

SHEILA GONÇALVES DE FARIA

**ESTIMATIVAS DE PARÂMETROS GENÉTICOS DE RESISTÊNCIA À
FERRUGEM EM *Coffea Canephora* USANDO MODELOS LINEARES
GENERALIZADOS MISTOS**

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*.

Orientador: Ana Carolina Campana Nascimento

Coorientadores: Camila Ferreira Azevedo
Moisés Nascimento

**VIÇOSA - MINAS GERAIS
2024**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

F224e
2024

Faria, Sheila Gonçalves, 1999-

Estimativas de parâmetros genéticos de resistência à ferrugem em *Coffea Canephora* usando modelos lineares generalizados mistos / Sheila Gonçalves Faria. – Viçosa, MG, 2024.

1 dissertação eletrônica (38 f.): il.

Orientador: Ana Carolina Campana Nascimento.

Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Estatística, 2024.

Referências bibliográficas: f. 33-38.

DOI: <https://doi.org/10.47328/ufvbbt.2024.417>

Modo de acesso: World Wide Web.

1. Distribuição binomial. 2. Modelos multiníveis (Estatística). 3. Café - Melhoramento genético. I. Nascimento, Ana Carolina Campana, 1983-. II. Universidade Federal de Viçosa. Departamento de Estatística. Programa de Pós-Graduação em Estatística Aplicada e Biometria. III. Título.

CDD 22. ed. 519.24

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317


SHEILA GONÇALVES DE FARIA

**ESTIMATIVAS DE PARÂMETROS GENÉTICOS DE RESISTÊNCIA À
FERRUGEM EM *Coffea Canephora* USANDO MODELOS LINEARES
GENERALIZADOS MISTOS**


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: 26 de março de 2024

Assentimento:

Documento assinado digitalmente
 **SHEILA GONCALVES DE FARIA**
Data: 30/07/2024 17:52:02-0300
Verifique em <https://validar.iti.gov.br>

Sheila Gonçalves de Faria
Autora

Documento assinado digitalmente
 **ANA CAROLINA CAMPANA NASCIMENTO**
Data: 30/07/2024 14:33:49-0300
Verifique em <https://validar.iti.gov.br>

Ana Carolina Campana Nascimento
Orientadora

Dedico aos meus pais.

AGRADECIMENTOS

Expresso minha profunda gratidão a Deus pelo dom da vida e por sua soberana providência sobre todas as coisas. Além disso, manifesto minha sincera devoção à Virgem Maria, que, juntamente com seu Filho, concedeu-me a fortaleza necessária para perseverar até este momento.

Agradeço aos meus pais, José Morais e Aparecida, pela educação, por tanto carinho, amor e por nunca duvidarem de mim, me motivando em todos os momentos da minha vida.

Ao meu irmão Marcos, por sempre cuidar de mim, por todo apoio, carinho e amizade. Agradeço também a minha cunhada Miriam por sempre torcer e rezar por mim.

Agradeço ao Caique por toda paciência, amor, incentivo e companheirismo, você sem dúvidas, fez dessa caminhada mais leve!

Aos meus avós, José, Antônia e Ester, por sempre rezarem por mim e se preocuparem tanto.

Agradeço carinhosamente à minha orientadora Ana Carolina, por todos os ensinamentos compartilhados, apoio, disponibilidade e por sempre me incentivar em todos os momentos, você foi muito mais que uma orientadora.

Aos meus coorientadores Moysés Nascimento e Camila Azevedo Ferreira pelos ensinamentos, pela disponibilidade e conselhos.

Agradeço aos meus amigos, os presentes que a Matemática me deu: Jonas e Gabriela, por todo companheirismo, atenção e carinho durante todo esse tempo. Vocês foram essenciais!

À família estatística que ganhei durante esse tempo, Brenda, Mariana, Sara e Lisandra, por sempre se fazerem tão presentes e me apoiarem.

Ao meu afilhado Thiago, por todo amor e por sempre torcer e me apoiar.

As minhas primas, Carol e Marina, por toda amizade, conselhos, carinho e por sempre se fazerem presente.

Agradeço ao Laboratório de Inteligência Computacional e Aprendizado Estatístico (LICAE), por ter contribuído significativamente para o meu desenvolvimento, tanto no âmbito profissional quanto pessoal.

Ao meu colega Deivisson pela amizade e ajuda com as rotinas computacionais.

À Universidade Federal de Viçosa, pelo acolhimento e oportunidade. Ao Programa de Pós-Graduação em Estatística Aplicada e Biometria por todo conhecimento.

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

Ademais, expresso minha gratidão a todos que, de forma direta ou indireta, colaboraram com minha jornada até este momento. A todos, o meu sincero agradecimento!

“Deus não poderia inspirar em mim desejos irrealizáveis.”
(Santa Teresa do Menino Jesus)

RESUMO

FARIA, Sheila Gonçalves, M.Sc Universidade Federal de Viçosa, março de 2024. **Estimativas de parâmetros genéticos de resistência à ferrugem em *Coffea Canephora* usando modelos lineares generalizados mistos**. Orientadora: Ana Carolina Campana Nascimento. Coorientadores: Camila Ferreira Azevedo e Moysés Nascimento.

O café é uma das bebidas mais consumidas e uma das *commodities* mais comercializadas globalmente. O *Coffea Canephora*, é uma espécie perene conhecida pelo alto potencial produtivo, alto teor de sólidos solúveis, e se sobressai por ser uma espécie que se adapta melhor a diversos estresses ambientais e por ser mais resistente. Nos programas de melhoramento, estimativas precisas de parâmetros genéticos são primordiais para desenvolver objetivos de criação eficazes e abrangentes. Os métodos para obtenção de parâmetros genéticos são mais bem desenvolvidos para características normalmente distribuídas. Porém, é comum que sejam coletadas informações advindas de características fenotípicas que possuem distribuição não normal. Nestes casos, os Modelos Lineares Generalizados Mistos (*Mixed Generalized Linear Models* - GLMM) apresentam-se como uma ferramenta útil pois permitem o ajuste do modelo quando a variável de interesse possui qualquer distribuição da chamada “família exponencial de distribuições”. Os GLMMs permitem que a relação entre a combinação linear das variáveis explicativas e a variável resposta seja mais geral do que no Modelo Linear (ML). Desta forma, ao estimar parâmetros genéticos com GLMMs, é crucial, além de considerar corretamente a distribuição da variável resposta Y, expressar os resultados na escala original das características medidas. Isso porque os GLMMs fornecem as estimativas dos parâmetros em uma escala latente, exigindo transformações para interpretações adequadas. Diante do exposto, este trabalho teve por objetivo a obtenção de parâmetros genéticos da característica relacionada a resistência à ferrugem da variedade Conilon de *Coffea Canephora*, mensurada de forma binária (1 = resistente e 0 = susceptível). Os resultados obtidos para as variâncias genética e fenotípica, herdabilidade e ganho de seleção foram comparados com os resultados advindos de modelos usuais e, com aqueles advindos de modelos GLMMs sem as correções de escala devidas. Para tanto, considerou-se informações referentes à 51 genótipos do grupo varietal Conilon. Como resultados, observou-se que, desconsiderar a não-normalidade dos caracteres avaliados (ou seja, utilizar o modelo tradicional, baseado em normalidade) fez com que as variâncias genéticas e fenotípicas fossem subestimadas. No entanto, a herdabilidade e o ganho de

seleção apresentaram-se maior que aquelas advindas dos GLMMs com a correção da escala. Para a situação em que os GLMMs foram ajustados, mas omitiu-se as correções de escala, isto é, interpretou-se os resultados na escala latente observou-se superestimação de todos os parâmetros genéticos. A superestimação dos ganhos de seleção podem levar a conclusões equivocadas sobre a eficácia da seleção genética, fazendo parecer que a seleção para determinadas características resultará em maior progresso genético do que realmente seria possível. Neste caso, pode-se ocorrer a alocação indevida de recursos em pesquisas para essa característica, em detrimento de outras com maior potencial real de aprimoramento. Além disso, programas de melhoramento baseados em dados superestimados da herdabilidade e do ganho de seleção podem ser ineficazes e dispendiosos, gerando resultados abaixo das expectativas. Desta forma, destaca-se a importância da utilização de modelos GLMMs com correção da escala dos parâmetros, a fim de obter resultados mais precisos e confiáveis, auxiliando no desenvolvimento de programas de melhoramento genético mais eficientes e eficazes.

Palavras-chave: Café; Distribuição Binomial; Modelos Mistos; Melhoramento Genético.

ABSTRACT

FARIA, Sheila Gonçalves, M.Sc., Universidade Federal de Viçosa, March, 2024. **Estimation of genetic parameters of rust resistance in *Coffea Canephora* using mixed generalized linear models.** Adviser: Ana Carolina Campana Nascimento. Co-advisers: Camila Ferreira Azevedo and Moysés Nascimento.

