p. 205–217. 1986.

SOLLER, M.; BECKMANN, J.S. Genetic polymorphism in varietal identification and genetic-improvement. **Theoretical and Applied Genetics**, vol. 67. p. 25–33. 1983.

SPINDEL, J.; BEGUM, H.; AKDEMIR, D.; VIRK, P.; COLLARD, B.; REDONA, E.; ATLIN, G.; JANNINK, J.L.; MCCOUCH, S.R. Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. **PLoS Genet**. vol. 11. n. 2. 2015.

SQUILLACE, A. Average genetic correlations among offspring from open-pollinated forest trees. **Silvae Genet**. vol. 23. p. 149–156. 1974.

SU, G.; CHRISTENSEN, O. F.; OSTERSEN, T.; HENRYON, M.; LUND, M.S. Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. **PLoS ONE**. vol. 7. n. 9. 2012.

SUONTAMA, M.; KLÁPSTE, J.; TELFER, E.; GRAHAM, N.; STOVOLD, T.; LOW, C.; MCKINLEY, R.; DUNGEY, H. Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. **Heredity**. vol. 122. p. 370-379. 2019.

SWAIN, T. L. **A status report on the trials in the ICFR cold tolerant Eucalyptus breeding program**. Scottsville: Institute for Commercial Forestry Research, 1996. 100 p. (Bulletin series 2/96).

TAMBARUSSI, E.V.; PEREIRA, F.B.; DA SILVA, P.H.M.; LEE, D.; BUSH, D. Are tree breeders properly predicting genetic gain? A case study involving *Corymbia* species. **Euphytica**. vol. 214. n. 150. 2018.

TAN, B.; GRATTAPAGLIA, D.; MARTINS, G. S.; FERREIRA, K. Z.; SUNDBERG, B.; INGVARSSON, P. K. Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F₁ hybrids. **BMC Plant Biology**. vol. 17 n. 110. 15 p. 2017.

TAN, B.; GRATTAPAGLIA, D.; WU, H.X.; INGVARSSON, P.K. Genomic relationship reveal significant dominance effect for growth in hybrid *Eucalyptus*. **Plant Science**. vol. 267. p. 84-93. 2018.

Technical Association of the Pulp and Paper Industry-TAPPI. T280 pm-99 standard - Acetone extractives of wood and pulp. **TAPPI Press**, 2000.

THAVAMANIKUMAR, S.; LUKE, J.M.; TIBBITS, J.F.G.; GERD, B. The Significance of Single Nucleotide Polymorphisms (SNPS) in *Eucalyptus globulus* Breeding Programs. **Aust. For.** vol. 74. p. 23–29.

THUMMA, B.R.; NOLAN, M.R.; EVANS, R.; MORAN, G.F. Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. **Genetics**. vol. 171. n. 3. p. 1257-1265. 2005.

THUMMA, B.R.; SOUTHERTON, S.G.; BELL, J.C.; OWEN, J.V.; HENERY, M.L.; MORAN, G.F. Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens*. **Tree Genetics & Genomes**. vol. 6. n. 2. p. 305-317. 2010.

TORRES-DINI, D.G. Detection of QTLs associated to DBH in a *Eucalyptus grandis* x *Eucalyptus globulus* monoprogeny. 2017. Tese (Doutorado). Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira.

VAN RADEN, P. 2008. Efficient methods to compute genomic predictions. **Journal of Dairy Science**. vol. 91, p. 4414-4423.

VELAZCO, J.G.; MALOSETTI, M.; HUNT, C.H.; MACE, E.S.; JORDAN, D.R.; VAN EUWJIK, F.A. Combining pedigree and genomic information to improve prediction quality: an example in sorghum. **Theoretical and Applied Genetics**. vol. 132. n. 7. p. 2055-2067. 2019.

VELLO, N.A.; NAZATO, F.M. Recurrent selection. In: SILVA, F.L; BORÉM, A.; SEDIYAMA, T.; LUDKE, W.H. **Soybean Breeding**. Springer International Publishing, 2017. p. 171-191.

VENCOVSKY, R. Tamanho efetivo populacional na coleta e preservação de germoplasmas de espécies alogamas. **IPEF**. p. 79-84. 1987.

VENCOVSKY, R.; CROSSA, J. Variance Effective Population Size under Mixed Self and Random Mating with Applications to Genetic Conservation of Species. **Crop Science**. vol. 39. p. 1282-1294. 1999.

