

RENAN GARCIA MALIKOUSKI

**MÉTODOS ESTATÍSTICOS NA AVALIAÇÃO DA REPETIBILIDADE
GENOTÍPICA EM LIMA ÁCIDA ‘TAHITI’**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

Orientador: Leonardo Lopes Bhering

**VIÇOSA - MINAS GERAIS
2023**

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

T

M251m
2023

Malikouski, Renan Garcia, 1993-

Métodos estatísticos na avaliação da repetibilidade
genotípica em lima ácida 'Tahiti' / Renan Garcia Malikouski. –
Viçosa, MG, 2023.

1 tese eletrônica (79 f.): il. (algumas color.).

Texto em português e inglês.

Orientador: Leonardo Lopes Bhering.

Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Biologia Geral, 2023.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2023.183>

Modo de acesso: World Wide Web.

1. *Citrus latifolia*. 2. Lima ácida taiti - Melhoramento
genético. 3. Modelos multiníveis (Estatísticas). 4. Inferência
estatística. I. Bhering, Leonardo Lopes, 1980-. II. Universidade
Federal de Viçosa. Departamento de Biologia Geral. Programa
de Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed.

RENAN GARCIA MALIKOUSKI

**MÉTODOS ESTATÍSTICOS NA AVALIAÇÃO DA REPETIBILIDADE
GENOTÍPICA EM LIMA ÁCIDA ‘TAHITI’**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 30 de março de 2023.

Assentimento:

Documento assinado digitalmente
 RENAN GARCIA MALIKOUSKI
Data: 13/04/2023 10:31:30-0300
Verifique em <https://validar.iti.gov.br>

Renan Garcia Malikouski
Autor

Documento assinado digitalmente
 LEONARDO LOPES BHERING
Data: 13/04/2023 10:42:01-0300
Verifique em <https://validar.iti.gov.br>

Leonardo Lopes Bhering
Orientador

A Deus e minha família.
DEDICO

AGRADECIMENTOS

A Deus, por me guiar e dar sabedoria em toda minha vida.

Aos meus pais, Lucimar Malikouski e Maria da Penha Garcia Malikouski, por serem meu alicerce e me apoiarem em todos os passos tomados em minha trajetória.

A minha noiva Andréia Lopes de Moraes, pela parceria e companheirismo nesses últimos cinco anos.

Ao meu irmão Henrique Garcia Malikouski, pela amizade e companheirismo.

A Universidade Federal de Viçosa, pela oportunidade de realização de um sonho, que era ser aluno desta Instituição.

Ao programa de Pós-Graduação em Genética e Melhoramento, pela excelência no ensino e na formação de profissionais.

Ao laboratório de Biometria da Universidade Federal de Viçosa, pela infraestrutura para realização dos estudos.

Ao meu orientador Leonardo Lopes Bhering, que foi um grande mestre, pelas orientações e ensinamentos fornecidos na pós-graduação.

Ao professor Kaio Olímpio das Graças Oliveira Dias, pela parceria, ensinamentos passados e coorientação.

Aos professores Edilson Romais Schmildt, Paulo Eduardo Teodoro e Evellyn Couto por participarem da banca avaliadora com contribuições valiosas para o trabalho.

Aos meus colegas do laboratório de Biometria, Marco Antônio Peixoto, Igor Ferreira Coelho, Filipe Manoel Ferreira, Arthur Elizeu, Jennifer Santana, Saulo Chaves, entre outros pela parceria e amizade durante a pós-graduação.

Aos meus colegas da República da Kika pela convivência durante os anos morando em Viçosa.

A Universidade Federal do Espírito Santo, pelo curso de mestrado realizado anteriormente e possibilitou a geração de dados utilizados neste estudo.

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

MUITO OBRIGADO!

RESUMO

MALIKOUSKI, Renan Garcia, D.Sc., Universidade Federal de Viçosa, março de 2023.
Métodos estatísticos na avaliação da repetibilidade genotípica em lima ácida ‘Tahiti’.
Orientador: Leonardo Lopes Bhering.

Frutíferas perenes como a lima ácida ‘Tahiti’ tiveram sua área de cultivo aumentada nos últimos anos devido ao acréscimo no consumo dos seus frutos na preparação de alimentos e bebidas. Para atender a demanda pela produção, a utilização de variedades com alto potencial produtivo recebe destaque como um método de potencializar o cultivo pela alta eficiência e sustentabilidade. A perenidade da lima ácida ‘Tahiti’, assim como em outras espécies, requer a utilização de métodos estatísticos de seleção que isolem efeitos ambientais e possibilitem a identificação apenas da fração genética entre os candidatos. Portanto, a busca de métodos de análise que possam corroborar para aumentar a eficiência e a confiabilidade na avaliação dos dados em experimentos é de suma importância para o progresso no melhoramento genético. Diferentes métodos estatísticos foram aplicados em um conjunto de dados de lima ácida ‘Tahiti’ a fim de investigar a repetibilidade genética na cultura. Sendo assim, 24 genótipos, constituídos de 12 variedades de copa enxertados em 2 híbridos como porta enxerto foram avaliados ao longo de 4 anos para características produtivas, vegetativas e de qualidade de frutos. Em um primeiro artigo, objetivou-se estimar parâmetros genéticos como o coeficiente de repetibilidade através de um modelo linear misto, a fim de determinar o número ótimo de medidas para se avaliar os genótipos com acurácia e precisão. Em resumo, quatro colheitas foi o número recomendado para identificar combinações com base em características produtivas. A seleção com base em várias características simultaneamente é um processo importante e necessário de ser realizado, porém desafiador, dado a diversidade de genes que controlam essas características com variadas magnitudes de efeitos e a interação destes com o ambiente. Deste modo, em um segundo capítulo, aplicou-se a metodologia de regressão aleatória no conjunto de dados e propôs-se um índice de seleção com as áreas abaixo das curvas dos valores genéticos preditos, obtidos pelos coeficientes de regressão aleatórios para características produtivas e vegetativas. Constatou-se que modelos de regressão aleatória lidam adequadamente com medidas repetidas, dados desbalanceados e são recomendados para lidar com interações ambientais. A metodologia aplicada permitiu a predição de valores genotípicos para medições não avaliadas e recomendação de genótipos superiores com base em caracteres simultaneamente. Ao selecionar ou recomendar genótipos superiores, a utilização de conceitos de probabilidade,

advindos da inferência bayesiana podem aumentar a confiabilidade, permitindo a identificação de genótipos superiores e estáveis, aumentando assim a eficiência de um programa de melhoramento. Em um terceiro estudo, testou-se a aplicabilidade de um modelo probabilístico bayesiano para a recomendação de lima ácida ‘Tahiti’ através de parâmetros de performance e estabilidade. Ajustou-se um modelo probabilístico por meio de um algoritmo amostrador de Monte Carlo Hamiltoniano. Calculou-se a probabilidade de superioridade do valor genético de cada genótipo no contexto geral e em cada colheita, bem como a probabilidade da inferioridade da interação genótipos x colheitas. Os resultados mostraram a aplicabilidade do modelo probabilístico bayesiano através de componentes de variância acurados, valores de probabilidade para as comparações de performance e estabilidade e intervalos de credibilidade para os parâmetros obtidos.

Palavras-chave: *Citrus latifolia*. Dados longitudinais. Modelos mistos. Inferência bayesiana.

ABSTRACT

MALIKOUSKI, Renan Garcia, D.Sc., Universidade Federal de Viçosa, March 2023.
Statistical methods in the evaluation of the genotypic repeatability of 'Tahiti' acid lime.
Advisor: Leonardo Lopes Bhering.

Perennial fruit trees such as the 'Tahiti' lime have seen an increase in cultivation in recent years due to the rise in consumption of its fruits in food and beverage preparation. To meet the demand for production, the use of varieties with high productivity potential is highlighted as a method to enhance cultivation for high efficiency and sustainability. The perennity of the 'Tahiti' lime, as with other species, requires the use of statistical selection methods that isolate environmental effects and allow for the identification of only the genetic fraction among the candidates. Therefore, the search for analytical methods that can increase efficiency and reliability in data evaluation in experiments is of utmost importance for progress in genetic breeding. Different statistical methods were applied to a set of 'Tahiti' lime data to investigate genetic repeatability in the culture. Thus, 24 genotypes, consisting of 12 graft varieties on 2 hybrids as rootstock, were evaluated for productive, vegetative, and fruit quality characteristics over 4 years. Initially, the objective was to estimate genetic parameters such as the repeatability coefficient through a mixed linear model, to determine the optimal number of measurements for accurate and precise genotype evaluation. In summary, four harvests were recommended to identify combinations based on productive characteristics. Selection based on multiple characteristics simultaneously is an important and necessary process to perform, but challenging given the diversity of genes that control these traits with varying magnitudes of effects and their interaction with the environment. Thus, in a second paper, the random regression methodology was applied to the data set and a selection index was proposed with the areas under the curves of the predicted genetic values, obtained by the random regression coefficients for productive and vegetative traits. It was found that random regression models handle repeated measures and unbalanced data adequately and are recommended for dealing with environmental interactions. The applied methodology allowed the prediction of genotypic values for unmeasured traits and the recommendation of superior genotypes based on multiple traits simultaneously. When selecting or recommending superior genotypes, the use of probability concepts, derived from Bayesian inference, can increase reliability, allowing for the identification of superior and stable genotypes, thus increasing the efficiency of a breeding program. In a third study, the applicability of a Bayesian probabilistic model was tested for the

recommendation of 'Tahiti' lime through performance and stability parameters. A probabilistic model was fitted using a Hamiltonian Monte Carlo sampling algorithm. The probability of superiority of the genetic value of each genotype in the overall context and in each harvest, as well as the probability of inferiority of the genotype x harvest interaction, were calculated. The results showed the applicability of the Bayesian probabilistic model through accurate variance components, probability values for performance and stability comparisons, and credibility intervals for obtained parameters.

Keywords: *Citrus latifolia*. Longitudinal data. Mixed models. Bayesian inference.

SUMÁRIO

INTRODUÇÃO GERAL	11
REFERÊNCIAS	14
ARTIGO 1.....	17
Repeatability coefficient estimates and optimum number of harvests of graft/rootstock combinations for 'Tahiti' acid lime.....	17
Abstract.....	17
Introduction.....	17
Material and methods.....	19
<i>Experimental network.....</i>	<i>19</i>
<i>Statistical analyses.....</i>	<i>20</i>
Results and discussion	22
Conclusions.....	29
Acknowledgements	29
Referendes	29
ARTIGO 2.....	35
Selection index based on random regression model in 'Tahiti' acid lime.....	35
Abstract.....	35
Introduction.....	35
Material and methods.....	37
<i>Experimental design and genetic assessment</i>	<i>37</i>
<i>Statistical analyses.....</i>	<i>38</i>
<i>Softwares.....</i>	<i>42</i>
Results	42
<i>Genotypic potential and variability.....</i>	<i>42</i>
<i>Genotype performance over time</i>	<i>45</i>
Discussion	47
Conclusion	50
References.....	50
Statements & Declarations.....	55
Figures.....	56
ARTIGO 3.....	58
Aplicação de modelos probabilísticos bayesianos para recomendação de lima ácida 'Tahiti' em dados longitudinais.	58

Resumo.....	58
Introdução	58
Material e Métodos	60
<i>Rede experimental e material vegetal.....</i>	<i>60</i>
<i>Análises estatísticas</i>	<i>61</i>
<i>Softwares e tempo computacional</i>	<i>64</i>
Resultados.....	65
<i>Probabilidade de performance superior dos genótipos.....</i>	<i>65</i>
<i>Probabilidade de persistência genotípica</i>	<i>66</i>
Discussão.....	68
Conclusão.....	70
Referências	71
Figuras	75

INTRODUÇÃO GERAL

A lima ácida ‘Tahiti’ é uma das espécies citrícolas com maior importância no Brasil e no mundo, e apresenta crescentes índices de exportação para países europeus devido ao aumento de utilização para a alimentação humana (BOTINA et al., 2019). O aumento do consumo da lima ácida ‘Tahiti’ requer uma otimização contínua do seu cultivo que envolvem práticas de manejo agrônômico cada vez mais eficientes (RIOS-ROJAS et al., 2020). Dentre as formas de aprimorar o manejo de uma espécie cultivada, o uso de variedades com elevado potencial através do melhoramento genético destaca-se por ser uma técnica eficiente, de amplo espectro de aplicação e sustentável (COOPER e MESSINA, 2021).

O melhoramento de plantas está relacionado à seleção de genótipos superiores, os quais podem ser utilizados em cruzamentos ou recomendados para o plantio, seja via propagação vegetativa ou seminal (RESENDE et al., 2012). Neste contexto, melhoristas de espécies perenes, como a lima ácida ‘Tahiti’, devem lidar com um grande desafio, que é a permanência das plantas no campo de produção durante anos. Indivíduos nesta situação estão sujeitos à diferentes condições climáticas ou formas de manejo, o que causa expressão diferencial do fenótipo e dificulta a separação das frações da variância fenotípica que possuem causas ambientais daquelas de causas genéticas (VIANA e RESENDE, 2014). Além de incidir sobre as plantas e dificultar a identificação de seu potencial genético, o efeito ambiental no contexto de medidas repetidas pode ser de dois tipos, o que leva a diferentes considerações e formas de contorná-lo (ANNICCHIARICO, 2002). O efeito ambiental permanente, presente durante todo o ciclo da cultura, e o efeito ambiental temporário, que é uma ação pontual que afeta o desenvolvimento da cultura em certa fase durante seu estágio de desenvolvimento (ALLARD e BRADSHAW, 1964).

Para que a seleção seja feita de forma eficiente a utilização de ferramentas estatísticas é preponderante, pois a partir destas é possível identificar os efeitos ambientais, isolá-los, e permitir a captura apenas da fração genética. Tradicionalmente, o método mais empregado no melhoramento vegetal é a análise de variância (ANOVA). No entanto, situações corriqueiras nos programas de melhoramento, tais como a presença de desbalanceamento no conjunto de dados (STOCHERO et al., 2020) a não ortogonalidade e não-equidistância entre as avaliações ao longo do tempo (MATHEWS e CROSSA, 2022) prejudicam a confiabilidade dos seus resultados (CELLON et al., 2018). Além disso, a inclusão de informações complementares aos experimentos de campo, tais como as relações de parentesco, não podem ser eficientemente empregadas no contexto da ANOVA (BERNARDO, 2020).

Como alternativa, surgem os modelos lineares mistos, capazes de contornar a situação de não-ortogonalidade e desbalanceamento, e permitir a inclusão do parentesco no efeito genético. Além disso, na metodologia de modelos, matrizes de covariância podem ser ajustadas para outros efeitos aleatórios do modelo além do genético, possibilitando a explicação mais fidedigna da correlação entre os níveis, como por exemplo no caso de medidas ao longo do tempo (CHAVES et al, 2022a).

A partir da aplicação da metodologia de modelos mistos são obtidos componentes de variância e valores genéticos, que serão utilizados para a seleção e entendimento da herança das características avaliadas. Os componentes de variância são utilizados para obtenção de parâmetros genéticos que permitem inferir sobre a eficiência da seleção e a acurácia dos valores utilizados para identificação dos indivíduos superiores.

Na avaliação de medidas repetidas, o coeficiente de repetibilidade é um desses parâmetros obtidos. A repetibilidade é uma quantificação da expectativa de desempenho de um dado genótipo. Por meio dela, é possível estimar o número de medições que devem ser realizadas para que a seleção dos genótipos seja eficiente, possibilitando a otimização de recursos (CHAVES et al 2022a, FERREIRA et al., 2020). Este coeficiente é estimado pela proporção da variância total que é explicada pelas variações devidas ao genótipo e pelas variações permanentes, normalmente atribuídas a um ambiente comum e constante (CRUZ et al., 2012). A determinação do número mínimo de colheitas a partir do coeficiente de repetibilidade tem sido empregada em outras fruteiras, considerando níveis de eficiência e acurácia pré-definidos pela literatura (ALVES et al 2022; CHAVES et al., 2021; SILVA et al., 2023).

A não-ortogonalidade das avaliações refere-se à correlação ou dependência entre as medidas (MATHEWS e CROSSA, 2022). Em algumas plantas perenes, como a lima ácida ‘Tahiti’, o desenvolvimento das plantas e a maturação dos frutos apresenta fortemente esse comportamento (SALES et al., 2018). Associado a não-ortogonalidade, a lima ácida ‘Tahiti’ apresenta colheitas não equidistantes, devido à desuniformidade do seu desenvolvimento. Esses fenômenos fazem com que o resíduo, que é obtido através do ajuste de um modelo estatístico, induza a violação de pressupostos de normalidade, homoscedasticidade e independência (COCHRAN, 1974). Além disso, a resposta dos efeitos genéticos, e de ambiente permanente, não necessariamente apresentam um comportamento linear ao longo do tempo (LI et al., 2020). Deste modo, torna-se necessário a modelagem dos efeitos com estruturas de covariâncias capazes de contemplar as nuances do contexto de medidas repetidas (CHAVES et al., 2022b). Os modelos de regressão aleatória, por exemplo, são amplamente utilizados em análises de

dados longitudinais e constituem uma ferramenta poderosa para identificar e quantificar a interação de genótipos entre medidas (ROCHA et al., 2018a). Nestes modelos, polinômios ortogonais são empregados para obtenção da curva dos efeitos aleatórios ao longo de um intervalo de tempo, permitindo o entendimento do comportamento da característica em estudo. A área abaixo da curva do comportamento genético pode descrever o desempenho dos genótipos, ou seja, quanto maior a área, maior o valor da característica (ALVES et al., 2020).

Dada a importância de se considerar outras características além de produtividade em lima ácida ‘Tahiti’, a seleção simultânea passa a ser uma ótima alternativa. Nestes casos, a seleção simultânea por meio de índices de seleção configura-se como uma alternativa multivariada conservacionista, permitindo ganhos conjuntos que, apesar de apresentarem-se em menor escala se comparados com seleções univariadas, evita que haja ganhos negativos indiretos em uma segunda variável (BAKER, 2020). Deste modo, associando-se as áreas obtidas de cada característica pelo método de regressão aleatória em um índice de seleção, é possível identificar materiais que de fato atendem todas as demandas produtivas e agronômicas de mercado, associando potencial produtivo com facilidade no manejo (ROCHA et al., 2018b).

À medida que o programa de melhoramento avança para as fases finais, a seleção deve ser feita com maior rigor possível. O melhorista não deve medir esforços em diminuir o efeito de agentes perturbadores, como a interação genótipos x ambientes, por exemplo (COOPER et al., 2022). No caso de plantas perenes, esta interação está relacionada aos efeitos ambientais diferenciais ao longo dos anos em que as plantas estão sujeitas. Neste contexto, é importante selecionar genótipos que possuam alelos de estabilidade, isto é, que sejam capazes de capitalizar condições ambientais favoráveis, e resistir às condições ambientais desfavoráveis. Em outras palavras, genótipos com desempenho previsível. Além disso, a busca pela combinação do alto desempenho e estabilidade em um único genótipo é bem difícil de ser encontrada. Deste modo, deve-se utilizar de metodologias estatísticas que aumentem o máximo a probabilidade de se encontrar estes dois comportamentos no mesmo genótipo. Uma abordagem inovadora baseada em estatística bayesiana auxilia na seleção de genótipos com estas características, proporcionando a diminuição dos riscos relacionados à seleção (DIAS et al., 2022). A abordagem bayesiana possibilita a utilização de informações prévias incorporadas na distribuição de parâmetros do modelo, além de apresentar componentes de variância e valores genéticos com intervalos de credibilidade inferiores quando comparados a modelos frequentistas (SILVA JÚNIOR et al., 2022).

Diante do exposto, este trabalho foi dividido em artigos que objetivam i) avaliar a repetibilidade em lima ácida ‘Tahiti’ determinando o número ótimo de avaliações necessárias

para recomendação de genótipos superiores; ii) propor um índice de seleção com base em regressão aleatória modelando efeitos genético e ambiental permanente por meio de polinômios ortogonais e; iii) recomendar genótipos superiores e persistentes por meio de um modelo probabilístico bayesiano.

REFERÊNCIAS

ALLARD, R.W.; BRADSHAW, A.D. Implications of genotype-environment interactions in applied plant breeding 1. **Crop science**, v.4, p.503–508, 1964.

ALVES, R.S.; RESENDE, M.D.V. DE; ROCHA, J.R. DO A.S. DE C.; PEIXOTO, M.A.; TEODORO, P.E.; SILVA, F.F. E; BHERING, L.L.; SANTOS, G.A. DOS. Quantifying individual variation in reaction norms using random regression models fitted through Legendre polynomials: application in eucalyptus breeding. **Bragantia**, v.79, p.360–376, 2020. DOI: 10.1590/1678-4499.20200125.

ANNICCHIARICO, P. Genotype x environment interactions: challenges and opportunities for plant breeding and cultivar recommendations. 2002.

BAKER, R. J. **Selection indices in plant breeding**. CRC Press, 2020.

BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. **Heredity**, v.125, p.375–385, 2020.

BOTINA A, B.L.; GARCÍA M, M.C.; ROMERO B, Y. Pre- and post-harvest factors that affect the quality and commercialization of the Tahiti lime. **Scientia Horticulturae**, v.257, p.108737, 2019. DOI: 10.1016/j.scienta.2019.108737.

CELLON, C.; AMADEU, R.R.; OLMSTEAD, J.W.; MATTIA, M.R.; FERRAO, L.F. V.; MUNOZ, P.R. Estimation of genetic parameters and prediction of breeding values in an autotetraploid blueberry breeding population with extensive pedigree data. **Euphytica**, v.214, p.87, 2018. DOI: 10.1007/s10681-018-2165-8.

CHAVES, S.F. DA S.; ALVES, R.M.; ALVES, R.S.; SEBBENN, A.M.; RESENDE, M.D.V. DE; DIAS, L.A. DOS S. Theobroma grandiflorum breeding optimization based on repeatability, stability and adaptability information. **Euphytica**, v.217, p.211, 2021.

CHAVES, S.F. DA S.; DIAS, L.A.S.; ALVES, R.S.; ALVES, R.M.; JOSÉ A.R.M.; ALMEIDA, C.M.V.V. Number of harvest years and selection for productivity, witches' broom resistance, stability, and adaptability in cacao. **Agronomy Journal**, v. 114, n. 6, p. 3234-3245, 2022a.

CHAVES, S.F. DA S.; EVANGELISTA, J.S.P.C.; ALVES, R.S.; FERREIRA, F.M.; ALVES R.M.; DIAS, K.O.G; BHERING, L.L. Application of linear mixed models for multiple harvest/site trial analyses in perennial plant breeding. **Tree Genetics & Genomes** v.18, n.44, 2022b.

COCHRAN, W.G. Some consequences when the assumptions for the analysis of variance are

not satisfied. **Biometrics**, v. 3, n. 1, p. 22-38, 1947.

COOPER, M.; MESSINA, C.D. Can We Harness “Enviromics” to Accelerate Crop Improvement by Integrating Breeding and Agronomy? **Frontiers in Plant Science**, v.12, 2021. DOI: 10.3389/fpls.2021.735143.