Coffee is one of the most consumed beverages and globally traded commodities. *Coffea canephora*, a perennial species, stands out for its high productivity, soluble solids content, adaptability to diverse environmental stresses, and resistance. Accurate genetic parameter estimates are crucial for developing effective breeding objectives. However, methods for obtaining genetic parameters are better developed for normally distributed traits. Phenotypic data often exhibits non-normal distributions, making Mixed Generalized Linear Models (GLMMs) a valuable tool. GLMMs allow model adjustment when the response variable follows any distribution from the "exponential family of distributions." They enable a more general relationship between the linear combination of explanatory variables and the response variable compared to the Linear Model (ML). When estimating genetic parameters with GLMMs, it is crucial to consider the distribution of the response variable (Y) and express results in the original scale of the measured traits. GLMMs provide parameter estimates on a latent scale, requiring transformations for appropriate interpretations. This study aimed to estimate genetic parameters for rust resistance in the Conilon variety of *Coffea canephora*, measured binarily (1 = resistant and 0 = susceptible). Results for genetic and phenotypic variances, heritability, and selection gain were compared with those from usual models and GLMMs without proper scale corrections. Data from 51 Conilon genotypes were analyzed. As a result, disregarding non-normality (using the traditional normality-based model) underestimated genetic and phenotypic variances. However, heritability and selection gain were higher than those from GLMMs with scale correction. Fitting GLMMs without scale corrections misinterpreted results on the latent scale and overestimated all genetic parameters. Overestimated selection gains can lead to erroneous conclusions about breeding efficiency, suggesting greater genetic progress than possible. This could lead to misallocation of research resources and ineffective, costly breeding programs based on overestimated heritability and selection gain data. Therefore, using GLMMs with parameter scale correction is crucial for obtaining more accurate and reliable results, enabling the development of more efficient and effective breeding programs.

Keywords: Coffee; Binomial Distribution; Mixed Models; Genetic Improvement.

SUMÁRIO

1. INTRODUÇÃO.....	12
2. REFERENCIAL TEÓRICO.....	15
2.1. Modelos Lineares Mistos.....	15
2.1.1. Parâmetros Genéticos nos Modelos Lineares Mistos.....	15
2.2. Modelos Lineares Generalizados.....	17
2.3. Modelos Lineares Generalizados Mistos.....	18
2.3.1. As escalas do Modelo Linear Generalizado Misto.....	19
2.3.2. Parâmetros Genéticos nos GLMMs.....	20
3. MATERIAL E MÉTODOS.....	23
3.1. Material genético.....	23
3.2. Avaliações Fenotípicas.....	23
3.3. Modelos Estatísticos.....	24
3.3.1. Componentes de variância, herdabilidade e ganho de seleção.....	25
4. RESULTADOS E DISCUSSÃO.....	26
5. CONCLUSÃO.....	31
REFERÊNCIAS BIBLIOGRÁFICAS.....	33

1. INTRODUÇÃO

O café é uma bebida globalmente importante, sendo uma das mais consumidas e uma das *commodities* mais comercializadas no mundo (FAO, 2023). O Brasil é o maior produtor e exportador de café, responsável por cerca de um terço da produção mundial (CECAFÉ, 2023). Pertencentes ao gênero *Coffea*, destacam-se as espécies *Coffea arabica* e *C. canephora* como as de maior importância no mercado mundial de cafés (VEGA et al. 2003; DENOEUDE et al. 2014).

Ao longo dos anos, os programas de melhoramento do cafeeiro no Brasil aspiram acrescentar valor ao produto final e reduzir os custos de produção a partir da seleção de cultivares resistentes a doenças o que, por consequência, resultam em maior produtividade. Segundo Talhinhos et al. (2017) a ferrugem do café, causada pelo fungo *Hemileia vastatrix Berk. and Broome*, é uma doença importante para o gênero *coffea*, podendo causar perdas de produtividade de até 35% e causar impactos epidemiológicos persistentes nos anos seguintes.

Dentre as espécies de café, o café *canephora* se sobressai pela rusticidade, alto potencial produtivo, alto teor de sólidos solúveis e resistência genética à ferrugem do cafeeiro causada pelo fungo *Hemileia vastatrix* (ZAMBOLIM, 2016). Segundo Tran et al. (2016) essa espécie é um bom ponto de partida para estudar o gênero *Coffea* por razões econômicas e genéticas, incluindo ploidia ($2n = 2x$) e a alta variabilidade genética.

A presença de variabilidade genética é fator preponderante para a obtenção de ganhos genéticos com seleção (BABOVA et al., 2016). Nesse contexto, visto que a informação sobre tal quantidade é advinda da estimação de componentes de variâncias torna-se crucial obter tais estimativas de forma acurada. Além disso, os componentes de variância permitem a obtenção de outros parâmetros genéticos, como as correlações genéticas ou fenotípicas, essenciais para definir estratégias de seleção mais eficientes (FALCONER E MACKAY, 1996).

O café por ser uma espécie perene, possui caracteres se expressando ao longo de vários anos, e ficando exposta a diversos tipos de problemas ambientais. Isso faz com que ocorram perdas de dados, acarretando um desbalanceamento dos mesmos que, por sua vez geram a necessidade de modelagens mais robustas para a estimação dos parâmetros genéticos (SERA, 2001).

Devido, principalmente ao desbalanceamento, segundo Oliveira et al. (2014) e Resende et al. (2014), os métodos tradicionais de estimação, que se baseiam na análise de

variância, têm sido substituídos pelos modelos mistos do tipo REML/BLUP (*Residual or Restricted Maximum Likelihood/Best Linear Unbiased Prediction*) (HENDERSON, 1975) para a análise de dados nos programas de melhoramento do cafeeiro (apud ALKIMIN, 2017). Os métodos de estimação permitem a obtenção dos componentes de variância e, a partir desses é possível acessar os parâmetros genéticos, que são extremamente importantes para o melhoramento. Spinelli et al. (2015), Carias et al. (2016) e Sousa et al. (2019) são exemplos de alguns autores que trabalharam com estimação de parâmetros genéticos a partir de modelos lineares mistos do tipo REML/BLUP.

Em geral, os métodos utilizados para a estimação das variâncias e covariâncias genéticas aditivas bem como a teoria genética quantitativa para prever respostas à seleção, assumem pressuposto de normalidade (HENDERSON, 1950; LYNCH e WALSH, 1998; MORRISSEY, 2015 apud VILLEMEREUIL et al., 2016). Entretanto, é comum em programas de melhoramento, que o interesse recaia sobre fenótipos não gaussianos (que não possuem distribuição Normal).

Fenótipos que não seguem distribuição normal vêm sendo relatados há algum tempo em programas de melhoramento e, a análise realizada deve levar em conta o fato da não normalidade dos mesmos (GIANOLA, 1982; GIANOLA e FOULLEY, 1983; FOULLEY et al., 1987). Alguns destes fenótipos se apresentam de diferentes formas, podendo ser, dentre outros, do tipo binário (ex.: presença/ausência de doenças), categórico nominal ou ordinal (ex.: notas) ou ainda, na forma de contagem (ex.: número de dias para o florescimento).

A fim de lidar corretamente com estes dados, foi proposto o Modelo Linear Generalizado Misto (*Generalized Linear Mixed Models – GLMM*, GILMOUR, 1985), que possibilita a utilização de diversas distribuições distintas, usando uma estrutura hierárquica que parte de uma característica latente normalmente distribuída (hipotética) até os dados observados. Os GLMMs são uma combinação de duas estruturas estatísticas, os Modelos Lineares Mistos (*Linear Mixed Models - LMM*) com os Modelos Lineares Generalizados (*Generalized Linear Models - GLM*) que são uma extensão do modelo linear clássico (NELDER e WEDDERBURN, 1972).

O uso de modelos lineares mistos generalizados (GLMM) para analisar características não gaussianas tem se tornado cada vez mais frequente, como demonstrado em estudos como os de Villemereuil et al. (2013), Wilson et al. (2011), Morrissey et al. (2012), Ayers et al. (2013), Chaves et al. (2022) e Azevedo et al. (2024). Especificamente

em café, pode-se citar os trabalhos de Beyene et al. (2022), que exploraram como a variação genética em locais de café Arábica no sudoeste da Etiópia estava relacionada com a incidência de quatro doenças fúngicas utilizando GLMM. Delly et al. (2022) que avaliaram o papel da formiga cortadeira *Atta cephalotes* nas plantações de café e, Carvalho et al. (2023), que utilizaram o modelo de limiar para otimizar modelos de predição genômica para resistência do café à ferrugem e ao bicho-mineiro.

A análise de parâmetros genéticos em modelos com fenótipos categóricos exige cautela, pois apresenta características distintas da análise via LMM. Diferentemente do LMM, que estima diretamente parâmetros genéticos aditivos na escala original das características, nos GLMMs as quantidades são obtidas em uma escala latente (VILEMEREUIL et al., 2016). Essa diferença impacta a interpretação e aplicação dos resultados.

A literatura propõe diferentes métodos para converter os parâmetros genéticos latentes dos GLMMs para a escala dos dados observados, especialmente para características binárias, que representam estados "presente" ou "ausente", "verdadeiro" ou "falso", geralmente codificados como 0 ou 1. Por exemplo, Nakagawa e Schielzeth (2010) propõem fórmulas fechadas para o cálculo dos parâmetros para diferentes distribuições e funções de ligação, aprimorando a compreensão e aplicação de modelos estatísticos em contextos diversos. Mais recentemente, Villemereuil et al. (2016) aperfeiçoaram a teoria de Morrissey (2015) e apresentaram fórmulas para estimar parâmetros genéticos na escala dos dados observados, preservando a escala original. Essa abordagem garante a interpretação direta dos resultados e evita distorções nas estimativas. Pois, a negligência no tratamento adequado de características não gaussianas na modelagem pode levar a superestimação ou subestimação dos parâmetros genéticos, impactando significativamente as inferências e, conseqüentemente, as decisões nos programas de melhoramento genético.

