

TAIANA LOPES RANGEL MIRANDA

**AVALIAÇÃO DE UMA NOVA MODELAGEM PROPORCIONAL DOS
EFEITOS ADITIVOS E DE DOMINÂNCIA E REFLEXOS NA
GENÉTICA QUANTITATIVA E NA SELEÇÃO GENÔMICA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2019

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

M672a
2019
Miranda, Taiana Lopes Rangel, 1995-
Avaliação de uma nova modelagem proporcional dos
efeitos aditivos e de dominância e reflexos na genética
quantitativa e na seleção genômica / Taiana Lopes Rangel
Miranda. – Viçosa, MG, 2019.
ix, 28 f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Marcos Deon Vilela de Resende.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 17-19.

1. Genética quantitativa. 2. Genômica. 3. BLUP.
4. Modelos genéticos. 5. Dominância. 6. Máxima
verossimilhança. I. Universidade Federal de Viçosa.
Departamento de Estatística. Programa de Pós-Graduação em
Estatística Aplicada e Biometria. II. Título.

CDD 22. ed. 576.5

TAIANA LOPES RANGEL MIRANDA

**AVALIAÇÃO DE UMA NOVA MODELAGEM PROPORCIONAL DOS EFEITOS
ADITIVOS E DE DOMINÂNCIA E REFLEXOS NA GENÉTICA QUANTITATIVA E
NA SELEÇÃO GENÔMICA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 24 de julho de 2019.

Assentimento:



Taiana Lopes Rangel Miranda

Autora



Marcos Deon Vilela de Resende

Orientador

Aos meus pais, Paulo e Tatiana;

Aos meus irmãos;

À minha avó Anna Amélia.

AGRADECIMENTOS

Agradeço primeiramente aos meus pais por estarem sempre ao meu lado me apoiando e pelo amor incondicional. Aos meus irmãos, pelo amor, carinho e amizade.

À minha avó, pessoa que é meu exemplo de vida, meu agradecimento especial. A todos os meus familiares, tios, tias, primos e primas, pelo incentivo e torcida.

Aos meus amigos, em especial: ao Fred por ser essa pessoa cheia de amor, pela amizade sincera e por estar comigo desde o início da minha jornada na Estatística. À Cristina e ao Pedro pelas risadas, carinho e por alegrar até os meus dias mais difíceis. À Alice pelo cuidado, pelos conselhos e por estar sempre disposta a ajudar.

Ao doutor Marcos Deon Vilela de Resende pela orientação, confiança e pelos ensinamentos.

Aos meus coorientadores Camila Ferreira Azevedo e Andrei Caíque Pires Nunes, pela disponibilidade e atenção.

Ao Guilherme Ferreira Simiqueli e aos professores Fabyano Fonseca e Silva e Luiz Alexandre Peternelli pelas contribuições.

À Elizabete Takahashi pelos dados cedidos.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, pela oportunidade concedida para realização do curso.

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

Muito obrigada a todos que de alguma forma contribuíram para a realização deste trabalho.

BIOGRAFIA

Taiana Lopes Rangel Miranda, filha de Tatiana Lopes de Moura Rangel e Paulo Sérgio Miranda, nasceu em Ouro Preto, Minas Gerais, em 13 de julho de 1995.

Em outubro de 2013, ingressou no curso de Bacharelado em Estatística na Universidade Federal de Ouro Preto, Ouro Preto-MG, graduando-se em julho de 2017.

Em agosto do mesmo ano, deu início ao curso de Mestrado no Programa de Pós-Graduação em Estatística Aplicada e Biometria na Universidade Federal de Viçosa, Viçosa-MG.

RESUMO

MIRANDA, Taiana Lopes Rangel, M.Sc., Universidade Federal de Viçosa, julho de 2019. **Avaliação de uma nova modelagem proporcional dos efeitos aditivos e de dominância e reflexos na genética quantitativa e na seleção genômica.** Orientador: Marcos Deon Vilela de Resende. Coorientadores: Andrei Caíque Pires Nunes e Camila Ferreira Azevedo.

A Seleção Genômica (*Genomic Wide Selection – GWS*) é uma abordagem muito utilizada nos programas de melhoramento e foi desenvolvida com o intuito de utilizar as informações diretas do DNA no processo de identificação de indivíduos geneticamente superiores, através da predição do valor genético genômico (*Genomic Estimated Breeding Value – GEBV*). A GWS baseia-se na análise de um grande número de marcadores moleculares SNPs (*Single Nucleotide Polymorphisms*) extensamente distribuídos no genoma. O modelo genético básico tradicionalmente utilizado na Genética Quantitativa e na GWS (modelo infinitesimal de Fisher, 1918), atribui à variância aditiva uma grande proporção de explicação da variância genética, mesmo sob ação gênica de dominância. Isto se deve ao fato de, no processo de derivação de expressões biométricas, a variância aditiva ser maximizada e a variância de dominância ser o resíduo da variação genética total. Recentemente um modelo genético alternativo foi proposto, no qual a variância de dominância é priorizada usando uma parametrização em que heterozigotos e um dos homozigotos são codificados com valores equivalentes. Nessa proposta o componente genético aditivo ao nível de locos é introduzido no modelo depois da variância de dominância ter sido maximizada. O objetivo desse trabalho foi avaliar essa nova parametrização dos efeitos aditivos e de dominância na seleção genômica e genética quantitativa em geral, e compará-la à parametrização tradicionalmente aplicada utilizando o método G-BLUP (*Genomic Best Linear Unbiased Predictor*). Adicionalmente essas comparações foram estendidas contemplando diferentes MAFs (*Minor Allele Frequency*). O procedimento de validação cruzada foi utilizado para avaliar as estimativas dos parâmetros usados nas comparações. Estimativas dos componentes da variação genética e das herdabilidades aditiva, devida à dominância e genotípica total, bem como da capacidade preditiva e do coeficiente da regressão entre o fenótipo e o GEBV foram obtidos, visando a comparação dos modelos. Dois índices combinando as estimativas dos componentes de variância obtidos pelos dois modelos foram propostos e avaliados. Os resultados indicaram que a nova modelagem, bem como a combinação de ambas são alternativas para melhorar as estimativas das variâncias genética aditiva e de dominância e da proporção entre elas.

ABSTRACT

MIRANDA, Taiana Lopes Rangel, M.Sc., Universidade Federal de Viçosa, July, 2019. **Evaluation of a new proportional modeling of additive and dominance effects in quantitative genetics and genomic selection.** Adviser: Marcos Deon Vilela de Resende. Co-advisers: Andrei Caíque Pires Nunes and Camila Ferreira Azevedo.

Genomic Wide Selection (GWS) is a widely used approach in breeding programs and was developed to use direct DNA information in the process of identifying genetically superior individuals by predicting Genomic Estimated Breeding Value (GEBV). GWS is based on the analysis of a large number of SNPs (Single Nucleotide Polymorphisms) widely distributed in the genome. The basic genetic model traditionally used in Quantitative Genetics and GWS (Fisher's infinitesimal model, 1918) attributes to additive variance a large proportion of the explanation of genetic variance, even under dominance gene action. This is because, in the process of derivation of biometric expressions, additive variance is maximized and dominance variance is the residue of total genetic variation. Recently an alternative genetic model has been proposed, in which the dominance variance is prioritized using a parameterization in which heterozygotes and one of the homozygotes are encoded with equivalent values. In this proposal the additive genetic component at the locus level is introduced in the model after the dominance variance has been maximized. The objective of this work was to evaluate this new parameterization of additive and dominance effects in genomic selection and quantitative genetics in general, and to compare it to the traditionally applied parameterization using the G-BLUP (*Genomic Best Linear Unbiased Predictor*) method. Additionally these comparisons were extended to include different MAFs (*Minor Allele Frequency*). The cross validation procedure was used to evaluate the estimates of the parameters used in the comparisons. Estimates of the components of genetic variation and additive, dominance and total genotypic heritability, as well as the predictive capacity and regression coefficient between phenotype and GEBV were obtained, aiming to compare the models. Two indices combining the variance component estimates obtained by the two models were proposed and evaluated. The results indicated that the new modeling as well as the combination of both are alternatives to improve the estimates of additive and dominance genetic variances and the proportion between them.

LISTA DE FIGURAS

Figura 1- Valores de variância para IMA em função da MAF.....	20
Figura 2- Valores de variância para densidade básica em função da MAF.....	20
Figura 3- Valores de variância para rendimento em função da MAF.....	21
Figura 4- Valores de herdabilidade para IMA em função da MAF.....	21
Figura 5- Valores de herdabilidade para densidade básica em função da MAF.....	22
Figura 6- Valores de herdabilidade para rendimento em função da MAF.....	22

LISTA DE TABELAS

Tabela 1 – Notações e definições de componentes de variância ao nível de locos neste trabalho.....	7
Tabela 2 – Descrição das variâncias aditiva e de dominância para as modelagens deste trabalho.....	8
Tabela 3 – Valores de herdabilidade, capacidade preditiva e coeficientes de regressão de acordo com as modelagens para os efeitos genéticos aditivas e de dominância para MAF 1%.....	11
Tabela 4 – Valores de variância para IMA de acordo com as duas modelagens para os efeitos genéticos aditivos e de dominância.....	13
Tabela 5 – Valores de variância para densidade básica de acordo com as duas modelagens para os efeitos genéticos aditivos e de dominância.....	14
Tabela 6 – Valores de variância para rendimento de acordo com as duas modelagens para os efeitos genéticos aditivos e de dominância.....	14
Tabela 7 – Valores de herdabilidade de acordo com as duas modelagens para os efeitos genéticos aditivos e de dominância.....	15