COOPER, M.; MESSINA, C.D.; TANG, T.; GHOSH, C.; POWELL, O.M.; PODLICH, D.W.; TECHNOW, F.; HAMMER, G.L. Predicting Genotype $\times$ Environment $\times$ Management ($G \times E \times M$) Interactions for the Design of Crop Improvement Strategies. Em: **Plant Breeding Reviews**. [s.l.] Wiley, 2022. p.467–585. DOI: 10.1002/9781119874157.ch8.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos Biométricos Aplicados ao Melhoramento Genético**. Viçosa: UFV, 2012. 514p.

DIAS, K.O.G.; SANTOS, J.P.R. DOS; KRAUSE, M.D.; PIEPHO, H.-P.; GUIMARÃES, L.J.M.; PASTINA, M.M.; GARCIA, A.A.F. Leveraging probability concepts for cultivar recommendation in multi-environment trials. **Theoretical and Applied Genetics**, v.135, p.1385–1399, 2022. DOI: 10.1007/s00122-022-04041-y.

FERREIRA, F.M.; CARVALHO ROCHA, J.R. DO A.S. DE; ALVES, R.S.; ELIZEU, A.M.; BENITES, F.R.G.; RESENDE, M.D.V. DE; SOUZA SOBRINHO, F. DE; BHERING, L.L. Estimates of repeatability coefficients and optimum number of measures for genetic selection of *Cynodon* spp. **Euphytica**, v.216, p.70, 2020. DOI: 10.1007/s10681-020-02605-x.

LI, J.; GAO, H.; MADSEN, P.; LI, R.; LIU, W.; BAO, P.; XUE, G.; GAO, Y.; DI, X.; SU, G. Impact of the order of Legendre polynomials in random regression model on genetic evaluation for milk yield in dairy cattle population. **Frontiers in Genetics**, v.11, p.586155, 2020.

MATHEWS, K.L.; CROSSA, J. Experimental design for plant improvement. Em: **Wheat Improvement: Food Security in a Changing Climate**. [s.l.] Springer International Publishing Cham, 2022. p.215–235.

PAULUS, D.; ZORZZI, I.C.; RANKRAPE, F.; PAULA, F.M.; MOURA, C.A. Harvest seasons and pruning management in pepper: production and pungency of the fruits. **Horticultura Brasileira**, v.35, p.434–439, 2017. DOI: 10.1590/s0102-053620170320.

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

RIOS-ROJAS, L.; CHAALI, N.; JARAMILLO-BARRIOS, C.I.; OUAZAA, S.; CORREA, J.F. Irrigation and nutrition as criteria for adequate management of Tahiti acid lime trees affected by a physiological disorder in tropical conditions. **Scientia Horticulturae**, v.270, p.109438, 2020. DOI: 10.1016/j.scienta.2020.109438.

RESENDE, M.D.V.; ALVES, R.S. Statistical significance, selection accuracy, and experimental precision in plant breeding. **Crop Breeding and Applied Biotechnology**, v. 22, 2022.

ROCHA, J.R.A.S. DE C.; MARÇAL, T. DE S.; SALVADOR, F.V.; SILVA, A.C. DA; MACHADO, J.C.; CARNEIRO, P.C.S. Genetic insights into elephantgrass persistence for bioenergy purpose. **PLOS ONE**, v.13, p.e0203818, 2018a. DOI:

10.1371/journal.pone.0203818.

ROCHA, J.R.A.S.C.; MACHADO, J.C.; CARNEIRO, P.C.S. Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. **GCB Bioenergy**, v.10, p.52–60, 2018b. DOI: 10.1111/gcbb.12443.

SALES, G.N.B.; SILVA, L. DE A.; ALMEIDA, J.F.; NOBRE, R.G.; COSTA, F.B. DA; BRITO, M.E.B.; SILVA, M.S. DA; QUEIROGA, A.X.M. DE; NASCIMENTO, A.M. DO; SILVA, J.L. DA; FERNANDES, P.D. Quality of Fruits from Grafted Tahiti Lime (*Citrus latifolia* Tan) Irrigated with Waters of Different Salinities. **Journal of Experimental Agriculture International**, v.27, p.1–10, 2018. DOI: 10.9734/JEAI/2018/44302.

SILVA, F.A. DA; VIANA, A.P.; CORRÊA, C.C.G.; LEAL, L.S. DA S.; GLÓRIA, L.S. Repeatability estimates in longitudinal data on guava trees. **Scientia Agricola**, v.80, 2023.

SILVA JÚNIOR, A.C. DA; SANT'ANNA, I. DE C.; SILVA SIQUEIRA, M.J.; CRUZ, C.D.; AZEVEDO, C.F.; NASCIMENTO, M.; SOARES, P.C. Multi-trait and multi-environment Bayesian analysis to predict the G x E interaction in flood-irrigated rice. **PLOS ONE**, v.17, p.e0259607, 2022. DOI: 10.1371/journal.pone.0259607.

STOCHERO, E.L.M.; JACOBI, L.F.; LÚCIO, A.D. Imputação de dados na análise de variância em experimentos no Delineamento Inteiramente Casualizado. **Ciência e Natura**, v.42, p.e37, 2020. DOI: 10.5902/2179460X40447.

VIANA, A.P.; RESENDE, M.D.V. DE. **Genética quantitativa no melhoramento de fruteiras**. [s.l.] Interciência, 2014. 296p.

ARTIGO 1

Repeatability coefficient estimates and optimum number of harvests of graft/rootstock combinations for 'Tahiti' acid lime.

Abstract

Combining longitudinal data and statistical models from perennial crops enabled us to estimate the optimum number of measures (harvests), implying accurate discrimination of superior genotypes in those crops. Herein, the goal of this study was to determine the optimum number of harvests based on yield traits and recommend a superior graft/rootstock combination (GRC) for *Citrus latifolia* Tanaka. Twenty-four GRCs of 'Tahiti' acid lime were evaluated from July 2017 to August 2018 for fruit yield per plant (FYP), number of fruits per plant (NFP), and longitudinal (LFD) and transversal fruit diameter (TFD). The experimental design was a randomized complete block with 4 replications. The experimental unit consisted of three individuals, totalling 244 individuals. The GRCs were composed of (i) two hybrids that were used as rootstock, *Citrumelo swingle* (*Citrus paradisi* x *Poncirus trifoliata*) and *Cintrandarin riverside* (*Citrus sunki* x *Poncirus trifoliata*); and (ii) 12 different *C. latifolia* genotypes that were used as grafts: Bello Fruit, Eledio, Iconha, Itarana, Santa Rosa, Bearss lime, CNPMF 01, CNPMF 02, CNPMF 2001, CNPMF 5059, BRS Passos, and Persian 58. Mixed models were employed to estimate the variance components. The optimum number of harvests was determined based on selective efficiency values above 0.9. The estimated repeatability coefficients presented values of 0.14 (LFD), 0.16 (TFD), 0.36 (FYP), and 0.38 (NFD). Based on the results, four harvests were able to choose genotypes based on FYP and NFP, whereas LFD and TFD were considered inefficient traits for recommending superior GRCs.

Key words

Citrus latifolia Tanaka, mixed models, REML-BLUP, longitudinal data, perennial crop.

Introduction

The socioeconomic importance of citriculture as an area of fruit growing stands out in a worldwide scenario (Sales et al., 2018). From the genus *Citrus*, the species *Citrus latifolia* (Tanaka), which is popularly known as the 'Tahiti' acid lime, has shown commercial potential in the increasing the Brazilian internal consumer market (Castricini, Silva, Silva, & Rodrigues, 2017). This fruit is prominent in the citrus market due to its availability throughout the year,

and nearly 50 thousand hectares of the crop exist in Brazil (Kist et al. 2017; IBGE, 2018).

The characteristics of the 'Tahiti' acid lime that grab attention from a commercial perspective are the fruit shape, fruit size, fruit shine, fruit flavour, nutritional value of the fruit, external fruit green coloration, and turgidity of the fruit (Bassan et al., 2015). However, a factor that can influence the quality of these characteristics is the graft/rootstock combination (GRC). It is worth mentioning that GRC also has a greater impact on giving greater vigour, productivity and longevity to the crop (Cerqueira et al., 2004; Soares et al., 2015). Due to the fact that in a breeding program of 'Tahiti' acid lime, much time is spent in experimental evaluation along with numerous harvests, the use of appropriate statistical methods to evaluate experiments is important in the decision making by breeders and allows the maximization of selection efficiency.

Another way to increase the chance of success in a breeding program is with the simultaneous selection index (SI). Thus, the SI is characterized as a linear function that combines different traits and assigns different weights according to the importance of a trait (Ramalho, Abreu, Santos, & Nunes, 2012). In this sense, the use of the SI allows various traits to be worked with to obtain genetic gains by combining multiple types of information contained in the experimental unit with the aim of selecting based on a complex of traits of economic importance (Cruz, Regazzi, & Carneiro, 2012).

The ability to predict genetic parameters, such as heritability and repeatability, is an important contribution of quantitative genetics to plant breeding (Resende, 2015). The estimation of repeatability coefficients provides an approximation of the maximum value that the heritability, in the broad sense, of a trait can reach (Lessa, Ledo, Amorim, & Silva, 2014). Alternative methods that consider these topics are the restricted maximum likelihood (REML), which estimates the variance components, and the best unbiased linear prediction (BLUP), which predicts the genetic values (Resende, 2015; Resende & Rosa-Perez, 2002).

Experimental analyses considering repeated measurements allow us to predict genetic values for each harvest individually and for all harvests simultaneously and to estimate the structure of variance among them. The repeatability coefficient (ρ) is defined as the correlation between repeated measurements of the same individual over time and space, as well as its associated coefficient of determination (R^2), which measures the accuracy in predicting the true value of an individual genotype (Cruz et al., 2012). Knowledge about trait repeatability is important in an experimental design. Understanding trait deviation over time or space to determine the number of measurements required for each individual allows genetic selection to be performed to obtain an appropriate level of accuracy to reduce labour time and budgets of

the breeding program (Resende, 2007). Thus, studying ρ estimation allows the necessary number of measures to be obtained to estimate the individual's real genetic value, taking into account the permanent, genetic and environmental effects and optimizing the evaluation time of the trait of interest in the field (Lessa et al., 2014).

Repeatability using the REML/BLUP methodology has been used in other fruit trees to determine the optimal number of measurements without compromising accuracy (Fonseca, Morais, Gonçalves, Aquino, & Rocha, 2018; Lessa et al., 2014; Negreiros, Neto, Miqueloni, & Lessa, 2014). Sánchez et al. (2017) evaluated that the number of measures to evaluate the productivity of *Annona muricata* L. could be reduced from sixteen to eight, maintaining a repeatability and accuracy coefficient above 80%. Herein, we aimed to determine the optimum number of harvests used in *C. latifolia* based on productive traits to recommend superior 'Tahiti' acid lime GRCs.

Material and methods

Experimental network

The experimental network was conducted at José Guarete Farm, which is owned by Bello Fruit Company, in São Mateus, Espírito Santo state, Brazil (18°48'21"S, 39°53'30" W), between July 2017 and August 2018. The climate of the region according to the Koppen classification system is tropical warm and humid (Aw), with rainy summers and dry winters (Alvares, Stape, Sentelhas, & Gonçalves, 2013). The average temperature during the experiment was 25 °C, and the precipitation was 140 mm (weather information in Appendix S1).

Twenty-four GRCs of 'Tahiti' acid lime, which were arranged in a spacing of six metres between rows and three metres between plants, were used as the variation factors. The experimental design was a randomized complete block with 4 replications. The experimental unit consisted of three plants, totalling 244 individuals. The GRCs were composed of (i) two hybrids that were used as rootstock, *Citrumelo swingle* (*Citrus paradisi* x *Poncirus trifoliata*) and *Cintrandarin riverside* (*Citrus sunki* x *Poncirus trifoliata*); (ii) 12 different *C. latifolia* genotypes that were used as grafts: Bello Fruit, Eledio, Iconha, Itarana, Santa Rosa, Bearss lime, CNPMF 01, CNPMF 02, CNPMF 2001, CNPMF 5059, BRS Passos, and Persian 58.

Limes were harvested in July, November, and October 2017 and January, March, July, and October 2018 for a total of seven harvests. All mature fruits were harvested. Maturity is considered when the fruits present a minimum diameter of 47 mm, external rough texture and dark green to light colour (HORTIBRASIL, 2000). The fruit yield per plant (FYP in kg plant⁻¹

¹⁾ was taken by measuring the total weight using a digital scale, and the number of fruits per plant (NFP) was counted at each harvest. Ten fruits were randomly separated from each GRC in each harvest. Then, the longitudinal fruit diameter (LFD in mm) and transversal fruit diameter (TFD in mm) were measured using digital callipers.

Statistical analyses

The REML/BLUP procedure was adopted to estimate the variance components and to predict the genotypic values following the procedures in Patterson and Thompson (1971) and Henderson (1975). The statistical model was as follows:

$$y = Xm + Zg + Wp + Ti + \varepsilon$$

where y is the data vector, m is the vector of the effects of the measuring-repetition combinations (assumed to be fixed), g is the vector of the genotypic effects (assumed to be random), p is the permanent effects vector (assumed to be random), i is the vector of the genotype x measuring (GxM) interaction effects (assumed to be random), and ε is the vector of errors (random). The capital letters X , Z , W , and T represent the incidence matrices for those effects (Resende & Duarte, 2007).

The significance of the random effects was tested using the likelihood ratio test (LRT) (Rao, 1973), which was calculated with the following equation:

$$\lambda = 2[\text{Log}_e L_{p+1} - \text{Log}_e L_p]$$

where Log_e is the logarithm of the REML function and L_{p+1} and L_p are the likelihood peaks associated with the complete model and the reduced model, respectively. Thus, λ was compared with the probability density function (χ^2) for a given number of degrees of freedom and probability of error (Dobson & Barnett, 2018). The number of degrees of freedom was defined as the difference between the numbers of parameters between the evaluated models (one degree of freedom in this case).

The phenotypic individual variance (σ_{phen}^2), average phenotypic variance ($\hat{\sigma}_{phen}^2$), heritability in a broad sense ($\hat{h}_g^2$), repeatability (ρ), determination coefficient of permanent environmental effects (c_{perm}^2), determination coefficient of GxM effects (c_{gm}^2), phenotypic correlation throughout measurements (r_{gm}), average heritability of genotypes ($\hat{h}_{mg}^2$), and coefficient of experimental variation (CV_e) were obtained with the following expressions:

$$\sigma_{phen}^2 = \sigma_g^2 + \sigma_{perm}^2 + \sigma_e^2 + \sigma_{gm}^2,$$

$$\hat{\sigma}_{phen}^2 = \sigma_g^2 + \frac{\sigma_{perm}^2}{r} + \frac{\sigma_e^2}{r \times m} + \frac{\sigma_{gm}^2}{m},$$

$$\begin{aligned}\hat{h}_g^2 &= \frac{\sigma_g^2}{\sigma_{phen}^2}; \\ \rho &= \frac{\sigma_g^2 + \sigma_{perm}^2}{\sigma_{phen}^2}; \\ c_{perm}^2 &= \frac{\sigma_{perm}^2}{\sigma_{phen}^2}; \\ c_{gm}^2 &= \frac{\sigma_{gm}^2}{\sigma_{phen}^2}; \\ r_{gm} &= \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{gm}^2}; \\ \hat{h}_{mg}^2 &= \frac{\sigma_g^2}{\sigma_{phen}^2}; \\ CV_e &= \frac{\sqrt{\sigma_e^2}}{\mu} \times 100.\end{aligned}$$

where $\hat{\sigma}_g^2$ is the genotypic variance; $\hat{\sigma}_{perm}^2$ is the permanent environmental variance; $\hat{\sigma}_e^2$ is the residual variance; $\hat{\sigma}_{gm}^2$ is the GxM interaction variance; r is the number of repetitions; and m is the number of measures.

The determination for the permanent phenotypic effects ($r_{\hat{f}\hat{p}fp}^2$), selective accuracy ($r_{\hat{g}g}$), and selective efficiency (E), all based on m measurements, were given by the following equations:

$$\begin{aligned}r_{\hat{f}\hat{p}fp}^2 &= \frac{m \rho}{1 + (m - 1) \rho} \\ r_{\hat{g}g} &= \left(\frac{\hat{\sigma}_g^2}{\sigma_g^2 + \frac{\sigma_{perm}^2}{r} + \frac{\sigma_e^2}{r \times m} + \frac{\sigma_{gm}^2}{m}} \right)^2 \\ E &= \left[\frac{m}{[1 + (m - 1) \rho]} \right]^{1/2}\end{aligned}$$

The optimum number of GRCs for recommendation was estimated, though the accuracy values were superior to 0.9 (Resende, Silva, & Azevedo, 2014).

The additive index (Resende, 2007), which was used to identify the superior GRC for use in the acid lime breeding program, was given by the following expression:

$$AI = \sum_{h=1}^4 w_t \frac{vg}{\sigma_t}$$

where w_t is the weight assigned for trait t , vg is the overall mean for trait t added to the predicted

genetic value, and σ_t is the standard deviation for v_g . For the additive index, the weights were set as equal to the genetic coefficient of variation (CV_g) using the following equation:

$$CV_g(\%) = \frac{100\sqrt{\hat{\sigma}_g^2}}{\bar{X}}$$

where the selection direction was "higher" for all traits. Two scenarios were explored: the first considered the seven harvests, and the second scenario took into consideration the number of optimum harvests obtained in this study. The predicted gain with the selection and the genotype ranking were obtained directly by the additive index output.

All analyses were carried out in the software Selegen-REML/BLUP (Resende, 2016).

Results and discussion

The traits FYP, NFP, and TFD were statistically significant for all effects (genotype, permanent, and GxM) (Table 1), while the LFD trait indicated no significant difference for permanent and GxM effects but only for the genetic effect. This implied differences between the values of the productive traits evaluated among the GRCs and that they were influenced throughout the crops. Similar studies by Shrestha, Dhakal, Gautum, Paudyal & Shrestha (2012) and Khankahdani, Rastegar, Golein, Golmohammadi, & Jahromi (2017) also indicated the existence of genetic variability in the acid lime culture. Regarding the experimental quality, the FYP and NFP traits presented CV_e values of 50.38% and 48.83%, respectively, which were higher than those found for LFD and TFD (4.33% and 3.40%, respectively). The observed large difference evidenced the low precision of NFP and FYP traits and the strong influence suffered by the environment since these traits are quantitative traits (Viana & Resende, 2014).

Table 1 - Likelihood ratio test (LRT) for the traits: fruit yield per plant (FYP), number of fruits per plant (NFP), longitudinal fruit diameter (LFD), and transversal fruit diameter (TFD) of the statistical analyses of 24 GRC of "Tahiti" acid lime.

Effect	Traits			
	FYP	NFP	LFD	TFD
Genotype	58.95**	59.63**	20.12**	12.75**
Mensurement	5.5*	7.37**	0.36 ^{ns}	5.11*
Genotype x Mensurement	35.37**	34.89**	2.5 ^{ns}	16.85**

^{ns}, *, **, not significant, significance at 5% and 1% of probability by the LRT, respectively.

The statistical significance of the GxM interaction effect confirmed the existence of genotypes with different performances throughout harvests in terms of yield. This fact was expected since the recurrent significant interactions between genotypes and crops in the evaluation of perennial plants is commonly caused by the effect of the environment on the expression of traits (Resende, 2009). Heritability is estimative and allows for evaluating the proportion of the total variance that results from the genetic origin. The results of the analyses indicated that the $\hat{h}_{mg}^2$ of the traits ranged from 0.66 to 0.88 (Table 2), which were interpreted as elevated values, indicating strong genetic control of the traits. However, heritability estimates in young perennial plants are frequently biased by the GxM effect (Resende, Furlani-Júnior, Moraes, & Fazuoli, 2001). Those plants have lower competitive effects and consequently have lower estimates for the variance components (Araújo, Dias, Scarpinati, & Paula, 2015). Thus, the parameters obtained, among them ρ , which refers to the heritability limits, tended to be higher in the early stages but reached normality as the plants develop. In the experiment, the plants were approximately 36 months old, while a citrus orchard has (in the Southeastern Region, Brazil) a life cycle of 25 years (Santana, Souza, Souza, & Fontes, 2006). Thus, the plants assessed in this study do not achieve commercial age, implying that high values of parameters were properly estimated. New experimental evaluations that account for more harvests should address the analysed genotypes. They would allow a more precise parameter estimation for the ‘Tahiti’ acid lime crop.

Table 2- Estimates of variance components and genetic parameters for fruit yield per plant (FYP), number of fruits per plant (NFP), longitudinal fruit diameter (LFD), and transversal fruit diameter (TFD) of 24 GRC of ‘Tahiti’ acid lime.

Estimate ⁽¹⁾	Traits			
	FYP	NFP	LFD	TFD
σ_g^2	317.42	2.56	0.95	0.55
σ_{perm}^2	34.30	0.31	0.11	0.26
σ_e^2	461.43	3.49	6.25	3.42
σ_{gm}^2	142.64	1.08	0.42	0.68
σ_{phen}^2	955.79	7.44	7.79	4.91
σ_{phen}^2	362.86	2.91	1.26	0.83
h_{mg}^2	0.87	0.88	0.75	0.66

ρ	0.36	0.38	0.14	0.16
c_{perm}^2	0.035	0.041	0.015	0.052
c_{gm}^2	0.15	0.14	0.054	0.14
r_{gm}	0.69	0.70	0.69	0.45
μ	42.63	3.83	57.72	54.23
$CV_e(\%)$	50.38	48.83	4.33	3.40

⁽¹⁾ genotypic variance (σ_g^2), permanent environment variance (σ_{perm}^2), temporary residual variance (σ_e^2), genotype x measurement interaction variance (σ_{gm}^2), individual phenotypic variance (σ_{phen}^2), mean phenotypic variance (σ_{phen}^2), repeatability coefficient at parcel level (ρ), determination coefficient of permanent environmental effects (c_{perm}^2), determination coefficient of the effects of interaction genotype x measurement (c_{gm}^2), genotypic correlation through measurements (r_{gm}), heritability of genotype average (h_{mg}^2), overall average (μ) and experimental coefficient of variation (CV_e).

The traits LFD and TFD presented ρ values of 0.14 and 0.16, respectively, while FYP and NFD presented higher values (0.36 and 0.38, respectively). These results were reasonable, considering a perennial culture and the complexity of the production traits, which are usually coordinated by many genes (Cruz et al. 2012, Laviola et al. 2013). There are no studies that have obtained ρ for the 'Tahiti' acid lime, which highlights the relevance of this study. However, in data from other citrus species, such as *Citrus aurantium*, the authors found similar values for total fruit number (0.262) and total fruit weight per plant (0.291) with variance analysis methodology (Negreiros, Saraiva, Oliveira, Álvares, & Roncatto, 2008). Pompeu Junior, Blumer, & Resende (2013) used the mixed model methodology and found ρ values of 0.35 for 'Valencia' sweet orange (*Citrus sinensis*), which is like the values found in this study.