O impacto da escolha do modelo na obtenção de algumas medidas de interesse prático no melhoramento já foi avaliado por Makouanzi et al. (2014), que investigaram os efeitos genéticos aditivos e de dominância da capacidade de propagação vegetativa em *Eucalyptus*. Especificamente, os autores investigaram o desempenho do modelo misto linear generalizado (GLMM) e o compararam com o do modelo misto clássico (LMM) para verificarem a influência da modelagem no ganho genético. No entanto, os autores concluem sobre a dificuldade de interpretar os parâmetros genéticos, especialmente,

quando o experimento é de desenho complexo e os dados são desbalanceados. Sugerindo serem necessários estudos de simulação para investigar o potencial dos GLMMs.

Diante do exposto, este trabalho teve por objetivo a obtenção de parâmetros genéticos para a característica relacionada a resistência à ferrugem em *Coffea Canephora*, com base em modelos lineares generalizados mistos (GLMM). Ademais, busca-se apresentar as fórmulas para a obtenção dos parâmetros genéticos para as diferentes escalas nos GLMMs; avaliar o impacto nos resultados ao analisar os parâmetros genéticos em escalas inapropriadas para o conjunto de dados; comparar as estratégias de obtenção dos parâmetros na escala original e, por fim, comparar os resultados obtidos com aqueles advindos de modelos tradicionais (LMM).

2. REFERENCIAL TEÓRICO

2.1. Modelos Lineares Mistos

Um modelo linear misto (*Linear Mixed Model* – LMM), pode ser representado como a seguir (HENDERSON, 1975):

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

em que \mathbf{y} é um vetor de observações fenotípicas de dimensão $(n \times 1)$; \mathbf{X} é uma matriz de incidência de efeitos fixos de dimensão $(n \times p)$; $\boldsymbol{\beta}$ é um vetor de efeitos fixos desconhecido de dimensão $(p \times 1)$ e, \mathbf{Z} é uma matriz de incidência de efeitos aleatórios de dimensão $(n \times q)$, \mathbf{u} é um vetor de dimensão $(q \times 1)$ de parâmetros de efeitos aleatórios desconhecidos, tal que $\mathbf{u} \sim N(0, \sigma_u^2)$ e, \mathbf{e} é um vetor de dimensão $(n \times 1)$ de erros aleatórios não observáveis tal que $\mathbf{e} \sim N(0, \sigma_e^2)$.

2.1.1. Parâmetros Genéticos nos Modelos Lineares Mistos

Variância genética

Segundo de los Campos et al. (2015), a variância genética de uma característica (σ_g^2) é a média dos quadrados dos desvios dos valores genéticos em relação à média populacional. Além de expressar a diversidade entre indivíduos de uma população, pode ser tratada no contexto biométrico e subdividida como aditiva, dominante e epistática, de tal forma que pode ser descrita como:

$$\sigma_g^2 = \sigma_a^2 + \sigma_d^2 + \sigma_i^2, \quad (2)$$

sendo que, σ_a^2 é a variância aditiva; σ_d^2 é a variância atribuída aos desvios da dominância e σ_i^2 é a variância atribuída aos efeitos epistáticos resultantes das interações inter-alélicas (CRUZ, 2005).

Variância fenotípica

A variância fenotípica (σ_f^2) é composta pela variância genotípica (σ_g^2) e pela variância atribuída aos desvios proporcionados pelo ambiente (σ_e^2), e pode ser expressa como:

$$\sigma_f^2 = \sigma_g^2 + \sigma_e^2. \quad (3)$$

Ou, de outra forma, como:

$$\sigma_f^2 = \sum_{i=1}^N f_i y_i^2 - \left(\sum_{i=1}^N f_i y_i \right)^2, \quad (4)$$

em que y_i é o valor fenotípico do i -ésimo indivíduo, f_i é a frequência com que o valor fenotípico y_i foi observado e N é o número de indivíduos da população (CRUZ, 2005).

Herdabilidade

A herdabilidade tem como finalidade medir a proporção da variação fenotípica na população devido à causa genética, quantificando o grau em que as diferenças observadas em uma característica, como: altura, peso, produção, dentre outras, são influenciadas por diferenças nos genes dos indivíduos dentro de uma população. A herdabilidade pode ser estimada por (CRUZ, 2005):

$$h^2 = \frac{\sigma_g^2}{\sigma_f^2}, \quad (5)$$

em que, σ_g^2 é a variância genotípica e σ_f^2 é a variância fenotípica.

Ganho de seleção

O ganho de seleção, conceito fundamental em genética e melhoramento, consiste em uma medida do aprimoramento gradual de características desejáveis em uma população, a partir da seleção de indivíduos que expressam tais características (dos ANJOS et. al., 2019). Os ganhos de seleção diretos para uma população podem ser estimados usando a seguinte equação (CRUZ, 2005):

$$GS = (\overline{X}_s - \overline{X}_o) h^2 = DS \times h^2 \quad (6)$$

em que DS é o diferencial de seleção realizada na população, ou seja, a diferença entre a média dos valores fenotípicos do conjunto de indivíduos selecionados (\overline{X}_s) e a média dos valores fenotípicos da população original (\overline{X}_o) e, h^2 é a herdabilidade da característica.

2.2. Modelos Lineares Generalizados

Os Modelos Lineares Generalizados (*Generalized Linear Models* - GLM) foram introduzidos em 1972, por Nelder e Wedderburn com a finalidade de generalizar algumas pressuposições do Modelo Linear e criar um marco teórico que englobasse vários modelos já conhecidos dos quais podemos citar: Probit, Logit, Poisson, Log-linear e o próprio Modelo Linear (CORDEIRO; DEMÉTRIO; MORAL, 2018).

De modo geral, os GLMs possibilitam a consideração de outras distribuições para a variável resposta, além da Distribuição Normal - desde que a distribuição seja qualquer uma pertencente à família exponencial de distribuições -, além de dar maior flexibilidade para a relação funcional entre média da variável resposta e as variáveis predictoras. Por exemplo, para dados de contagem, em vez de aplicar a transformação \sqrt{y} para buscar a normalidade dos dados, pode-se supor que a distribuição de Y segue Poisson, e que a relação funcional entre a média de Y e o preditor linear é dada por $\log \mu = \boldsymbol{\eta}$ (NELDER E WEDDERBURN, 1972).

Para a formulação de um GLM, conforme Demétrio (2002), é necessário a definição de três componentes: i) Componente aleatória; ii) Componente sistemática e iii) Função de ligação.

A componente aleatória consiste na distribuição de probabilidade adotada para a variável dependente Y. A componente sistemática, $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^T$, também chamada de preditor linear, é uma função linear dos parâmetros desconhecidos $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ e das variáveis explicativas, representada por $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, sendo \mathbf{X} uma matriz modelo $n \times (p + 1)$.

A função de ligação é a função da média que será modelada em função das variáveis explicativas adotadas ou, em função da componente sistemática. Esta função deve ser uma função diferenciável e monótona g, que irá associar as componentes aleatória e sistemática, através de uma relação do seguinte tipo:

$$g(E[\mathbf{Y}]) = g(\mu) = \mathbf{X}\boldsymbol{\beta}$$

em que \mathbf{Y} é o vetor com as n observações de Y , μ é o valor esperado de Y , \mathbf{X} é matriz de observações das variáveis preditoras e $\boldsymbol{\beta}$ é o vetor de coeficientes. Se g for invertível, temos:

$$g(\mu) = g(E[\mathbf{Y}]) = \mathbf{X}\boldsymbol{\beta} \Leftrightarrow \mu = E[\mathbf{Y}] = g^{-1}(\mathbf{X}\boldsymbol{\beta}),$$

em que, a função g^{-1} é responsável pela transformação da escala latente para a escala esperada. Vale ressaltar que, deste modo, a função de ligação estrutura as variações na escala latente para variações compatíveis com a distribuição utilizada.

A estimação dos parâmetros em GLMs é realizada com o método de máxima verossimilhança. Com respostas que não seguem distribuição normal, o uso de mínimos quadrados ordinários (OLS) para estimar os parâmetros não é viável devido à natureza não linear da verossimilhança em relação a esses parâmetros. Portanto, métodos iterativos como Newton-Raphson ou Fisher Scoring são empregados para resolver a função de verossimilhança não linear e obter estimativas para $\boldsymbol{\beta}$ (AGRESTI, 2015).

Para um estudo aprofundado dos GLMs, sugerem-se a consulta aos trabalhos de Dobson e Barnett (2008), Hardin e Hilbe (2007), Demétrio (2001), McCullagh e Nelder (1989), Cordeiro (1986) (apud CORDEIRO; DEMÉTRIO; MORAL, 2018) e Agresti (2015).

2.3. Modelos Lineares Generalizados Mistos

Os Modelos Lineares Generalizados Mistos (*Generalized Linear Mixed Models - GLMM*) combinam duas estruturas estatísticas: os Modelos Lineares Mistos (LMM) e os Modelos Lineares Generalizados (GLM).

Os GLMMs são uma ferramenta para analisar variáveis normais e não normais na presença de efeitos aleatórios e, para a sua utilização é necessário, em princípio, determinar a distribuição de probabilidade adotada para a variável dependente, a função de ligação e a estrutura dos efeitos aleatórios (BOLKER et al., 2009).