SUMÁRIO

1. INTRODUÇÃO GERAL.....	1
2. REVISÃO DE LITERATURA.....	2
2.1. Seleção Genômica Ampla.....	2
2.1.1. Definição e importância.....	2
2.2. Efeito genético aditivo e efeito genético devido à dominância.....	2
2.3. Método G-BLUP.....	3
2.4. Validação cruzada.....	5
3. MATERIAL E MÉTODOS.....	5
3.1. Descrição dos dados.....	5
3.2. Modelagens e parametrizações da matriz de incidência genotípica.....	6
3.2.1. Parametrização tradicional.....	6
3.2.2. Parametrização alternativa.....	6
3.3. Novas composições das matrizes de parentesco genômicas.....	7
3.4. Validação do modelo.....	8
3.5. Comparação das metodologias.....	9
3.6. Recursos Computacionais.....	10
4. RESULTADOS E DISCUSSÕES.....	10
5. CONCLUSÕES.....	16
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	17
Apêndice A: Gráficos com resultados das Tabelas 4, 5, 6 e 7.....	20
Apêndice B: Script elaborado no software R para implementação da nova modelagem.....	23

1. INTRODUÇÃO GERAL

Os avanços nas tecnologias de sequenciamento de DNA possibilitaram a geração de um grande número de dados genômicos. Essa evolução viabilizou o uso direto das informações oriundas do DNA na identificação de indivíduos geneticamente superiores através de marcadores moleculares, aumentando assim a eficiência do melhoramento genético. O SNP (*Single Nucleotide Polymorphism*) é um tipo de marcador molecular que se destaca na genômica atual. Assim, a Seleção Genômica Ampla (*Genome Wide Selection – GWS*), idealizada por Meuwissen et al. (2001), é baseada no uso de milhares de marcadores SNPs dispostos ao longo de todo o genoma, para prever o valor genético genômico dos indivíduos (*Genomic Estimated Breeding Value - GEBV*), que representa o valor predito de um modelo de regressão no qual o fenótipo é regredido nos marcadores. A GWS permite a identificação de indivíduos geneticamente superiores antes dos seus dados fenotípicos serem coletados, o que aumenta a eficiência e acelera o melhoramento genético.

O modelo genético básico tradicionalmente utilizado na Genética Quantitativa e na GWS (modelo infinitesimal de Fisher, 1918), atribui à variância aditiva uma grande proporção de explicação da variância genética, mesmo sob ação gênica de dominância. Isto se deve ao fato de, no processo de derivação de expressões biométricas, a variância aditiva ser maximizada e a variância de dominância ser o resíduo da variação genética total. Em um modelo genético alternativo (Huang e Mackay, 2016), a variância de dominância é priorizada usando uma parametrização em que heterozigotos e um dos homozigotos têm a mesma ênfase (código 2 para a dosagem alélica, em ambas as classes). O componente aditivo é introduzido no modelo depois da variância de dominância ter sido maximizada.

Introduzido por Habier (2007) e VanRaden (2008), o G-BLUP (*Genomic Best Linear Unbiased Prediction*) é um dos métodos utilizados na Seleção Genômica e tem se revelado apropriado para predição de valores genotípicos, assim como para estimação de efeitos aditivos e de dominância (Azevedo et al., 2015).

Este trabalho teve como objetivo analisar e avaliar as diferentes modelagens básicas para a Genética Quantitativa, propostas por Fisher (1918) e por Huang e Mackay (2016), quanto à eficiência na estimação de componentes de variância, herdabilidades e valores genéticos genômicos aditivos e devido à dominância por meio do método G-BLUP e de dados reais envolvendo a experimentação com Eucalipto.

2. REVISÃO DE LITERATURA

2.1. Seleção Genômica Ampla

2.1.1. Definição e importância

Uma das principais contribuições da genética molecular para o melhoramento genético é o uso de informações oriundas do DNA. Por meio das informações presentes nos marcadores moleculares é possível aumentar a eficiência na predição de valores genéticos e identificar indivíduos geneticamente superiores com mais rapidez e eficiência. A seleção genômica ampla (*Genome Wide Selection – GWS*), proposta por Meuwissen et al. (2001), é definida pelo uso de marcadores moleculares vastamente distribuídos no genoma, que capturam os genes que afetam o caráter quantitativo de interesse. Dentre os marcadores moleculares, aqueles que mais se destacam são os SNPs (*Single Nucleotide Polymorphisms*).

De acordo com Hayes et al. (2009), a seleção genômica se refere a decisões baseadas no valor genético genômico dos indivíduos (*Genomic Estimated Breeding Value – GEBV*). O GEBV é calculado por meio dos efeitos dos marcadores, onde esses marcadores estão em desequilíbrio de ligação (*Linkage Disequilibrium - LD*) com os locos de características quantitativas (*Quantitative Trait Loci - QTL*). Para obtenção do GEBV é necessário o uso de métodos estatísticos para estimação dos parâmetros. Contudo, devido ao grande número de marcadores que, na maior parte dos casos, é maior que o número de indivíduos e por esses marcadores estarem altamente correlacionados, surgem os problemas de alta dimensionalidade e multicolinearidade, o que inviabiliza a utilização de métodos tradicionais como o de mínimos quadrados. Em razão disso, surgiram alguns métodos para obtenção do GEBV. Neste trabalho será abordado o método G-BLUP.

2.2. Efeito genético aditivo e efeito genético devido à dominância

O valor fenotípico de um indivíduo pode ser particionado em componentes relativos ao genótipo e a parte relativa ao ambiente, ou seja, é a soma do valor genético e do valor ambiental. O valor genético pode ainda ser decomposto em valor genético aditivo e valor genético correspondente aos desvios de dominância.

A soma dos efeitos aditivos dos genes de um mesmo loco é definida como o valor genético aditivo (fração herdável). Já o efeito devido à dominância é determinado quando os

efeitos dos alelos não são puramente aditivos, ou seja, há a interação de forma que o valor genético do heterozigoto se distancia da média dos valores genéticos dos homozigotos. Por ser parte herdável, os efeitos aditivos são tradicionalmente de maior interesse na literatura, porém o estudo dos efeitos devido à dominância vem aumentando e mostrando cada vez mais a atuação desse efeito sobre o fenótipo (Varona et al., 2018; Muñoz et al., 2014; Nishio e Satoh, 2014; Wang e Da, 2014; Su et al., 2012), o que leva a análises mais acuradas na seleção genômica.

2.3. Método G-BLUP

O método G-BLUP (*Genomic Best Linear Unbiased Predictor*) faz a predição do valor genético genômico dos indivíduos por meio de um modelo linear misto. Os primeiros a utilizar esse método no âmbito da Seleção Genômica foram Habier (2007), VanRaden (2008), Goddard (2009) e Hayes et al. (2009).

“Se a matriz de parentesco G for computada via informação de marcadores e utilizada no método BLUP fenotípico tradicional, tem-se o método denominado G-BLUP ou BLUP genômico” (Resende et. al, 2012, p. 176).

O modelo aditivo é dado por:

$$y = Xb + Zu_a + \epsilon$$

em que:

y é o vetor de fenótipos ($N \times 1$, em que N é o número de indivíduos); b é o vetor de efeitos fixos ($p \times 1$, em que p é o número de efeitos fixos) com matriz de incidência X ($N \times p$); u_a é o vetor de efeitos genéticos aditivos dos indivíduos ($N \times 1$) com matriz de incidência Z ($N \times N$), assumindo que todos os indivíduos possuem observações; ϵ é o vetor de resíduos do modelo.

As estruturas de variâncias são dadas por $u_a \sim N(0, G_a \sigma_a^2)$ em que σ_a^2 é a variância aditiva do caráter e G_a ($N \times N$) é a matriz de parentesco genômica entre indivíduos para os efeitos aditivos e $\epsilon \sim N(0, I \sigma_e^2)$ em que σ_e^2 é a variância residual.

A matriz de parentesco genômica aditiva é dada por:

$$G_a = \frac{WW'}{\sum_{i=1}^n 2p_i q_i}$$

em que W contempla uma generalização para a codificação genotípica e p_i e q_i são frequências alélicas. As equações de modelos mistos para predição de u_a através do método G-BLUP, são mostradas a seguir:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G_a^{-1} \frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u}_a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

em que os componentes de variância são estimados via Método da Máxima Verossimilhança Restrita (*Restricted Maximum Likelihood* - REML). Os efeitos aditivos de marcadores podem ser estimados pela expressão $\hat{m}_a = (W'W)^{-1}W\hat{u}_a$.

A disponibilidade de estimativas de efeitos não-aditivos é interessante pois contribuem para um aumento na precisão da predição dos valores genéticos e da resposta genética (Varona et al., 2018). Modelos aditivo-dominantes para encontrar essas estimativas de efeitos não-aditivos já foram empregados no método G-BLUP (Da et al., 2014; Nishio e Satoh, 2014; Su et al., 2012).

O modelo linear misto para predição dos efeitos de indivíduos para o modelo aditivo-dominante, é descrito a seguir:

$$y = Xb + Zu_a + Zu_d + \epsilon$$

em que:

y , X , b e ϵ foram definidos no modelo aditivo; u_a e u_d são os vetores de efeitos genéticos aditivos e devido a dominância dos indivíduos ($N \times 1$) com matriz de incidência Z ($N \times N$), sendo as estruturas de variâncias dadas por $u_a \sim N(0, G_a \sigma_a^2)$ em que σ_a^2 é a variância aditiva do caráter e G_a ($N \times N$) é a matriz de parentesco genômica entre indivíduos para os efeitos aditivos e $u_d \sim N(0, G_d \sigma_d^2)$ em que σ_d^2 é a variância de dominância do caráter e G_d ($N \times N$) é a matriz de parentesco genômica entre indivíduos para os efeitos devido à dominância.