According to Resende (2015), ρ is classified as high when it has a value equal to or greater than 0.6, moderate when ρ is in the interval of 0.3-0.6, and low when it is equal to or less than 0.3. In the results of this study, the estimated repeatability was classified as moderate for NFP and FYP and low for TFD and LFD. Similar findings have been reported for other fruit crops, such as banana (Lessa et al., 2014), sweet orange (Negreiros et al., 2008), cashew (Maia et al., 2016), guava (Quintal, Viana, Campos, Vivas, & Amaral Junior, 2017), mangaba (Fonseca et al., 2018), grapes (Leão, Nunes, & Souza, 2018), and cupuassu (Alcoforado, Pedrozo, Mayer, & Lima-Primo, 2019).

The c_{perm}^2 estimates the proportion of individual phenotypic variations that directly

results from measurements performed in an individual. Low values of 0.035, 0.041, 0.015, and 0.052 were observed for the FYP, NFP, LFD and TFD traits, respectively. The low values presented revealed that the permanent environmental variation from one harvest to another did not represent influences on the expression of the traits. Pompeu Junior et al. (2013) also found that low proportions of individual phenotypic variance were explained by the permanent environmental effects (0.004).

The determination coefficient of the GxM interaction (c_{gm}^2) showed similar values for FYP, NFP, and TFD (0.15, 0.14, and 0.14, respectively), whereas the LFD presented a lower value of c_{gm}^2 (0.054) when compared to the other three traits. These results showed that the FYP, NFP, and TFD traits collaborated more to express individual phenotypic variation than the LFD. Regarding magnitude, the FYP, NFP, and TFD values of c_{gm}^2 were considered high and the LFD values of c_{gm}^2 were considered low (Resende, 2009), while Mariguele et al. (2012) studied models of longitudinal data in *Annona squamosa* and found results of similar magnitudes (0.099).

The r_{gm} values obtained were similar for the traits FYP, NFP, and LFD (0.69, 0.70, and 0.69, respectively), inasmuch as the trait LFD presented an inferior value (0.45). The magnitude of this correlation, which was moderate to high for the characteristics FYP, NFP and LFD, indicated that at an approximate 70% coincidence of the values repeating over the harvests. Conversely, the LFD trait did not demonstrate the same expression among harvests (Resende, 2007).

Our results stated that seven measurements for NFP and FYP were necessary to obtain determination higher than 0.8, which is considered good for selecting superior individuals in perennial cultures (Sánchez et al., 2017) (Fig. 1A). The LFD and TFD traits did not reach the desired minimum values even with the maximum number of estimated measurements (ten). The values of determination in the four harvests reached 0.7 for NFP and FYP, whereas the maximum number of harvests (ten) did not return appropriate values for genotype recommendations in LFD and TFD. When considering accuracy as a parameter for determining the optimal number of harvests, the result was different. Taking this estimate into account, it was observed that the ideal number of harvests was four (Fig. 1B). Resende (2002) stated that a few measurements with accuracy values above 0.9 were sufficient to indicate superior genotypes. The selection efficiency based on the average of several measures in relation to a single measure was also inferred in this study (Fig. 1C). A more imminent increment was observed by increasing the number of harvests for LFD and TFD traits, which reached

efficiency values of higher than two with ten harvests. In contrast, the traits FYP and NFP were less responsive to the increase in the number of harvests, denoting higher precision.

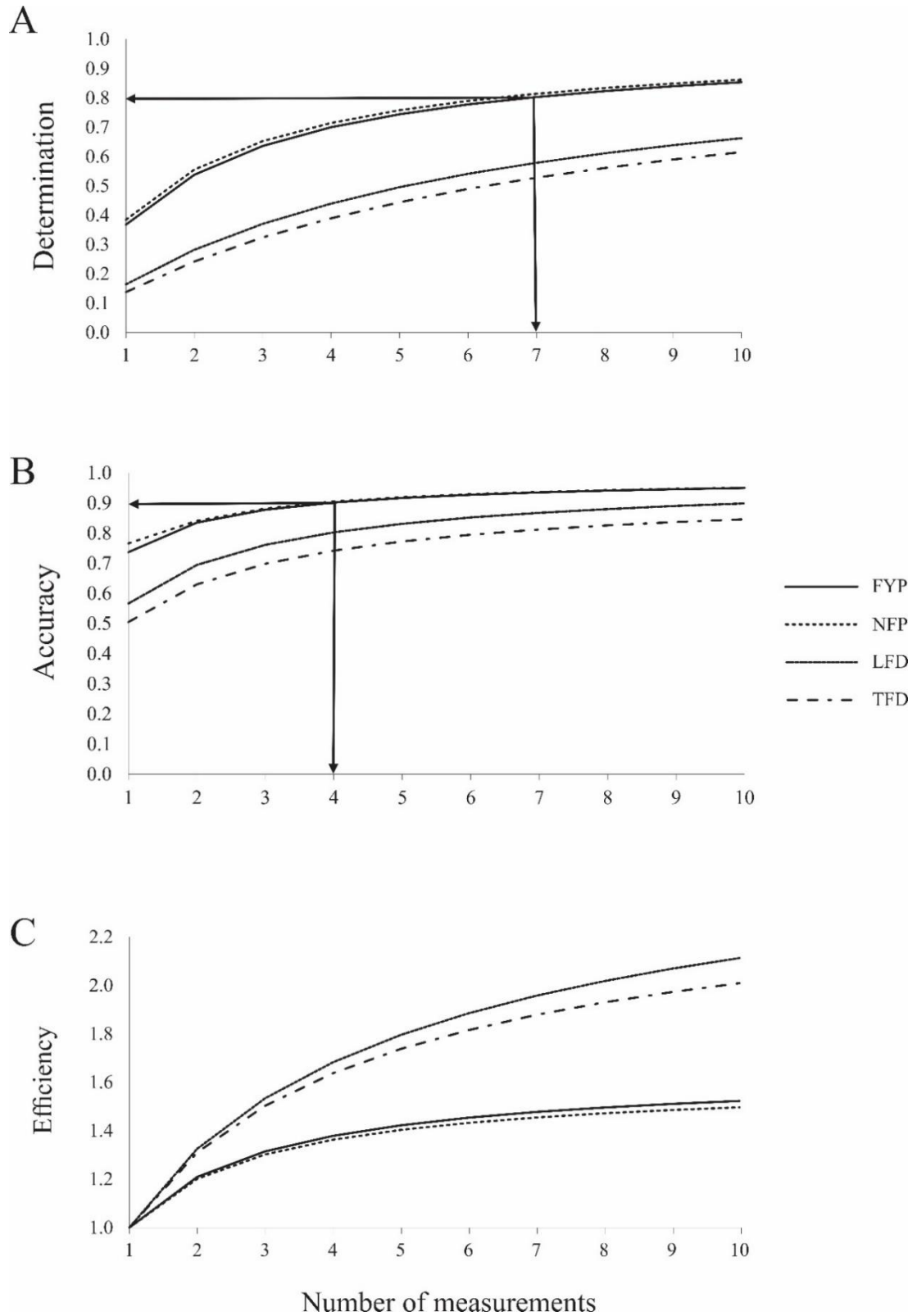


Figure 1 - Phenotypic permanent effects determination (A), selection accuracy (B) and C) selective efficiency as a function of the number of measurements by fruit yield per plant (FYP),

number of fruits per plant (NFP), longitudinal (LFD), and transversal fruit diameter (TFD) evaluated in 24 GRC of 'Tahiti' acid lime.

Overall, the determination, accuracy, and selective efficiency increased as the number of measurements increased. For the FYP and NFP traits, with the evaluation of seven harvests, it was possible to reach a maximum determination coefficient of 0.8 with high precision and high selective efficiency. However, with four harvests evaluated, it may be appropriate to indicate superior genotypes for 'Tahiti' acid lime cultivation. Because the values did not reach at least 0.8 for determination and 0.9 for accuracy with the maximum yields estimated in this study, the LFD and TFD traits should not be considered in the selection of superior genotypes. In summary, four harvests were the optimum number selected without losing accuracy and achieved values higher than 0.9. For this reason, four harvests were considered in the second scenario of the SI. However, the plants analysed in this study were younger, at only 36 months of age. This fact implies reduced variation among the genotypes and, as a result, differentiation of them. Most likely, with more experimental evaluation, i.e., including more harvests, those traits tended to reach the reliability needed to recommend in this crop.

The additive selection index revealed no differences in the results of the GRC selection in the two scenarios considered (Table 3). The top GRCs selected for the seven harvests were the same as those selected for the four harvests, while the ranking presented slight differences. Thus, the results of the two scenarios showed certain similarities and precision. Overall, the worst GRC (CNPMF 2001 grafted onto the 'Citrumelo Swingle') was the same under both conditions. The differences in ranking genotypes were probably due to GxM interaction, which had different genotype productions throughout the evaluations. In this case, the most productive genotype in one harvest was not necessarily the most productive in the other harvest. This may be due to the fact that they have multiple blooms throughout the year, have different fruit maturation mechanisms and are influenced by a large number of abiotic factors that can contribute to the increase in this GxM interaction (Iglesias et al., 2007).

Table 3 - Additive index and selection gain for four productive traits of 24 GRC of 'Tahiti' acid lime.

Rank (°)#	Four harvests (measures)		Seven harvests (measures)	
	Genotype	Selection Gain (%)	Genotype	Selection Gain (%)

1	18	28.35	15	30.60
2	3	27.44	4	25.94
3	15	26.47	18	24.28
4	4	24.15	3	22.72
5	23	22.39	22	21.68
6	1	20.52	23	20.68
7	22	18.83	11	19.71
8	11	17.51	19	18.76
9	19	16.34	1	17.81
10	2	15.07	14	16.92
11	14	13.79	13	16.12
12	13	12.74	16	15.32
13	16	11.72	2	14.52
14	17	10.76	7	13.54
15	7	9.88	17	12.51
16	6	9.11	6	11.07
17	10	8.02	10	9.63
18	24	6.92	24	8.30
19	5	5.84	5	7.03
20	12	4.71	20	5.84
21	20	3.56	8	4.59
22	21	2.44	12	3.23
23	8	1.39	21	1.65
24	9	0	9	0

[#] Genotypes numbers 1 to 12 refer to Bello Fruit, Éledio, Iconha, Itarana, Santa Rosa, Bearss Lime, CNPMF 01, CNPMF 02, CNPMF 2001, CNPMF 5059, BRS Passos and Persian 58 grafts grafted to the rootstock Citrumelo 'Swingle', and of 13 to 24 the same grafts grafted in the Citrandarin 'Riverside' rootstock, respectively.

The importance of doing more studies like this is emphasized, given the low amount of repeatability in the 'Tahiti' acid lime. In addition, the detection of superior genotypes is extremely important for breeding programs that can be used as parents in the development of new cultivars.

Conclusions

In summary, four harvests were the recommended number for identifying superior acid lime ‘Tahiti’ GRCs based on FYP and NFP. On the other hand, to accurately recommend superior genotypes of acid lime ‘Tahiti’, the LFD and TFD require further evaluations. The best GRCs recommended based on this study for 4 harvests and for 7 harvests were Bearss lime grafted onto citrandarin ‘Riverside’, Iconha grafted onto citrumelo ‘Swingle’, Iconha grafted onto citrandarin ‘Riverside’, and Itarana grafted onto citrumelo ‘Swingle’.

Acknowledgements

The authors are grateful to the Bello Fruit company, which provided the location for the development of the experiment, and to the Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) and the Frucafé company, which were the sources of our vegetal material. We thank the Universidade Federal do Espírito Santo and the Universidade Federal de Viçosa for their support in the analyses and contributions to human resource training. We appreciate the financial support from the National Council for Scientific and Technological Development (CNPq), the Coordination for the Improvement of Higher Education Personnel (CAPES), and the Espírito Santo Research and Innovation Support Foundation (FAPES). This study was financed in part by the CAPES - Finance Code 001.

Referendes

- Alcoforado, A. T. W., Pedrozo, C. Â., Mayer, M. M., & Lima-Primo, H. E. de. (2019). Repeatability of morpho-agronomic characters of *Theobroma grandiflorum* fruits. *Revista Brasileira de Fruticultura*, 41(2). <https://doi.org/10.1590/0100-29452019142>
- Alvares, C. A., Stape, J. L., Sentelhas, P. C., & Gonçalves, J. L. de M. (2013). Modeling monthly mean air temperature for Brazil. *Theoretical and Applied Climatology*, 113(3–4), 407–427. <https://doi.org/10.1007/s00704-012-0796-6>
- Araújo, M. J. de, Dias, D. da C., Scarpinati, E. A., & Paula, R. C. de. (2015). Número de repetições, de plantas por parcela e de avaliações para testes clonais de eucalipto. *Pesquisa Agropecuária Brasileira*, 50(10), 923–931. <https://doi.org/10.1590/S0100-204X2015001000008>
- Bassan, M. M., Filho, F. de A. A. M., Alves, R. F., Bezerra, D. F., Couto, H. T. Z., &

- Jacomino, A. P. (2015). Beneficiamento pós-colheita de lima ácida “Tahiti” afeta sua qualidade e conservação. *Ciência Rural*. <https://doi.org/10.1590/0103-8478cr20141492>
- Castricini, A., Silva, J. T. A. DA, Silva, I. P. DA, & Rodrigues, M. G. V. (2017). Quality of ‘Tahiti’ acid lime fertilized with nitrogen and potassium in the semiarid region of Minas Gerais. *Revista Brasileira de Fruticultura*, 39(2), 1–10. <https://doi.org/10.1590/0100-29452017288>
- Cerqueira, E. C., Neto, M. T. de C., Peixoto, C. P., Filho, W. dos S. S., Ledo, C. A. da S., & Oliveira, J. G. de. (2004). Resposta de porta-enxertos de citros ao déficit hídrico. *Revista Brasileira de Fruticultura*, 26(3), 515–519.
- Cruz, C. D., Regazzi, A. J., & Carneiro, P. C. S. (2012). *Modelos Biométricos Aplicados ao Melhoramento Genético*. Viçosa: UFV.
- Dobson, A. J., & Barnett, A. G. (2018). *An introduction to generalized linear models* (4°). <https://doi.org/10.1201/9780367807849>
- Fonseca, C. E. L. da, Moraes, F. M. de, Gonçalves, H. M., Aquino, F. de G., & Rocha, F. S. (2018). Repeatability of fruit traits from two *Hancornia speciosa* populations from the core region of the Brazilian Cerrado. *Pesquisa Agropecuária Brasileira*, 53(6), 710–716. <https://doi.org/10.1590/s0100-204x2018000600007>
- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31(2), 423–447.
- Iglesias, D. J., Cercós, M., Colmenero-Flores, J. M., Naranjo, M. A., Ríos, G., Carrera, E., ... Tadeo, F. R. (2007). Physiology of citrus fruiting. *Brazilian Journal of Plant Physiology*, 19(4), 333–362. <https://doi.org/10.1590/S1677-04202007000400006>
- Instituto Brasileiro de Geografia e Estatística [IBGE]. (2018). *Produção Agrícola Municipal*. Retrieved November 15, 2019, from <https://sidra.ibge.gov.br/pesquisa/pam/tabelas>
- Instituto Brasileiro de Qualidade em Horticultura [HORTIBRASIL]. (2000). *Normas de Classificação, Padronização e Identidade do Limão (Lima Ácida) Tahiti (Citrus latifolia*

), para o Programa Brasileiro para a Melhoria dos Padrões Comerciais e Embalagens de Hortigranjeiros. Retrieved October 28, 2019, from <http://www.hortibrasil.org.br/classificacao/limao/arquivos/norma.html>

Khankahdani, H. H., Rastegar, S., Golein, B., Golmohammadi, M., & Jahromi, A. A. (2017). Genetic diversity in Persian Lime (*Citrus latifolia* Tanaka) accessions using morphological and molecular markers. *The Journal "Agriculture and Forestry,"* 63(3), 221–231. <https://doi.org/10.17707/AgricultForest.63.3.22>

Kist, B. B., Santos, C. E. dos, Carvalho, C. de, Treichel, M., & Filter, C. F. (2017). *Anuário Brasileiro da Fruticultura*. Editora Gazeta.

Laviola, B. G., Oliveira, A. M. C. e, Bhering, L. L., Alves, A. A., Rocha, R. B., Gomes, B. E. L., & Cruz, C. D. (2013). Estimates of repeatability coefficients and selection gains in *Jatropha* indicate that higher cumulative genetic gains can be obtained by relaxing the degree of certainty in predicting the best families. *Industrial Crops and Products,* 51, 70–76. <https://doi.org/10.1016/j.indcrop.2013.08.016>

Leão, P. C. de S., Nunes, B. T. G., & Souza, E. M. C. de. (2018). Repeatability coefficients and genetic gains in table grape progenies for the Brazilian semi-arid region. *Scientia Agricola,* 75(5), 387–392. <https://doi.org/10.1590/1678-992x-2017-0046>

Lessa, L. S., Ledo, C. A. da S., Amorim, E. P., & Silva, S. de O. (2014). Estimativas de repetibilidade de híbridos diploides (AA) de bananeira. *Pesquisa Agropecuária Brasileira,* 49(2), 109–117. <https://doi.org/10.1590/S0100-204X2014000200005>

Maia, M. C. C., Almeida, A. da S., Macedo, L. M., Resende, M. D. V. de, Lacerda, M. N. de, Vasconcelos, L. F. L., ... Silva, R. B. da. (2016). Heritability, repeatability and genetic gains in a improvement population of cajuzeiro. *Revista Árvore,* 40(4), 715–722. <https://doi.org/10.1590/0100-67622016000400015>

Mariguele, K. H., Resende, M. D. V. de, Viana, J. M. S., Silva, F. F. e, Silva, P. S. L. de, & Knop, F. de C. (2011). Métodos de análise de dados longitudinais para o melhoramento genético da pinha. *Pesquisa Agropecuária Brasileira,* 46(12), 1657–1664. <https://doi.org/10.1590/S0100-204X2011001200011>

- Negreiros, J. R. da S., Neto, R. de C. A., Miqueloni, D. P., & Lessa, L. S. (2014). Estimativa de repetibilidade para caracteres de qualidade de frutos de laranja-doce. *Pesquisa Agropecuária Brasileira*, 49(1), 40–48. <https://doi.org/10.1590/S0100-204X2014000100006>
- Negreiros, J. R. da S., Saraiva, L. L., Oliveira, T. K. de, Álvares, V. de S., & Roncatto, G. (2008). Estimativas de repetibilidade de caracteres de produção em laranjeiras-doces no Acre. *Pesquisa Agropecuária Brasileira*, 43(12), 1763–1768. <https://doi.org/10.1590/S0100-204X2008001200017>
- Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3), 545–554.
- Pompeu Junior, J., Blumer, S., & Resende, M. D. V. de. (2013). Avaliação genética de seleções e híbridos de limões cravo, volkameriano e rugoso como porta-enxertos para laranjeiras Valência na presença da morte súbita dos citros. *Revista Brasileira de Fruticultura*, 35(1), 199–209. <https://doi.org/10.1590/S0100-29452013000100023>
- Quintal, S. S. R., Viana, A. P., Campos, B. M., Vivas, M., & Amaral Junior, A. T. do. (2017). Analysis of structures of covariance and repeatability in guava segregating population. *Revista Caatinga*, 30(4), 885–891. <https://doi.org/10.1590/1983-21252017v30n408rc>
- Ramalho, M. A. P., Abreu, Â. de F. B., Santos, J. B. dos, & Nunes, J. A. R. (2012). Aplicações da Genética Quantitativa no Melhoramento de Plantas Autógamas. In *Biologia* (1st ed.). Lavras: UFLA.
- Rao, C. R. (1973). *Linear statistical inference and its applications* (Vol. 2). Wiley New York.
- Resende; Marcos Deon Vilela de, & Rosa-Perez, J. O. H. (2002). Genética e melhoramento de ovinos. In UFPR (Ed.), *Curitiba. Ed. UFPR* (1st ed.). Curitiba.
- Resende, M. D. V. de. (2007). *Matemática e estatística na análise de experimentos e no melhoramento genético* (1st ed.; Embrapa, Ed.). Colombo: Embrapa Florestas.
- Resende, M. D. V. de. (2009). *Genética biométrica e estatística no melhoramento de plantas*

perenes (1st ed.; Embrapa, Ed.). Colombo: Embrapa Florestas.

Resende, M. D. V. de. (2015). *Genética quantitativa e de populações* (1st ed.). Viçosa: UFV.

Resende, M. D. V. de. (2016). Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breeding and Applied Biotechnology*, 16(4), 330–339.
<https://doi.org/10.1590/1984-70332016v16n4a49>

Resende, M. D. V. de, & Duarte, J. B. (2007). Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesquisa Agropecuária Tropical*, 37(3), 182–194.

Resende, M. D. V. de, Furlani-Júnior, E., Moraes, M. L. T. de, & Fazuoli, L. C. (2001). Estimativas de parâmetros genéticos e predição de valores genotípicos no melhoramento do cafeeiro pelo procedimento REML/BLUP. *Bragantia*, 60(3), 185–193.
<https://doi.org/10.1590/S0006-87052001000300005>

Resende, M. D. V. de, Silva, F. F. e, & Azevedo, C. F. (2014). *Estatística Matemática, Biométrica e Computacional*. Viçosa: UFV.

Sales, G. N. B., Silva, L. de A., Almeida, J. F., Nobre, R. G., Costa, F. B. da, Brito, M. E. B., ... Fernandes, P. D. (2018). Quality of Fruits from Grafted Tahiti Lime (*Citrus latifolia* Tan) Irrigated with Waters of Different Salinities. *Journal of Experimental Agriculture International*, 27(2), 1–10. <https://doi.org/10.9734/JEAI/2018/44302>

Sánchez, C. F. B., Alves, R. S., Garcia, A. d. P., Teodoro, P. E., Peixoto, L. A., Silva, L. A., ... Resende, M. D. V. (2017). Estimates of repeatability coefficients and the number of the optimum measure to select superior genotypes in *Annona muricata* L. *Genetics and Molecular Research*, 16(3), 1–8. <https://doi.org/10.4238/gmr16039753>

Santana, M. B., Souza, L. da S., Souza, L. D., & Fontes, L. E. F. (2006). Atributos físicos do solo e distribuição do sistema radicular de citros como indicadores de horizontes coesos em dois solos de tabuleiros costeiros do estado da Bahia. *Revista Brasileira de Ciência Do Solo*, 30(1), 1–12. <https://doi.org/10.1590/S0100-06832006000100001>

- Shrestha, R. L., Dhakal, D. D., Gautum, D. M., Paudyal, K. P., & Shrestha, S. (2012). Genetic Diversity Assessment of Acid Lime (*Citrus aurantifolia* Swingle Swingle) Landraces in Nepal, Using SSR Markers. *American Journal of Plant Sciences*, 03(12), 1674–1681. <https://doi.org/10.4236/ajps.2012.312204>
- Soares, L. A. do. A., Brito, M. E. B., Fernandes, P. D., De Lima, G. S., Filho, W. dos S. S., & De Oliveira, E. S. (2015). Crescimento de combinações copa - Porta-enxerto de citros sob estresse hídrico em casa de vegetação. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 19(3), 211–217. <https://doi.org/10.1590/1807-1929/agriambi.v19n3p211-217>
- Viana, A. P., & Resende, M. D. V. de. (2014). *Genética quantitativa no melhoramento de fruteiras*. Interciência.