Um GLMM pode ser definido a partir das seguintes expressões (VILLEMEREUIL et al., 2016):

$$\mathbf{l} = \mu + \mathbf{X}\mathbf{b} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_1\mathbf{u}_1 + \dots + \mathbf{Z}_k\mathbf{u}_k + \mathbf{o}, \quad (7)$$

$$\boldsymbol{\eta} = g^{-1}(\mathbf{l}), \quad (8)$$

$$\mathbf{y} \sim D(\boldsymbol{\eta}, \boldsymbol{\theta}). \quad (9)$$

Na Equação 7, \mathbf{l} é o vetor de variáveis latentes associado ao vetor de valores fenotípicos \mathbf{y} , μ é o intercepto do modelo, \mathbf{X} é a matriz de incidência de efeitos fixos, \mathbf{b} é o vetor de efeitos fixos, \mathbf{Z}_a é a matriz de incidência de efeitos aleatórios \mathbf{a} é o vetor de

efeitos aleatórios genéticos aditivos de indivíduos com $\mathbf{a}|\sigma_a^2 \sim \mathbf{N}(\mathbf{0}, \sigma_a^2 \mathbf{I})$ ou *breeding values*, com σ_a^2 sendo a variância genética aditiva, \mathbf{Z}_i são matrizes de incidência de outros efeitos aleatórios \mathbf{u}_i ($i = 1, \dots, k$) que podem ser considerados. Assume-se que todos os efeitos aleatórios seguem distribuição normal com variâncias a serem estimadas na escala latente. O termo o se refere ao parâmetro de sobredispersão (com variância igual a σ_o^2) que pode ser ou não considerado no modelo.

Na Equação 8, tem-se que, $\eta = E(Y)$ e, como no GLM, g representa a função de ligação. Logo, função g^{-1} é a função inversa de g , que é responsável pela transformação da escala latente para a escala esperada.

Por fim, a Equação 9 especifica a distribuição de \mathbf{y} que tem parâmetros denotados por $\boldsymbol{\eta}$ e $\boldsymbol{\theta}$. É importante notar que, diferentemente dos modelos lineares mistos, aqui não são explicitados erros aleatórios. Toda a variação de \mathbf{y} em torno da média será devido às probabilidades associadas que serão baseadas na distribuição de probabilidades adotada. Aqui, define-se uma distribuição para a variável resposta que representa as observações e não uma distribuição para o erro aleatório, e essa escolha da distribuição vai depender da natureza dos dados (se contínuos ou discretos) e também do intervalo de variação destes dados (VILLEMEREUIL et al., 2016).

2.3.1. As escalas do Modelo Linear Generalizado Misto

As equações apresentadas para o ajuste de um GLMM (Equações 7 – 9) definem três escalas de variação. A escala da variável latente l (Equação 7); a escala de dados esperada (Equação 8), que é a transformação da escala latente para a escala esperada e é dada pelo inverso da função de ligação g ; e, escala de dados observada (Equação 9), que é derivada da distribuição de probabilidades de \mathbf{y} que está sendo considerada para modelar o “ruído” dos valores observados em torno do valor esperado.

As escalas citadas estão ilustradas na Figura 1, abaixo.

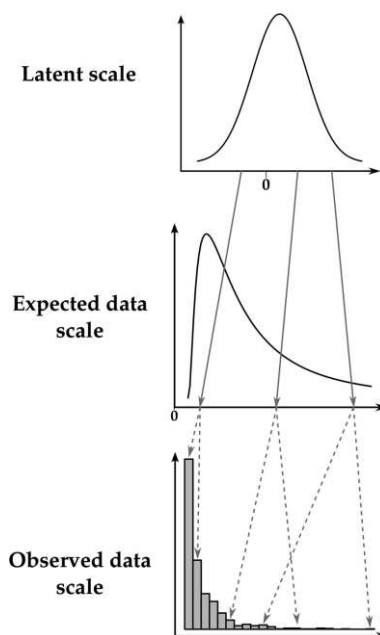


Figura 1: As três escalas do Modelo Linear Misto Generalizado. Nesta imagem representa-se uma distribuição Poisson com função de ligação logarítmica (VILLEMEREUIL, 2016).

2.3.2. Parâmetros Genéticos nos GLMMs

Villemereuil et al. (2016) argumentam que, apesar de ser possível utilizar os modelos GLMMs para inferência de parâmetros genéticos quantitativos, a distinção entre as diferentes escalas é importante para a correta interpretação dos resultados obtidos. Segundo os autores, devido ao fato de que os valores genéticos são valores individuais intrínsecos (moldados tanto pelos genes como pelo ambiente), a variância genética aditiva é a mesma para ambas as escalas, mas, devido ao ruído adicionado nos dados observados (pela escolha da distribuição), as herdabilidades não o são.

Ademais, ainda segundo Villemereuil et al. (2016), os efeitos fixos causam complicações adicionais e importantes para a interpretação dos GLMMs. Assim, ao contrário de um modelo linear misto ou de um GLMM na escala latente, os componentes de variância na escala observada num GLMM dependem dos efeitos fixos. Por exemplo, em um GLMM com uma função de ligação log tem-se que, como a exponencial é uma função convexa, a influência de efeitos fixos e aleatórios faz com que se observe maior variação nas escalas de dados esperadas e observadas para valores que são maiores do que para aqueles que são menores.

Assim, algumas equações são essenciais para inferir parâmetros genéticos quantitativos em GLMMs. São elas (VILLEMEREUIL et al., 2016):

Média populacional

Quando não se tem efeitos fixos no modelo, o fenótipo médio nas escalas esperada e observada são iguais, e pode ser obtido por:

$$\bar{y} = \int g^{-1}(l) f_N(l, \mu, \sigma_{g,l}^2 + \sigma_{RE}^2 + \sigma_o^2) dl, \quad (10)$$

em que $f_N(l, \mu, \sigma_{g,l}^2 + \sigma_{RE}^2 + \sigma_o^2)$ é a densidade de probabilidade de uma distribuição normal com média μ e $\sigma_{g,l}^2 + \sigma_{RE}^2 + \sigma_o^2$ corresponde à variância fenotípica da escala latente.

Quando o modelo possui efeitos fixos, em modelos lineares generalizados mistos (GLMM), não há pressupostos específicos sobre como as variáveis independentes estão distribuídas nos efeitos fixos. Portanto, é possível marginalizar esses efeitos, o que significa calcular a média dos valores previstos (considerando a aleatoriedade nos efeitos, ou seja, $X\hat{b}$, onde \hat{b} são as estimativas dos efeitos fixos). É importante notar que, ao fazer isso, está sendo assumido que a amostra está representando a população de interesse.

Sendo assim, para um modelo com efeitos fixos, a média pode ser calculada da seguinte forma:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N \int g^{-1}(l) f_N(l, \mu + \hat{l}_i, \sigma_{g,l}^2 + \sigma_{RE}^2 + \sigma_o^2) dl, \quad (11)$$

em que N é o número de observações. Além disso, neste caso, pressupõe-se que todos os efeitos fixos representam variações biologicamente relevantes, em vez de serem correções para o processo de observação ou condição experimental.

Variância fenotípica em escala esperada

A variância fenotípica em escala esperada pode ser calculada de forma análoga ao que foi realizado para a média, tanto para o modelo sem efeito fixo, quanto para o modelo com efeitos fixos. Para o modelo sem efeitos fixos, ela pode ser calculada da seguinte forma:

$$\sigma_{f,exp}^2 = \int (g^{-1}(l) - \bar{y})^2 f_N(l, \mu, \sigma_{g,l}^2 + \sigma_{RE}^2 + \sigma_o^2) dl. \quad (12)$$

Por outro lado, para o modelo com efeitos fixos, é calculada da seguinte forma:

$$\sigma_{f,exp}^2 = \frac{1}{N} \sum_{i=1}^N \int (g^{-1}(l) - \bar{y})^2 f_N(l, \mu + \hat{l}_i, \sigma_{g,l}^2 + \sigma_{RE}^2 + \sigma_o^2) dl. \quad (13)$$

Variância fenotípica em escala observada

Para passar da Equação 8 para a Equação 9, é necessário levar conta uma função de variância $\sigma^2(l, \theta)$ que descreve a variância da distribuição utilizada. Assim, no modelo sem efeitos fixos, a variância fenotípica na escala observada é a soma da variância dos valores esperados e da variância resultante da função de distribuição:

$$\sigma_{fo}^2 = \sigma_{f,exp}^2 + \int \sigma^2(l, \theta) f_N(l, \mu, \sigma_{g,l}^2 + \sigma_{RE}^2 + \sigma_o^2) dl, \quad (14)$$

e, para o modelo com efeitos fixos:

$$\sigma_{fo}^2 = \sigma_{f,exp}^2 + \frac{1}{N} \sum_{i=1}^N \int \sigma^2(l, \theta) f_N(l, \mu + \hat{l}_i, \sigma_{g,l}^2 + \sigma_{RE}^2 + \sigma_o^2) dl. \quad (15)$$

Vale ressaltar que, além das fórmulas apresentadas por Villemereuil et al. (2016), a literatura apresenta, fórmulas “fechadas” da variância fenotípica para modelos cuja variável é binária – aqui denominada variância fenotípica fechada (σ_{ff}^2) – e que pode ser calculada como (NAKAGAWA e SCHIELZETH, 2010).

$$\sigma_{ff}^2 = \delta + \varphi \times \frac{\pi^2}{3}, \quad (16)$$

para o GLM cuja variável é binária e a função de ligação é a logit.

E,

$$\sigma_{ff}^2 = \delta + \varphi \times 1, \quad (17)$$

para o GLM cuja variável é binária e a função de ligação é a probit.

Nos dois casos, δ representa os componentes de variância de todos os efeitos aleatórios do modelo, φ é o parâmetro de dispersão que está relacionado à variância da distribuição e, os valores $\frac{\pi^2}{3}$ (Equação 16) e 1 (Equação 17) se referem à variação da função de ligação que representará a variância residual (NAKAGAWA e SCHIELZETH, 2010).

