

MÉTODOS FREQUENTISTAS E BAYESIANO NO MELHORAMENTO
GENÉTICO DE *Psidium guajava* L.

FLAVIA ALVES DA SILVA

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE
DARCY RIBEIRO – UENF

CAMPOS DOS GOYTACAZES – RJ
FEVEREIRO – 2020

MÉTODOS FREQUENTISTAS E BAYESIANO NO MELHORAMENTO
GENÉTICO DE *Psidium guajava* L.

FLAVIA ALVES DA SILVA

Tese apresentada ao Centro de Ciências e
Tecnologias Agropecuárias da Universidade
Estadual do Norte Fluminense Darcy Ribeiro
como parte das exigências para obtenção do
título de Doutora em Genética e Melhoramento
de Plantas

Orientador: Prof. Dr. Alexandre Pio Viana

CAMPOS DOS GOYTACAZES – RJ
FEVEREIRO – 2020

FICHA CATALOGRÁFICA

UENF - Bibliotecas

Elaborada com os dados fornecidos pela autora.

S586

Silva, Flavia Alves da.

MÉTODOS FREQUENTISTAS E BAYESIANO NO MELHORAMENTO GENÉTICO DE *Psidium guajava* L. / Flavia Alves da Silva. - Campos dos Goytacazes, RJ, 2020.

106 f. : il.

Bibliografia: 70 - 88.

Tese (Doutorado em Genética e Melhoramento de Plantas) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Ciências e Tecnologias Agropecuárias, 2020.

Orientador: Alexandre Pio Viana.

1. Fruteiras. 2. Bayesiana. 3. Covariância. 4. Ajuste de modelos. I. Universidade Estadual do Norte Fluminense Darcy Ribeiro. II. Título.

CDD - 631.5233

MÉTODOS FREQUENTISTAS E BAYESIANO NO MELHORAMENTO
GENÉTICO DE *Psidium guajava*

FLAVIA ALVES DA SILVA

“Tese apresentada ao Centro de Ciências e
Tecnologias Agropecuárias da Universidade
Estadual do Norte Fluminense Darcy Ribeiro,
como parte das exigências para obtenção do
título de Doutor em Genética e Melhoramento de
Plantas”

Aprovada em 18 de fevereiro de 2020.

Comissão Examinadora:



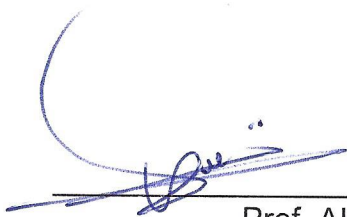
Prof. Paulo Eduardo Teodoro (D.Sc., Genética e Melhoramento) - UFMS



Prof. Antonio Teixeira do Amaral Júnior (D.Sc., Genética e Melhoramento) - UENF



Prof. Leonardo Siqueira Glória (D.Sc., Genética e Melhoramento) - UENF



Prof. Alexandre Pio Viana (D.Sc., Produção Vegetal) - UENF
(Orientador)

DEDICATÓRIA

Dedico a Francisca Alves, Lilian Alves e Caio Guedes, meus maiores
incentivadores.

AGRADECIMENTOS

Primeiramente a Deus, por sua presença constante em minha vida, me proporcionando força e me guiando em todos os meus passos;

À minha mãe Francisca Alves, exemplo de garra e determinação, minha fonte de inspiração, muito obrigada por estar sempre ao meu lado e nunca medir esforços para que eu pudesse realizar meus sonhos;

À minha irmã Lilian, pelo companheirismo, apoio e por torcer tanto pelo meu sucesso;

Ao meu namorado Caio Cezar Guedes Correa, pelo apoio incondicional, carinho, cuidado, atenção e, principalmente, pela paciência;

À Universidade Estadual do Norte Fluminense Darcy Ribeiro, por tornar possível a obtenção do título de Doutora em Genética e Melhoramento de Plantas;

À Capes e Faperj, pelo financiamento do trabalho através da bolsa de estudo e recursos financeiros para implantação e manutenção dos experimentos;

Ao meu orientador, professor Dr. Alexandre Pio Viana, muito obrigada por ter acreditado em mim, por estar sempre disponível, pelo incentivo e aprendizagem;

Aos professores do curso de Genética e Melhoramento de Plantas da Universidade Estadual do Norte Fluminense Darcy Ribeiro, que foram fundamentais para minha formação acadêmica e profissional, especialmente os professores Leonardo e Amaral, pelas sugestões no trabalho, pela orientação;