ARTIGO 2

Selection index based on random regression model in ‘Tahiti’ acid lime.

Abstract

An accurate and efficient statistical method for repeated measures analysis and selection of scion/rootstock combinations (SRC) is a fundamental prerequisite for genetic progress in ‘Tahiti’ acid lime breeding. Thus, this work aimed: i) to evaluate the applicability and efficiency of random regression models in the analysis of repeated measures and ii) to apply a selection index to recommend SRC of ‘Tahiti’ acid lime, associating vegetative and productive traits. Between July 2017 and January 2019, 24 SRC of ‘Tahiti’ acid lime were evaluated for vegetative and productive traits over eight measurements. The random regression model (RRM) was compared with a repeatability model (REP). The variance components were estimated via residual maximum likelihood. The predicted genotypic values were predicted via the best linear unbiased prediction. The genotypic values (for REP) and the areas below the genotypic trajectories (for RRM) were used as input for the additive index aiming for simultaneous selection. The three best-performing genotypes were selected, considering a selection percentage equal to 12.5%. Models based on random regression and repeatability showed the same selected genotypes, even though, there were changes in the ranking. The random regression model allows the prediction of genotypic values for genotypes that were not evaluated in the field being advantageous for genetic evaluation of ‘Tahiti’ acid lime. The selection index based on the REP and RRM models indicated that the recommended SRC for propagation and planting are Iconhal/‘Riverside’ *citrandarin*, CNPMF5059/‘Riverside’ *citrandarin* and Itarana/*Swingle citrumelo*.

Keywords

Citrus latifolia Tanaka L., repeated measures, Legendre polynomials, wood perennial crops, mixed models.

Introduction

With large areas cultivated in tropical and subtropical regions, the genus *Citrus* represents one of the most important fruit trees in the world’s agricultural scenario (Denaro et al. 2020; Santis and Frangipane 2020; Ladaniya et al. 2020). Worldwide, the production of

lemons and limes has reached over 21.3 million tons in 2020, of which Brazil accounted for roughly 1.6 million tons (7.5%) of the global production, representing the fifth-largest producer (FAOSTAT, 2020). Among the species belonging to the genus *Citrus*, the ‘Tahiti’ acid lime (*Citrus latifolia* Tanaka L.) has brought attention to the Brazilian fruit industry due to its wide adaptation to edaphoclimatic conditions and high economic value (Bastos et al. 2017; Machado et al. 2017). Given the growing internal and external demand for fresh fruit in the citrus market, especially for ‘Tahiti’ acid lime, it is essential to maximize the productivity of the orchards (Morais et al. 2020). In this scenario, ‘Tahiti’ acid lime breeding programs have devoted great efforts to identifying more efficient genetic selection strategies to improve yield and quality traits of agronomical and commercial interest.

The differential genotypic responses to different environmental conditions, referred to as genotype-by-environment (G×E) interaction, can drastically change the rankings of genotypes across environments, affecting the selection of superior genotypes. The definition of the G×E interaction can consider different locations, or even different harvests or measurements over time, being called genotypes x measurement (G×M) interaction (Malikouski et al. 2021). Therefore, the application of statistical methods capable of overcoming G×E and G×M interactions effects in a more parsimonious way is essential for the progress of any breeding program (Rocha et al. 2018; Ferreira et al. 2020).

More usual statistical methodologies have been used to deal with data from different measurements in woody perennial crops, such as repeatability models (Alves et al. 2019). The repeatability models are mixed models that consider the genetic and permanent environment as random effects (Paulus et al. 2017; Oliveira et al. 2021). However, this methodology has some limitations, such as the requirement of homogeneous correlations between measurements as well as the normality of residuals (Viana and Resende 2014). Working with repeated measures, in perennial crops such as ‘Tahiti’ acid lime, these assumptions are frequently not met (Huynh and Feldt 1970; Ketelaere et al. 2003). Neglecting these assumptions can bias the estimates of variance components and mislead the selection of superior genotypes across repeated measures (Cellon et al. 2018). In addition, crops such as ‘Tahiti’ acid lime present a large variation in productivity over time (Machado et al. 2017; Sales et al. 2018). This fact implies in differences in fruit ripening among genotypes. This difference leads to a non-standardization of the fruits at the time of harvest if this is carried out at the same time for all genotypes.

In this context, random regression models have been successfully adopted for analyzing multi-environment and repeated measures trials (Schaeffer 2004; Rocha et al. 2018a; Alves et

al. 2020b). Random regression models capture the change of a trait continuously over space and/or time with great parsimony (Sun et al. 2017; Coelho et al. 2020). By characterizing the covariance structures, random regression models count the trend over space and/or time, representing an efficient alternative for repeated measures analysis in perennial crops (Resende et al. 2001; Rocha et al. 2018a). In addition, random regression models with different genetic, permanent environmental, and residual covariance structures can be tested to choose the best-fitted model (Rocha et al. 2018a). The random regression model was initially applied to animal breeding (Meyer and Hill 1997; Schaeffer 2004; Kranis et al. 2007; Jamrozik et al. 2010). Recently, this technique has been applied to plant breeding providing excellent results (Ly et al. 2018; Campbell et al. 2019; Momen et al. 2019; Baba et al. 2020). Despite its great potential, this methodology remains unexplored for the analysis of repeated measures in ‘Tahiti’ acid lime trials.

The selection of superior ‘Tahiti’ acid lime genotypes is a complex process, as the traits of economic importance are controlled by many genes and highly influenced by environmental conditions. Moreover, the correlation between traits can make it difficult to obtain satisfactory genetic gains (Ribeiro et al. 2018; Silva et al. 2020). Therefore, the use of selection indices constitutes an important strategy for the simultaneous selection of multiple desired traits (Resende et al. 2014a; Ribeiro et al. 2018).

In accordance with the mentioned, this study aimed i) to evaluate the applicability and efficiency of random regression models in the analysis of repeated measures for ‘Tahiti’ acid lime, and ii) to recommend ‘Tahiti’ acid lime scion/rootstock combinations (SRC) based on selection index for vegetative and productive traits.

Material and methods

Experimental design and genetic assessment

The experiment was carried out at Fazenda José Guarete (18°48'21"S, 39°53'30"W, 35 m.a.s.l), belonging to the company Bello Fruit®, located in the municipality of São Mateus, state of Espírito Santo, Brazil. According to Koppen (Alvares et al. 2013), the region's climate is classified as Aw, characterized by a rainy season in summer and a dry season in winter (detailed information on precipitation and maximum, average, and minimum temperatures are presented in the Supplementary Information – Fig S1).

From July 2017 to January 2019, twenty-four scion/rootstock combinations (SRC) of ‘Tahiti’ acid lime were grown at spacing of 6 m between trees and 3 m between rows. The

design experimental used was a randomized complete block design with 4 replications. The experimental unit consisted of 3 trees per SRC, totaling 288 individuals. The ‘Tahiti’ acid lime SRC consisted of two hybrids used as rootstock, *Swingle citrumelo* (*Citrus paradisi* x *Poncirus trifoliata*) and ‘Riverside’ *citrandarin* (*Citrus sunki* x *Poncirus trifoliata*) and as scions, twelve ‘Tahiti’ acid lime selections [numbers 1 to 12 identify the scion selections grafted on *Swingle citrumelo*, and 13 to 24 identify the selections grafted on ‘Riverside *citrandarin*’] (Bello Fruit [1, 13], Elédio [2, 14], Iconha [3, 15], Itarana [4, 16], Santa Rosa [5, 17], Bearss Lime [6, 18], CNPMF01 [7, 19], CNPMF02 [8, 20], CNPMF2001 [9, 21], CNPMF5059 [10, 22], BRS Passos [11, 23] and Persian 58 [12, 24]). All scion selections came from Citrus Active Germplasm Bank at Embrapa Mandioca e Fruticultura, information’s about their origins are available in Santos et al. (2013, 2016).

Four traits were measured divided into two groups, two traits from vegetative development group: trunk diameter (TD, in mm), measured with a digital caliper, 5 centimeters above the graft region and canopy volume (CV, in m³), obtained through the equation proposed by Zekri (2000): $CV = \left(\frac{\pi}{6}\right) x TH x DR x DS$, where TH is tree height, DR and DS are canopy projection diameter in the row and in-between rows, respectively. The other group was productive group, with also two traits measured: fruit yield (FY, in kg tree⁻¹) through weighting the fruits from each tree with a minimum diameter of 47 mm, and the number of fruits per tree (NF) assessed through simple counting. Eight measurements were taken for FY and NF (at 736 (M1), 808 (M2), 861 (M3), 918 (M4), 972 (M5), 1083 (M6), 1200 (M7), and 1249 (M8) days after planting) and four measurements were taken for TD and CV traits (736 (M1), 918 (M4), 1083 (M6) and 1249 (M8) days after planting).

Statistical analyses

Repeatability model

Data were subjected to mixed-model analysis using a statistical repeatability model (REP) for experiments in randomized complete block design, as follows:

$$y_{[768,1]} = X_{[768,32]}s_{[32,1]} + Z_{[768,24]}g_{[24,1]} + W_{[768,96]}p_{[96,1]} + e_{[768,1]},$$

where y is the vector of phenotypic data, s is the vector of the effects of measurement-replication combinations (assumed to be fixed) added to the overall mean, g is the vector of genotypic effects (assumed to be random) [$g \sim N(0, I\sigma_g^2)$], p is the vector of the permanent plot effects (assumed to be random) [$p \sim N(0, I\sigma_p^2)$], and e is the vector of errors [$e \sim N(0, I\sigma_e^2)$] (random), where I is an identity matrix, σ_g^2 , σ_p^2 and σ_e^2 are genotypic, permanent plot and residual

variances, respectively. The capital letters X, Z, and W represent the incidence matrices for these effects. The dimension of the matrices is in subscripted square brackets.

The variance components were estimated via residual maximum likelihood – REML (Patterson and Thompson 1971), and genotypic values were predicted via best linear unbiased prediction - BLUP (Henderson 1975).

The significance of the random effects of the model (genotypic and permanent plot effects) were tested using the Likelihood Ratio Test (LRT) (Wilks 1938), given by:

$$LRT = -2 * (LogL_{r. mod.} - LogL_{c. mod.}),$$

where $LogL_{r. mod.}$ is the logarithm of the restricted likelihood function of the reduced model (without the effect under test) and $LogL_{c. mod.}$ is the logarithm of the likelihood function of the complete model (with all effects included). Thus, the significance was tested by the LRT using the chi-square (χ^2) probability density function with a degree of freedom with 1 and 5% error probability (Dobson and Barnett 2018).

The phenotypic variance ($\hat{\sigma}_{phen}^2$), mean heritability of genotypes in the broad sense ($\hat{h}_{mg}^2$), genotypic coefficient of variation (CV_g) and the selective accuracy ($r_{\hat{g}g_REP}$) were estimated according to the following equations (Resende et al., 2014b):

$$\hat{\sigma}_{phen}^2 = \hat{\sigma}_g^2 + \frac{\hat{\sigma}_{perm}^2}{r} + \frac{\hat{\sigma}_{res}^2}{r*m},$$

$$\hat{h}_{mg}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_{phen}^2 \text{ or } \hat{\sigma}_{phen \text{ joint}}^2}$$

$$CV_g(\%) = \frac{\sqrt{\hat{\sigma}_g^2}}{\mu} * 100$$

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

where r and m are the number of replications and measurements, respectively, $\hat{\sigma}_x^2$ is the estimated variance component (where x refers to any effect mentioned above), and PEV is the prediction error variance extracted from the coefficients on the diagonal of the generalized inverse matrix of the mixed model equations (Henderson 1975).

Random Regression Model

For each trait evaluated, 48 random regression models (RRM) were fitted via Legendre polynomials and tested to identify the best fit to the data.

Time measurements were scaled to range from -1 to 1 to allow the use of Legendre polynomials. The sizing formula is given below:

$$t_x = -1 + 2\left[\frac{m_x - m_{min}}{m_{max} - m_{min}}\right]$$

where m_x refers to the x measurement time, m_{min} is the first measurement time ($m = 1$), and m_{max} is the last measurement time ($m = 8$).

The following random regression model (RRM) was used:

$$Y_{ijk} = R_k + \beta_T \varphi_{ijT} + \sum_{t=0}^T g_{it} \varphi_{ijt} + \sum_{t=0}^T p_{ikt} \varphi_{ijt} + \varepsilon_{ijk},$$

where Y_{ijk} is the phenotypic value of the i^{th} genotype ($i = 1, 2, \dots, 24$), in the j^{th} harvest ($j = 1, 2, \dots, 8$), of the k^{th} replication ($k = 1, 2, 3, 4$). R_k is the fixed replication effect, β_T is the third-degree regression coefficient (maximum order - 4) of the Legendre polynomials, assumed to be fixed. The random effects g_{it} and p_{ikt} are random regression coefficients of the Legendre polynomials for the genotypic and permanent plot effects, respectively, in the i^{th} genotype, k^{th} replication in the t^{th} degree of adjustment of the polynomial ($t = 0, 1, 2$, and 3). φ_{ijt} is the degree of the Legendre polynomial in the i^{th} genotype, k^{th} replication and t^{th} degree of adjustment of the polynomial, with T being the maximum degree order considered ($T = 4$) for the genotypic and permanent plot effects, ε_{ijk} is the random residual effect associated with the individual Y_{ijk} .

In matrix notation, the above RRM is given by:

$$y_{[768,1]} = X_{[768,32]} b_{[32,1]} + Z_{[768,24]} g_{[24,8]} + W_{[768,96]} p_{[96,8]} + e_{[768,1]},$$

where y is the vector of phenotypic data, b is the vector of effects of measurement-replication combinations added to the overall mean (assumed to be fixed), g is the vector of regression coefficients of genotypic effects (assumed to be random), p is the vector of regression coefficients of permanent plot effects (assumed to be random), and e is the vector of residual (random) effects. X , Z and W refer to the incidence matrices for these effects. The dimensions of the matrices are in subscripted square brackets.

In this model, $g \sim N(0, K_g \otimes I)$, $p \sim N(0, K_p \otimes I)$, and $e \sim N(0, R)$; where K_g and K_p are the covariance matrices for genotypic and permanent plot effects, respectively; $\otimes$ denotes the Kronecker product; I is an identity matrix with order adequate to the respective random effect; and R refers to the residual covariance matrix. Different models considering distinct residual covariance structures were tested based on the number of measurements (homogeneous, diagonal, and unstructured).

After obtaining the 48 models for each trait, the one that presented the best fit was chosen. For this, we used the Akaike Information Criteria (AIC) (Akaike 1974), given by:

$$AIC = -2\log L + 2p,$$

where $\text{Log}L$ is the logarithm of the likelihood function and p is the number of estimated parameters. The model with the lowest AIC value is considered the model with the best adjustment.

As in the repeatability model, the significance of the random effects of the best-fit model was performed using LRT. Estimates of the variance components ($\hat{\sigma}_x^2$) for the genetic and permanent plot effects were estimated via reparametrized REML (Meyer and Hill 1997), as well as the predicted genotypic values ($\hat{g}_{ij}$) obtained via BLUP (Kirkpatrick et al. 1990), according to the expressions below:

$$\hat{\sigma}_x^2 = \varphi_{ijt} K_x \varphi'_{ijt} \text{ and}$$

$$\hat{g}_{ij} = \sum_{t=0}^T \hat{g}_{it} \varphi_{ijt},$$

where K_x is the estimated coefficient covariance matrix for the random effect (genotypic or permanent plot) and $\hat{g}_{it}$ is the genotypic value of order t for the genotypic effect of the i^{th} genotype.

The phenotypic variance ($\hat{\sigma}_{phen}^2$), heritability of mean genotypes in the broad sense ($\hat{h}_{mg}^2$), genotypic coefficients of variation (CV_g) were obtained in a similar way to those mentioned above. The selective accuracy of genotypic values ($r_{\hat{g}g_RRM}$) was obtained according to the equation below (Kirkpatrick et al. 1990):

$$r_{\hat{g}g_RRM} = \sqrt{1 - \frac{\varphi_{ijt} PEV \varphi'_{ijt}}{\varphi_{ijt} \hat{K}_g \varphi'_{ijt}}}$$

For genotype ranking purposes, the genotypic trajectories were obtained from the interpolation of genotypic values at 200 points within the parametric range of the harvests. The areas below the genotypic trajectories (AUGT) were obtained through the following expression (Alves et al. 2020a):

$$AUGT = \mu + \int_{-1}^1 \sum_{t=0}^T \hat{g}_{it} \varphi_{ijt} x^m dx;$$

where μ represents the phenotypic mean and x^m is the measurement scale of the trait.

Selection index

The predicted genotypic values (BLUP), and the areas below the genotypic trajectories (AUGT) of all traits were used as input for the index to promote the selection of superior genotypes while considering the four traits simultaneously. The Additive Index (AI) given by (Resende 2016):

$$AI = \sum_{t=1}^{nt} w_t \frac{v_i}{\sigma_t}$$

where w_t is the weight for each trait t , defined by the genotypic coefficient of variation (CV_g) of the trait in REP and RRM, v_{ij} is the input according to the model used in the i^{th} genotype for a target trait, σ_t represents the standard deviation to the trait t and nt the total number of traits (four). We considered as the direction of selection of the traits as “higher” for FY and NF, as “lower” for CV, and as “null” for TD.

The genotypes were classified according to their position in each index, and the ranking correlation was obtained using Spearman's coefficient. The three best performing genotypes were considered as selected, with a selection intensity of 12.5%.

Softwares

The repeatability model and random regression models were fitted using Selegen REML/BLUP (Resende 2016) and ASReml (Gilmour et al. 2015) softwares, respectively. The areas under genotypic trajectories were obtained by ‘integrate’ function in R (R Development Core Team 2020).

Results

Genotypic potential and variability

For the REP, the genotypic variances were significant for all traits via the LRT. The permanent plot effect was significant for CV and TD traits, while non-significant for FY and NF trait (Table 1). Heritability estimates ranged from 0.81 (TD) to 0.87 (FY and NF). The genotypic coefficients of variation were also higher for the productive traits, with 42.17% and 41.43% for FY and NF respectively, while TD and CV had values of 7.03% and 15.44%, respectively (Table 1). The selective accuracy by the repeatability model was 0.88 for TD and CV and 0.94 for FY and NF (Table 1).

Table 1. Variance components and genetic parameters through repeatability model (REP), for the traits trunk diameter (TD), canopy volume (CV), fruit yield (FY), and number of fruits (NF), evaluated in 24 combinations of ‘Tahiti’ acid lime scion/rootstock.

Estimates	Trait			
	TD	CV	FY	NF
$\hat{\sigma}_g^2$	22.204 ^{*(1)}	0.734**	3.256**	380.251**

$\hat{\sigma}_{perm}^2$	1.718*	0.0795*	0.0395 ^{ns}	1.839 ^{ns}
$\hat{\sigma}_{res}^2$	3.199	0.0825	0.4175	50.762
$\hat{\sigma}_{phen}^2$	27.121	0.896	3.713	432.853
h_{mg}^2	0.81	0.82	0.87	0.87
r_{gg}	0.88	0.88	0.94	0.94
CV_g	7.03	15.44	42.17	41.43
μ	66.94	5.55	4.27	47.06

$\hat{\sigma}_g^2$: genotypic variance; $\hat{\sigma}_{perm}^2$: permanent plot variance; $\hat{\sigma}_{res}^2$: residual variance; $\hat{\sigma}_{phen}^2$: phenotypic variance; h_{mg}^2 : average heritability of genotypes in the broad sense; r_{gg} : selective accuracy; CV_g : genotypic coefficient of variation, and μ : mean. ⁽¹⁾ ns, * and **, non-significant, significant at 5, and at 1% of error probability, respectively, through in REP model.

The best fitted model in RRM was given by the lowest AIC values. Thus, the model with order two for genotypic effects, order zero for permanent plot effect, and with heterogeneous residual variance (2.0D), presented the best fit for the TD trait (Table 2). For the CV trait, the model with order three for genotypic effect and for permanent environmental effect, with homogeneous residual variance (3.3H), obtained the lowest AIC value (Table 2). For the productive traits, FY and NF, the model with order three for genotypic effect, order zero for permanent plot effects, with heterogeneous residual structure (3.0D), presented the best fit (Table 2). The order of fixed effects in the model has not been changed, keeping degree 3 for all models. The convergence parameters for all analyzed models for the four traits are available in the supplementary material (Table A.1 to Table A.4).

Table 2. Some of the 48 models fitted using Legendre Polynomials, with their polynomial degree of random effects, convergence, number of parameters and Akaike Information Criteria (AIC) for the traits trunk diameter (TD), canopy volume (CV), fruit yield (FY), and number of fruits (NF), evaluated in 24 combinations of ‘Tahiti’ acid lime scion/rootstock.

Model ⁽¹⁾	Polynomial degree*			LogL	Number of parameters			AIC
	Fix	Gen	Perm		Gen	Perm	Res	
	TD							
2.0H	3	2	0	Converged	6	1	1	1955.63
2.1H	3	2	1	Converged	6	3	1	1938.23

2.0D	3	2	0	Converged	6	1	4	1878.02
2.1D	3	2	1	Converged	6	3	4	1880.35
2.0U	3	2	1	Converged	6	3	10	1993.94
2.1U	3	2	2	Converged	6	6	10	2005.09
CV								
3.2H	3	3	2	Converged	10	6	1	488.206
3.3H	3	3	3	Converged	10	10	1	436.228
3.2D	3	3	2	Not converged	10	6	4	-
3.3D	3	3	3	Not converged	10	10	4	-
3.2U	3	3	2	Converged	10	6	10	688.986
3.3U	3	3	3	Converged	10	10	10	697.066
FY								
3.0H	3	3	0	Converged	10	1	1	2106.52
3.1H	3	3	1	Converged	10	3	1	2082.18
3.0D	3	3	0	Converged	10	1	8	1874.02
3.1D	3	3	1	Not converged	10	3	8	-
3.0U	3	3	0	Converged	10	1	36	2183.94
3.1U	3	3	1	Converged	10	3	36	2194.36
NF								
3.0H	3	3	0	Converged	10	1	1	5502.86
3.1H	3	3	1	Converged	10	3	1	5495.12
3.0D	3	3	0	Converged	10	1	8	5331.42
3.1D	3	3	1	Not converged	10	3	8	-
3.0U	3	3	0	Converged	10	1	36	5594.6
3.1U	3	3	1	Converged	10	3	36	5603.6

* Fix: fixed effect; Gen: genotypic effect; Perm: permanent plot effect; and Res: residual effect.

⁽¹⁾: tested models are referred as N.nR, where N and n represent Legendre polynomial adjusted order for genotypic and permanent environmental effects, respectively, and the residue R can assume different covariance structures such as homogeneous (H), heterogeneous (D), and unstructured (U).