Variância genética aditiva na escala dos dados observados

A variância genética aditiva na escala dos dados observados, será determinada, a partir de dois elementos: a variância genética aditiva latente (σ_{gl}^2) e um parâmetro denominado Ψ , que está relacionado à derivada média do valor esperado em relação à variável latente l (MORISSEY, 2015; VILLEMEREUIL et al., 2016).

Deste modo, a variância genética aditiva na escala dos dados observados (σ_{go}^2) pode ser calculada da seguinte maneira:

$$\sigma_{go}^2 = \Psi^2 \sigma_{gl}^2. \quad (18)$$

Herdabilidade

Com os valores já obtidos na escala observada, para a variância genética (σ_{go}^2) e fenotípica (σ_{fo}^2), a herdabilidade na escala dos dados observados (h_o^2) pode ser calculada por:

$$h_o^2 = \frac{\sigma_{go}^2}{\sigma_{fo}^2}. \quad (19)$$

Considerando as fórmulas “fechadas” da variância fenotípica para modelos cuja variável é binária, tem-se, portanto, a herdabilidade “fechada” que pode ser calculada como:

$$h_f^2 = \frac{\sigma_{gf}^2}{\sigma_{ff}^2}. \quad (20)$$

Podendo, σ_{ff}^2 ser calculada a partir da Equação 16 ou Equação 17, a depender do modelo adotado. Assim, a herdabilidade se altera dependendo da respectiva distribuição, pois o parâmetro de dispersão φ se altera visto que está relacionado a variância da distribuição utilizada (NAKAGAWA e SCHIELZETH, 2010).

3. MATERIAL E MÉTODOS

3.1. Material genético

Os dados utilizados neste estudo foram coletados de uma população composta por 51 clones pertencentes ao grupo varietal Conilon. A população utilizada compõe o programa de melhoramento genético da Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig), em parceria com a Universidade Federal de Viçosa (UFV) e da Empresa Brasileira de Pesquisa Agropecuária-Café (Embrapa Café), localizada em Oratórios/MG e Viçosa/MG. O material genético do Conilon foi obtido no Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper).

3.2. Avaliações Fenotípicas

A população foi testada em um delineamento de blocos incompletos com até 35 repetições e parcelas de árvore única, incluindo híbridos e pais. As avaliações foram realizadas durante três anos consecutivos (2014-2016) (ALKIMIM et al., 2020).

Neste estudo, foi considerado o caractere não normal associado à incidência de ferrugem (Fer), sendo que as avaliações foram realizadas no momento em que havia a maturação fisiológica dos frutos do café. Inicialmente, as incidências de ferrugem foram medidas por atribuição de “notas”, variando de 1 a 5, sendo a nota 1 atribuída a genótipos sem sintomas do patógeno e a nota 5 atribuída a genótipos altamente suscetíveis (ALKIMIM et al, 2020). Assim como realizado em Sousa et al. (2021), as notas de 1 a 2 foram codificadas como sendo 1 (1= resistente) e as notas de 3 a 5 foram codificadas como sendo iguais a 0 (0 = susceptível), deste modo a mesma foi analisada como uma variável binária que se refere à resistência à ferrugem do cafeeiro.

3.3. Modelos Estatísticos

Para estimação de parâmetros genéticos, um dos métodos estatísticos que se destaca é o REML/BLUP (*Residual or Restricted Maximum Likelihood/Best Linear Unbiased Prediction* – PATTERSON; THOMPSON, 1971). E, isso se dá, pois o mesmo permite a previsão de valores genéticos associados a observações fenotípicas, ajustando os dados para efeitos fixos e informações desiguais para cada fração (RESENDE, 2002).

Neste trabalho, foram ajustados três diferentes modelos: i) M1 – Normal: modelo linear misto que considera a variável resposta resistência à ferrugem (y , 0 e 1) como tendo distribuição Normal; ii) M2 – Binomial/Logit: modelo linear generalizado misto que considera a variável resposta y como seguindo distribuição binomial e função de ligação logit; e, iii) M3 – Binomial/Probit: modelo linear generalizado misto que considera a variável resposta y como seguindo distribuição binomial e função de ligação probit. Os modelos ajustados estão apresentados a seguir.

$$\text{M1 - Normal: } y_{ijk} = \mu + a_j + b_k + g_i + ab_{jk} + e_{ijk},$$

em que y_{ijk} é o vetor de dados fenotípicos relacionado a resistência à ferrugem mensurada de forma binária (1 = resistente e 0 = susceptível), μ é a média geral, a_j é o efeito fixo do j -ésimo ano, b_k é o efeito aleatório do k -ésimo bloco, g_i é o efeito aleatório dos genótipos, ab_{jk} é o efeito aleatório da interação entre blocos e anos e, e_{ijk} é o efeito aleatório residual.

M2 – Binomial/Logit: Y | efeitos fixos, efeitos aleatórios \sim Binomial (n, p)

$$\log\left(\frac{p_i}{1-p_i}\right) = \mu + a_j + b_k + g_i + ab_{jk},$$

em que n e p_i são os parâmetros da distribuição binomial; p representa a probabilidade do genótipo ser resistente à ferrugem do cafeeiro, μ é a média geral; a_j é o efeito fixo do j -ésimo ano; b_k é o efeito aleatório do k -ésimo bloco; g_i é o efeito aleatório dos genótipos; ab_{jk} é o efeito aleatório da interação entre blocos e anos.

M3 – Binomial/Probit: $Y \mid$ efeitos fixos, efeitos aleatórios \sim Binomial (n, p)

$$\Phi^{-1}(\mu) = \mu + a_j + b_k + g_i + ab_{jk},$$

em que Φ^{-1} representa a função distribuição acumulada da normal, μ é a média geral; a_j é o efeito fixo do j -ésimo ano; b_k é o efeito aleatório do k -ésimo bloco; g_i é o efeito aleatório dos genótipos; ab_{jk} é o efeito aleatório da interação entre blocos e anos.

3.3.1. Componentes de variância, herdabilidade e ganho de seleção

Nos GLMMs é necessário definir o parâmetro de dispersão φ que está relacionado à variância da distribuição adotada. Neste estudo, as estimativas foram obtidas considerando $\varphi = 1$ ou seja, sem a presença de sub ou superdispersão. Assim, as variâncias fenotípicas na escala fechada, de acordo com Nakagawa e Schielzeth (2010), são:

$$\text{M2 – Binomial/Logit: } \sigma_{ff}^2 = b_k + g_i + ab_{jk} + \varphi \frac{\pi^2}{3},$$

em que $\varphi = 1$, b_k é o efeito aleatório do k -ésimo bloco, g_i é o efeito aleatório dos genótipos, ab_{jk} é o efeito aleatório da interação entre blocos e anos e $\frac{\pi^2}{3}$ é a variância específica da função de ligação logit.

$$\text{M3 – Binomial/Probit: } \sigma_{ff}^2 = b_k + g_i + ab_{jk} + \varphi \times 1,$$

em que $\varphi = 1$, b_k é o efeito aleatório do k -ésimo bloco, g_i é o efeito aleatório dos genótipos, ab_{jk} é o efeito aleatório da interação entre blocos e anos e 1 é a variância específica da função de ligação probit.

Entretanto, para a escala observada, a variância fenotípica utilizada é:

$$\sigma_{fo}^2 = \sigma_{f,exp}^2 + \frac{1}{N} \sum_{i=1}^N \int \sigma^2(l, \theta) f_N(l, \mu + \hat{l}_i, \sigma_{g,l}^2 + \sigma_{RE}^2 + \sigma_o^2) dl, \quad (22)$$

visto que o modelo possui efeitos fixos.

A variância genética na escala observada é:

$$\sigma_{go}^2 = \Psi^2 \sigma_{gl}^2. \quad (23)$$

A herdabilidade para o modelo M1 em que a variável resposta é considerada Normal é denominada herdabilidade latente:

$$h_l^2 = \frac{\sigma_{gl}^2}{\sigma_{fl}^2}; \quad (23)$$

para o caso que foi considerado as variâncias genética e fenotípica da escala observada, a herdabilidade obtida será a observada:

$$h_o^2 = \frac{\sigma_{go}^2}{\sigma_{fo}^2}. \quad (24)$$

De forma análoga, utilizando as variâncias da escala fechada, obtém-se a herdabilidade fechada:

$$h_f^2 = \frac{\sigma_{gf}^2}{\sigma_{ff}^2}; \quad (25)$$

E, conseqüentemente, o ganho de seleção também será específico para cada escala. Isto é, para a escala latente, tem-se:

$$GS_l = (\bar{X}_s - \bar{X}_o) h_l^2 = DS \times h_l^2; \quad (26)$$

Para a escala observada:

$$GS_o = (\bar{X}_s - \bar{X}_o) h_o^2 = DS \times h_o^2; \quad (27)$$

e, para a escala fechada:

$$GS_f = (\bar{X}_s - \bar{X}_o) h_f^2 = DS \times h_f^2. \quad (28)$$

As análises foram realizadas no software R versão 4.3.1 (RCORE TEAM, 2023), utilizando o pacote ASReml-R, versão 4.1 (BUTLER et al., 2018) para ajuste dos modelos, e o pacote QGglmm (de VILLEMEREUIL, 2020) para computar os parâmetros genéticos na escala observada.