A matriz de parentesco genômica devido à dominância é dada por:

$$G_d = \frac{SS'}{\sum_{i=1}^n (2p_i q_i)^2}$$

em que S contempla o tipo de parametrização adotada para os efeitos de dominância.

Empregando as equações de modelos mistos é possível predizer u_a e u_d através do método G-BLUP aditivo-dominante, conforme a seguir:

$$\begin{bmatrix} X'X & X'Z & X'Z \\ Z'X & Z'Z + G_a^{-1} \frac{\sigma_e^2}{\sigma_a^2} & Z'Z \\ Z'X & Z'Z & Z'Z + G_d^{-1} \frac{\sigma_e^2}{\sigma_d^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u}_a \\ \hat{u}_d \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Z'y \end{bmatrix}$$

em que os componentes de variância são estimados via Método da Máxima Verossimilhança Restrita (*Restricted Maximum Likelihood* - REML). Sendo assim, o valor genômico predito

total para cada indivíduo é dado por:

$$\widehat{GEBV} = \hat{u}_a + \hat{u}_d.$$

2.4. Validação cruzada

De acordo com Resende et. al (2012), a população de interesse pode ser decomposta em população de estimação e população de validação. A população de estimação (treinamento) tem o fim de estimar os efeitos dos marcadores sobre a característica de interesse, e a população de validação tem o intuito de avaliar a eficiência da estimação desses efeitos e obter os valores genéticos genômicos. Essas duas populações são compostas por indivíduos genotipados e fenotipados. Feita a estimação e validação, estes serão utilizados na população de seleção, que é composta apenas de indivíduos genotipados.

De forma geral, a validação cruzada é um processo dividido em ciclos, onde são definidos k grupos e a cada ciclo esses grupos são designados como população de estimação ou população de treinamento, ou seja, um grupo é definido como população de validação e os demais grupos como população de estimação. Esse procedimento é repetido até que todos os grupos sejam utilizados como população de validação. Assim, a cada etapa são obtidas as medidas de eficiência e com o término do processo pode-se obter a média e o desvio-padrão dos k valores. Esse procedimento tem como objetivo validar as estimativas obtidas pelo modelo estatístico.

3. MATERIAIS E MÉTODOS

3.1. Descrição dos dados

Os dados utilizados nesse estudo são provenientes da empresa Cenibra. Foram avaliadas três características do eucalipto: incremento médio anual (IMA) em volume de madeira, densidade básica da madeira e rendimento de celulose. Informações mais detalhadas sobre os dados podem ser encontradas em Resende et al. (2016) e Resende et al. (2017).

Os 47069 marcadores SNPs passaram pelo processo de controle de qualidade, onde aqueles marcadores cuja call rate (taxa de atendimento na genotipagem) foram menor que 95% e MAF (*Minor Allele Frequency*) menor que 1% foram retirados da análise, posto que são pouco informativos. Após o controle de qualidade do banco de dados genômico foi totalizado 23129

marcadores SNPs. Foram avaliados 756 fenótipos referentes a IMA, 752 referentes a densidade básica e 749 para rendimento.

3.2. Modelagens e parametrizações da matriz de incidência genotípica

3.2.1. Parametrização Tradicional

Existem diversas formas de parametrizações para a matriz de incidência genotípica (Varona et al., 2018; Da et al., 2014; Vitezica et al., 2013). No entanto, a parametrização descrita a seguir é a mais tradicional.

A matriz de incidência W , para efeitos aditivos, recebe os valores 0, 1 e 2, para o número de alelos do marcador em um indivíduo diplóide. Ou seja,

$$W = \begin{cases} 2, & \text{genótipo} = AA \\ 1, & \text{genótipo} = Aa \\ 0, & \text{genótipo} = aa \end{cases}$$

Para a matriz de incidência devido a dominância S , têm-se a seguinte parametrização:

$$S = \begin{cases} 2(p - q), & \text{genótipo} = AA \\ 2p, & \text{genótipo} = Aa \\ 0, & \text{genótipo} = aa \end{cases}$$

3.2.2. Parametrização Alternativa

Na parametrização proposta por Huang e Mackay (2016), a matriz de incidência aditiva W é dada por:

$$W = \begin{cases} \frac{-2q}{1+q}, & \text{genótipo} = AA \\ \frac{1-q}{1+q}, & \text{genótipo} = Aa \\ 0, & \text{genótipo} = aa \end{cases}$$

Para a matriz de incidência S , o heterozigoto e o homozigoto dominante são codificados com o valor 2, enquanto o homozigoto recessivo recebe 0. Assim, S é definida como:

$$S = \begin{cases} 2, \text{genótipo} = AA \\ 2, \text{genótipo} = Aa \\ 0, \text{genótipo} = aa \end{cases}$$

3.3. Novas composições das matrizes de parentesco genômicas

A nova abordagem proposta utiliza o método G-BLUP com uma parametrização alternativa da matriz de incidência genotípica (Huang e Mackay, 2016) e uma nova forma de calcular as matrizes de parentesco genômica. Sendo assim, o modelo linear misto para predição dos efeitos de indivíduos para a nova modelagem é descrito a seguir:

$$y = Xb + Zu_a + Zu_d + \epsilon$$

A matriz de parentesco genômica aditiva e devido à dominância são dadas por:

$$G_a = \frac{WW'}{\sum_{i=1}^n \frac{2p_i^2 q_i}{1+q}} \text{ e } G_d = \frac{SS'}{\sum_{i=1}^n \frac{4p_i q_i^2}{1+q}}$$

em que W e S contemplam a parametrização proposta por Huang e Mackay (2016) para os efeitos aditivos e de dominância. Os denominadores das matrizes de parentesco genômica foram determinados com base nos coeficientes dos componentes de variância obtidos com a codificação alternativa.

Tabela 1 – Notações e definições de componentes de variância ao nível de locos neste trabalho.

Notação	Componente de variância	Codificação do genótipos (aa,Aa,AA)
σ_a^2	$2pq[a + d(q - p)]^2$	0,1,2
σ_d^2	$(2pqd)^2$	0,2p, 2(p - q)
$\sigma'_a{}^2$	$\frac{2p^2 q}{1+q} (a - d)^2$	$0, \frac{1-q}{1+q}, \frac{-2q}{1+q}$
$\sigma'_d{}^2$	$\frac{4pq^2}{1+q} (a + dq)^2$	0,2,2

a e d: valores genotípicos do homocigoto e do heterocigoto, respectivamente; p e q: frequências dos alelos A e a, respectivamente; σ_a^2 e σ_d^2 : variância aditiva e devido à dominância, respectivamente, pela modelagem tradicional; $\sigma'_a{}^2$ e $\sigma'_d{}^2$: variância aditiva e devido à dominância, respectivamente, pela modelagem alternativa.

Os efeitos genéticos u_a e u_d são preditos por meio das equações de modelos mistos descritas a seguir:

$$\begin{bmatrix} X'X & X'Z & X'Z \\ Z'X & Z'Z + G_a^{-1} \frac{\sigma_e^2}{\sigma_a^2} & Z'Z \\ Z'X & Z'Z & Z'Z + G_d^{-1} \frac{\sigma_e^2}{\sigma_d^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u}_a \\ \hat{u}_d \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Z'y \end{bmatrix}$$

em que os componentes de variância são estimados via REML (*Restricted Maximum Likelihood*).

Com base nos resultados obtidos nas duas modelagens (tradicional e alternativa), foram calculadas outras duas estimativas para as variâncias: corrigida 1, que é a média das estimativas obtidas pela modelagem tradicional e pela modelagem alternativa; corrigida 2, que utiliza uma ponderação pelos desvios-padrão (medida obtida das várias amostras de validação) das estimativas obtidas pelas modelagens tradicional e alternativa. A Tabela 2 apresenta um resumo dos quatro estimadores utilizados nesse estudo.

Tabela 2 – Descrição das variâncias aditiva e de dominância para as modelagens deste trabalho.

Modelagem	Variância aditiva	Variância de dominância
Tradicional	σ_a^2	σ_d^2
Alternativa	$\sigma_a'^2$	$\sigma_d'^2$
Corrigida 1	$\frac{\sigma_a^2 + \sigma_a'^2}{2}$	$\frac{\sigma_d^2 + \sigma_d'^2}{2}$
Corrigida 2	$\frac{\sigma_a^2}{s_a} s_{a_{medio}} + \frac{\sigma_a'^2}{s_a'} s_{a_{medio}}$	$\frac{\sigma_d^2}{s_d} s_{d_{medio}} + \frac{\sigma_d'^2}{s_d'} s_{d_{medio}}$

s_a e s_d : desvios padrão dos valores de variância aditiva (σ_a^2) e devido à dominância (σ_d^2) obtidos nos grupos de validação pela modelagem tradicional; s_a' e s_d' : desvios padrão dos valores de variância aditiva ($\sigma_a'^2$) e devido à dominância ($\sigma_d'^2$) obtidos nos grupos de validação pela modelagem alternativa; $s_{a_{medio}}$: média entre s_a e s_a' ; $s_{d_{medio}}$: média entre s_d e s_d' .

3.4. Validação do modelo

O processo de validação cruzada consistiu em dividir uma população em k grupos. Aqueles indivíduos que pertenciam aos k-1 grupos foram usados como população de estimação e o grupo restante foi usado como população de validação. O modelo estatístico foi ajustado na população de estimação com o intuito de se obter as estimativas dos efeitos dos parâmetros contidos no modelo. Posteriormente os valores genéticos genômicos (GEBV) foram calculados na população de validação, utilizando os efeitos já obtidos na população de estimação. Esse

processo é repetido até que todos os grupos sejam a população de validação uma vez. No caso, como o número de indivíduos era diferente para cada característica de eucalipto, o número de grupos também foi diferente. Para IMA tomou-se $k=12$ (63 indivíduos em cada grupo), para densidade básica tomou-se $k=16$ (47 indivíduos em cada grupo) e para rendimento tomou-se $k=7$ (107 indivíduos em cada grupo). Após a validação cruzada, foram obtidas as medidas de capacidade preditiva, herdabilidades e coeficientes de regressão em cada uma das populações de validação e em seguida foram feitas as médias dessas medidas.