The estimation of variance components was performed using the model selected by the AIC. Genotypic and permanent plot effects were statistically significant, by LRT, for all traits. The LRT values for the genotypic variation factor were 11.85 for FY, 8.5 for NF, 190.6 for CV and 27.3 for TD, all higher than 3.84 (chi-square, 1 degree of freedom, 5% probability).

Genotype performance over time

The TD and CV traits showed variable genotypic variance (σ_g^2), with the performance increasing and decreasing over time (measurements) respectively (Table 3). For the productive traits (FY and NF), there was a progressive increase in genotypic variance (σ_g^2), over time. A similar performance was observed for the estimates of mean heritability of genotypes in the broad sense (h_{mg}^2), ranging from 0.12 (TD in M8) to 0.65 (CV in M2). The selective accuracy values (r_{gg}) were greater than 0.8 for all traits in all assessments, except for TD at M8 (0.75) (Table 3).

Table 3. Variance components and genetic parameters estimated by the random regression model (RRM) in each harvest for the characteristics: trunk diameter (TD), canopy volume (CV), fruit yield (FY) and number of fruits (NF) evaluated in 24 ‘Tahiti’ acid lime scion /rootstock combinations through eight harvests.

Estimates	Evaluations							
	M1	M2	M3	M4	M5	M6	M7	M8
	TD							
$\hat{\sigma}_g^2$	22.085	26.460	30.004	32.619	33.214	28.304	19.588	18.259
$\hat{\sigma}_{perm}^2$	11.283	11.283	11.283	11.283	11.283	11.283	11.283	11.283
$\hat{\sigma}_{res}^2$	15.409	15.409	15.409	18.463	18.463	27.339	27.339	120.210
$\hat{\sigma}_{phen}^2$	48.778	53.153	56.693	62.366	62.961	66.921	58.211	149.752
h_{mg}^2	0.45	0.50	0.53	0.52	0.53	0.42	0.34	0.12
r_{gg}	0.89	0.92	0.92	0.92	0.92	0.91	0.84	0.75
μ	66.9418							
CV								
$\hat{\sigma}_g^2$	0.231	1.643	2.300	2.149	1.520	0.442	0.598	1.457
$\hat{\sigma}_{perm}^2$	0.060	0.902	1.289	1.270	1.068	0.869	1.307	2.638
$\hat{\sigma}_{res}^2$	0.107							
$\hat{\sigma}_{phen}^2$	0.399	2.546	3.590	3.420	2.588	1.311	1.905	4.096

h_{mg}^2	0.58	0.65	0.64	0.63	0.59	0.34	0.31	0.36
$r_{\hat{g}g}$	0.91	0.92	0.92	0.91	0.91	0.89	0.85	0.84
μ	5.5564							
FY								
$\hat{\sigma}_g^2$	0.616	0.944	1.494	2.266	3.347	5.974	8.064	12.067
$\hat{\sigma}_{perm}^2$	0.229	0.229	0.229	0.229	0.229	0.229	0.229	0.229
$\hat{\sigma}_{res}^2$	1.165	1.120	0.983	6.096	3.453	4.365	9.242	15.350
$\hat{\sigma}_{phen}^2$	2.011	2.293	2.706	8.591	7.029	10.568	17.534	27.646
h_{mg}^2	0.31	0.41	0.55	0.26	0.48	0.57	0.46	0.44
$r_{\hat{g}g}$	0.88	0.93	0.93	0.94	0.94	0.93	0.93	0.90
μ	4.2790							
NF								
$\hat{\sigma}_g^2$	89.592	134.424	207.97	303.83	445.73	787.01	893.343	1338.48
$\hat{\sigma}_{perm}^2$	27.663	27.663	27.663	27.663	27.663	27.663	27.663	27.663
$\hat{\sigma}_{res}^2$	136.756	154.217	216.98	1034.78	540.17	542.84	772.322	1628.640
$\hat{\sigma}_{phen}^2$	254.01	316.305	452.62	1366.27	1013.5	1357.5	1693.329	2994.788
h_{mg}^2	0.35	0.42	0.46	0.22	0.44	0.58	0.53	0.45
$r_{\hat{g}g}$	0.89	0.92	0.92	0.93	0.93	0.93	0.93	0.90
μ	47.0678							

$\hat{\sigma}_g^2$: genotypic variance; $\hat{\sigma}_{perm}^2$: permanent plot variance; $\hat{\sigma}_{res}^2$: residual variance; $\hat{\sigma}_{phen}^2$: phenotypic variance; h_{mg}^2 : average heritability of genotypes in the broad sense; r_{gg} : selective accuracy; and μ : mean.

The genotypic trajectories were non-linear for the four traits evaluated in the 24 genotypes (Figure 1), crossing, diverging and/or converging over time (x-measurement axis). It was noticed that trajectories close to the general average only for the FY in the first three evaluations, from M4 to M8 it was noticed a great variation in the performance of the genotypes. The other traits showed different trajectories for TD, VC and NF over the eight assessments.

Ranking and selection of superior genotypes

The genotypes selected by the REP and RRM models were the same ones according to the adopted selection intensity. However, there were differences in the ranking based on the additive index (Figure 2). For REP, the three best performance genotypes in ascending order were 4, 15, and 22. For RRM, the three best performance genotypes were in ascending order were 15, 22, and 4. On the other hand, genotype nine was classified as twenty-fourth position

in all scenarios (Figure 2). The Spearman correlation between the REP and RRM rankings was 0.96.

Discussion

The repeatability model proved the existence of genetic variability among individuals for all the evaluated traits. According to Resende and Alves (2020), heritability estimates by the REP were classified as very high for all traits, as they presented $h_{mg}^2 \geq 0.80$. The selective accuracies obtained by the repeatability model was classified as high magnitude for TD and CV ($0.70 \leq r_{gg} < 0.90$) and very high magnitude for FY and NF ($r_{gg} \geq 0.90$) (Resende and Alves 2020).

The permanent plot effect, tested via LRT, indicated that the vegetative traits (CV and TD) are not influenced by the persistent microenvironmental effect among measurements. On the other hand, the productive traits (FY and NF) need information from more than one evaluation for assertive recommendation of superior genotypes given the presence of significant permanent plot effect. Repeatability studies aim to identify the permanent plot effect and indicate the optimum number of measurements were reported in ‘Tahiti’ acid lime plant breeding (Malikouski et al. 2021).

Except for the TD trait, the selected models fitted diagonal residuals (i.e., a residual variance for each measurement). The genotypic effects were modeled by degrees 2, 3, 3, and 3 Legendre polynomials for the TD, CV, FY, and NF traits, respectively, whereas the permanent plot effects were modeled by Legendre polynomials of degrees 0, 3, 0, and 0 for the traits TD, CV, FY and NF, respectively. This implies different number of covariance parameters estimated by each model (11, 19, 19, and 21 for TD, FY, NF, and CV, respectively). Only polynomials with degree equal to three or lower were considered, in a way to represent the most common biological effects that occur in plant development, in addition to the guarantee of adjusting less parameterized models (Meyer and Hill 1997; Meyer and Kirkpatrick 2005). The maximum degree adjustment for fixed effects was used once it explains better the effect, in addition to being usually adopted (Schaeffer 2016). In addition, testing the residual variance structure that was performed in RRM is considered relevant in the analysis of repeated measures, as it can directly affect the estimates of genetic parameters and the prediction of genotypic values (Coelho et al., 2020).

According to the LRT, there is genotypic variability between the SRC of ‘Tahiti’ acid lime for the evaluated traits by RRM. The presence of genotypic variability is an important

prerequisite in breeding programs, as it enables recombination and recommendation in the process of developing new varieties (Nielsen et al. 2014), as well as obtaining gains with selection.

The permanent plot variances, by RRM, were significant for all traits. However, the obtaining degree 0 of the Legendre polynomials for FY, NF, and TD traits in RRM indicates a permanently constant effect over the measurements, and the non-significance of the permanent plot effect in REP for FY and NF traits reinforce the behavior of the effect.

The absence of studies addressing genetic parameters in *Citrus latifolia* Tanaka is noteworthy. According to Resende and Alves (2020), the average heritability estimates in the broad sense had moderate magnitude ($0.15 < h_{mg}^2 < 0.50$) or high magnitude ($0.50 \leq h_{mg}^2 < 0.80$), except for TD in M2 which presented a low magnitude ($0.01 \leq h_{mg}^2 < 0.15$). Variation in the average heritability estimates of genotypes in the broad sense is indicative of differential gene expression across measurements. Differences in environmental and crop management conditions also interfere with heritability estimates, as phenotypic variance is influenced by these factors (Tucker et al. 2020).

All estimates of selective accuracy in the random regression model were classified as high or very high magnitudes, being high when $0.70 \leq r_{gg} < 0.90$ and very high when $r_{gg} \geq 0.90$ (Resende and Alves 2020). The high selective accuracy values for the CV and TD traits demonstrate the efficiency of RRM in predicting genotypic values and estimating variance components. According to Resende and Alves (2020), very high selective accuracies are desirable to recommend superior genotypes for planting.

The trajectories illustrated the variation of genotypic values over time (Alves et al. 2020a). The selected models evidenced the presence of $G \times M$ interaction for the evaluated traits, as the trajectories were not parallel along the measurements. The crossing of trajectories demonstrates the occurrence of $G \times M$ interaction of a complex nature (Van Eeuwijk et al., 2016). This type of interaction is problematic in breeding programs, as it can change the ranking of genotypes among measurements misleading the identification of superior genotypes (Resende 2015).

According to Schaeffer (2016), genetic parameters estimated via RRM allow the prediction of breeding values along the entire temporal gradient, mitigating the adverse $G \times M$ interaction effects. In addition, RRM allows the estimation of variance components and the prediction of breeding values at a specific time, within the temporal gradient, that no measurement was performed. As an example, the vegetative traits TD and CV, which were

evaluated only in M1, M4, M6 and M8, but estimates were obtained for the eight evaluations. This advantage allows for a more accurate selection of superior genotypes, either for inclusion in breeding programs, or for recommendation for planting, through vegetative propagation (Peixoto et al. 2021). In addition, the evaluation of traits over time demands resources and labor. The use of RRM allows reducing the number of evaluations, since it does not require a minimum number or linearity in the evaluated interval (Jamrozik and Schaeffer 1997).

The calculation of the AUGT enables the ranking of genotypes considering all crops simultaneously and considering the genotypic plasticity of genotypes over time (Peixoto et al. 2020). For the productive traits, FY and NF, the greater the area under the curve, the greater the productive capacity of the genotype. For the vegetative trait CV, smaller values of AUGT are of greater interest, since trees with a large production of tree biomass occupy larger spaces and their large size makes cultural treatments difficult (Morais et al. 2020). Due to its low genotypic variation and plasticity, in addition to the absence of correlation with productive traits (Peixoto et al. 2021), null gain was considered for TD in this study, it focused only on the estimation of variance components and prediction of genotypic values for a referred trait.

The use of selection indices is an alternative to selecting for several traits simultaneously (Ribeiro et al. 2018). The similarity in the classification of genotypes between the selection indices REP and RRM can be confirmed by the high Spearman correlation. However, the importance of the certain similarity of the results of REP and RRM should be highlighted, since REP is a model commonly used in the evaluation of perennial species. Regarding selected genotypes, the similarity between RRM and REP may be related to the low permanent plot effect for NF and FY, which was identified as non-significant by LRT in REP and of zero degree in the random regression model. However, the inability of the predictive and estimation repeatability models at unassessed ages is highlighted.

The use of RRM does not require a standardization of time in measuring traits, that is, a trait does not need to be evaluated with another set of trait and not in all individuals at the same time (Araújo et al. 2015). For the evaluation of ‘Tahiti’ acid lime, this is extremely important, as it allows the measurement of trait at their optimal time, which can overcome the inconveniences of genetic evaluation, such as, for example, the inequality of fruit maturation (Sales et al. 2018).

Unlike the REP model, the RRM allows the estimation of variance components and the prediction of breeding values along the entire time gradient (Coelho et al., 2020), and the

modeling of genetic and non-genetic effects makes selection of superior genotypes much more efficient (Melo et al. 2020).

Knowing the superiority of RRM for what has already been presented, the selection index based on RRM should be preferred. Furthermore, as it presents the same results with REP regarding selected genotypes, the selection of genotypes 15 (Iconha / '*Riverside*' *citrandarin*), 22 (CNPMF5059 / '*Riverside*' *citrandarin*) and 4 (Itarana / *Swingle citrumelo*) is the most pertinent. These combinations can be recommended for planting, since their propagation is completely asexual, via grafting to promote union of the scion with the rootstock (Carvalho et al., 2019).

It is worth to highlight the importance of this study due to the lack of information related to the estimation of genetic parameters in the species *C. latifolia* Tanaka. The application of the random regression technique for the selection and recommendation of superior genotypes was a reliable methodology for 'Tahiti' acid lime since this crop has uninterrupted and irregular production over the years. Furthermore, this work demonstrated the efficiency of the selection index once it selects high productivity and less laborious genotypes.

Conclusion

Random regression models, fitted using Legendre polynomials, adequately deal with repeated measures, unbalanced and non-standardized data and are recommended to deal with $G \times M$ interaction in 'Tahiti' acid lime aiming at genotype recommendation. The applied methodology allowed the prediction of genotypic values for not evaluated measurements, and therefore, should be preferred over models based on analysis of variance and repeatability models in the genetic evaluation of 'Tahiti' acid lime.

The use of the additive index, based on the AUGT, allowed an accurate recommendation for SRC of 'Tahiti' acid lime. The most indicate scion/rootstock combinations indicated for cultivation are Iconha/'*Riverside*' *citrandarin*, CNPMF5059/'*Riverside*' *citrandarin*, and Itarana/*Swingle citrumelo*.

References

- Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Alvares CA, Stape JL, Sentelhas PC, et al (2013) Köppen's climate classification map for Brazil. Meteorologische Zeitschrift 22:711–728. <https://doi.org/10.1127/0941-2948/2013/0507>

- Alves RS, Resende MDV, Rocha JRASC, et al (2020a) Quantifying individual variation in reaction norms using random regression models fitted through Legendre polynomials: application in eucalyptus breeding. *Bragantia* 79:360–376. <https://doi.org/10.1590/1678-4499.20200125>
- Alves RS, Resende MDV, Azevedo CF, et al (2020b) Optimization of Eucalyptus breeding through random regression models allowing for reaction norms in response to environmental gradients. *Tree Genet Genomes* 16:38. <https://doi.org/10.1007/s11295-020-01431-5>
- Alves RS, Teodoro PE, Peixoto LA, et al (2019) Multiple-trait BLUP in longitudinal data analysis on *Jatropha curcas* breeding for bioenergy. *Ind Crops Prod* 130:558–561. <https://doi.org/10.1016/j.indcrop.2018.12.019>
- Araújo MJ, Dias DC, Scarpinati EA, Paula RC (2015) Número de repetições, de plantas por parcela e de avaliações para testes clonais de eucalipto. *Pesqui Agropecu Bras* 50:923–931. <https://doi.org/10.1590/S0100-204X2015001000008>
- Baba T, Momen M, Campbell MT, et al (2020) Multi-trait random regression models increase genomic prediction accuracy for a temporal physiological trait derived from high-throughput phenotyping. *PLoS One* 15:1–17. <https://doi.org/10.1371/journal.pone.0228118>
- Bastos DC, Sombra KES, Loureiro FLC, et al (2017) Initial development of Tahiti acid lime trees on different rootstocks in the semiarid region of Ceará, Brazil. *Citrus Research & Technology* 38:77–82. <https://doi.org/10.4322/crt.icc045>
- Campbell M, Momen M, Walia H, Morota G (2019) Leveraging Breeding Values Obtained from Random Regression Models for Genetic Inference of Longitudinal Traits. *Plant Genome* 12:180075. <https://doi.org/10.3835/plantgenome2018.10.0075>
- Carvalho SA, Girardi EA, Mourão Filho FAA, et al (2019) Advances in citrus propagation in Brazil. *Rev Bras Frutic* 41:. <https://doi.org/10.1590/0100-29452019422>
- Cellon C, Amadeu RR, Olmstead JW, et al (2018) Estimation of genetic parameters and prediction of breeding values in an autotetraploid blueberry breeding population with extensive pedigree data. *Euphytica* 214:87. <https://doi.org/10.1007/s10681-018-2165-8>
- Denaro M, Smeriglio A, Xiao J, et al (2020) New insights into Citrus genus: From ancient fruits to new hybrids. *Food Front* 1:305–328. <https://doi.org/10.1002/fft2.38>
- Dobson AJ, Barnett AG (2018) An introduction to generalized linear models, 4°. Chapman and Hall/CRC, New York
- Ferreira Coelho I, Peixoto MA, Evangelista JSPC, et al (2020) Multiple-trait, random regression, and compound symmetry models for analyzing multi-environment trials in maize breeding. *PLoS One* 15:e0242705. <https://doi.org/10.1371/journal.pone.0242705>
- Ferreira FM, Rocha JRASC, Alves RS, et al (2020) Estimates of repeatability coefficients and optimum number of measures for genetic selection of *Cynodon* spp. *Euphytica* 216:70. <https://doi.org/10.1007/s10681-020-02605-x>

- Gilmour AR, Gogel BJ, Cullis BR, et al (2015) ASReml user guide release 4.1 structural specification. Hemel hempstead: VSN international ltd
- Henderson CR (1975) Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics* 31:423. <https://doi.org/10.2307/2529430>
- Huynh H, Feldt LS (1970) Conditions Under Which Mean Square Ratios in Repeated Measurements Designs Have Exact F-Distributions. *J Am Stat Assoc* 65:1582. <https://doi.org/10.2307/2284340>
- Jamrozik J, Bohmanova J, Schaeffer LR (2010) Relationships between milk yield and somatic cell score in Canadian Holsteins from simultaneous and recursive random regression models. *J Dairy Sci* 93:1216–1233. <https://doi.org/10.3168/jds.2009-2585>
- Jamrozik J, Schaeffer LR (1997) Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. *J Dairy Sci* 80:762–770
- Ketelaere B De, Lammertyn J, Molenberghs G, et al (2003) Statistical models for analyzing repeated quality measurements of horticultural products. *Math Biosci* 185:169–189. [https://doi.org/10.1016/S0025-5564\(03\)00092-0](https://doi.org/10.1016/S0025-5564(03)00092-0)
- Kirkpatrick M, Lofsvold D, Bulmer M (1990) Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124:979–993. <https://doi.org/10.1093/genetics/124.4.979>
- Kranis A, Su G, Sorensen D, Woolliams JA (2007) The application of random regression models in the genetic analysis of monthly egg production in turkeys and a comparison with alternative longitudinal models. *Poult Sci* 86:470–475. <https://doi.org/10.1093/ps/86.3.470>
- Ladaniya MS, Marathe RA, AK, et al (2020) High density planting studies in acid lime (*Citrus aurantifolia* Swingle). *Sci Hortic* 261:108935. <https://doi.org/10.1016/j.scienta.2019.108935>
- Ly D, Huet S, Gauffreteau A, et al (2018) Whole-genome prediction of reaction norms to environmental stress in bread wheat (*Triticum aestivum* L.) by genomic random regression. *Field Crops Res* 216:32–41. <https://doi.org/10.1016/j.fcr.2017.08.020>
- Machado DLM, Siqueira DL, Salomão LCC, et al (2017) Evaluation of rootstocks for ‘Tahiti’ acid lime in northern state of Minas Gerais. *Rev Bras Frutic* 39:. <https://doi.org/10.1590/0100-29452017790>
- Malikouski RG, Peixoto MA, Moraes AL, et al (2021) Repeatability coefficient estimates and optimum number of harvests in graft/rootstock combinations for ‘tahiti’ acid lime. *Acta Sci Agron* 43:e51740–e51740 <https://doi.org/10.4025/actasciagron.v43i1.51740>
- Melo VL de, Marçal T de S, Rocha JRAS de C, et al (2020) Modeling (co)variance structures for genetic and non-genetic effects in the selection of common bean progenies. *Euphytica* 216:77. <https://doi.org/10.1007/s10681-020-02607-9>

- Meyer K, Hill WG (1997) Estimation of genetic and phenotypic covariance functions for longitudinal or “repeated” records by restricted maximum likelihood. *Livest Prod Sci* 47:185–200. [https://doi.org/10.1016/S0301-6226\(96\)01414-5](https://doi.org/10.1016/S0301-6226(96)01414-5)
- Meyer K, Kirkpatrick M (2005) Up hill, down dale: quantitative genetics of curvaceous traits. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360:1443–1455. <https://doi.org/10.1098/rstb.2005.1681>
- Momen M, Campbell MT, Walia H, Morota G (2019) Predicting longitudinal traits derived from high-throughput phenomics in contrasting environments using genomic Legendre polynomials and B-splines. *G3: Genes, Genomes, Genetics* 9:3369–3380. <https://doi.org/10.1534/g3.119.400346>
- Morais AL de, Zucoloto M, Malikouski RG, et al (2020) Vegetative development and production of ‘Tahiti’ acid lime clone selections grafted on different rootstocks. *Rev Bras Frutic* 42:. <https://doi.org/10.1590/0100-29452020585>
- Nielsen NH, Backes G, Stougaard J, et al (2014) Genetic Diversity and Population Structure Analysis of European Hexaploid Bread Wheat (*Triticum aestivum* L.) Varieties. *PLoS One* 9:e94000. <https://doi.org/10.1371/journal.pone.0094000>
- Oliveira LGM, Marques A, Lopes ED, et al (2021) Produtividade, adaptabilidade e estabilidade genotípica de clones de *Eucalyptus* spp. e *Corymbia* spp. em diferentes espaçamentos de plantio. *Sci For* 49:e3664
- Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554. <https://doi.org/10.1093/biomet/58.3.545>
- Paulus D, Zorzi IC, Rankrape F, et al (2017) Harvest seasons and pruning management in pepper: production and pungency of the fruits. *Hortic Bras* 35:434–439. <https://doi.org/10.1590/s0102-053620170320>
- Peixoto MA, Alves RS, Coelho IF, et al (2020) Random regression for modeling yield genetic trajectories in *Jatropha curcas* breeding. *PLoS One* 15:e0244021. <https://doi.org/10.1371/journal.pone.0244021>
- Peixoto MA, Malikouski RG, Evangelista JSPC, et al (2021) Multitrait and multiharvest analyses for genetic assessment and selection of Tahiti acid lime genotypes through Bayesian inference. *Sci Hortic* 290:110536. <https://doi.org/10.1016/j.scienta.2021.110536>
- Resende MAV, Freitas JA, Lanza MA, et al (2014a) Divergência genética e índice de seleção via BLUP em acessos de algodoeiro para características tecnológicas da fibra. *Pesqui Agropecu Trop* 44:334–640. <https://doi.org/10.1590/S1983-40632014000300006>
- Resende MDV (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breeding and Applied Biotechnology* 16:330–339. <https://doi.org/10.1590/1984-70332016v16n4a49>
- Resende MDV (2015) *Genética quantitativa e de populações*, 1st edn. UFV, Viçosa