4. RESULTADOS E DISCUSSÃO

Este trabalho teve como objetivo a estimação de parâmetros genéticos da característica relacionada a resistência à ferrugem da variedade Conilon de *Coffea Canephora*, avaliada de forma binária (1 = resistente e 0 = susceptível). Para o conjunto de clones avaliados, 74% das observações foram classificadas como resistentes à ferrugem do cafeeiro e o restante 26%, como susceptíveis à esta doença. A maior porcentagem de indivíduos resistentes está possivelmente relacionada ao fato desta variedade ser considerada como mais resistente à ferrugem em relação à variedade arábica (MAIA et al. 2017).

Os resultados obtidos para variância genética, herdabilidade e ganho de seleção pelos diferentes modelos ajustados estão apresentados na Tabela 1.

Tabela 1. Estimativas dos parâmetros genéticos da variável resistência à ferrugem para a variedade Conilon pelas diferentes abordagens avaliadas

Parâmetros Genéticos	Modelos		
	Normal (M1)	Binomial/Logit (M2)	Binomial/Probit (M3)
σ_{gl}^2	0,02275	0,90807	0,29635
σ_{fl}^2	0,16453	1,94869	1,30977
h_l^2	0,13825	0,46599	0,22626
GS_l	0,01579	0,33794	0,08940
σ_{go}^2	0,02275	0,03020	0,02037
σ_{fo}^2	0,16453	0,24960	0,24964
h_o^2	0,13825	0,12102	0,08161
GS_o	0,01579	0,08776	0,03225
σ_{gf}^2	0,02275	0,90807	0,29635
σ_{ff}^2	0,16453	4,23856	1,30977
h_f^2	0,13825	0,21424	0,22626
GS_f	0,01579	0,15537	0,08940

σ_g^2 – variância genética; σ_f^2 – variância fenotípica; h^2 – herdabilidade; GS – ganho de seleção. Os índices l referem-se aos resultados em escala latente, os com índices o referentes aos resultados em escala observada e os com índice f são referentes aos parâmetros obtidos usando as fórmulas fechadas, ou seja, já calculados com as respectivas variâncias específicas de cada distribuição.

Devido às distintas escalas empregadas - por se considerar diferentes distribuições de probabilidade e funções de ligação - como previsto, observou-se grande variação nos valores dos parâmetros genéticos em função dos modelos utilizados e da distribuição de probabilidade assumida para a variável em estudo (Tabela 1).

Especificamente, observou-se que as variâncias genéticas na escala latente (σ_{gl}^2) variaram de 0,02275 (M1) à 0,90807 (M2). A discrepância dos valores observados deve-se, principalmente à diferença entre as escalas admitidas (devido a diferenças nas distribuições e funções de ligação consideradas). Isto é, no modelo M1, a escala latente é a mesma dos dados observados, uma vez que a função de ligação no modelo Normal é a função identidade. Por outro lado, em M2, a função de ligação é definida em escala logarítmica, o que faz com que não seja adequada a interpretação direta deste valor. Neste caso, ao utilizar um modelo GLMM e, erroneamente, interpretar o valor obtido para a variância genética na escala latente, tem-se que este valor estará superestimando o verdadeiro valor da variância genética, e conseqüentemente, os efeitos dos fatores genéticos sobre a característica em estudo.

Dito isto, a avaliação da variância genética no modelo GLMM deve ser realizada na escala dos dados observados, podendo-se utilizar as fórmulas propostas por Villemeuril et al. (2016) (σ_{go}^2). Neste caso, observa-se que houve uma menor variação entre as variâncias fornecidas pelos diferentes modelos avaliados, tendo os valores variado de 0,02037 (M3) à 0,03020 (M2) (Tabela 1). Estando, o valor de variância genética apresentado pelo modelo Normal (M1) entre aqueles observados para os GLMM (Tabela 1).

Para esta mesma variedade, Alkimin et al. (2017), utilizando um modelo LMM encontrou uma variância genética 0,10, aproximadamente cinco vezes maior do que a observada neste trabalho para o modelo M1 ($\sigma_{gl}^2 = 0,02275$). No entanto, destaca-se que, essa diferença pode estar associada à forma de mensuração da característica, que aqui foi avaliada de forma binária e, no estudo de Alkimin et al. (2017), foi realizada com base em notas, que variaram de 1 a 5, onde a nota 1 foi atribuída para genótipos sem nenhum sintoma do patógeno e nota 5 para genótipos altamente suscetíveis. A maior variabilidade genética na escala observada (σ_{go}^2) foi aquela obtida pelo modelo M2 (0,03020), que considera a função de ligação logit.

Assim como as variâncias genéticas, nos GLMMs as variâncias fenotípicas são obtidas na escala latente. Assim, para uma correta interpretação dos valores obtidos, a literatura apresenta duas abordagens distintas para transformação na escala original dos dados. A primeira, baseada nas fórmulas propostas por Villemeuril et al. (2016) (σ_{fo}^2) e, a segunda, baseada nas fórmulas fechadas apresentadas por Nakagawa e Schielzeth (2010) e que levam em conta a variância da distribuição assumida (σ_{ff}^2).

Para a variância fenotípica, observou-se ainda mais discrepância entre os valores dos modelos avaliados, destacando a importância das diferenças nas escalas admitidas e distribuições consideradas. Especificamente, a variância fenotípica (σ_f^2) no modelo normal M1 foi igual 0,16453. Enquanto, nos modelos M2 e M3, foram, respectivamente, na escala latente (σ_{fl}^2), iguais a 1,94869 e 1,30977; na escala observada (σ_{fo}^2) foram 0,24960 e 0,24964 e, por fim, na escala fechada σ_{ff}^2 foram iguais a 4,23856 e 1,30977 (Tabela 1). Quando se compara o valor da σ_f^2 em M1 com os valores dos demais modelos na escala observada (σ_{fo}^2), observa-se que, assim como ocorreu com a variância genética, os modelos GLMM apresentam valor de variância fenotípica superior ao LMM (M1) e, aqui os valores apresentados pelos modelos GLMM foram menos discrepantes entre si,

quando se considerou a transformação baseada na proposta de Villemeuril et al. (2016). Mais especificamente, neste caso, a $\sigma_{f_o}^2$ nos modelos GLMM (M2 e M3) apresentaram estimativas muito semelhantes para a variância fenotípica e estes valores foram, aproximadamente, 51,73% maiores que no modelo Normal (Tabela 1). Por outro lado, quando se utiliza as variâncias fenotípicas calculadas com base nas fórmulas apresentadas por Nakagawa e Schielzeth (2010), observa-se uma maior discrepância dos valores entre os dois modelos GLMM e, estes sendo, no mínimo, quase 10 vezes maiores que os do modelo Normal (M1). É razoável esperar que os valores encontrados tanto para a variância genotípica quanto para a fenotípica sejam inferiores aos encontrados na literatura, visto que neste trabalho a variável foi discretizada e, assim a sua variabilidade consequentemente foi reduzida. Neste caso, os valores encontrados no modelo normal (M1) para a variância genotípica e fenotípica foram respectivamente 0,02275 e 0,16453.

Assim como para os demais parâmetros genéticos, as herdabilidades variaram de acordo com o modelo e a escala adotados, entretanto, neste caso, a variação entre os valores observados foi menor. Como esperado, a maior variação entre as herdabilidades dos diferentes modelos ocorre quando se compara os valores na escala latente (Tabela 1). Para o modelo M1 não há distinção entre a herdabilidade na escala latente (h_l^2), a escala original (h_o^2) e a escala fechada (h_f^2), uma vez que o LMM é equivalente ao GLMM com distribuição Normal e função de ligação identidade, onde as escalas são as mesmas. Neste estudo, a herdabilidade da característica resistência à ferrugem na variedade Conilon foi de 0,13825, considerada baixa, segundo Resende (1997). Por outro lado, a herdabilidade para o M2 que na escala latente ($h_l^2 = 0,46599$) foi, aproximadamente, 385% maior em comparação à medida na escala observada ($h_o^2 = 0,12102$) e, 217,5% maior em comparação à medida na escala fechada ($h_f^2 = 0,21424$). E, no M3, a herdabilidade na escala latente ($h_l^2 = 0,22626$) é, aproximadamente, 277,24% maior em comparação à medida na escala observada (h_o^2) e, quando comparada com a escala fechada (h_f^2), ambas apresentaram o mesmo valor ($h_o^2 = h_f^2 = 0,22626$). Não interpretar o valor obtido na escala latente evita que resultados sejam superestimados e/ou subestimados, além de estar garantindo uma maior precisão nos parâmetros levando assim a resultados mais realistas.

Comparando as herdabilidades na escala original dos dados, o valor estimado em M1 ($h_o^2 = 0,13825$) foi superior aos encontrados em ambos os modelos (M2 e M3), em contrapartida, as herdabilidades em escala fechada (h_f^2) nos dois modelos (M2 e M3) foram superiores aquela observada no M1.

Quando comparamos os parâmetros entre os modelos M1 e M2 ou M1 e M3, estamos avaliando, principalmente a distribuição assumida para a variável resposta. No entanto, ao comparar os modelos M2 e M3, a análise incide principalmente sobre o efeito da escolha da função de ligação, já que ambos os modelos consideram a distribuição binomial. Os valores de herdabilidade na escala original, para os modelos M1 ($h_o^2 \cong 0,138$) e M2 ($h_o^2 \cong 0,121$ e $h_f^2 \cong 0,214$) avaliados, são próximo àqueles observados por Alkimim et al. (2017) foi $h^2 = 0,14$. E, para o modelo M3, o valor foi inferior ($h_o^2 \cong 0,082$ e $h_f^2 \cong 0,226$) àquele relatado por estes autores. Vale ressaltar que Nakagawa e Schielzeth, (2010) sugerem que em dados binários e de proporção, o ideal é usar GLMMs com função de ligação logit, pelo fato de a ligação logit estabilizar a variância, fazendo assim com que a estimativa na escala vinculada seja adequada, tal como na escala original.