VENCOVSKY, R.; CROSSA, J. Measurements of representativeness used in genetic resources conservation and plant breeding. **Crop Science**. vol. 43. p. 1912-1921. 2003.

VILHJALMSSON, B.J.; NORDBORG, M. The nature of confounding in genome-wide association studies. **Nat Rev Genet**. vol. 14. n. 1. p. 1-2. 2013.

VISSCHER, P.M.; HEMANI, G.; VINKHUYZEN, A.A.E.; CHEN, G.B.; LEE, S.H.; WRAY, N.R.; GODDARD, M.E.; YANG, J. Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples. **PLoS Genet**. vol. 10. n. 4. 2014.

VISSCHER, P.M.; WRAY, N.R.; ZHANG, Q.; SKLAR, P.; MCCARTHY, M.I.; BROWN M.A.; YANG, J. 10 years of GWAS discovery: biology, function, and translation. **American Journal of Human Genetics**. vol. 101. p. 5–22. 2017.

VITEZICA, Z. G.; VARONA, L.; LEGARRA, A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. **Genetics**. vol. 195. n. 4. p. 1223-1230. 2013.

VITEZICA, Z.G.; LEGARRA, A.; TORO, M.A.; VARONA, L. Orthogonal Estimates of Variances for Additive, Dominance and Epistatic Effects in Populations. **Genetics**. vol. 206. n. 3. p. 1297-1307. 2017.

VOLKER, P.W.; POTTS, B.M.; BORRALHO, N.M.G. Genetic parameters of intra and interspecific hybrids of *Eucalyptus globulus* and *E. nitens*. **Tree Genet. Genomes**. vol. 4. p. 445- 460. 2008.

WALLIS, A.F.A.; WEARNE, R.H. and WRIGHT, P.J. Analytical characteristics of plantation eucalypt woods relating to Kraft pulp yields. **Appita Journal**. vol. 49. 1996.

WANG, J. Effective size and F-statistics of subdivided populations. I. Monoecious species with partial selfing. **Genetics**. vol. 146. p. 1453-1463. 1997.

WANG, D.G.; FAN, J.B.; SIAO, C.J.; BERNO, A.; YOUNG P.; SAPOLSKY, R. et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. **Science**. vol. 280. n. 5366. p. 1077-1082. 1998.

WANG, X.; XING, E.P.; SCHAID, D.J. Kernel method for large-scale genomic data analysis. **Briefings in Bioinformatics**. vol. 16. n. 2. p. 183-192. 2014.

WANG, S.; FENG, J.; REN, W.; HUANG, B.; ZHOU, L.; WEN, Y.; ZHANG, J.; DUNWELL, J.M.; XU, S.; ZHANG, Y. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. **Scientific Reports**. vol. 6. 2016a.

WANG, J.; SANTIAGO, E.; CABALLERO, A. Prediction and estimation of effective population size. **Heredity**. vol. 117. p. 193-206. 2016b.

- WAPLES, R.S. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. **Conservation Genetics**, Netherlands, vol. 7, n. 2, p. 167-184. 2006.
- WAPLES, R.S.; DO, C. LDNE: a program for estimating effective population size from data on linkage disequilibrium. **Molecular Ecology Resources**. vol. 8. p. 753-756. 2008.
- WAPLES, R.S.; DO, C. Linkage disequilibrium estimated of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution **Evol. Appl.** vol. 3. n. 3. p. 244-262. 2010.
- WARNES, G.R.; BOLKER, B.; BONEBAKKER, L.; GENTLEMAN, R.; LIAW, W.; LUMLEY, T. et al. **gplots: Various R Programming Tools for Plotting Data**. R package version 3.0.1.1. 2019.
- WEIR, B. S. Inferences about linkage disequilibrium. **Biometrics**, Washington, vol. 35, n. 1, p. 235-254. 1979.
- WELLMANN, R.; BENNEWITZ, J. Bayesian models with dominance effects for genomic evaluation of quantitative traits. **Genet Res.** vol. 94. p. 21-37. 2012.
- WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016.
- WILLIAMS, K.S.; CUMMINGS, M.R. **Concepts of genetics**, 5th edn. Englewood Cliffs, NJ: Prentice-Hall. 1997.
- WRIGHT, S. Evolution in Mendelian populations. **Genetics**. vol. 16. p. 97–159. 1931.
- WRIGHT, S. Inbreeding and homozygosis. **Proc Natl Acad Sci USA**. vol. 19. p. 411–420. 1933.
- XIANG, T.; CHRISTENSEN, O.F.; VITEZICA, Z.G., LEGARRA, A. Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. **Genet. Sel. Evol.** vol. 48. n. 98. 2016.
- XIANG, T.; CHRISTENSEN, O.F.; VITEZICA, Z.G., LEGARRA, A. Genomic model with correlation between additive and dominance effects. **Genetics**. vol. 209. n. 3. p. 711-723. 2018.
- XIAO, Y.; LIU, H.; WU, L.; WARBURTON, M.; YAN, J. Genome-wide association studies in maize: Praise and stargaze. **Molecular Plant**. vol. 10. p. 359–374. 2017.
- XU, S. Genetic mapping and genomic selection using recombination breakpoint data. **Genetics**. Vol. 195, p. 1103-1115. 2013.
- XU, YU; LIU, X.; FU, J.; WANG, H.; WANG, J.; HUANG, C.; PRASANNA, B.M.; OLSEN, M.S. WANG, G.; ZANG, A. Enhancing genetic gain through genomic selection: from livestock to plants. **Plant Communications**. vol. 1. n. 1 2020.
- XUE, A.; WU, Y.; ZHU, Z.; ZHANG, F.; KEMPER, K.E.; ZHILI, Z.; YENGO, L.; LLOYD-JONES, L.R. et al. Genome-wide association analysis identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. **Nature Communications**. vol. 9. n. 2941. 2018.