- Resende MDV, Alves RS (2020) Linear, generalized, hierarchical, bayesian and random regression mixed models in genetics/genomics in plant breeding. *Functional Plant Breeding Journal* 2:1–31. <https://doi.org/10.35418/2526-4117/v2n2a1>
- Resende MDV, Resende GDSP, Fernandes JSC (2001) Regressao aleatória e funções de covariância na análise de medidas repetidas. *Revista de matematica e estatistica* 21–40
- Resende MDV, Silva FF, Azevedo CF (2014b) Estatística matemática, biométrica e computacional: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-QWAS, estatística espacial e temporal, competição, sobrevivência. UFV, Viçosa
- Ribeiro LP, Carvalho LP, Farias FJC, et al (2018) Genetic gains in agronomic and technological traits of elite cotton genotypes. *Bragantia* 77:466–475. <https://doi.org/10.1590/1678-4499.2017329>
- Rocha JRASC, Marçal TS, Salvador FV, et al (2018a) Genetic insights into elephantgrass persistence for bioenergy purpose. *PLoS One* 13:e0203818. <https://doi.org/10.1371/journal.pone.0203818>
- Rocha JRAS, Machado JC, Carneiro PCS (2018b) Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. *GCB Bioenergy* 10:52–60. <https://doi.org/10.1111/gcbb.12443>
- Sales GNB, Silva LA, Almeida JF, et al (2018) Quality of Fruits from Grafted Tahiti Lime (*Citrus latifolia* Tan) Irrigated with Waters of Different Salinities. *Journal of Experimental Agriculture International* 27:1–10. <https://doi.org/10.9734/JEAI/2018/44302>
- Santis D, Frangipane MT (2020) *Citrus aurantium* L.: Cultivar impact on sensory profile. *Int J Gastron Food Sci* 20:100203. <https://doi.org/10.1016/j.ijgfs.2020.100203>
- Santos MG, Passos OS, Soares Filho WS, et al (2013) Variability analysis of “Persian” acid lime tree selections using agronomic and molecular markers. *Genetics and Molecular Research* 12:4604–4614. <https://doi.org/10.4238/2013.October.17.3>
- Santos MG, Soares Filho WS, Girardi EA, et al (2016) Initial horticultural performance of nine ‘Persian’ lime selections grafted onto Swingle citrumelo. *Sci Agric* 73:109–114. <https://doi.org/10.1590/0103-9016-2015-0058>
- Schaeffer LR (2004) Application of random regression models in animal breeding. *Livest Prod Sci* 86:35–45. [https://doi.org/10.1016/S0301-6226\(03\)00151-9](https://doi.org/10.1016/S0301-6226(03)00151-9)
- Schaeffer LR (2016) Random regression models. University of Guelph. Available in <http://animalbiosciences.uoguelph.ca/~lrs/BOOKS/rrmbook.pdf>
- Silva MF, Maciel GM, Finzi RR, et al (2020) Selection indexes for agronomic and chemical traits in segregating sweet corn populations. *Hortic Bras* 38:71–77. <https://doi.org/10.1590/s0102-053620200111>
- Sun J, Rutkoski JE, Poland JA, et al (2017) Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic

Prediction for Wheat Grain Yield. *Plant Genome* 10:.
<https://doi.org/10.3835/plantgenome2016.11.0111>

Tucker SL, Dohleman FG, Grapov D, et al (2020) Evaluating maize phenotypic variance, heritability, and yield relationships at multiple biological scales across agronomically relevant environments. *Plant Cell Environ* 43:880–902.
<https://doi.org/10.1111/pce.13681>

van Eeuwijk FA, Bustos-Korts D, Malosetti M (2016) What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype × Environment Interactions? *Crop Sci* 56:2119–2140. <https://doi.org/10.2135/cropsci2015.06.0375>

Viana AP, Resende MDV de (2014) Genética quantitativa no melhoramento de fruteiras. *Interciência*

Zekri M (2000) 594 Citrus Rootstocks Affect Scion Nutrition, Fruit Quality, Growth, Yield, and Economical Return. *HortScience* 35:499C – 499.
<https://doi.org/10.21273/HORTSCI.35.3.499C>

Statements & Declarations

Funding

This work had the financial support of the National Council for Scientific and Technological Development (CNPq), Espírito Santo Foundation for Research and Innovation Support (FAPES), the Research Support Foundation of the State of Minas Gerais (Fapemig) and Coordination for the Improvement of Higher Education Personnel (CAPES) financing code 001.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Data Availability

The ASReml software outputs are available in the Supplementary Information (SI) – Table S1 to Table S4. The other datasets generated during the current study are available from the corresponding author on reasonable request.

Supplementary information

The supplementary information is available in the link below:

<https://link.springer.com/article/10.1007/s10681-022-03105-w#Sec15>

Figures

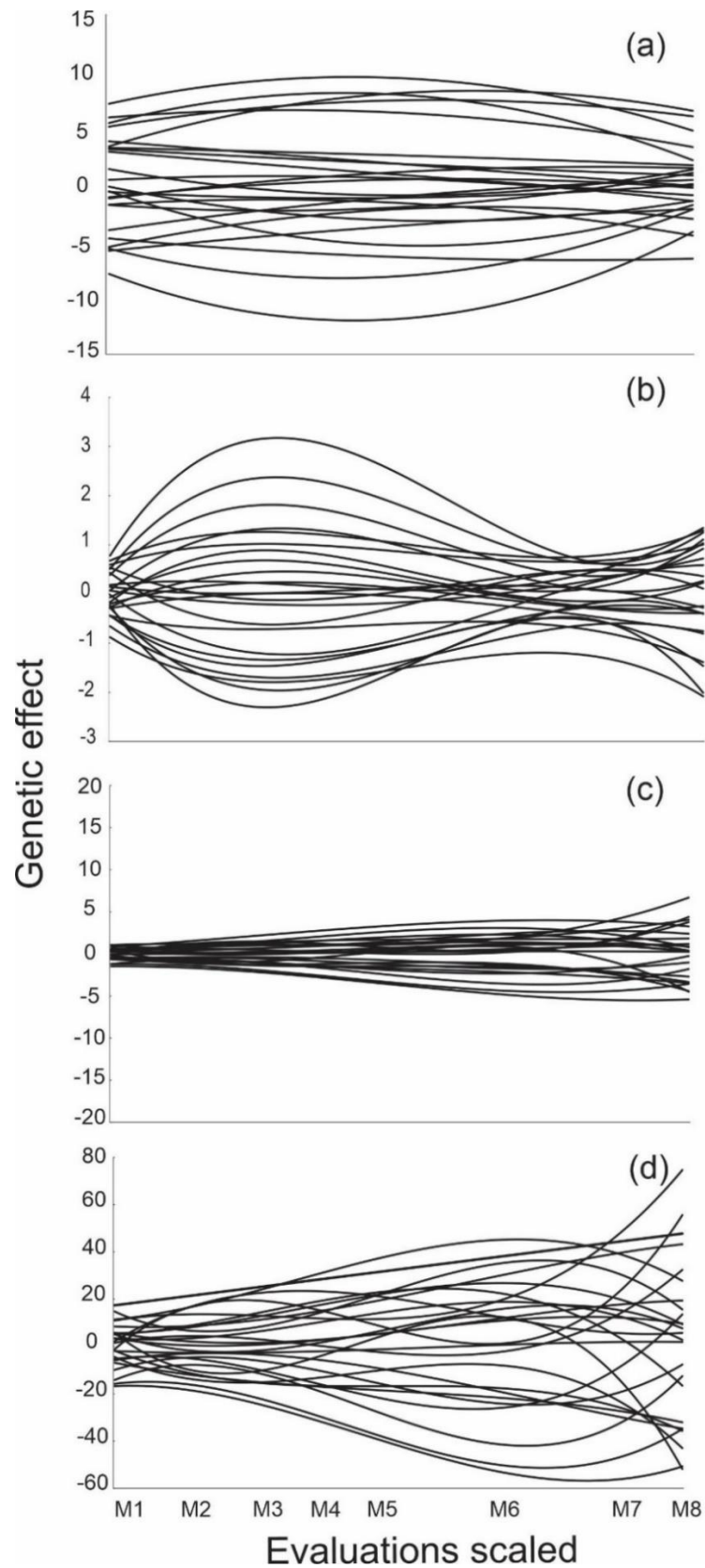


Fig 1. Genotypic trajectories for trunk diameter (a), canopy volume (b), fruit yield (c) and number of fruits (d) on the measurement scale (M1-M8) of 24 ‘Tahiti’ acid lime scion/rootstock combinations.

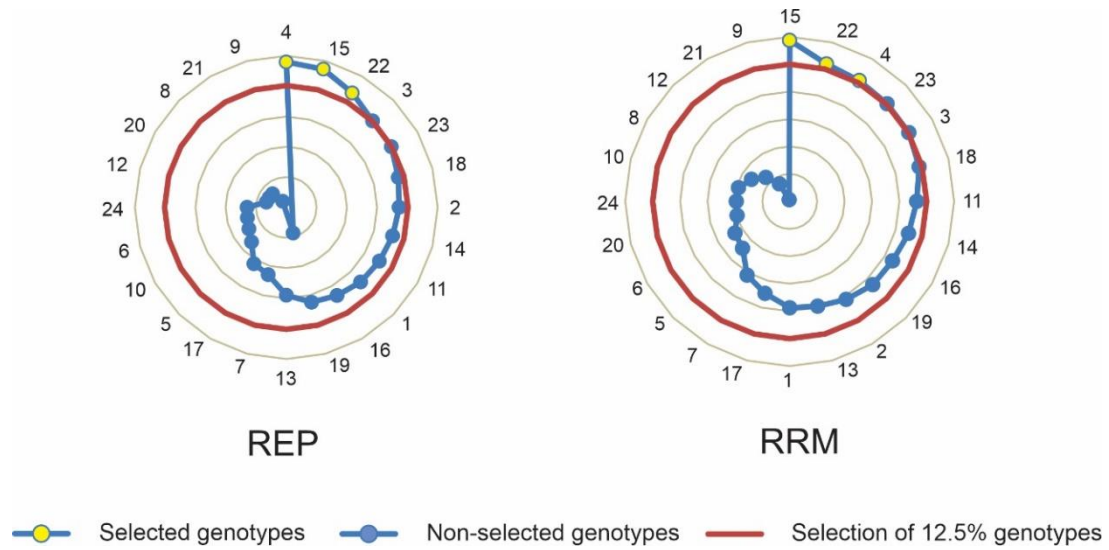


Fig 2. Ranking of the 24 ‘Tahiti’ acid lime scion/rootstock combinations according to the additive selection index by predicted genotypic values via repeatability model (REP) and random regression model (RRM).

ARTIGO 3

Aplicação de modelos probabilísticos bayesianos para recomendação de lima ácida ‘Tahiti’ em dados longitudinais.

Resumo

Modelos probabilísticos capazes de prever a superioridade e a persistência de genótipos podem potencializar programas de melhoramento de espécies perenes, dada sua propriedade de indicar indivíduos superiores e persistentes com um nível de probabilidade associado, além de fornecer uma medida de incerteza em torno dos parâmetros estimados através dos intervalos de credibilidade. Neste estudo, testamos a aplicabilidade de um modelo probabilístico bayesiano para a recomendação de 24 genótipos de lima ácida ‘Tahiti’. Dados provenientes de 12 colheitas nos genótipos avaliados foram ajustados em um modelo probabilístico por meio de um algoritmo amostrador de Monte Carlo Hamiltoniano. Calculou-se a probabilidade de performance superior do valor genético de cada genótipo no contexto geral e em cada colheita, bem como a probabilidade de reduzida variância da interação genótipos x colheitas como forma de avaliação da performance e persistência dos genótipos, respectivamente. O ajuste do modelo bayesiano com estrutura de variância residual heterogênea foi validado e comparado utilizando um modelo frequentista com os mesmos parâmetros para avaliação da persistência dos genótipos. Nossos resultados mostraram a aplicabilidade do modelo probabilístico bayesiano através de parâmetros semelhantes com os do modelo frequentista, porém com níveis de probabilidade associados a performance e estabilidade dos genótipos. Os genótipos G15, G4, G18 e G11 foram considerados os mais superiores, enquanto G24, G7, G13 e G3 foram os mais persistentes.

Palavras-chave: *Citrus latifolia* Tanaka, culturas perenes, modelagem estatística, Monte Carlo Hamiltoniano (MCH).

Introdução

Fruteiras perenes possuem características que dificultam seu melhoramento, como o longo período juvenil e a expressão diferencial de caracteres quantitativos ao longo dos anos (Gray and Brady, 2016). O período juvenil impede que caracteres de produtividade sejam avaliados precocemente, atrasando a geração de resultados e consequentemente a seleção dos

genótipos superiores. A variação da performance no decorrer do tempo é causada pela interação genótipos $\times$ colheitas (IGC), e exige que estas plantas sejam avaliadas repetidas vezes, aumentando os custos e a duração dos ciclos de melhoramento (Chaves et al., 2021). Devido a presença da IGC, os objetivos dos melhoristas não devem se restringir somente à avaliação de desempenho, mas também devem contemplar a estabilidade dos genótipos (Malosetti et al., 2013; Rai and Shekhawat, 2014; Zhang et al., 2018). Além disso, deve-se lidar com a não-ortogonalidade e avaliações não equidistantes ao longo do tempo, visto que são realizadas no mesmo indivíduo (Piepho and Eckl, 2014). Estes fatores prejudicam os ganhos genéticos por ciclo de melhoramento e a recomendação de indivíduos superiores.

Em lima ácida ‘Tahiti’ (*Citrus latifolia* Tanaka L.), algumas soluções foram propostas para superar tais problemas. A enxertia, processo bem estabelecido no melhoramento de espécies citrícolas, influencia no desenvolvimento vegetativo e reduz a duração do período juvenil, permitindo avaliações mais precoces no programa de melhoramento. Estes e outros efeitos podem variar em intensidade, a depender da constituição genotípica do porta-enxerto (García-Muñoz et al., 2021; Hassanzadeh Khankahdani et al., 2019). Unindo a enxertia com a proposta de Malikouski et al. (2021) de que a avaliação de somente quatro colheitas é suficiente para selecionar de forma acurada, os ciclos de melhoramento de lima ácida podem ser ainda mais encurtados. Métodos para lidar com dados longitudinais não-ortogonais de lima ácida também foram propostos, a exemplo de modelos de regressão aleatória utilizando polinômios ortogonais de Legendre, que permitem estimar a trajetória genotípica dos tratamentos em avaliação (Malikouski et al., 2022).

Apesar dos avanços, ainda há necessidade de estabelecer metodologias para lidar com a IGC na seleção dos melhores candidatos. A IGC possui duas principais causas que, agindo de maneira conjunta, concorrem para a sua maior ou menor intensidade. A primeira é extrínseca à planta e está relacionada aos diferentes padrões climáticos e práticas culturais dispendidas a partir do momento em que o indivíduo é implantado no campo (Dunham et al., 2018). Tais padrões e práticas podem agir de maneira benéfica ou maléfica à performance dos genótipos em avaliação (Lambert et al., 2015; Leisner, 2020). A segunda causa é fisiológica, e tem relação com o estágio de maturação da planta. Em outras palavras, haverá expressão diferencial dos genes a depender da idade fisiológica do indivíduo.

Genótipos que possuam alelos que confirmam plasticidade para se adequarem às variadas condições ambientais de cada ano devem ser priorizados. Para isto, a estabilidade dos candidatos à seleção deve ser aferida. Um conceito análogo à estabilidade chamado

“persistência” é utilizado por melhoristas de forrageiras perenes, e consiste na capacidade de sobreviver, e de produzir matéria seca por longos períodos (Parsons et al., 2011; Rocha et al., 2018). Em fruteiras perenes, este conceito pode ser readequado como a capacidade de manter um alto rendimento de frutos por vários anos (Chaves et al. 2022). Uma segunda possibilidade é guiar a seleção usando metodologias baseadas em conceitos de probabilidade, tendo como ideia central a redução do risco associado à seleção de um dado genótipo. Esta vertente aproxima o melhoramento do dilema diário do produtor rural, que procura guiar suas ações de modo a minimizar os riscos de ocorrência de baixa produção em uma determinada safra (Eskridge et al. 1991). Dias et al. (2022) partiram desta ideia para propor um método embasado em estatística bayesiana que auxilia na seleção de genótipos que reúnam alelos favoráveis para a característica de interesse e que sejam, simultaneamente, previsíveis, proporcionando a diminuição dos riscos relacionados à seleção. A base bayesiana proporciona ao método características particulares à esta vertente estatística, como por exemplo, o aproveitamento de informações *a priori*, isto é, informações prévias incorporadas na distribuição de parâmetros do modelo, além de apresentar componentes de variância e valores genéticos com intervalos de confiabilidade inferiores quando comparados a modelos frequentistas (Covre et al., 2022).

Apesar de ter sido proposta e implementada para o contexto multi-locais, a metodologia proposta por Dias et al (2022) pode ser readequada ao contexto multi-colheitas. Ainda não há registros de tal readequação, tão pouco de sua aplicação em espécies perenes como a lima ácida ‘Tahiti’, através de múltiplas colheitas. Desse modo, objetivou-se neste estudo, avaliar a aplicabilidade de um modelo probabilístico bayesiano para a recomendação de genótipos superiores e persistentes de lima ácida ‘Tahiti’.

Material e Métodos

Rede experimental e material vegetal

Os dados de lima ácida ‘Tahiti’ são provenientes de um experimento implantado em julho de 2015, em São Mateus, Espírito Santo, Brasil (18°48'21"S, 39°53'30"W, 35 m.a.s.l). Esta região possui uma estação chuvosa no verão e uma estação seca no inverno, sendo classificado como Aw, seguindo a classificação de Köppen (Alvares et al., 2013). Os dados climáticos de precipitação e temperatura durante o período de realização do experimento podem ser consultados na Figura 1.

Para este estudo, 24 combinações de copa/porta enxerto de lima ácida ‘Tahiti’ foram avaliadas para a característica de rendimento de frutos, expresso em kg/árvore. Tais

combinações contemplaram dois híbridos como porta-enxerto, *Citrumelo swingle* (*Citrus paradisi* X *Poncirus trifoliata*) e *Citrandarin 'riverside'* (*Citrus sunki* X *Poncirus trifoliata*), e 12 clones de lima ácida 'Tahiti' como enxerto (Tabela 1). Todos os materiais utilizados no experimento são provenientes do Banco Ativo de Germoplasma da Embrapa Mandioca e Fruticultura (Santos et al., 2013). Deste ponto em diante, as combinações serão denominadas genótipos. O período de avaliação do experimento ocorreu entre julho de 2017 e setembro de 2020, e consistiu em 12 colheitas realizadas nos seguintes dias após a implantação do experimento: 736, 808, 861, 918, 972, 1083, 1200, 1249, 1415, 1568, 1633 e 1867.

Tabela 3. Codificação dos 24 genótipos de lima ácida 'Tahiti' utilizadas no experimento.

Copa	Porta-enxerto	
	<i>Citrumelo swingle</i>	<i>Citrandarin riverside</i>
Bello Fruit	G1	G13
Elédio	G2	G14
Iconha	G3	G15
Itarana	G4	G16
Santa Rosa	G5	G17
Bearss Lime	G6	G18
CNPMF01	G7	G19
CNPMF02	G8	G20
CNPMF2001	G9	G21
CNPMF5059	G10	G22
BRS Passos	G11	G23
Persian 58	G12	G24

Análises estatísticas

Dois modelos bayesianos foram utilizados para a avaliação da superioridade dos genótipos neste estudo. O primeiro modelo (B-ID) apresentou a seguinte probabilidade normal condicional:

$$y_{ijk} \sim N(E[y_{ijk}], \sigma^2)$$

em que:

$$E[y_{ijk}] = \mu + g_i + r_j + h_k + gh_{ik} + p_{ij} + e$$

onde $E[y_{ijk}]$ é o valor fenotípico do $i^{\text{ésimo}}$ genótipo, avaliado no $j^{\text{ésimo}}$ bloco, na $k^{\text{ésima}}$ colheita. μ é a média geral, g_i é o efeito genotípico, r_j é o efeito de bloco, h_k é o efeito de colheita, gh_{ik} é a interação genótipos x colheitas, p_{ij} é o efeito ambiental permanente.

A distribuição de probabilidades *a priori* para cada parâmetro do modelo foi definida como:

$$\begin{aligned}\mu &\sim N(0, \sigma_{[\mu]}) \\ r &\sim N(0, \sigma_{[r]}) \\ h &\sim N(0, \sigma_{[h]}) \\ g &\sim N(0, \sigma_{[g]}) \\ gh &\sim N(0, \sigma_{[gh]}) \\ p &\sim N(0, \sigma_{[p]}) \\ e &\sim HalfCauchy(0, \sigma_{[e]})\end{aligned}$$

onde $N(0, \sigma_{[\alpha]})$ e $HalfCauchy(0, \sigma_{[\alpha]})$ representam as distribuições normal e half-Cauchy, respectivamente, com média zero e diferentes escalas de hiperparâmetros $\sigma_{[\alpha]}^2$. Por ser um modelo hierárquico, é conveniente adotar essas distribuições (Gelman, 2006). Também consideraram-se as seguintes hiperprioris para os respectivos hiperparâmetros:

$$\begin{aligned}\sigma_{[\mu]} &\sim HalfCauchy(0, \varphi) \\ \sigma_{[r]} &\sim HalfCauchy(0, \varphi) \\ \sigma_{[h]} &\sim HalfCauchy(0, \varphi) \\ \sigma_{[g]} &\sim HalfCauchy(0, \varphi) \\ \sigma_{[gh]} &\sim HalfCauchy(0, \varphi) \\ \sigma_{[p]} &\sim HalfCauchy(0, \varphi)\end{aligned}$$

onde φ é o hiperparâmetro global conhecido, sendo definido como $\varphi = \max(y) \times 10$, definido de tal forma que resulta em hiperprioris de segundo nível, pouco informativos, e que permitem aos dados dominar a distribuição a posteriori se a probabilidade é forte (Gelman et al., 2013).

O segundo modelo (B-DG) foi ajustado com as mesmas considerações de B-ID, diferindo apenas por adotar efeitos residuais heterogêneos. A seleção do modelo de melhor ajuste se deu pelo menor valor do Critério de Informação Watanabe-Akaïke 2 (WAIC2) (Gelman et al., 2013). O modelo selecionado foi utilizado para calcular as probabilidades e realizar as inferências a respeito do desempenho e estabilidade dos genótipos de lima ácida ‘Tahiti’. A partir deste modelo, também foram estimados os componentes de variância e os

valores genéticos via algoritmo Monte Carlo Hamiltoniano, com 4000 iterações, 4 cadeias e 4 núcleos declarados (Nishio and Arakawa, 2019).