Ferrão et al. (2019) ao calcularem a herdabilidade ampla da incidência de ferrugem do café, usando notas em uma escala de 1 a 9, em dois locais (FEM e FES) e duas populações de seleção recorrente de *Coffea canephora* (Intermediário e Prematuro) obtiveram uma variação de herdabilidade de 0,56 a 0,89, utilizaram uma estrutura de modelo misto com integração de estruturas de variância-covariância espacial e temporal e, não fizeram o cálculo dos parâmetros em diferentes escalas como neste trabalho. Além destes, Sousa et al. (2019) avaliaram genótipos de café arábica com modelos mistos e, observaram o valor de 0,31 como a menor herdabilidade para a resistência à ferrugem, sendo utilizada uma escala de pontuação que variou de 1 a 5, em que 1 correspondia a genótipos sem sintomas e 5 referia-se a genótipos altamente suscetíveis. As diferenças entre os processos – mensuração das variáveis e escolha dos modelos - podem justificar as discrepâncias nos valores encontrados.

O ganho de seleção na escala latente (GS_l), no M2 é, aproximadamente, 385% maior em comparação à medida na escala observada (GS_o) e, 217,5% maior em comparação à medida na escala fechada (GS_f). E, no M3, o ganho de seleção na escala latente (GS_l) é, aproximadamente, 277,2% maior em comparação à medida na escala observada (GS_o) e, quando comparada com a escala fechada (GS_f), ambas apresentaram o mesmo valor (0,0894). Ao comparar o ganho de seleção (GS) do modelo M1 com o GS_o e GS_f em ambos os modelos (M2 e M3), observa-se um aumento significativo nos valores corrigidos. Isso indica que calcular o parâmetro com a devida correção resulta em um valor de ganho genético maior do que aquele observado quando avaliado erroneamente (ao desconsiderar a não normalidade dos caracteres avaliados). Um valor

de ganho genético maior, indica que há uma maior variação genética entre os indivíduos dessa população, e que também essa variação é herdada de uma geração para outra de forma mais marcante.

Assim, é crucial exercer cautela ao lidar com dados não normais. Nesses casos, é imperativo ajustar a escala dos dados para garantir maior precisão nos parâmetros e evitar a superestimação ou subestimação nos resultados. Isso ajuda a prevenir conclusões equivocadas que poderiam ser erroneamente alcançadas.

5. CONCLUSÃO

A estimação dos parâmetros genéticos de variáveis não normais utilizando modelos que assumem normalidade pode gerar resultados subestimados para as variâncias e superestimados para herdabilidade e ganho de seleção. Por outro lado, é fundamental utilizar a escala adequada para interpretar os resultados dos modelos GLMMs (Binomial/Logit e Binomial/Probit), pois os resultados obtidos na escala latente, superestimam aqueles na escala observada para todos os parâmetros genéticos, que é a escala em que os dados foram coletados. Superestimar o potencial de resposta à seleção de uma característica pode levar à alocação indevida de recursos em pesquisas para essa característica, em detrimento de outras com maior potencial real de aprimoramento. Além disso, programas de melhoramento baseados em dados superestimados da herdabilidade e do ganho de seleção podem ser ineficazes e dispendiosos, gerando resultados abaixo das expectativas. Ao comparar os resultados dos diferentes Modelos Lineares Generalizados Mistos, observa-se que, de modo geral, a função de ligação logit produziu estimativas mais elevadas em comparação com a função de ligação probit. Com relação ao tipo de correção de escala, observou-se que, a utilização das fórmulas fechadas produziu valores mais elevados do que a correção baseada nas integrações propostas por Villemeuril et al. (2016), o que pode ter implicações importantes na interpretação dos resultados e na tomada de decisões em programas de melhoramento genético. Os parâmetros genéticos da ferrugem do café indicam que a resistência a essa doença pode ser significativamente melhorada por meio de programas de melhoramento genético, focando na seleção de variedades com genes de resistência específicos.

Portanto, este estudo reforça a importância da utilização de abordagens estatísticas adequadas, como os GLMMs com correção de escala, para garantir resultados mais confiáveis e precisos. Essa prática não apenas contribui para um entendimento mais preciso da variação genética subjacente às características estudadas, mas também

fortalece a base para o desenvolvimento de estratégias de melhoramento genético mais eficientes e eficazes no cultivo de *Coffea Canephora*.

REFERÊNCIAS BIBLIOGRÁFICAS

- A. R. GILMOUR, R. D. ANDERSON, A. L. RAE, The analysis of binomial data by a generalized linear mixed model, *Biometrika*, Volume 72, Issue 3, December 1985, Pages 593–599, <https://doi.org/10.1093/biomet/72.3.593>
- ALKIMIM, E. R.; CAIXETA, E. T.; SOUSA, T. V.; RESENDE, M. D. V.; DA SILVA, F. L.; SAKIYAMA, N. S.; ZAMBOLIM, L. Selective efficiency of genome-wide selection in *Coffea canephora* breeding. **Tree Genetics & Genomes**, v. 16, n. 3, p. 1-11, 2020
- ALKIMIM, E. R., DIVERSIDADE GENÉTICA, GANHOS COM SELEÇÃO E SELEÇÃO GENÔMICA AMPLA NA ESPÉCIE *Coffea canephora*. Tese, Doctor Scientiae, Universidade Federal de Viçosa, Viçosa, 2017.
- AZEVEDO, C.F., FERRÃO, L.F.V., BENEVENUTO, J. *et al.* Using visual scores for genomic prediction of complex traits in breeding programs. *Theor Appl Genet* **137**, 9 (2024). <https://doi.org/10.1007/s00122-023-04512-w>
- AYERS D R, PEREIRA R J, BOLIGON A A, SILVA F F, SCHENKEL F S *et al.* , 2013 Linear and Poisson models for genetic evaluation of tick resistance in cross-bred Hereford x Nellor cattle. *J. Anim. Breed. Genet.* 130: 417–424
- BABOVA O, OCCHIPINTI A, MAFFEI ME (2016) Chemical partitioning and antioxidant capacity of green coffee (*Coffea arabica* and *Coffea canephora*) of different geographical origin. *Phytochemistry* 123:33–39. doi: 10.1016/j.phytochem.2016.01.016
- BEYENE ZEWDIE, YVES BAWIN, AYCO J. M. Tack, Sileshi Nemomissa, Kassahun Tesfaye, Steven B. Janssens, Sabine Van Glabeke, Isabel Roldán-Ruiz, Tom Ruttink, Olivier Honnay, Kristoffer Hylander. *Molecular Ecology*, 04 April 2022, <https://doi.org/10.1111/mec.16458>
- BOLKER, B. M.; BROOKS, M. E.; CLARK, C. J.; GEANGE, S. W.; POULSEN, J. R.; STEVENS, M. H. H.; WHITE, J. S. S. Generalized linear mixed models: a practical guide for ecology and evolution. **Trends in ecology & evolution**, v. 24, n. 3, p. 127-135, 2009.

- BUTLER, D. G., CULLIS, B. R., GILMOUR, A. R., GOGEL, B. J., & THOMPSON, R. (2018). *ASReml-R reference manual Version 4*. VSN International. <https://mmade.org/wp-content/uploads/2019/01/asremlRMfinal.pdf>
- CARVALHO, H.F., FERRÃO, L.F.V., GALLI, G. *et al.* On the accuracy of threshold genomic prediction models for leaf miner and leaf rust resistance in arabica coffee. *Tree Genetics & Genomes* **19**, 11 (2023). <https://doi.org/10.1007/s11295-022-01581-8>
- CARIAS CM de OM, GRAVINA G do A, FERRÃO MAG, et al (2016) Predição de ganhos genéticos via modelos mistos em progênies de café Conilon. *Coffee Sci* 11:39–45.
- CHAVES, S.F.S., DIAS, L.A.S., ALVES, R.S. *et al.* Leveraging multi-harvest data for increasing genetic gains per unit of time for fruit yield and resistance to witches' broom in *Theobroma grandiflorum*. *Euphytica* 218, 171 (2022). <https://doi.org/10.1007/s10681-022-03126-5>
- CORDEIRO, G.; DEMÉTRIO, C.; MORAL, R. Modelos Lineares Generalizados e Extensões. Piracicaba, 2018.
- CORDEIRO, G. M. (1986). Modelos lineares generalizados. VII SINAPE, UNICAMP.
- COSTA, R.B., RESENDE, M.D.V., SILVA, V.S.M. (2007). Experimentação e seleção no melhoramento genético de TECA (*Tectona grandis* L.f.). **Floresta e ambiente**, v.14, n.1, 76-92p.
- CRUZ, C. D. Princípios de genética quantitativa. 1. Ed. Viçosa: UFV, 2005. 394p.
- De LOS CAMPOS, G., SORENSEN, D., e GIANOLA, D. (2015). Genomic heritability: what is it?. *PLoS Genet*, 11(5). DOI: <http://dx.doi.org/10.1371/journal.pgen.1005048>
- DELLY ROCÍO GARCÍA-CÁRDENAS, INGE ARMBRECHT, WILMAR TORRES, MARTHA L. BAENA, JAMES MONTOYA-LERMA, Functional significance of the leaf-cutting ant *Atta cephalotes* (Formicidae) in coffee plantations: An enemy or an ally?, *Pedobiologia*, Volumes 93–94, 2022, 150825, ISSN 0031-4056, <https://doi.org/10.1016/j.pedobi.2022.150825>.
- DEMÉTRIO, C. G. B. (2001). Modelos Lineares Generalizados em Experimentação Agrônômica. ESALQ/USP, Piracicaba.