3.5. Comparação das metodologias

As duas modelagens foram avaliadas através da predição dos valores genéticos genômicos (GEBV) aditivos e devido a dominância obtidos pelo método G-BLUP. A partir disso, foram calculadas as medidas de eficiência para a predição genômica, sendo elas: capacidade preditiva (CP_a, CP_d, CP_g), coeficiente de regressão ($b_{y\hat{a}}, b_{y\hat{d}}, b_{y\hat{g}}$) e herdabilidades genômicas aditiva, de dominância e genotípica (h_a^2, h_d^2 e h_g^2), das estimativas de cada uma das características, que foram usadas para comparar as duas metodologias.

As medidas de eficiência que foram empregadas são detalhadas em seguida: a capacidade preditiva genotípica (CP_g) é calculada através da correlação entre as estimativas dos efeitos genotípicos totais e os fenótipos; a capacidade preditiva aditiva (CP_a) e de dominância (CP_d) são calculadas através da correlação entre os fenótipos e as estimativas dos efeitos aditivos e devido a dominância; o coeficiente da regressão entre o fenótipo e o GEBV é definido como $\hat{b}_{y\hat{g}} = \frac{cov(y,\hat{g})}{var(\hat{g})}$; os coeficientes da regressão para a parte aditiva e de dominância são definidos como $b_{y\hat{a}} = \frac{cov(y,\hat{a})}{var(\hat{a})}$ e $b_{y\hat{d}} = \frac{cov(y,\hat{d})}{var(\hat{d})}$, respectivamente; coeficientes de regressão abaixo de 1 indicam que os valores preditos são superestimados e, acima de 1, indicam que os valores preditos são subestimados.

A herdabilidade genotípica total (sentido amplo) é um parâmetro que representa a proporção da variabilidade fenotípica (observada) que é capaz de ser explicada pela regressão linear em um grupo de marcadores. Essa fração da variabilidade é dada por $h_g^2 = \frac{\sigma_a^2 + \sigma_d^2}{\sigma_a^2 + \sigma_d^2 + \sigma_e^2}$, onde σ_a^2 é a variância genética aditiva, σ_d^2 é a variância genética devido à dominância e σ_e^2 é a variância residual. A herdabilidade aditiva (sentido restrito) e a herdabilidade devido à

dominância expressam a proporção da variância fenotípica que é explicada pelos efeitos aditivos e pelos efeitos devido à dominância, respectivamente. A herdabilidade aditiva é definida como $h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_d^2 + \sigma_e^2}$ e a herdabilidade devido à dominância é dada por $h_d^2 = \frac{\sigma_d^2}{\sigma_a^2 + \sigma_d^2 + \sigma_e^2}$.

3.6. Recursos Computacionais

Todas as análises foram executadas no software R (R Core Team, 2019), visto que se trata de um software livre e por suportar a utilização de amplos conjuntos de dados, assim como os arquivos de marcadores SNPs. O pacote “sommer” (Covarrubias-Pazarán, 2016) na versão 3.4 e a função “mmer” foram usados na implementação do método G-BLUP.

4. RESULTADOS E DISCUSSÃO

Os valores das herdabilidades, capacidade preditiva e coeficientes da regressão para as características de eucalipto IMA, Densidade básica e Rendimento, de acordo com a modelagem usual e a modelagem alternativa, são apresentados na Tabela 3. Além disso, os resultados dos valores de variâncias são mostrados na Tabela 4, Tabela 5 e Tabela 6. Essas medidas foram obtidas em cada grupo da população de validação no procedimento de validação cruzada e, em seguida, obteve-se uma média.

Dentre as características, a que apresentou maior estimativa de herdabilidade aditiva, quando utilizado o modelo usual (Tabela 3), foi a densidade básica. Esse resultado é invertido quando a modelagem alternativa é aplicada, ou seja, a herdabilidade aditiva se torna menor. Os valores de herdabilidade devido à dominância com a modelagem alternativa aumentaram consideravelmente quando comparadas com a modelagem tradicional, conforme esperado teoricamente.

Os valores de capacidade preditiva genotípica total foram satisfatórios para todos os casos (modelagens) e para todas as características, variando de 0.54 a 0.62, sendo esses valores pouco mais altos para a modelagem tradicional. O mesmo acontece para o coeficiente de regressão, que varia entre 0.97 a 1.06, indicando que as predições são efetivas em prever as verdadeiras amplitudes das diferenças entre os indivíduos em análise (Resende et. al, 2012).

Tabela 3 – Valores de herdabilidade, capacidade preditiva e coeficientes de regressão de acordo com as modelagens para os efeitos genéticos aditivos e de dominância para MAF 1%.

Caract.	Modelagem	h_a^2	h_d^2	h_g^2	CP_g	CP_a	CP_d	$b_{y\hat{a}}$	$b_{y\hat{d}}$	$b_{y\hat{g}}$
IMA	Tradicional	0,47	0,08	0,55	0,58	0,56	0,04	0,92	0,38	0,97
	Alternativa	0,28	0,13	0,41	0,57	0,56	0,57	2,52	1,84	0,97
	Corrigida 1	0,38	0,11	0,48	0,58	0,56	0,31	1,72	1,11	0,97
	Corrigida 2	0,48 ($\sigma_{a_{c2}}^2$)	0,11 ($\sigma_{d_{c2}}^2$)	0,59 ($\sigma_{g_{c2}}^2$)	-	-	-	-	-	-
Densidade básica	Tradicional	0,56	0,00	0,56	0,62	0,62	-	1,00	-	1,00
	Alternativa	0,00	0,31	0,31	0,58	-	0,58	-	0,99	0,99
	Corrigida 1	0,28	0,16	0,44	0,60	0,62	0,58	1,00	0,99	1,00
	Corrigida 2	0,22 ($\sigma_{a_{c2}}^2$)	0,09 ($\sigma_{d_{c2}}^2$)	0,31 ($\sigma_{g_{c2}}^2$)	-	-	-	-	-	-
Rendimento	Tradicional	0,42	0,03	0,45	0,55	0,54	0,18	1,07	3,86	1,05
	Alternativa	0,00	0,24	0,25	0,54	-	0,54	-	1,06	1,06
	Corrigida 1	0,21	0,14	0,35	0,55	0,54	0,36	1,07	2,46	1,00
	Corrigida 2	0,16 ($\sigma_{a_{c2}}^2$)	0,13 ($\sigma_{d_{c2}}^2$)	0,29 ($\sigma_{g_{c2}}^2$)	-	-	-	-	-	-

h_a^2 : herdabilidade aditiva; h_d^2 : herdabilidade devido à dominância; h_g^2 : herdabilidade genotípica total; CP_g : capacidade preditiva genotípica; CP_a : capacidade preditiva aditiva; CP_d : capacidade preditiva de dominância; $b_{y\hat{a}}$: coeficiente da regressão para a parte aditiva; $b_{y\hat{d}}$: coeficiente da regressão para a parte de dominância; $b_{y\hat{g}}$: coeficiente da regressão entre o fenótipo e o GEBV; $\sigma_{a_{c2}}^2, \sigma_{d_{c2}}^2$ e $\sigma_{g_{c2}}^2$: variância aditiva, devido à dominância e genética total, respectivamente, para a modelagem corrigida 2.

Os resultados da Tabela 3 indicam que a nova modelagem se mostrou eficiente. No entanto, para o caso em que não existe variação de dominância, essa modelagem leva a variância aditiva a valores muito baixos. É importante ressaltar que a nova modelagem somente é adequada quando existe algum grau de dominância atuando no controle genético do caráter. Isto ocorre para a variável IMA, mas não ocorre para densidade da madeira e rendimento de celulose, os quais não apresentam dominância alélica (Fonseca et al. 2010; Rezende et al. 2014). Isto explica os valores zero para variância aditiva obtidos na Tabela 3, para densidade da madeira e rendimento de celulose. Portanto, é essencial verificar através da literatura ou por análises prévias dos dados, se o caráter em questão apresenta qualquer grau de dominância alélica. Em populações naturais, se uma dessas modelagens é mais eficaz do que outra depende dos verdadeiros modelos genéticos.

A modelagem corrigida 1 utiliza valor médio e obteve bons resultados para todos os caracteres, indicando ser uma boa opção.

A parametrização tradicional para os efeitos aditivos e de dominância conduz ao fato de a variância aditiva (σ_a^2) explicar a maioria da variação genética mesmo sob ação gênica de dominância. Nesse modelo, efeitos aditivos e de dominância, ambos contribuem para a variação aditiva (Huang e Mackay, 2016). Assim, a variância de dominância não é adequada para inferir sobre a respectiva ação gênica de dominância. Isto ocorre porque a variância aditiva é maximizada inicialmente e a variância de dominância é o resíduo da variação genética total. A variância aditiva (σ_a^2) é a soma de quadrados Tipo 1 da regressão dos valores genotípicos no número de cópias dos alelos, enquanto σ_d^2 é a variância residual. Assim, ao componente aditivo é dada prioridade para explicar a variação genética. Se a prioridade for dada ao componente de dominância usando uma parametrização na qual heterozigotos (Aa) e homozigotos (AA) são codificados identicamente, tem-se a definição de uma variância de dominância alternativa, $\sigma'_d{}^2$ (Huang e Mackay, 2016). Nesse caso, à ação gênica de dominância é permitido explicar a variância primeiro, enquanto o componente aditivo ($\sigma'_a{}^2$) somente entra no modelo após $\sigma'_d{}^2$ ter sido maximizada. Assim, $\sigma'_a{}^2$ passa a ser a variância genética residual. Portanto, conforme Resende (2015), uma maneira de verificar a importância relativa de efeitos aditivos e de dominância é comparar $(\sigma_a^2 + \sigma'_d{}^2)/2$ com $(\sigma'_a{}^2 + \sigma_d^2)/2$, sendo que o maior deles indicará qual ação gênica é mais importante. O modelo para $\sigma_a^2 + \sigma_d^2$ visa maximizar σ_a^2 e minimizar σ_d^2 . O modelo para $\sigma'_a{}^2 + \sigma'_d{}^2$ visa maximizar $\sigma'_d{}^2$ e minimizar $\sigma'_a{}^2$.