Aos funcionários, ex-funcionários e bolsistas da Universidade Estadual do Norte Fluminense Darcy Ribeiro, Geraldo, Durval, Luzimara e Valdinei, que tanto colaboraram com conhecimento técnico e apoio na condução dos experimentos;

Ao Pedro Henrique Dias dos Santos e Beatriz Murizini Carvalho, pela amizade sincera, apoio e momentos de descontração;

À Dra. Eileen e “DD”, por todo apoio e motivação. Aos demais amigos, todos foram fundamentais na minha caminhada profissional, em especial, Sandra, Natan, Rodrigo, Daniele, Misael, Bruno, Odimar, Paulo, Fernando, Géssica, Raiane, Valquíria, Mariana, Júnior, Débora, Letícia e Ravena.

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

Muito obrigada!

SUMÁRIO

RESUMO	VII
ABSTRACT	IX
1. INTRODUÇÃO	1
2. OBJETIVOS	4
2.1. GERAL	4
2.2. ESPECÍFICOS.....	4
3. CAPÍTULOS	5
3.1. IMPACTO DA INFERÊNCIA BAYESIANA NA SELEÇÃO DE <i>Psidium guajava</i> L.	5
3.1.1. INTRODUÇÃO.....	5
3.1.2. REVISÃO BIBLIOGRÁFICA.....	7
3.1.3. MATERIAL E MÉTODOS.....	14
3.1.4. RESULTADOS E DISCUSSÃO.....	17
3.1.5. CONCLUSÃO.....	27
3.2. ESTIMATIVAS DE REPETIBILIDADE EM DADOS LONGITUDINAIS DE GOIABEIRA	28
3.2.1. INTRODUÇÃO.....	28
3.2.2. REVISÃO.....	30
3.2.3. MATERIAL E MÉTODOS.....	34
3.2.4. RESULTADOS E DISCUSSÃO.....	38

3.2.5. CONCLUSÃO.....	45
3.3. REGRESSÃO DE CUMEEIRA BAYESIANA APRESENTA MELHOR AJUSTE PARA MARCADORES SSR EM <i>Psidium Guajava</i> L. ENTRE OS MODELOS BAYESIANOS.....	47
3.3.1. INTRODUÇÃO.....	47
3.3.2. REVISÃO BIBLIOGRÁFICA.....	49
3.3.3. MATERIAL E MÉTODOS.....	53
3.3.4. RESULTADOS.....	56
3.3.5. DISCUSSÃO.....	64
3.3.6. CONCLUSÃO.....	69
REFERÊNCIAS BIBLIOGRÁFICAS.....	70
ANEXO.....	89

RESUMO

SILVA, Flavia Alves da; DSc Universidade Estadual do Norte Fluminense Darcy Ribeiro; fevereiro de 2020; MÉTODOS FREQUENTISTAS E BAYESIANO NO MELHORAMENTO GENÉTICO DE *Psidium guajava* L.; Orientador: Prof. Alexandre Pio Viana; Conselheiros: Prof. Leonardo Siqueira Glória e Prof. Antônio Teixeira do Amaral.

A goiaba (*Psidium guajava* L.) tem destaque na fruticultura brasileira pela sua importância econômica, social e alimentar. Pesquisas agronômicas referentes ao melhoramento genético dessa espécie apresentam diversas peculiaridades por ser uma espécie perene, com ciclo reprodutivo longo, oscilação anual de produção, precocidade e redução da taxa de sobrevivência dos indivíduos. Esse desbalanceamento pode também ocorrer porque algumas plantas não produzem em determinada medição/colheita. Diante disso, o objetivo deste trabalho foi a aplicação e a seleção de métodos estatísticos na análise genética de indivíduos em uma população de goiabeira. O primeiro capítulo abordou a comparação entre a metodologia frequentista (REML/BLUP) e a abordagem bayesiana, usando uma *priori* não informativa e uma *priori* informativa, observando o desvio desses métodos em relação à média fenotípica, para a seleção de genótipos superiores. No segundo capítulo, foram avaliadas as estruturas de matrizes de covariância na modelagem da resposta produtiva e estimado o coeficiente de repetibilidade para cada característica no sentido de determinar o número mínimo de avaliações