YANG, J.; BEBEN, B.; MCEVOY, B.P.; GORDON, S.; HENDERS, A.K.; NYHOLT, D.R.; MADDEN, P.F.; HEATH, A.C.; MARTIN, N.G.; MONTGOMERY, G.W.; GODDARD, M.E.; VISSCHER, P.M. Missing heritability of human height explained by genomic relationships. **Nat. Genet.** vol. 42. p. 565–569. 2010.

YANG, J.; LEE, S.H.; GODDARD, M.E.; VISSCHER, P.M. Genome-Wide Complex Trait Analysis (GCTA): Methods, Data Analyses, and Interpretations. **Methods Mol Biol.** vol. 1039. p. 215–236. 2013.

YANG, J.; ZENG, J.; GODDARD, M.E.; WRAY, N.R.; VISSCHER, P.M. Concepts, estimation and interpretation of SNP-based heritability. **Nature Genetics.** vol. 49. n. 9. p. 1304–1310. 2017.

YANG, Y.; WANG, L.; LI, L.; LI, W.; ZHANG, Y.; CHANG, H.; XIAO, X.; LI, M.; LV, L. Genetic association and meta-analysis of a schizophrenia GWAS variant rs10489202 in East Asian populations. **Translational Psychiatry.** vol. 8. n. 144. 2018.

YIN, T.; KÖNIG, S. Genome-wide associations and detection of potential candidate genes for direct genetic and maternal genetic effects influencing dairy cattle body weight at different ages. **Genetics Selection Evolution.** vol. 51. n. 4. 2019.

YONQI, Z.; HUORAN, W.; RONGGUI, Z. QUINGSHENG, M. Trials of *Eucalyptus smithii* and other eucalypt species in Yunnan province, China. In: BROWN, A. G. (Ed.). **Australian Tree Species Research in China: Proceedings of an International Workshop held at Zhangzhou, Fujian Province.** Canberra: ACIAR, 1994. n. 48.

YOUNG, A.L Solving the missing heritability problem. **PLoS Genet.** vol. 15. n. 6. 2019.

YU, J.; PRESSOIR, G.; BRIGGS, W.H.; VROH BI, I.; YAMASAKI, M.; DOEBLEY, J.F.; MCMULLEN, M.D. et al unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nat Genet.** vol. 38. p. 203–208. 2006.

YU, A.; GALLAGHER, T. Analysis on the Growth Rhythm and Cold Tolerance of Five Year Old *Eucalyptus benthamii* Plantation for Bioenergy. **Open Journal of Forestry.** vol. 5. p. 585–592. 2015.

ZEGEYE, H.; RASHEED, A.; MAKDIS, F.; BADEBO, A.; OGBONNAYA, F.C. Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic Hexaploid wheat. **PLoS One.** vol. 9. n. 8. 2014.

ZHANG, F.; WANG, Y. P.; DENG, H. W. Comparison of population-based association study methods correcting for population stratification, **PLoS One.** vol. 3 n. 10. 2008.