As estimativas das herdabilidades para a modelagem corrigida 2 somente são válidas para o caráter IMA, o qual apresenta dominância, situação essa em que o modelo aditivo-dominante se adequa. Essas herdabilidades são as mais adequadas estatisticamente e revelam que a inferência prática para h_a^2 , h_d^2 e h_g^2 devem ser 0.48, 0.11 e 0.59, respectivamente. Pela abordagem tradicional essas inferências seriam 0.47, 0.08 e 0.55, respectivamente. Assim, a abordagem conjunta (tradicional e alternativa) permite capturar melhor a variância e herdabilidade de dominância.

As Tabelas 4, 5 e 6, em função da MAF, revelam um aumento da variância devido a dominância, quando comparados os resultados da modelagem tradicional (σ_d^2) com a modelagem alternativa ($\sigma'_d{}^2$).

Tabela 4 – Valores de variância, em função da MAF, para IMA de acordo com as duas modelagens para os efeitos genéticos aditivos e de dominância.

MAF	m*	Modelagem tradicional maximizando σ_a^2				Modelagem alternativa maximizando $\sigma'_a{}^2$			
		σ_a^2	σ_d^2	σ_e^2	σ_f^2	$\sigma'_a{}^2$	$\sigma'_d{}^2$	σ_e^2	σ_f^2
1%	23129	34,66	6,12	32,95	73,73	15,91	7,08	32,79	55,77
5%	18240	35,15	6,15	33,50	74,80	10,43	9,82	33,58	53,84
10%	16778	37,29	5,91	33,73	76,93	5,41	13,03	34,00	52,44
15%	15075	37,82	5,71	34,02	77,55	4,34	13,57	34,37	52,28
20%	13698	41,55	5,36	33,99	80,90	2,32	15,83	34,36	52,50
25%	12486	42,65	5,19	34,18	82,02	1,28	16,54	34,64	52,45
30%	11114	45,09	4,84	34,19	84,12	1,13	17,38	34,56	53,07
35%	9485	47,98	4,18	34,88	87,03	0,46	18,36	35,21	54,04
40%	7619	49,89	3,63	35,66	89,18	0,00	19,30	35,78	55,08
45%	4691	61,60	2,39	37,03	101,02	0,00	22,37	36,97	59,34

m*: número de marcadores retidos na análise; σ_a^2 e σ_d^2 : variância aditiva e devido à dominância, respectivamente, pela modelagem tradicional; $\sigma'_a{}^2$ e $\sigma'_d{}^2$: variância aditiva e devido à dominância, respectivamente, pela modelagem alternativa; σ_e^2 : variância devido à fatores ambientais; σ_f^2 : variância fenotípica.

A inferência da estrutura da população pode ser fortemente influenciada pela escolha do limiar da MAF (Linck e Battey, 2019). Com o aumento da MAF mais marcadores são eliminados e, com isso, nota-se que os valores da variância aditiva (σ_a^2) crescem enquanto os valores de variância devido à dominância (σ_d^2) diminuem, para a modelagem tradicional. O contrário acontece no modelo alternativo: quanto maior a MAF, maiores os valores de $\sigma'_a{}^2$ e menores os valores de $\sigma'_d{}^2$. Esses resultados podem ser observados para todas as variáveis (Tabelas 4, 5 e 6 e Figuras 1, 2 e 3). Esse resultado é esperado teoricamente, visto que a variância de dominância cresce em relação a variância aditiva, à medida que a frequência alélica na população cresce (Resende, 2015). Assim, a dominância tende a ser mais importante para o melhoramento e seleção em populações mais melhoradas, as quais apresentam maiores frequências dos alelos favoráveis.

Tabela 5 – Valores de variância, em função da MAF, para Densidade básica de acordo com as duas modelagens para os efeitos genéticos aditivos e de dominância.

MAF	m*	Modelagem tradicional maximizando σ_a^2				Modelagem alternativa maximizando $\sigma'_a{}^2$			
		σ_a^2	σ_d^2	σ_e^2	σ_f^2	$\sigma'_a{}^2$	$\sigma'_d{}^2$	σ_e^2	σ_f^2
1%	23129	545,96	0,20	427,87	974,04	0,00	187,56	422,52	610,08
5%	18228	557,07	0,18	430,33	987,58	0,00	190,41	425,75	616,16
10%	16777	561,34	0,16	431,42	992,92	0,00	190,83	428,27	619,10
15%	15073	567,88	0,07	433,70	1001,65	0,00	192,81	431,88	624,69
20%	13709	575,57	0,10	435,06	1010,73	0,00	193,43	435,53	628,96
25%	12485	587,37	0,10	434,44	1021,91	0,00	195,88	436,37	632,25
30%	11107	586,06	0,64	439,23	1025,93	0,00	202,40	435,69	638,08
35%	9490	586,65	0,37	445,54	1032,55	0,00	204,48	442,19	646,67
40%	7615	608,97	0,60	450,13	1059,69	0,00	213,04	446,43	659,47
45%	4696	629,05	1,40	465,96	1096,41	0,00	215,50	468,40	683,90

m*: número de marcadores retidos na análise; σ_a^2 e σ_d^2 : variância aditiva e devido à dominância, respectivamente, pela modelagem tradicional; $\sigma'_a{}^2$ e $\sigma'_d{}^2$: variância aditiva e devido à dominância, respectivamente, pela modelagem alternativa; σ_e^2 : variância devido à fatores ambientais; σ_f^2 : variância fenotípica.

Tabela 6 – Valores de variância, em função da MAF, para Rendimento de celulose, de acordo com as duas modelagens para os efeitos genéticos aditivos e de dominância.

MAF	m*	Modelagem tradicional maximizando σ_a^2				Modelagem alternativa maximizando $\sigma'_a{}^2$			
		σ_a^2	σ_d^2	σ_e^2	σ_f^2	$\sigma'_a{}^2$	$\sigma'_d{}^2$	σ_e^2	σ_f^2
1%	23129	1,21	0,08	1,56	2,85	0,00	0,49	1,51	2,00
5%	18264	1,22	0,09	1,56	2,87	0,00	0,50	1,51	2,01
10%	16781	1,25	0,09	1,56	2,90	0,00	0,51	1,51	2,02
15%	15077	1,31	0,08	1,57	2,95	0,00	0,52	1,53	2,05
20%	13698	1,35	0,08	1,57	3,00	0,00	0,54	1,53	2,07
25%	12478	1,40	0,07	1,58	3,05	0,00	0,55	1,55	2,09
30%	11109	1,45	0,08	1,58	3,12	0,00	0,57	1,55	2,12
35%	9490	1,68	0,08	1,60	3,36	0,00	0,64	1,58	2,22
40%	7611	1,80	0,09	1,63	3,51	0,00	0,64	1,65	2,28
45%	4648	2,05	0,09	1,71	3,84	0,00	0,68	1,74	2,42

m*: número de marcadores retidos na análise; σ_a^2 e σ_d^2 : variância aditiva e devido à dominância, respectivamente, pela modelagem tradicional; $\sigma'_a{}^2$ e $\sigma'_d{}^2$: variância aditiva e devido à dominância, respectivamente, pela modelagem alternativa; σ_e^2 : variância devido à fatores ambientais; σ_f^2 : variância fenotípica.

Na Tabela 5 observa-se grandes valores de variância aditiva e pequenos valores de variância devido à dominância na modelagem em que σ_a^2 é maximizada. Quando analisada a modelagem em que $\sigma'_a{}^2$ é maximizada, percebe-se que os valores da variância aditiva

(σ'_d) foram levados à zero. Esse mesmo comportamento é percebido na Tabela 6. Esses resultados levam à conclusão de que o novo modelo não parece se adequar aos casos em que a variação de dominância não é relevante, o que já era previsto teoricamente.

Os resultados obtidos para os valores de herdabilidade são mostrados na Tabela 7. O mesmo padrão que os encontrados nas tabelas anteriores (Tabela 4, Tabela 5 e Tabela 6) ocorre quando analisados os valores de herdabilidade aditiva e devido à dominância. Quando σ'_d é maximizada (modelagem alternativa), há um aumento da herdabilidade devido a dominância e diminuição da herdabilidade aditiva, quando comparados com os resultados da modelagem tradicional.

Tabela 7 – Valores de herdabilidade, em função da MAF, de acordo com as duas modelagens para os efeitos genéticos aditivos e de dominância.