- DEMÉTRIO, C. G. P., (2002). Modelos lineares generalizados em experimentação agrônômica. (Apostila) Escola Superior de Agricultura Luiz de Queiroz Departamento de Ciências Exatas LCE USP, Piracicaba, SP. Available at: <<http://ce.esalq.usp.br/clarice/Apostila.pdf>>.
- DENOEUDE F, CARRETERO-PAULET L, DEREPPER A, et al (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* (80-) 345:1181–1184. doi: 10.1126/science.1255274
- de VILLEMEREUIL, PIERRE (2022). Estimate Quantitative Genetics Parameters from Generalised Linear Mixed Models. **Package ‘QGglmm’** version 0.7.4
- de VILLEMEREUIL P, GIMENEZ O, DOLIGEZ B, 2013. Comparing parent–offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits. *Methods Ecol. Evol.* 4(3): 260–275.
- de VILLEMEREUIL P, 2020. How to use the QGglmm package?
- DOBSON, A. J.; BARNETT, A. G. (2008). An Introduction to Generalized Linear Models. Chapman & Hall/CRC, London, third edition.
- dos ANJOS, R.S.R., MARÇAL, T.D.S., CARNEIRO, P.C.S. AND CARNEIRO, J.E.d.S. (2019), New Proposals to Estimate Unbiased Selection Gain and Coefficient of Variation in Traits Evaluated Using Score Scales. *Crop Science*, 59: 937-944. <https://doi.org/10.2135/cropsci2018.08.0481>
- Exportação. CECAFÉ (Conselho dos Exportadores de Café do Brasil),2023. Disponível em: <https://www.cecafe.com.br/sobre-o-cafe/exportacao/>. Acesso em 04/05/2023
- FALCONER DS, MACKAY TFC. Introduction to Quantitative Genetics. 4th edition. Harlow, UK: Pearson Education Limited; 1996
- FAO (Organização das Nações Unidas para Agricultura e Alimentação), 2023. Disponível em: <https://www.fao.org/brasil/pt/>. Acesso em 03/05/2023
- FERRÃO, L.F.V., FERRÃO, R.G., FERRÃO, M.A.G. *et al.* Accurate genomic prediction of *Coffea canephora* in multiple environments using whole-genome statistical models. *Heredity* **122**, 261–275 (2019). <https://doi.org/10.1038/s41437-018-0105-y>

- FISHER R.A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 52: 399–433. HARDIN, J. W.; HILBE, J. M. (2007). *Generalized Linear Models and Extensions*. Stata, Texas, 2^{ed}ition.
- FOULLEY, J. L.; GIANOLA, D.; IM, S. (1987). Genetic evaluation of traits distributed as Poisson-binomial with reference to reproductive characters. **Theoretical and Applied Genetics**, v. 73, n. 6, p. 870-877, 1987.
- GIANOLA, D. Theory and analysis of threshold characters. **Journal of Animal Science**, v. 54, n. 5, p. 1079-1096, 1982.
- GIANOLA, D.; FOULLEY, J. L. Sire evaluation for ordered categorical data with a threshold model. **Génétique, Sélection, Évolution**, v. 15, n. 2, p. 1, 1983.
- HENDERSON C.R. (1950). Estimation of genetic parameters. *Annals of Mathematical Statistics* 21: 309–310.
- HENDERSON, C. R. “Best Linear Unbiased Estimation and Prediction under a Selection Model.” *Biometrics*, vol. 31, no. 2, 1975, pp. 423–47. *JSTOR*, <https://doi.org/10.2307/2529430>. Accessed 20 Nov. 2023.
- LYNCH M, WALSH B, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- OLIVEIRA EJ, SANTANA FA, OLIVEIRA LA, SANTOS VS (2014) Genetic parameters and prediction of genotypic values for root quality traits in cassava using REML/BLUP. *Genet Mol Res* 13:6683–6700. doi: 10.4238/2014.August.28.13
- PATTERSON HD, THOMPSON R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554. doi: 10.1093/biomet/58.3.545
- PIERRE DE VILLEMEREUIL, HOLGER SCHIELZETH, SHINICHI NAKAGAWA, MICHAEL MORRISSEY, General Methods for Evolutionary Quantitative Genetic Inference from Generalized Mixed Models, *Genetics* , Volume 204, Issue 3, 1 de novembro de 2016, páginas 1281–1294, <https://doi.org/10.1534/genetica.115.186536>
- MAIA T, BADEL JL, FERNANDES MB, ET al (2017) Variation in Aggressiveness Components in the *Hemileia vastatrix* Population in Brazil. *J Phytopathol* 165:174–188. doi: 10.1111/jph.12548

- MAKOUANZI, G., BOUVET, JM., DENIS, M. *et al.* Assessing the additive and dominance genetic effects of vegetative propagation ability in *Eucalyptus*—influence of modeling on genetic gain. *Tree Genetics & Genomes* 10, 1243–1256 (2014). <https://doi.org/10.1007/s11295-014-0757-6>
- MCCULLAGH, P.; NELDER, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, London, second edition.
- McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*. Wiley.
- MORRISSEY M B, WALLING C A, WILSON A J, PEMBERTON J M, CLUTTON-BROCK T H *et al.* , 2012 Genetic analysis of life-history constraint and evolution in a wild ungulate population. *Am. Nat.* 179(4): E97–E114.
- MORRISSEY M B, 2015 Genética quantitativa evolutiva de sistemas de desenvolvimento não lineares. *Evolução* 69:2050–2066.
- NAKAGAWA, S. AND SCHIELZETH, H. (2010), Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, 85: 935-956. <https://doi.org/10.1111/j.1469-185X.2010.00141.x>
- NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models. *Journal of the Royal Statistical Society, A* 135, p. 370-384, 1972.
- R CORE TEAM (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical
- RESENDE, M.D.V., FURLANI-JÚNIOR, E., MORAES, M.L.T., FAZUOLI, L.C. (2001). Estimativas de parâmetros genéticos e predição de valores genotípicos no melhoramento do cafeeiro pelo procedimento REML/BLUP. **Bragantia**, Campinas, 185-193p
- RESENDE, M. D. V. DE; SILVA, F. F.; 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**. Suprema ed. Viçosa, 2014. RESENDE, M. D. V. *Genética biométrica e estatística no melhoramento de plantas perenes*. Brasília: Embrapa Informação Tecnológica, 2002. 975p

- SERA, T. Coffee genetic breeding at IAPAR. **Crop Breeding and Applied Biotechnology**, v.1, n.2, p.179-190, jun. 2001. DOI: 10.13082/1984-7033.v01n02a08.
- SPINELLI VM, DIAS LAS, ROCHA RB, RESENDE MD V. (2015) Estimates of genetic parameters with selection within and between half-sib families of *Jatropha curcas* L. *Ind Crops Prod* 69:355– 361. doi: 10.1016/j.indcrop.2015.02.024
- SOUSA, I. C. DE ., NASCIMENTO, M., SILVA, G. N., NASCIMENTO, A. C. C., CRUZ, C. D., SILVA, F. F. E ., ALMEIDA, D. P. DE ., PESTANA, K. N., AZEVEDO, C. F., ZAMBOLIM, L., & CAIXETA, E. T.. (2021). Genomic prediction of leaf rust resistance to Arabica coffee using machine learning algorithms. *Scientia Agricola*, 78(4), e20200021. <https://doi.org/10.1590/1678-992X-2020-0021>
- SOUSA, T.V. et al. Early Selection Enabled by the Implementation of Genomic Selection in *Coffea arabica* Breeding .*Front. Plant Sci.*, 08 January 2019. Disponível em: <https://doi.org/10.3389/fpls.2018.01934>.
- TALHINHAS P, BATISTA D, DINIZ I, VIEIRA A, SILVA DN, LOUREIRO A, TAVARES S, PEREIRA AP, AZINHEIRA HG, GUERRA-GUIMARÃES L, VÁRZEA V, SILVA MDC. The coffee leaf rust pathogen *Hemileia vastatrix*: one and a half centuries around the tropics. *Mol Plant Pathol*. 2017 Oct;18(8):1039-1051. doi: 10.1111/mpp.12512. Epub 2017 Jan 25. PMID: 27885775; PMCID: PMC6638270.
- TRAN HTM, LEE LS, FURTADO A, SMYTH H, HENRY RJ (2016) Avanços em genômica para a melhoria da qualidade do café. *J Sci Food Agric* 96(10):3300–3312 <https://doi.org/10.1002/jsfa.7692>
- VEGA FE, ROSENQUIST E, COLLINS W (2003) Global project needed to tackle coffee crisis. *Nature* 425:343. doi: 10.1038/425343a
- WILSON A J, MORRISSEY M B, ADAMS M J, WALLING C A, GUINNESS F E ET AL. , 2011 Indirect genetics effects and evolutionary constraint: an analysis of social dominance in red deer, *Cervus elaphus*. *J. Evol. Biol.* 24(4): 772–783.
- ZAMBOLIM L (2016) Current status and management of coffee leaf rust in Brazil. *Trop Plant Pathol* 41:1–8. <https://doi.org/10.1007/s40858-016-0065-9>