necessárias para estimativas fidedignas dos valores genéticos individuais que representem a população. No terceiro capítulo, foi avaliada a eficiência dos modelos bayesianos na predição dos valores genéticos genômicos. O experimento foi conduzido em delineamento em blocos casualizados com duas repetições, sendo avaliadas 17 famílias de irmãos completos e 12 plantas por parcela. A metodologia bayesiana sugere estimativas mais precisas dos componentes de variância bem como melhores resultados no ajuste do modelo. A *priori* informativa para o modelo bayesiano é muito importante para a convergência das cadeias, principalmente para pequenos conjuntos de observações. Mesmo com uma *priori* pouco informativa, o modelo bayesiano foi melhor que a abordagem frequentista para este conjunto de observações. Empregando um conjunto de nove safras, a estrutura autorregressiva apresentou o menor valor em ambos os critérios de seleção, em todas as características. As variáveis número de frutos e produção total por planta apresentaram estimativas de repetibilidade superiores a 0,5, considerados valores moderados. A característica massa de fruto apresentou estimativa de repetibilidade de 0,24, mostrando baixa regularidade na repetição do caráter de um ciclo para o outro. Foi possível ainda estabelecer quatro colheitas como número mínimo aceitável de observações necessárias no mesmo indivíduo. Com a aplicação de marcadores moleculares (SSR), foram avaliados diferentes modelos com o uso de diferentes métodos bayesianos para selecionar o melhor modelo para prever valores genéticos genômicos em características produtivas e de qualidade de uma população de goiaba. O modelo de regressão de cumeeira bayesiana apresentou os melhores resultados e foi escolhido para prever os valores genômicos dos indivíduos nas variáveis relacionadas à produção e à herdabilidade, que apresentaram boa acurácia preditiva.

ABSTRACT

SILVA, Flavia Alves da; DSc Universidade Estadual do Norte Fluminense Darcy Ribeiro; february, 2020; FREQUENTIST AND BAYESIAN METHODS IN THE GENETIC BREEDING OF *Psidium guajava* L.; Advisor: Prof. Alexandre Pio Viana; Committe Membrrers: Prof. Leonardo Siqueira Glória and Prof. Antônio Teixeira do Amaral.

Guava (*Psidium guajava* L.) stands out in Brazilian fruit growing due to its economic, social and food importance. Agronomic research related to the genetic improvement of this species has peculiar peculiarities because it is a perennial species, such as the long reproductive cycle, the annual fluctuation of production, precocity, and the reduction in the survival rate of individuals. This imbalance can also occur because some plants do not produce in a certain measurement/harvest. Therefore, the objective of the work was the application and selection of statistical methods in the genetic analysis of individuals in a guava population. The first chapter addressed the comparison between the frequentist methodology (REML/BLUP) and the Bayesian approach using a non-informative priori and an informative priori, observing the deviation of these methods in relation to the phenotypic mean, for the selection of superior genotypes; in the second, the covariance matrix structures were evaluated in the modeling of the productive response and estimated the repeatability coefficient for each trait in order to determine the minimum number of evaluations necessary for reliable

estimates of the individual genetic values that represent the population and in the third chapter the efficiency of Bayesian models in predicting genomic genetic values. The experiment was conducted in a randomized block design with two replications, being evaluated 17 families of complete siblings and 12 plants per plot. The Bayesian methodology suggests more accurate estimates of the components of variance, as well as better results in adjusting the model. An informative priori for the Bayesian model is very important for the convergence of the chains, mainly for small sets of observations. Even with a little informative priori, the Bayesian model was better than the frequentist approach for this set of observations. Using a set of nine vintages, the autoregressive structure showed the lowest value in both selection criteria, in all characteristics. The variables number of fruits and total production per plant showed repeatability estimates greater than 0.5 considered moderate values. The fruit mass characteristic presented a repeatability estimate of 0.24, showing low regularity in the repetition of the character from one cycle to the next. It was also possible to establish four harvests as the minimum acceptable number of observations required in the same individual. With the application of molecular markers (SSR), different models were evaluated using different Bayesian methods in order to select the best model to predict genomic genetic values in productive and quality characteristics of a guava population. Where the Bayesian ridge regression model showed the best results and was chosen to predict the individuals' genomic values in variables related to production and heritability, they presented good predictive accuracy.

1. INTRODUÇÃO

A goiaba (*Psidium guajava* L.) pertence ao gênero *Psidium*, compreendendo cerca de 150 espécies de árvores pequenas e arbustos, com apenas 20 espécies produzindo frutos comestíveis. No entanto, apenas a goiabeira é cultivada a nível comercial (Oliveira et al., 2012). A goiabeira é originária do México, Ilhas Caribenhas e América do Sul, e está amplamente distribuída em diversos países, especialmente em países de clima tropical e subtropical (Mehmood et al., 2014).