Caract.	Paramet.	Herdab.	MAF									
			1%	5%	10%	15%	20%	25%	30%	35%	40%	45%
IMA	Tradicional	h_a^2	0,47	0,47	0,48	0,49	0,51	0,52	0,54	0,55	0,56	0,61
		h_d^2	0,08	0,08	0,08	0,07	0,07	0,06	0,06	0,05	0,04	0,02
		h_g^2	0,55	0,55	0,56	0,56	0,58	0,58	0,59	0,60	0,60	0,63
	Alternativa	h_a^2	0,28	0,19	0,10	0,08	0,04	0,02	0,02	0,01	0,00	0,00
		h_d^2	0,13	0,18	0,25	0,26	0,30	0,32	0,33	0,34	0,35	0,38
		h_g^2	0,41	0,38	0,35	0,34	0,35	0,34	0,35	0,35	0,35	0,38
Densidade básica	Tradicional	h_a^2	0,56	0,56	0,57	0,57	0,57	0,57	0,57	0,57	0,57	0,57
		h_d^2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
		h_g^2	0,56	0,56	0,57	0,57	0,57	0,57	0,57	0,57	0,58	0,58
	Alternativa	h_a^2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
		h_d^2	0,31	0,31	0,31	0,31	0,31	0,31	0,32	0,32	0,32	0,32
		h_g^2	0,31	0,31	0,31	0,31	0,31	0,31	0,32	0,32	0,32	0,32
Rend.	Tradicional	h_a^2	0,42	0,43	0,43	0,44	0,45	0,46	0,46	0,50	0,51	0,53
		h_d^2	0,03	0,03	0,03	0,03	0,03	0,02	0,03	0,02	0,02	0,02
		h_g^2	0,45	0,46	0,46	0,47	0,48	0,48	0,49	0,52	0,54	0,55
	Alternativa	h_a^2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
		h_d^2	0,24	0,25	0,25	0,25	0,26	0,26	0,27	0,29	0,28	0,28
		h_g^2	0,25	0,25	0,25	0,25	0,26	0,26	0,27	0,29	0,28	0,28

h_a^2 : herdabilidade aditiva; h_d^2 : herdabilidade devido à dominância; h_g^2 : herdabilidade genotípica total.

Por fim, é possível afirmar que a modelagem alternativa avaliada se mostrou eficiente para a estimação dos componentes de variância genética e predição dos valores genéticos genômicos, no caso em que existe dominância atuando no caráter. Após 100 anos de modelo

infinitesimal parece surgir agora alternativas de modelagem de caracteres quantitativos (Visscher e Goddard, 2019).

5. CONCLUSÕES

A modelagem alternativa avaliada apresentou resultados interessantes e se mostrou eficiente para a estimação dos componentes de variância genética e predição dos valores genéticos genômicos. A nova abordagem conduziu a maiores estimativas da variância de dominância em relação à abordagem tradicional, embora a estimativa da variância genética aditiva tenha se mostrado superior à de dominância para as três características avaliadas em eucalipto. O novo modelo mostrou-se não adequado para o caso em que não existe variação de dominância, conforme esperado teoricamente. Portanto, esse novo modelo é adequado para o caso em que ambos os efeitos, aditivo e de dominância, são importantes no controle genético do caractere em questão.

A escolha do modelo a ser adotado na prática deve se basear conjuntamente no conhecimento prático (expressão de heterose para os componentes de médias) que se tem sobre os caracteres e nas estimativas propiciadas pelos dois modelos e pelos dois índices avaliados, baseados em pesos dados as estimativas geradas pelos dois modelos. O uso do valor médio dos componentes de variância é uma boa opção pois combina duas fontes de informações teoricamente não correlacionadas. A modelagem alternativa corrigida 2 utiliza uma ponderação pelos desvios-padrão (precisão) das estimativas obtidas pelas modelagens tradicional e alternativa e é a mais adequada estatisticamente.

6. REFERÊNCIAS BIBLIOGRÁFICAS

AZEVEDO, C. F.; RESENDE, M. D. V.; SILVA, F. F.; VIANA, J. M. S.; VALENTE, M. S. F.; RESENDE JUNIOR, M. F. R.; MUÑOZ, P. Ridge, Lasso and Bayesian additive-dominance genomic models. **BMC Genetics**, v. 16, p. 105, 2015.

COVARRUBIAS-PAZARAN, G. Genome-assisted prediction of quantitative traits using the R package sommer. **PloS one**, v. 11, n. 6, p. e0156744, 2016.

DA, Y. et al. Mixed model methods for genomic prediction and variance component estimation of additive and dominance effects using SNP markers. **PloS one**, v. 9, n. 1, p. e87666, 2014.

FISHER, R. A. The correlation between relatives on the supposition of mendelian inheritance. **Trans. Earth and Environmental Science Transactions of the Royal Society of Edinburgh**. v. 52, n. 2, p. 399-433, 1918.

FONSECA, S. M.; RESENDE, M. D. V.; ALFENAS, A. C.; GUIMARAES, L. M. S.; ASSIS, T. F. GRATTAPAGLIA, D. **Manual Prático de Melhoramento Genético do Eucalipto**. 1. ed. Viçosa: Editora UFV, v. 1, 200p, 2010.

GODDARD, M. Genomic selection: prediction of accuracy and maximisation of long term response. **Genetica**, v. 136, n. 2, p. 245-257, 2009.

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v. 177, n. 4, p. 2389-2397, 2007.

HAYES, B. J et al. Invited review: Genomic selection in dairy cattle: Progress and challenges. **Journal of dairy science**, v. 92, n. 2, p. 433-443, 2009.

HUANG, W; MACKAY, T. F. C. The Genetic Architecture of Quantitative Traits Cannot Be Inferred from Variance Component Analysis. **PLOS Genetics**, v. 12, p. e1006421, 2016.

LINCK, E; BATTEY, C.J. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. **Molecular ecology resources**, v. 19, n. 3, p. 639-647, 2019.

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

MUÑOZ, P. R.; RESENDE, M. F.; GEZAN, S. A.; RESENDE, M. D. V.; DE LOS CAMPOS, G.; KIRST, M.; PETER, G. F. Unraveling additive from nonadditive effects using genomic relationship matrices. **Genetics**, v. 198, n. 4, p. 1759-1768, 2014.

NISHIO, M.; SATOH, M. Including Dominance Effects in the Genomic BLUP Method for Genomic Evaluation. **PLOS ONE**, v. 9, p. e85792, 2014.

RESENDE, M. D. V.; SILVA, F. F.; LOPES, P. S.; AZEVEDO, C. F. **Seleção genômica ampla (GWS) via modelos mistos (REML/BLUP), inferência bayesiana (MCMC), regressão aleatória multivariada (RRM) e estatística espacial**. Viçosa: UFV, 291p, 2012.

RESENDE, M. D. V. **Genética Quantitativa e de Populações**. 1. ed. Visconde do Rio Branco: Suprema, v. 1, 422p, 2015.

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**, v. 213, n. 3, 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 individuals and families in Eucalyptus breeding with an additive-dominant model. **Heredity**, v. 119, n. 4, p. 245, 2017.

REZENDE, G. D. S.; DE RESENDE, M. D. V.; DE ASSIS, T.F. **Eucalyptus Breeding for Clonal Forestry**. Forestry Sciences. 1ed: Springer Netherlands, p. 393-424, 2014.

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**, v.7, p.e45293, 2012.

VANRADEN, P. M. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science**, v. 91, p. 4414-4423, 2008.

VARONA, L.; LEGARRA, A.; TORO, M. A.; VITEZICA, Z. G. Non-additive Effects in Genomic Selection. **Frontiers in Genetics**, v. 9, p. 78, 2018.

VISSCHER, P. M.; GODDARD, M. E. From RA Fisher's 1918 Paper to GWAS a Century Later. **Genetics**, v. 211, n. 4, p. 1125-1130, 2019.

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

WANG, C.; DA, Y. Quantitative genetics model as the unifying model for defining genomic relationship and inbreeding coefficient. **PLoS One**, v. 9, n. 12, p. e114484, 2014.

Apêndice A: Gráficos com resultados das Tabelas 4, 5, 6 e 7.

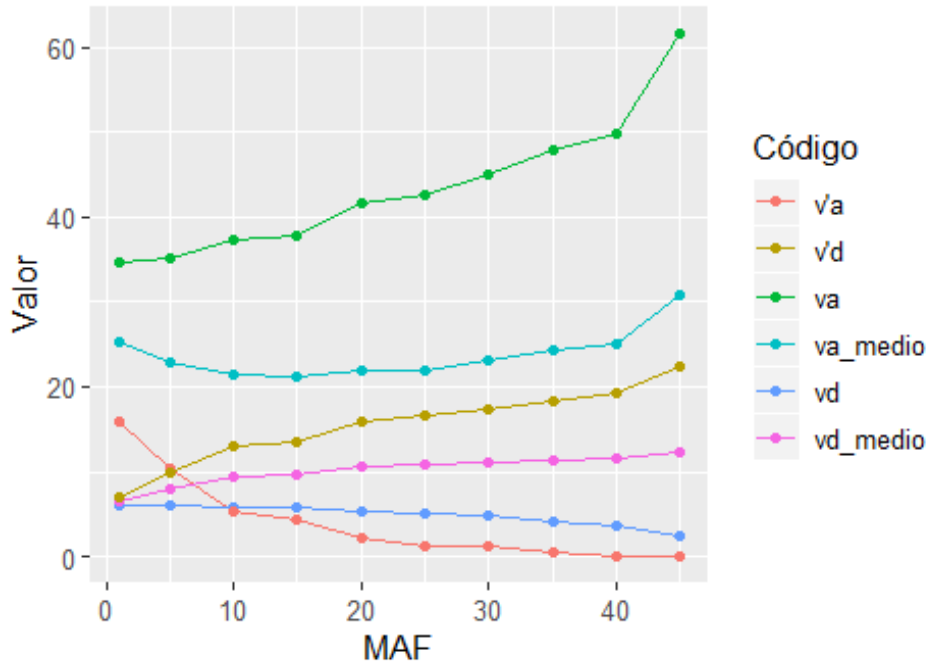


Figura 1 – Valores de variância para IMA em função da MAF.