Probabilidade de performance e persistência genotípica no contexto bayesiano

Ambas as medidas de desempenho e estabilidade utilizadas neste estudo, foram propostas por Dias et al. (2022). A probabilidade de performance dos genótipos de lima ácida ‘Tahiti’ foi calculada a partir da amostragem da distribuição a posteriori do valor genético marginal dos genótipos, dado o valor fenotípico observado $[p(g|y)]$. Os genótipos foram ranqueados a partir de valores de probabilidade dada a ocorrência de o certo evento. Os genótipos foram ranqueados a partir de valores decrescentes de probabilidade dada a ocorrência do seguinte evento: um dado genótipo apresente performance superior entre os quatro maiores valores genéticos (intensidade de 17%) em cada uma das colheitas ou de forma geral, considerando todas as colheitas, sendo representado por $T(g_i^s \in \Omega_s)$. Deste modo, a probabilidade de genótipo pertencer ao subconjunto de genótipos superiores (probabilidade de desempenho superior) foi calculada a partir da seguinte expressão:

$$Pr(g_i^s \in \Omega_s | y) = \frac{1}{S} \sum_{s=1}^S T(g_i^s \in \Omega_s | y)$$

em que S é o número de amostragens.

A estabilidade através das medidas foi calculada a partir dos efeitos da variância da interação genótipos x colheitas (IGC). Os genótipos que apresentam menores valores de variância de IGC tendem a ser mais estáveis, mostrando um desempenho mais previsível entre as colheitas. Dessa forma, o desempenho de cada genótipo pertencer ao grupo daqueles quatro de menor IGC (menor valor de σ_{gh}^2), representada por $T(g_i^s \in \ell V)$, foi computado a partir da seguinte expressão:

$$Pr(var(gh_{ik}) \in T(g_i^s \in \ell V) = \frac{1}{S} \sum_{s=1}^S T(g_i^s \in \ell V)$$

Persistência no contexto frequentista

Obteve-se também a estabilidade no contexto frequentista procedendo-se com a aplicação do seguinte modelo (F-DG):

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_1\mathbf{r} + \mathbf{X}_2\mathbf{h} + \mathbf{Z}_1\mathbf{g} + \mathbf{Z}_2\mathbf{gh} + \mathbf{Z}_3\mathbf{p} + \mathbf{e}$$

onde $\mathbf{y}$ é o vetor de dados fenotípicos, μ é o intercepto do modelo, $\mathbf{r}$ é o vetor de efeitos de repetição (assumidos como fixo), $\mathbf{h}$ é o vetor de efeitos de colheitas (assumidos como fixos), $\mathbf{g}$

é o vetor de efeitos genotípicos (aleatório) [$\mathbf{g} \sim N(0, \mathbf{I}\sigma_g^2)$], $\mathbf{gh}$ é o vetor de efeitos da interação genótipos x colheitas (aleatório) [$\mathbf{gh} \sim N(0, \mathbf{I}\sigma_{gh}^2)$], $\mathbf{p}$ é o vetor de efeitos permanentes de parcela (aleatório) [$\mathbf{p} \sim N(0, \mathbf{I}\sigma_p^2)$], e $\mathbf{e}$ é o vetor de resíduos associados as observações fenotípicas (aleatório) [$\mathbf{e} \sim N(0, \mathbf{R})$], onde $\mathbf{R}$ é uma matriz de covariância residual de dimensão $\mathbf{h} \times \mathbf{h}$. As letras maiúsculas $\mathbf{X}_1$ e $\mathbf{X}_2$ referem-se as matrizes de incidência de efeitos fixos, e $\mathbf{Z}_1$, $\mathbf{Z}_2$ e $\mathbf{Z}_3$ as matrizes de efeitos aleatórios dos respectivos efeitos. $\mathbf{1}$ é um vetor de uns.

Os componentes de variância foram estimados via máxima verossimilhança residual – REML (Patterson and Thompson, 1971), e os valores genéticos preditos via melhor preditor linear não viesado – BLUP (Henderson, 1975). A hipótese de significância dos efeitos aleatórios do modelo foi verificada via teste de razão de verossimilhança (em inglês, likelihood ratio test - LRT) (Wilks, 1938).

A persistência dos genótipos (P_i) no contexto frequentista foi obtida pela distância entre cada genótipo em relação ao ideótipo, considerando todos os valores genéticos estimados em cada uma das doze colheitas (g_{ij}). O ideótipo (g_{max}) foi definido como o valor genético máximo estimado em cada dia do período experimental. A persistência foi calculada utilizando a seguinte expressão (Rocha et al., 2018):

$$P_i = \frac{1}{\sum_{k=1}^{12} (g_{ij} - g_{max})^2} \bigg/ \sum_{i=1}^{24} \frac{1}{\sum_{k=1}^{12} (g_{ij} - g_{max})^2}$$

Estimou-se a correlação de Spearman no ranqueamento de persistência dos genótipos entre o método probabilístico bayesiano e frequentista.

Softwares e tempo computacional

As análises bayesianas foram realizadas usando a linguagem de programação probabilística Stan (Carpenter et al., 2017) através do pacote RStan (Stan Development, 2018) e as frequentistas por meio do pacote ASReml (Gilmour et al., 2015), ambas no software R (R Development Core Team, 2020).

O tempo computacional para o ajuste dos modelos bayesianos e frequentistas foi mensurado através da função base *Sys.time* em R. As análises foram executadas em um servidor Dell Power Edge R720 com 24 núcleos, 512GB de memória RAM, e 7TB de espaço em disco.

Resultados

Probabilidade de performance superior dos genótipos

Os modelos bayesianos B-ID e B-DG apresentaram a média da estatística $\hat{R}$ próximo a 1, indicando boa convergência para os parâmetros dos modelos (Tabela 2). No entanto, o modelo B-DG apresentou o melhor ajuste, devido o menor valor de WAIC2 (Tabela 2). A Figura 2A apresenta a densidade dos dados observados com a densidade dos dados gerados pelas cadeias no modelo B-DG. Nota-se que o modelo foi eficiente em replicar a distribuição dos dados observados por meio dos dados gerados.

Tabela 2. Estatística comparativa dos modelos bayesianos usando o conjunto de dados de lima ácida 'Tahiti'.

Estatísticas de verificação preditivas a posteriori	Modelo de variância residual homogêneo	Modelo de variância residual heterogêneo
WAIC2	8420,87	6899,61
$\hat{R}$ médio	1,00	0,99

Os genótipos G15, G4, G18 e G11 apresentaram os maiores valores genéticos com seus respectivos intervalos de densidade a posteriori, enquanto os menores valores foram encontrados para G9 e G21 (Figura 2B). A probabilidade de performance superior de um dado genótipo pertencer a intensidade de seleção predefinida de 4 ($Pr(g_i^s \in \Omega_s | y)$) indicou que G15, G4, G18, G11, G3, G23, G22, G1, G19, G14, G2, G13 e G16 possuem alguma chance de estarem dentre os selecionados. Os 11 genótipos restantes não apresentaram desempenho superior em nenhum cenário (Figura 2C). Vale destacar os genótipos G15 e G4, os quais apresentaram elevada performance e oferecem um baixo risco de insucesso se selecionados [$Pr(g_i^s \in \Omega_s | y) > 0,75$].

Os gráficos de calor ilustram uma comparação par-a-par da probabilidade de performance superior dos genótipos considerando todas as colheitas simultaneamente e em cada colheita de forma individual (Figuras 3A e 4). O gráfico de comparação par-a-par apresenta dois lados simétricos, que indicam a probabilidade de vitória (diagonal inferior) e derrota (diagonal superior) dos genótipos. O genótipo G15, por exemplo, possui altas probabilidades de vencer quase todos os genótipos, exceto G4, onde a probabilidade se aproxima de 0.75. Por outro lado, G9 é vencido por todos os seus pares em quase todas as situações, com exceção de G21, que vence em três terços das vezes. A coloração esverdeada indica genótipos que empatam

em seu desempenho (probabilidade de desempenho superior aproximadamente 50%), destacando-se G18 x G11, G2 x G14 e G10 x G24 (Figura 3A).

A probabilidade de desempenho dos materiais também pode ser analisada ao longo das colheitas (Figura 3B). Nota-se que a interação genótipos x colheitas foi considerável, provocando alterações nos desempenhos relativos dos genótipos ao longo das colheitas (Figura 3A e 4), ou seja, aqueles que apresentaram produção alta em uma colheita não necessariamente tiveram essa performance na seguinte. Os genótipos G15, G4, G18 e G11 apresentaram probabilidades superiores a 0.5 em quase todas as medidas, indicando que eles possuem alta performance produtiva, estando entre os quatro de melhor desempenho (Figura 3B). Os genótipos G9 e G21 apresentaram probabilidades nulas durante o período de avaliação de colheitas neste estudo, não sendo recomendados.

Probabilidades de performance superior dos genótipos foram inferiores ao se considerar as doze colheitas de forma isolada (Figura 4). O desempenho dos genótipos entre si também sofreu elevada variação ao longo das colheitas, ou seja, genótipos que apresentaram produção alta em uma colheita não necessariamente tiveram essa performance na seguinte. Na Figura 4 observou-se probabilidades de vitória próximos a 50% na maioria das comparações, indicando maior incerteza na recomendação de genótipos superiores quando comparado ao par-a-par simultâneo entre as colheitas.

Probabilidade de persistência genotípica

A probabilidade de se ter um genótipo de lima ácida ‘Tahiti’ com uma pequena variância na interação genótipos x colheitas apresentou baixos valores no contexto bayesiano, sendo que apenas G24 apresentou valores superiores a 0.3, que o caracteriza como o genótipo mais estável dentre os avaliados (Figura 2D).

A aplicabilidade das análises foi validada e comparada com análises tradicionais frequentistas. Ambos os modelos, o bayesiano de melhor ajuste (B-DG) e o frequentista (F-DG) apresentaram efeitos genéticos, de interação genótipos x colheitas e permanente de parcela, significativos pelo LRT. Os componentes de variância apresentaram valores similares (Tabela 3). A variância genética superou os efeitos da interação genótipos x colheitas e permanente de parcela em ambos os modelos, apresentando valor de 2.20 para o modelo frequentista e 2.44 para o modelo bayesiano. A variância residual apresentou valores heterogêneos ao longo das colheitas, que variaram de 0.88 (no modelo frequentista, na colheita 2) a 492.71 (no modelo bayesiano, na colheita 12).

Tabela 3. Componentes de variância e tempo computacional de ajuste do modelo frequentista e bayesiano nos dados de lima ácida ‘Tahiti’ com os respectivos intervalos de credibilidade a posteriori inferior (L-HPD) e superior (U-HPD) para o modelo bayesiano a 5% de probabilidade de erro.

Parâmetro	Modelo frequentista		Modelo bayesiano	
	Componente	L-HPD	Componente	U-HPD
σ_g^2	2,2087	1,3096	2.4441	4,2189
σ_{gh}^2	0,8102	0,4448	0.7566	1,0907
σ_p^2	0,2424	0,1073	0.2619	0,4224
σ_e^2	C1	1,1914	0,8870	1.2352
	C2	0,8834	0,6316	0.9087
	C3	0,8630	0,7004	0.9710
	C4	5,6030	4,4784	5.8836
	C5	3,0726	2,7240	3.6013
	C6	4,2182	3,3668	4.5158
	C7	10,2574	8,5301	11.2297
	C8	20,1693	20,5676	26.7143
	C9	412,1721	328,6592	418.6705
	C10	136,7357	113,6122	146.6260
	C11	404,4514	324,7274	413.0839
	C12	473,6126	384,1706	492.7147
Tempo computacional	0,0667 min		375,8 min	

σ_g^2 : variância genotípica, σ_{gh}^2 variância da interação genótipos x colheitas, σ_p^2 variância do efeito de permanente do ambiente, σ_e^2 variância residual nas 12 colheitas (C1 a C12).

O modelo bayesiano possibilita a obtenção de intervalos de credibilidade. Deste modo cada componente de variância possui um intervalo inferior (L-HPD) e superior (U-HPD). Observou-se que nenhum intervalo passou pelo valor zero, indicando significância estatística para todos os efeitos do modelo a 5% de probabilidade de erro.

Em ambos os modelos, frequentista e bayesiano, o comportamento no aumento da variância residual foi similar. O tempo computacional utilizado no ajuste dos modelos foi de

0.06 minutos para o modelo frequentista e de 375.8 minutos para o modelo bayesiano, considerando mesmo hardware de execução.

A persistência no sentido frequentista apresentou resultados diferentes da persistência no contexto bayesiano. No entanto, em ambos os contextos, os valores foram baixos. A persistência genotípica no contexto frequentista variou de 5.7 a 3.4, sendo estes valores referentes a G3 e G9, respectivamente. Com exceção de G3, que apresentou o maior valor e bem superior aos demais, os demais valores dos genótipos situaram-se em um intervalo de 1.1, evidenciando a baixa capacidade de distinção entre os genótipos de lima ácida ‘Tahiti’ quanto à persistência.

A correlação entre o ranqueamento da persistência no contexto bayesiano (Figura 2 D) e no contexto frequentista (Figura 5), apresentou valor de 0,69 (p-valor de 0,01758), indicando correlação estatisticamente significativa entre a classificação dos dois métodos discutidos a 5% de probabilidade de erro.

Discussão

A consideração da interação genótipos x colheitas é preponderante na avaliação de espécies perenes, uma vez que a expressão dos genes se dá de forma diferencial frente as influências ambientais ao longo das colheitas (Ferreira et al., 2020). A utilização de modelos com variância heterogênea, ou seja, uma variância específica para cada colheita tem apresentado melhores ajustes, seja para espécies anuais em ensaios multi-locais ou para espécies perenes (Nuvunga et al., 2015; Piepho and Eckl, 2014). A seleção do modelo com base no WAIC2, associado ao aumento na variância residual ao decorrer das colheitas reforçou que o modelo heterogêneo é mais adequado do que o homocedástico para ajustar aos dados. Esses resultados destacam a importância de modelar a estrutura de covariâncias residuais e contabilizar diferentes fontes de variação em um conjunto de dados longitudinais.

A comparação dos valores genéticos associados a uma probabilidade a posteriori assegura maior confiança na análise do desempenho dos genótipos de lima ácida ‘Tahiti’. Em medidas de probabilidades obtidas por metodologia bayesiana, o melhorista pode explorar a probabilidade de um determinado genótipo superar qualquer outro, como por exemplo um candidato à seleção ser superior a uma cultivar amplamente difundida (Dias et al., 2022). Esta probabilidade pode auxiliar na tomada de decisão em situações em que genótipos possuem desempenhos próximos. Adicionalmente, a probabilidade de desempenho superior marginal, isto é, do genótipo ser superior a qualquer um outro, pode trazer maior confiabilidade quanto a

recomendação. Em um programa de melhoramento, há intensidades de seleção pré-definidas para a seleção e recomendação (Silva et al., 2021), e a utilização das abordagens probabilísticas neste contexto fornece maior certeza na escolha dos candidatos. Segundo Dias et al (2022), a vantagem de explorar a probabilidade de genótipos vencedores através da utilização de mapas de calor traz uma interpretação amigável para as comparações, facilita o processo de tomada de decisão e permite a extrapolação para outros campos de aplicação.

Os genótipos de lima ácida ‘Tahiti’ com maior probabilidade de performance superior, G15, G4, G18 e G11, podem ser recomendadas em programas de melhoramento, ou mesmo propagadas vegetativamente por viveiristas, agricultores e cultivadas em pomares comerciais. Segundo os métodos probabilísticos, estes genótipos possuem alelos que conferem plasticidade para se adequarem às variadas condições ambientais de cada ano mantendo um bom desempenho. A variação de desempenho superior dos genótipos ao longo das safras enfatizou uma existência de IGC atuante e que a seleção com base em uma ou poucas safras pode não ser representativa na escolha fiel dos candidatos superiores geneticamente.

Uma outra forma eficiente para investigação da IGC é por meio da persistência. A persistência foi inicialmente utilizada em espécies forrageiras, e mede a sustentação de um nível de produtividade através dos cortes (Rocha et al., 2018). No entanto, sua interpretação também é válida em estudos de colheitas em espécies frutíferas, dada sua natureza perene. Assim, diante do IGC, o melhorista pode dar preferência a genótipos com alta persistência, que são aqueles que mantêm um alto desempenho independentemente das mudanças ambientais.

A persistência no contexto bayesiano destacou os quatro genótipos com menor probabilidade de apresentarem variância genótipos x colheitas. Os baixos valores de probabilidade (todos inferiores a 0,4) evidenciaram que de forma geral os genótipos de lima ácida ‘Tahiti’ não apresentaram elevada persistência, destacando o comportamento diferencial dos materiais dada a influência do ambiente. A persistência computada utilizando um modelo frequentista apresentou 50% de concordância ao se referir aos 4 genótipos superiores entre as duas metodologias. Em outras palavras, ao se identificar os 4 genótipos mais persistentes, G24 e G3 seriam selecionados em ambos os cenários. Este fato foi reforçado pela correlação de 0,69 no ranqueamento de persistência entre o método bayesiano e frequentista, que indicou certa similaridade entre as metodologias.

A comparação feita aqui foi com fins de validação do modelo bayesiano, no entanto salienta-se que o modelo probabilístico bayesiano no contexto do trabalho prioriza a invariância, já a persistência frequentista está comparando o desempenho dos genótipos com

base no desempenho de um ideótipo, e a variância é utilizada de forma indireta nessa segunda metodologia.

O modelo frequentista ajustado considerou variâncias residuais heterogêneas para tornar a comparação com o modelo probabilístico justa. Apesar dos resultados de persistência e componentes de variância convergentes com o modelo Bayesiano, algumas limitações do modelo frequentista podem ser destacadas. De forma geral, observou-se valores de persistência muito semelhantes, com exceção de G3, o que torna difícil diferenciação de genótipos com resultados iguais ou próximos (Figura 5). A pequena diferença do ranqueamento entre as duas metodologias, reitera a necessidade por modelos mais precisos e poderosos estatisticamente. A utilização das prioris sobre os parâmetros do modelo Bayesiano conferem maior segurança em selecionar materiais mais ou menos persistentes (Dias et al., 2022, Silva et al., 2020). Além disso, através da densidade a posteriori, modelos bayesianos permitem a obtenção de componentes de variância com intervalos de probabilidade associados. Os intervalos de credibilidade são mais intuitivos e fornecem uma maneira natural de quantificar incerteza dos componentes, e do ponto de vista assintótico, os intervalos de credibilidade, obtidos de modelos bayesianos, são melhores do que os intervalos de confiança frequentistas (Gelman et al, 2013). À medida que o tamanho da amostra aumenta, os intervalos de credibilidade se tornam cada vez mais estreitos em torno do verdadeiro valor do parâmetro. Por outro lado, os intervalos de confiança frequentistas podem ser imprecisos para tamanhos de amostra pequenos ou moderados e, em alguns casos, podem não convergir para o verdadeiro valor do parâmetro à medida que o tamanho da amostra aumenta (Kruschke, 2015; Wasserman, 2013).

Apesar dos pontos de destaque, as análises bayesianas apresentam a desvantagem de requerer maior tempo computacional para sua execução. Em situações com baixa infraestrutura de hardware, modelos bayesianos podem ser inviáveis. A inferência bayesiana voltou a ganhar força na ciência com o avanço da tecnologia e o aumento da capacidade de memória dos computadores (Csilléry et al., 2010). Acredita-se que a realização das análises irá demandar cada vez menos tempo ao decorrer dos próximos anos, e o tempo computacional não será um fator impeditivo no uso de modelos bayesianos (Technow et al., 2015).

Conclusão

Este trabalho abordou o uso de métodos probabilísticos baseados em modelos bayesianos para melhorar a precisão na recomendação de genótipos que sofrem influência do efeito permanente do ambiente, além do efeito da interação ao longo das colheitas. Utilizar

modelos probabilísticos torna as práticas de seleção e recomendação mais acuradas, pois permite comparar genótipos com um valor probabilidade associado. Tais modelos consideram uma intensidade pré-estabelecida de seleção, o que resulta em um processo de tomada de decisão mais seguro. Além disso, os modelos bayesianos possibilitam a obtenção de intervalos de credibilidade para os componentes de variância, tornando a obtenção de parâmetros genéticos e a interpretação destes mais precisa.

A persistência oferece informações a respeito das tendências de adaptação dos genótipos de lima ácida ‘Tahiti’, possibilitando a identificação do desempenho e estabilidade dos genótipos. As combinações copa/porta enxerto G15, G4, G18 e G11 foram consideradas as mais superiores, e G24, G7, G13 e G3 as mais persistentes, dentre os genótipos avaliados.