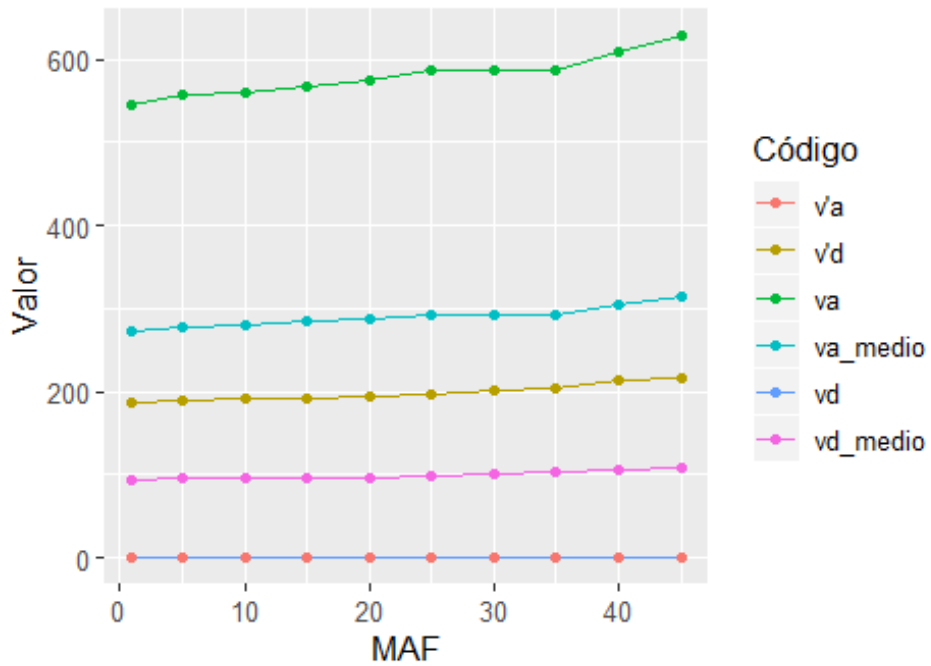


Figura 2 – Valores de variância para densidade básica em função da MAF.

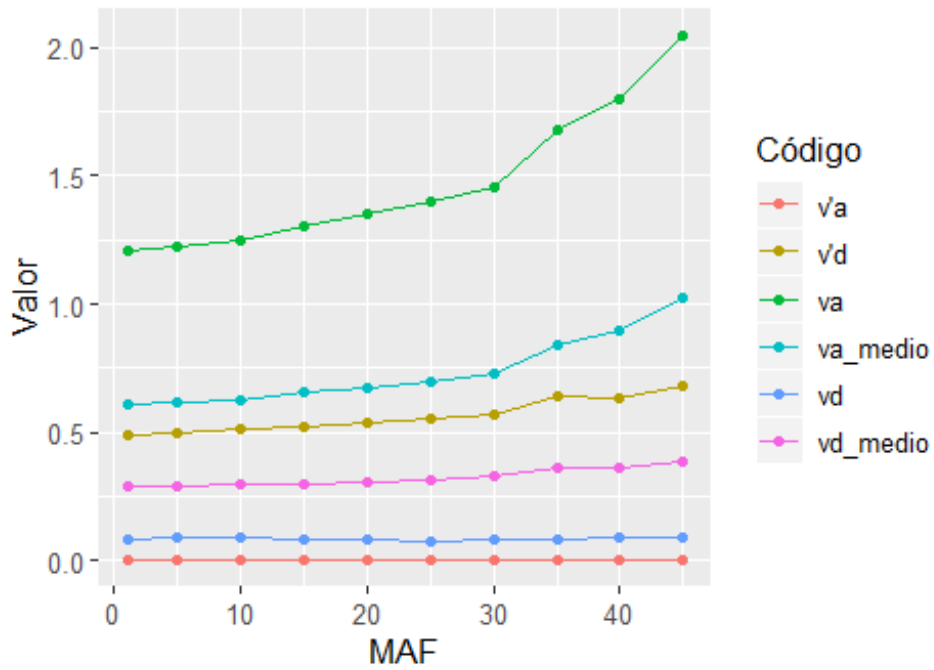


Figura 3 – Valores de variância para rendimento em função da MAF.

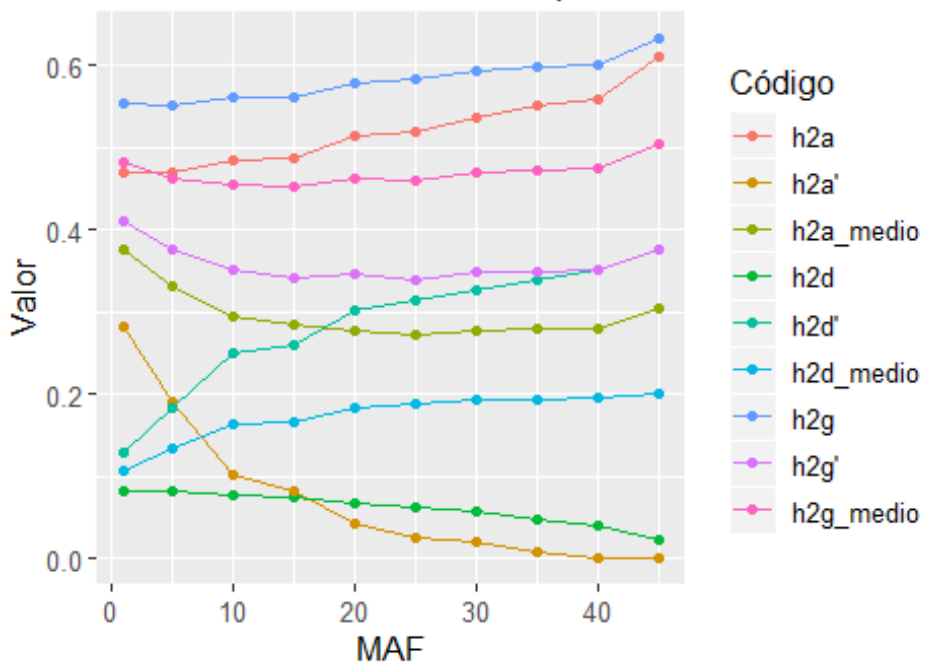


Figura 4 – Valores de herdabilidade para IMA em função da MAF.

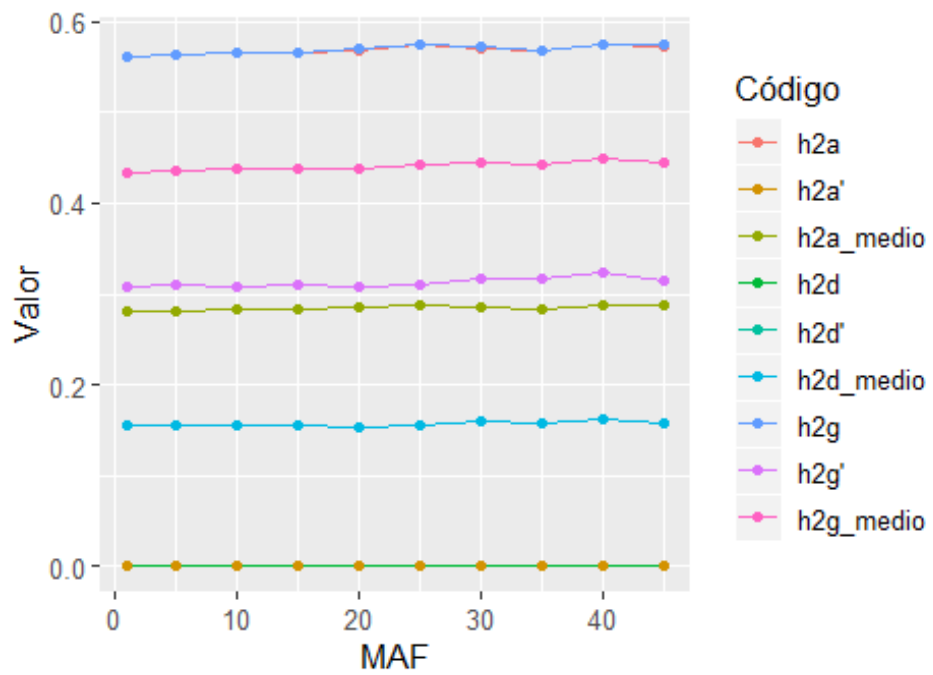


Figura 5 – Valores de herdabilidade para densidade básica em função da MAF.

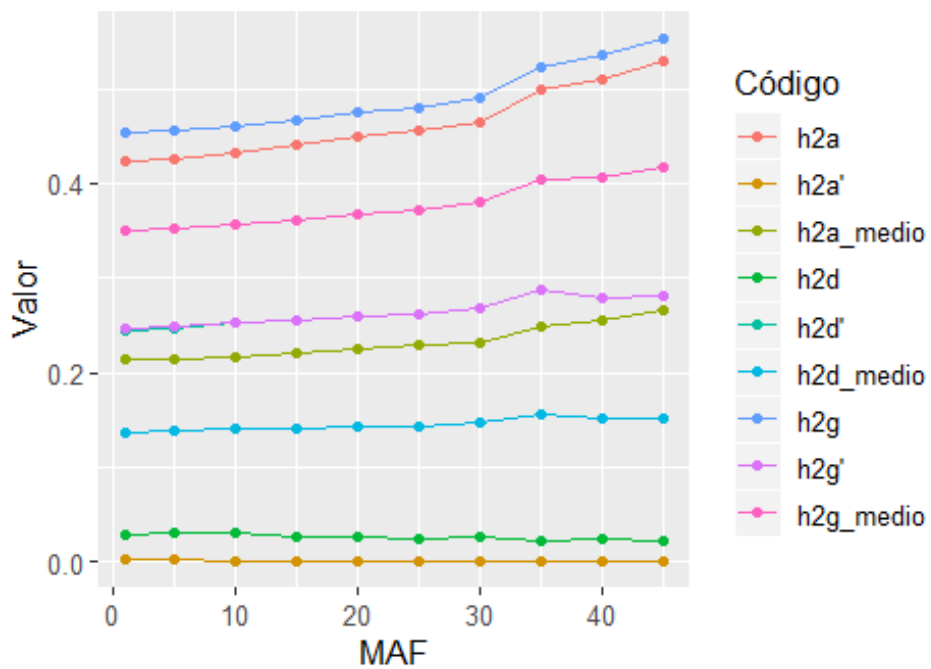


Figura 6 – Valores de herdabilidade para rendimento em função da MAF.

Apêndice B: Script elaborado no software R para implementação da nova modelagem.

```
# Calculando a MAF
#=====
library(HapEstXXR)
M[M=="1"]=3 #Padronizando genótipos (conforme manual HapEstXXR)
M[M=="0"]=1 #M: nome do arquivo de dados com os genótipos
SNP_name=matrix(unique(colnames(M)))
#minor.allele=1 quer dizer que a MAF calculada representa o q
#minor.allele=2 quer dizer que a MAF calculada representa o p
a=data.frame(maf(M, marker.label= SNP_name)) #função maf do pacote HapEstXXR
#Cálculo de frequências alélicas
p1=NULL
q1=NULL
for(i in 1:ncol(M)){
  if(a$minor.allele[i]==1){
    q1[i]= a$maf[i]
    p1[i]= 1-q1[i]}
  if(a$minor.allele[i]==2){
    p1[i]= a$maf[i]
    q1[i]= 1-p1[i]}
}
M1=M[,a$maf>=0.01] #novo arquivo com MAF 1%
p=p1[a$maf>=0.01]
q=q1[a$maf>=0.01]
dim(M1) #dimensão do arquivo após remoção de SNPs
dim(M) #dimensão arquivo original
```

```

# Fazendo a validação cruzada
#=====

CVal<-function(N,K,Ind){
  cvf<-cbind(rep_len(1:K, N),sample(as.character(Ind),N))
  foldx<-list()
  for(cv in 1:K){
    foldx[[cv]]<-subset(cvf[,2], subset = cvf[,1]==cv)
  }
  return(foldx)
}

N1<-756 #N1: total de indivíduos
k1<-12 #k1: número de grupos de validação (folds)
ind1<-feno$Ind #nomes dos indivíduos
res<-CVal(N = N1,K = k1,Ind = ind1) #res= lista com os grupos de validação

# GBLUP com a parametrização tradicional
#=====

W=M #M: é o nome da matriz de genótipos
S=M
p=matrix(0,ncol(M),1)
q=matrix(0,ncol(M),1)
for(z in 1:ncol(M))
{
  p[z,]=(2*length(which(M[,z]==2))+length(which(M[,z]==1)))/(2* length(M[,z]))
  q[z,]=(2*length(which(M[,z]==0))+length(which(M[,z]==1)))/(2* length(M[,z]))
  S[S[,z]==2,z]=2*(p[z,]-q[z,])
  S[S[,z]==1,z]=2*p[z,]
}

W=as.matrix(W)
S=as.matrix(S)

```

```

Ga = (W%*%t(W))/sum(2*p*q) #Ga: matriz de parentesco genômica aditiva
Gd = (S%*%t(S))/sum((2*p*q)^2) #Gd: matriz de parentesco devido a dominância
library(foreach)
library(doSNOW)
library(sommer)
y1=feno[,2] #feno: nome do arquivo de fenótipos
names(y1)=feno$Ind
Z = diag(nrow(Ga)) #Efeitos aleatórios de indivíduos
ff=function(n){
  y=y1
  y[res[[i]]=NA
  ETA=list(add=list(Z=Z,K=Ga),dom=list(Z=Z,K=Gd))
  gblup=mmer(Y=y, Z=ETA)
  return(gblup)
}
cl=makeCluster(4)
registerDoSNOW(cl)
resultado=foreach(i=1:12, .packages = "sommer") %dopar%
  ff(i)
stopCluster(cl)
final=NULL
for (j in 1:12) {
  gblup=resultado[[j]]
  VAR_A=as.numeric(gblup$var.comp[1])
  VAR_D=as.numeric(gblup$var.comp[2])
  VAR_E=as.numeric(gblup$var.comp[3])
  VAR_F=sum(VAR_A,VAR_D,VAR_E)
  h2_A=VAR_A/(VAR_A+VAR_D+VAR_E)
  h2_D=VAR_D/(VAR_A+VAR_D+VAR_E)
  h2_G=(VAR_A+VAR_D)/(VAR_A+VAR_D+VAR_E)
}

```

```

ua_hat_g=gblup$fitted.u
row.names(ua_hat_g)=feno$Ind
efad=gblup$u.hat$add[,1]; names(efad)=feno$Ind
efdom=gblup$u.hat$dom[,1]; names(efdom)=feno$Ind
cp=cor(ua_hat_g[res[[j]],],y1[res[[j]]],use = "na.or.complete")
cp_a=cor(efad[res[[j]]],y1[res[[j]]],use = "na.or.complete")
cp_d=cor(efdom[res[[j]]],y1[res[[j]]],use = "na.or.complete")
lm_a=lm(y1[res[[j]]]~efad[res[[j]])
lm_d=lm(y1[res[[j]]]~efdom[res[[j]])
lm_g=lm(y1[res[[j]]]~ua_hat_g[res[[j]],)
data=cbind(VAR_A,VAR_D,VAR_E,VAR_F,h2_A,h2_D,h2_G,cp,cp_a,cp_d,beta_a=lm_a$
coefficients[2],beta_d=lm_d$coefficients[2],beta_g=lm_g$coefficients[2])
final=rbind(final,cbind(j,data))
}
medias=apply(final,2,mean)
sd=apply(final,2,sd)

# GBLUP com a parametrização alternativa
#=====
W=M
S=M
p=matrix(0,ncol(M),1)
q=matrix(0,ncol(M),1)
for(z in 1:ncol(M))
{
p[z,]=(2*length(which(M[,z]==2))+length(which(M[,z]==1)))/(2* length(M[,z]))
q[z,]=(2*length(which(M[,z]==0))+length(which(M[,z]==1)))/(2* length(M[,z]))
W[W[,z]==2,z]=(-2*q[z,])/(1+q[z,])
W[W[,z]==1,z]=(1-q[z,])/(1+q[z,])
S[S[,z]==1,z]=2
}

```

```

W=as.matrix(W)
S=as.matrix(S)
Ga = (W%*%t(W))/sum((2*(p^2)*q)/(1+q)) #Matriz de parentesco Genômica aditiva
Gd = (S%*%t(S))/sum((4*p*(q^2))/(1+q)) #Matriz de parentesco devido a dominância
library(foreach)
library(doSNOW)
library(sommer)
y1=feno[,2]
names(y1)=feno$Ind
Z = diag(nrow(Ga)) # Efeitos aleatórios de indivíduos
ff=function(n){
  y=y1
  y[res[[i]]=NA
  ETA=list(add=list(Z=Z,K=Ga),dom=list(Z=Z,K=Gd))
  gblup=mmer(Y=y, Z=ETA)
  return(gblup)
}
cl=makeCluster(4)
registerDoSNOW(cl)
resultado=foreach(i=1:12, .packages = "sommer") %dopar%
  ff(i)
stopCluster(cl)
final=NULL
for (j in 1:12) {
  gblup=resultado[[j]]
  VAR_A=as.numeric(gblup$var.comp[1])
  VAR_D=as.numeric(gblup$var.comp[2])
  VAR_E=as.numeric(gblup$var.comp[3])
  VAR_F=sum(VAR_A,VAR_D,VAR_E)
  h2_A=VAR_A/(VAR_A+VAR_D+VAR_E)

```

```

h2_D=VAR_D/(VAR_A+VAR_D+VAR_E)
h2_G=(VAR_A+VAR_D)/(VAR_A+VAR_D+VAR_E)
ua_hat_g=gblup$fitted.u
row.names(ua_hat_g)=feno$Ind
efad=gblup$u.hat$add[,1]; names(efad)=feno$Ind
efdom=gblup$u.hat$dom[,1]; names(efdom)=feno$Ind
cp=cor(ua_hat_g[res[[j]],],y1[res[[j]]],use = "na.or.complete")
cp_a=cor(efad[res[[j]]],y1[res[[j]]],use = "na.or.complete")
cp_d=cor(efdom[res[[j]]],y1[res[[j]]],use = "na.or.complete")
lm_a=lm(y1[res[[j]]]~efad[res[[j]])
lm_d=lm(y1[res[[j]]]~efdom[res[[j]])
lm_g=lm(y1[res[[j]]]~ua_hat_g[res[[j]],)
data=cbind(VAR_A,VAR_D,VAR_E,VAR_F,h2_A,h2_D,h2_G,cp,cp_a,cp_d,beta_a=lm_a$
coefficients[2], beta_d=lm_d$coefficients[2],beta_g=lm_g$coefficients[2])
final=rbind(final,cbind(j,data))
write.table(final,paste("alternativa_IMA_maf1.txt"),quote=FALSE,row.names=FALSE,
col.names=TRUE,dec=",")
}
medias=apply(final,2,mean)
sd=apply(final,2,sd)

# Gráficos
#=====

library(ggplot2)
ggplot(data=dados, aes(x=maf, y=valor, group=var, colour=var)) +
geom_line() +
geom_point() + xlab("MAF") + ylab("Valor") +
ggtitle("Valores de variâncias para IMA") +
scale_colour_hue(name="Código") + theme(text = element_text(size = 12,family="TT Times
New Roman"))

```