Referências

- Alvares, C.A., Stape, J.L., Sentelhas, P.C., de Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift* 22, 711–728. <https://doi.org/10.1127/0941-2948/2013/0507>
- Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. Stan: A Probabilistic Programming Language. *J Stat Softw* 76. <https://doi.org/10.18637/jss.v076.i01>
- Chaves, S.F. da S., Alves, R.M., Dias, L.A. dos S., 2021. Contribution of breeding to agriculture in the Brazilian Amazon. I. Açaí palm and oil palm. *Crop Breeding and Applied Biotechnology* 21. <https://doi.org/10.1590/1984-70332021v21sa21>
- Chaves, S.F. da S.; Evangelista, J.S.P.C.; Alves, R.S.; Ferreira, F.M.; Alves R.M.; Dias, K.O.G; Bhering, L.L. 2022. Application of linear mixed models for multiple harvest/site trial analyses in perennial plant breeding. *Tree Genetics & Genomes* v.18, n.44. <https://doi.org/10.1007/s11295-022-01576-5>
- Covre, A. M., da Silva, F. A., Oliosi, G., Correa, C. C. G., Viana, A. P., & Partelli, F. L. (2022). Multi-Environment and Multi-Year Bayesian Analysis Approach in Coffee canephora. *Plants*, 11(23), 3274. <https://doi.org/10.3390/plants11233274>
- Cruz, C.D., Regazzi, A.J., Carneiro, P.C.S., 2012. Modelos Biométricos Aplicados ao Melhoramento Genético. UFV, Viçosa.
- Csilléry, K., Blum, M.G.B., Gaggiotti, O.E., François, O., 2010. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol Evol* 25, 410–418. <https://doi.org/10.1016/j.tree.2010.04.001>
- da Silva, F.A., Viana, A.P., Corrêa, C.C.G., Carvalho, B.M., de Sousa, C.M.B., Amaral, B.D., Ambrósio, M., Glória, L.S., 2020. Impact of Bayesian Inference on the Selection of *Psidium guajava*. *Sci Rep* 10, 1999. <https://doi.org/10.1038/s41598-020-58850-6>

- Dias, K.O.G., dos Santos, J.P.R., Krause, M.D., Piepho, H.-P., Guimarães, L.J.M., Pastina, M.M., Garcia, A.A.F., 2022. Leveraging probability concepts for cultivar recommendation in multi-environment trials. *Theoretical and Applied Genetics* 135, 1385–1399. <https://doi.org/10.1007/s00122-022-04041-y>
- Driedonks, N., Rieu, I., Vriezen, W.H., 2016. Breeding for plant heat tolerance at vegetative and reproductive stages. *Plant Reprod* 29, 67–79. <https://doi.org/10.1007/s00497-016-0275-9>
- Dunham, A.E., Razafindratsima, O.H., Rakotonirina, P., Wright, P.C., 2018. Fruiting phenology is linked to rainfall variability in a tropical rain forest. *Biotropica* 50, 396–404. <https://doi.org/10.1111/btp.12564>
- Eskridge, k. M.; Byrne, P. F.; Crossa, J. Selection of stable varieties by minimizing the probability of disaster. **Field Crops Research**, v. 27, n. 1-2, p. 169-181, 1991. [https://doi.org/10.1016/0378-4290\(91\)90029-U](https://doi.org/10.1016/0378-4290(91)90029-U)
- Evangelista, J.S.P.C., Peixoto, M.A., Coelho, I., Alves, R., Resende, M.D.V., Silva, F.F. e, Laviola, B., Bhering, L.L., 2022. Genetic evaluation and selection in *Jatropha curcas* through Frequentist and Bayesian inferences. *Bragantia* 81. <https://doi.org/10.1590/1678-4499.20210262>
- Ferreira, F.M., de Carvalho Rocha, J.R. do A.S., Alves, R.S., Elizeu, A.M., Benites, F.R.G., de Resende, M.D.V., de Souza Sobrinho, F., Bhering, L.L., 2020. Estimates of repeatability coefficients and optimum number of measures for genetic selection of *Cynodon* spp. *Euphytica* 216, 70. <https://doi.org/10.1007/s10681-020-02605-x>
- García-Muñoz, M.C., Henao-Rojas, J.C., Moreno-Rodríguez, J.M., Botina-Azain, B.L., Romero-Barrera, Y., 2021. Effect of rootstock and environmental factors on fruit quality of Persian lime (*Citrus latifolia* Tanaka) grown in tropical regions. *Journal of Food Composition and Analysis* 103, 104081. <https://doi.org/10.1016/j.jfca.2021.104081>
- Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal* 1. <https://doi.org/10.1214/06-BA117A>
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2013. *Bayesian Data Analysis*, Chapman & Hall/CRC Texts in Statistical Science Vol. 3.
- Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., Thompson, R., 2015. ASReml user guide release 4.1 structural specification. Hemel hempstead: VSN international ltd.
- Gray, S.B., Brady, S.M., 2016. Plant developmental responses to climate change. *Dev Biol* 419, 64–77. <https://doi.org/10.1016/j.ydbio.2016.07.023>
- Hassanzadeh Khankahdani, H., Rastegar, S., Golein, B., Golmohammadi, M., Aboutalebi Jahromi, A., 2019. Effect of rootstock on vegetative growth and mineral elements in scion of different Persian lime (*Citrus latifolia* Tanaka) genotypes. *Sci Hortic* 246, 136–145. <https://doi.org/10.1016/j.scienta.2018.10.066>

- Henderson, C.R., 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics* 31, 423. <https://doi.org/10.2307/2529430>
- Kruschke, J. K., 2015. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.
- Lambert, G.F., Lasserre, A.A., Azzaro-Pantel, C., Miranda-Ackerman, M.A., Vázquez, R.P., Salazar, M. del R.P., 2015. Behavior patterns related to the agricultural practices in the production of Persian lime (*Citrus latifolia* tanaka) in the seasonal orchard. *Comput Electron Agric* 116, 162–172. <https://doi.org/10.1016/j.compag.2015.06.007>
- Leisner, C.P., 2020. Review: Climate change impacts on food security- focus on perennial cropping systems and nutritional value. *Plant Science* 293, 110412. <https://doi.org/10.1016/j.plantsci.2020.110412>
- Malikouski, R.G., Alves, R.S., Peixoto, M.A., Ferreira, F.M., do Nascimento, E.F., de Moraes, A.L., Zucoloto, M., das Graças Dias, K.O., Bhering, L.L., 2022. Selection index based on random regression model in ‘Tahiti’ acid lime. *Euphytica* 218, 153. <https://doi.org/10.1007/s10681-022-03105-w>
- Malikouski, R.G., Peixoto, M.A., de Moraes, A.L., Elizeu, A.M., Zucoloto, M., Bhering, L.L., 2021. Repeatability coefficient estimates and optimum number of harvests in graft/rootstock combinations for ‘tahiti’ acid lime. *Acta Sci Agron* 43, e51740–e51740.
- Malosetti, M., Ribaut, J.-M., van Eeuwijk, F.A., 2013. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Front Physiol* 4. <https://doi.org/10.3389/fphys.2013.00044>
- Nishio, M., Arakawa, A., 2019. Performance of Hamiltonian Monte Carlo and No-U-Turn Sampler for estimating genetic parameters and breeding values. *Genetics Selection Evolution* 51, 73. <https://doi.org/10.1186/s12711-019-0515-1>
- Nuvunga, J.J., Oliveira, L.A., Pamplona, A.K.A., Silva, C.P., Lima, R.R., Balestre, M., 2015. Factor analysis using mixed models of multi-environment trials with different levels of unbalancing. *Genetics and Molecular Research* 14, 14262–14278. <https://doi.org/10.4238/2015.November.13.10>
- Parsons, A.J., Edwards, G.R., Newton, P.C.D., Chapman, D.F., Caradus, J.R., Rasmussen, S., Rowarth, J.S., 2011. Past lessons and future prospects: plant breeding for yield and persistence in cool-temperate pastures. *Grass and Forage Science* 66, 153–172. <https://doi.org/10.1111/j.1365-2494.2011.00785.x>
- Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554. <https://doi.org/10.1093/biomet/58.3.545>
- Piepho, H.-P., Eckl, T., 2014. Analysis of series of variety trials with perennial crops. *Grass and Forage Science* 69, 431–440. <https://doi.org/10.1111/gfs.12054>
- R Development Core Team, 2020. *R: A language and environment for statistical computing*.

- Rai, M.K., Shekhawat, N.S., 2014. Recent advances in genetic engineering for improvement of fruit crops. *Plant Cell, Tissue and Organ Culture (PCTOC)* 116, 1–15. <https://doi.org/10.1007/s11240-013-0389-9>
- Rocha, J.R. do A.S. de C., Marçal, T. de S., Salvador, F.V., da Silva, A.C., Machado, J.C., Carneiro, P.C.S., 2018. Genetic insights into elephantgrass persistence for bioenergy purpose. *PLoS One* 13, e0203818. <https://doi.org/10.1371/journal.pone.0203818>
- Rocha, J.R.A.S.C., Machado, J.C., Carneiro, P.C.S., 2018. Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. *GCB Bioenergy* 10, 52–60. <https://doi.org/10.1111/gcbb.12443>
- Santos, M.G., Passos, O.S., Soares Filho, W.S., Girardi, E.A., Gesteira, A.S., Ferreira, C.F., 2013. Variability analysis of “Persian” acid lime tree selections using agronomic and molecular markers. *Genetics and Molecular Research* 12, 4604–4614. <https://doi.org/10.4238/2013.October.17.3>
- Silva, É.D.B. da, Xavier, A., Faria, M.V., 2021. Impact of Genomic Prediction Model, Selection Intensity, and Breeding Strategy on the Long-Term Genetic Gain and Genetic Erosion in Soybean Breeding. *Front Genet* 12. <https://doi.org/10.3389/fgene.2021.637133>
- Stan Development, T., 2018. RStan: the R interface to Stan. R package version 2.17. 3.
- Technow, F., Messina, C.D., Totir, L.R., Cooper, M., 2015. Integrating Crop Growth Models with Whole Genome Prediction through Approximate Bayesian Computation. *PLoS One* 10, e0130855. <https://doi.org/10.1371/journal.pone.0130855>
- Van Etten, J., de Sousa, K., Aguilar, A., Barrios, M., Coto, A., Dell’Acqua, M., Fadda, C., Gebrehawaryat, Y., van de Gevel, J., Gupta, A., Kiros, A.Y., Madriz, B., Mathur, P., Mengistu, D.K., Mercado, L., Nurhisen Mohammed, J., Paliwal, A., Pè, M.E., Quirós, C.F., Rosas, J.C., Sharma, N., Singh, S.S., Solanki, I.S., Steinke, J., 2019. Crop variety management for climate adaptation supported by citizen science. *Proceedings of the National Academy of Sciences* 116, 4194–4199. <https://doi.org/10.1073/pnas.1813720116>
- Wasserman, L., 2013. *All of Statistics: A Concise Course in Statistical Inference*. Springer Science & Business Media.
- Wilks, S.S., 1938. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics* 9, 60–62. <https://doi.org/10.1214/aoms/1177732360>
- Zhang, S., Liang, M., Wang, N., Xu, Q., Deng, X., Chai, L., 2018. Reproduction in woody perennial Citrus: an update on nucellar embryony and self-incompatibility. *Plant Reprod* 31, 43–57. <https://doi.org/10.1007/s00497-018-0327-4>

Figuras

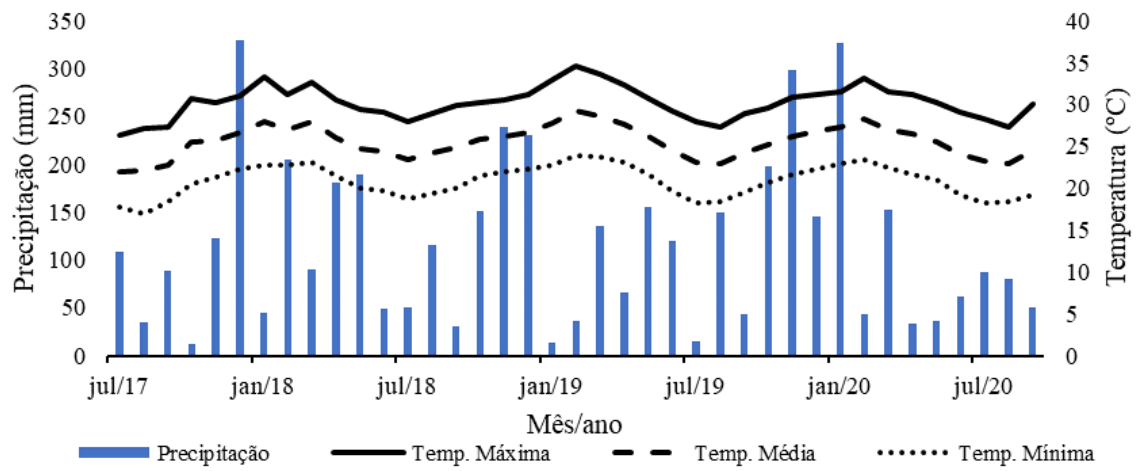


Figura 2. Dados climáticos de São Mateus, ES, Brasil, com informações de temperatura e precipitação de junho de 2017 a setembro de 2020. Fonte: Instituto Nacional de Meteorologia (INMET).

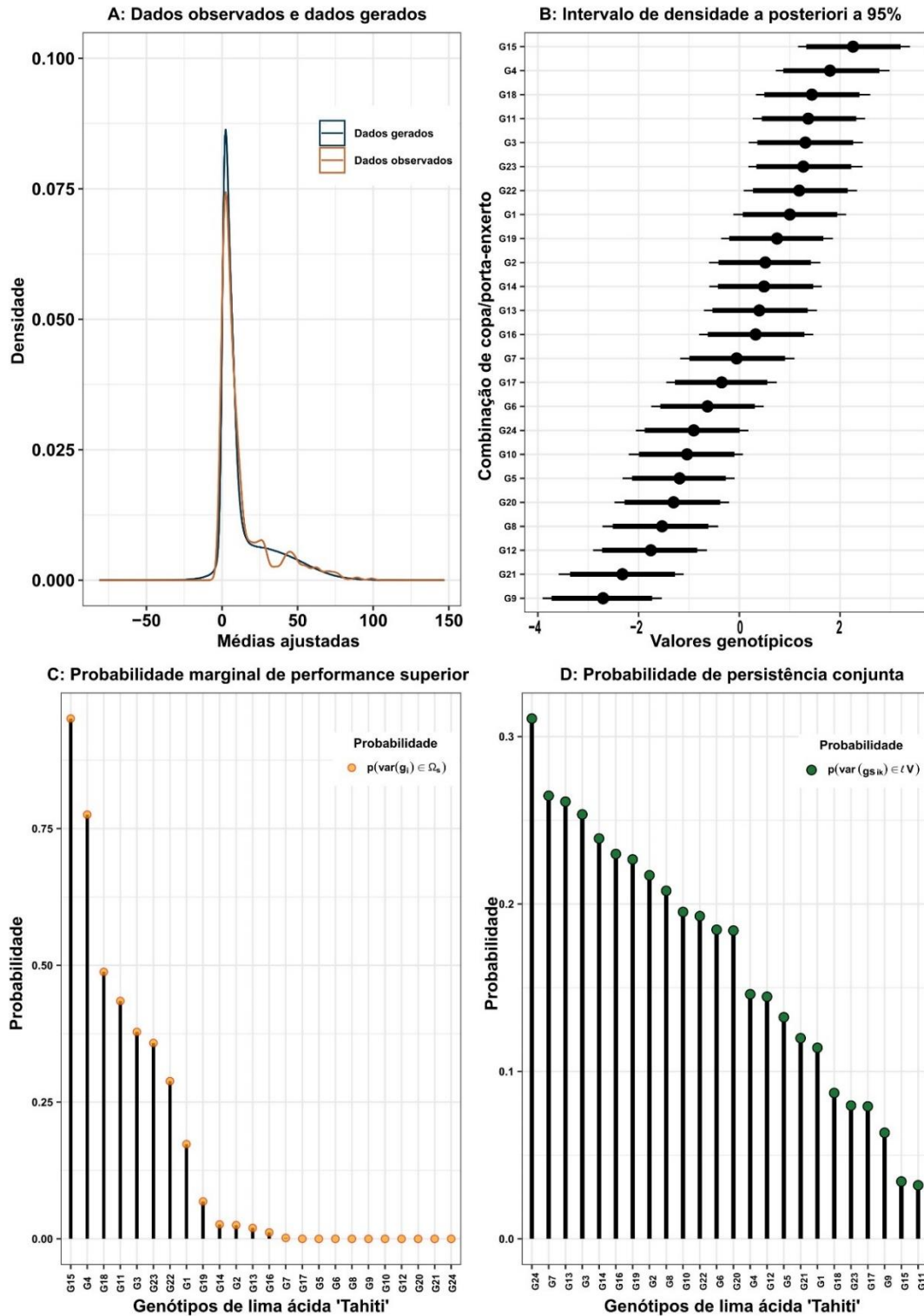


Figura 2. A) Distribuição dos dados observados e gerados pelo modelo bayesiano em lima ácida 'Tahiti'. B) Gráfico caterpillar para os 24 genótipos com seus intervalos de confiabilidade. C) Probabilidade de um genótipo ser superior a qualquer outro usando o efeito a posteriori da amostragem de Monte Carlo. D) Persistência através de probabilidades de os genótipos estarem entre os quatro com menor variância genótipos x colheitas.

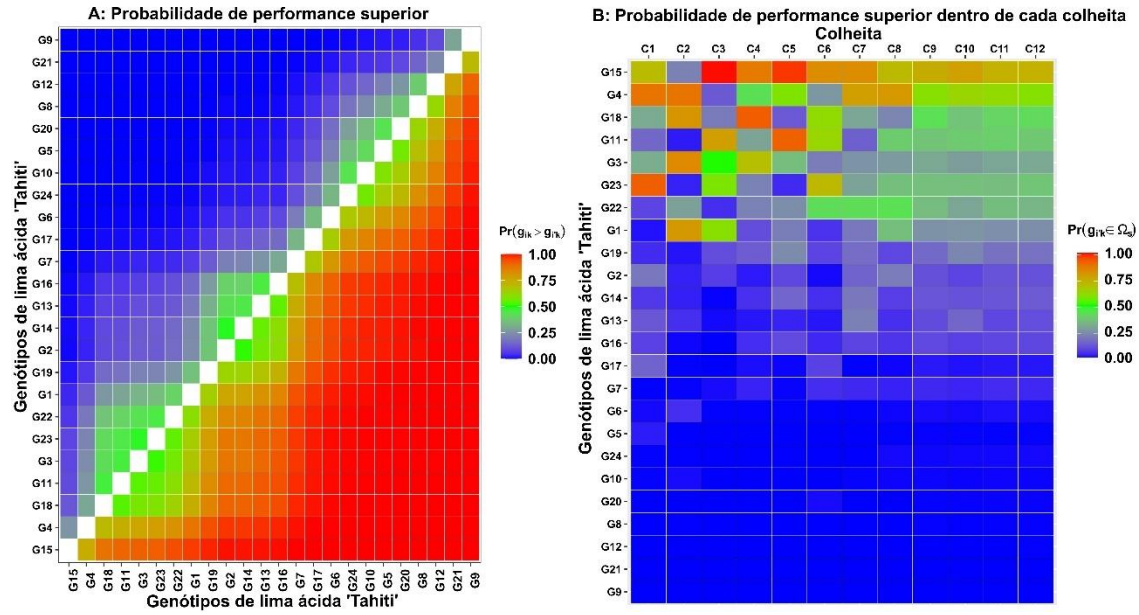


Figura 3. Probabilidades par-a-par de performance superior no contexto bayesiano dos 24 genótipos de lima ácida ‘Tahiti’ (A) e probabilidade de performance superior dos genótipos através das colheitas (B).

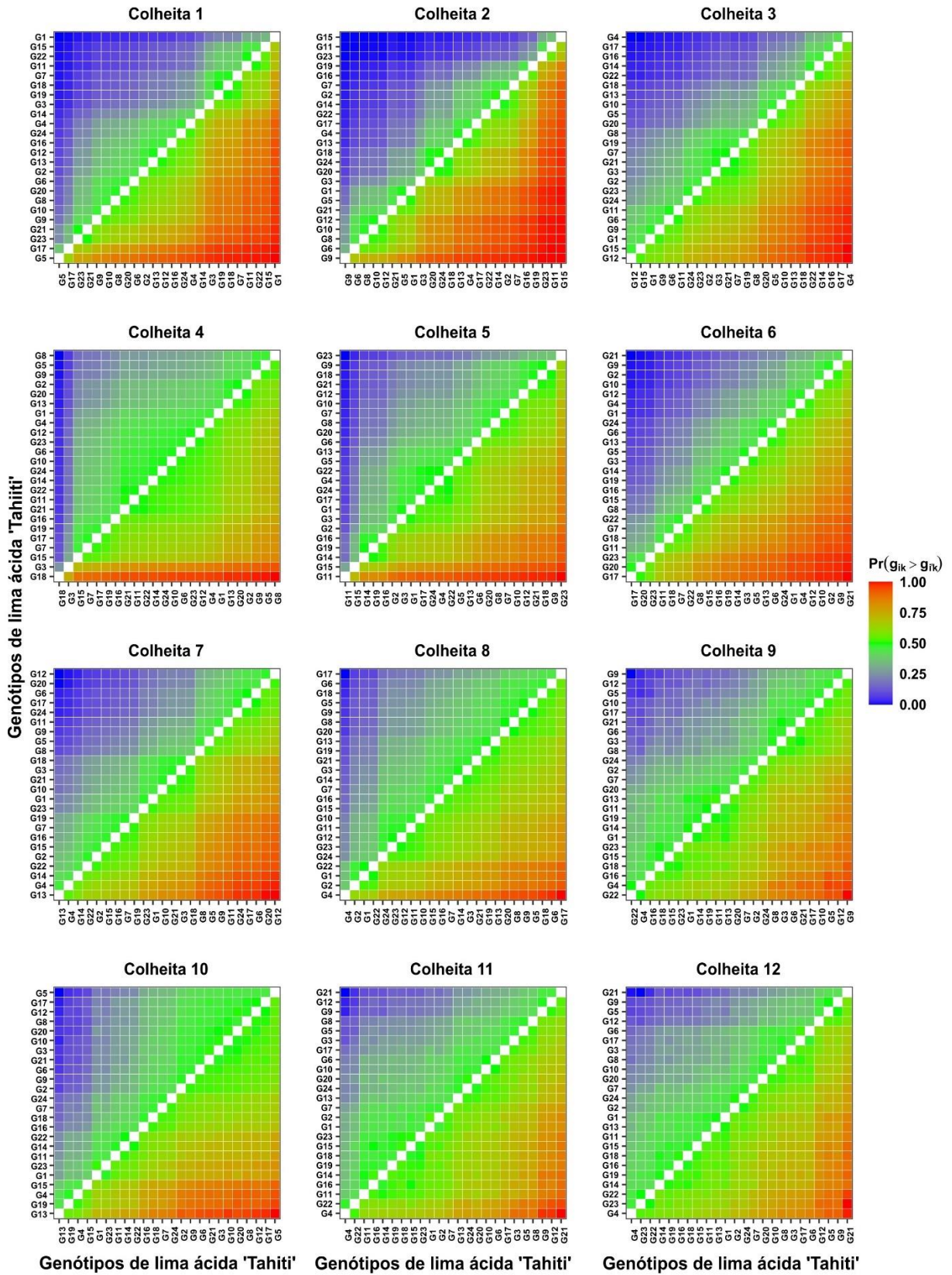


Figura 4. Comparação par-a-par da probabilidade no contexto bayesiano de performance superior dos genótipos em cada uma das doze colheitas.

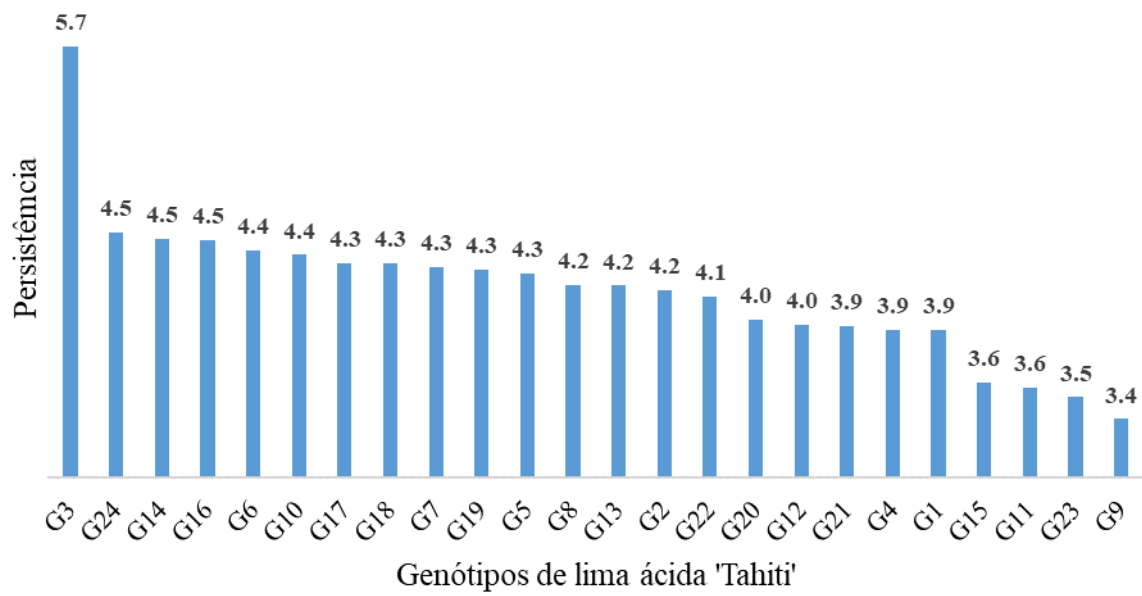


Figura 5. Persistência de 24 genótipos de lima ácida “Tahiti” estimadas utilizando o modelo frequentista.