Entre as fruteiras, ela tem grande importância econômica para o Brasil, terceiro produtor mundial. Na primeira posição, está a China, seguida pela Índia. Esses três países totalizam aproximadamente 43% da produção mundial (Boti et al., 2016). Em 2005, o Brasil produziu 345.533 toneladas de frutos. Já no ano de 2015, os pomares de goiaba produziram cerca de oitenta mil toneladas a mais (aproximadamente 23 % maior), com o montante de 424.305 toneladas de fruto fresco, sendo os estados de São Paulo, Bahia, Pernambuco e Rio de Janeiro os maiores produtores, totalizando 78% da produção nacional. A cadeia produtiva de frutas no Brasil, onde a goiaba está inserida, abrange 2,2 milhões de hectares, gerando 4 milhões de empregos diretos com uma demanda de mão de obra de duas a cinco pessoas por hectare, fixando o trabalhador no campo, relacionando-se com um PIB agrícola de 11 bilhões de reais (IBGE, 2017).

A crescente demanda do mercado pela fruta de mesa e a valorização do produto como matéria-prima para a indústria têm proporcionado mudanças no

sistema de produção e de comercialização, mudanças relacionadas com a programação da produção, padronização dos frutos, expansão da área destinada aos pomares, aumento da produtividade, entre muitas outras. Com isso, torna-se necessário o uso de variedades que atendam às exigências do mercado, tanto para mesa quanto para a indústria, tendendo os trabalhos de melhoramento a selecionar genótipos que gerem frutos com propriedades particulares, como, por exemplo, maior teor de ácido ascórbico, maior concentração de vitaminas e sais minerais, cultivares mais produtivas. Contudo, os pomares comerciais, em sua grande maioria, apresentam alta variabilidade genética, e uma parte considerável desta variabilidade é formada por genótipos suscetíveis a diversas intempéries (Pereira, 1995).

A alta variabilidade dos genótipos, aliada à escassez de estudos básicos no gênero, resultado da implantação de mudas obtidas de sementes (Santos et al., 1998), tem causado diminuição da produção frente à susceptibilidade às intempéries ambientais e bióticas. E essa alta variabilidade gera sérios problemas nos sistemas de produção, pois dificulta a execução racional de todas as práticas culturais de forma generalizada no pomar, desorganizando o sistema de manejo por parte do produtor (Oliveira et al., 2014). Por outro lado, a ocorrência dessa variabilidade genética possibilita estudos que visem à avaliação e caracterização de genótipos para a quantificação da divergência genética, tanto para a seleção de indivíduos superiores ou mais bem adaptados, bem como para indicação dos melhores cruzamentos a serem realizados, visando a uma maior ocorrência de efeito heterótico na geração híbrida, por conseguinte, maior probabilidade de recuperação de segregantes superiores em gerações avançadas, permitindo maior chance de sucesso nos programas de melhoramento nessa espécie (Amaral Júnior et al., 2010).

Alguns aspectos biológicos da goiabeira, como o ciclo reprodutivo longo, a oscilação anual de produção e a precocidade, geram algumas particularidades em termos de melhoramento, como a necessidade de medidas repetidas. Outra questão é a redução da taxa de sobrevivência dos indivíduos, fato que gera conjuntos de observações desbalanceados. O desbalanceamento pode também ocorrer porque algumas plantas não produzem em determinada medição/colheita. Na cultura da goiaba, também é importante a seleção em nível de planta, não somente com base na média de progênie (Resende, 2002).

Para a obtenção de novas plantas com características desejáveis, a adoção de métodos eficazes de seleção, além da variabilidade genética, é de suma importância, visto o sucesso e a agilidade de tais programas dependerem basicamente desses fatores (Resende et al., 2000). Há inúmeras maneiras de um melhorista utilizar a seu favor procedimentos biométricos para predição de parâmetros que podem fornecer subsídios para tomada de decisão nos programas e, geralmente, esses procedimentos biométricos utilizam metodologias de modelos mistos (Yates e Cochran, 1938; Eberhart e Russel, 1966; Freire Filho et al., 2002; Rocha et al., 2007; Barros et al., 2013; Teodoro et al., 2015). Outro método que tem proporcionado novas perspectivas relacionadas à estimação de componentes de variância e parâmetros genéticos é a metodologia bayesiana. Embora a inferência bayesiana tenha muitas aplicações ainda é um procedimento estatístico relativamente robusto.

A procura por genótipos adaptados às condições ambientais do Sudeste, em especial o estado do Rio de Janeiro, em conjunto com técnicas biométricas mais apuradas, justifica a execução de um programa de melhoramento genético desta espécie que vise à seleção de indivíduos superiores e mais bem adaptados à região do Norte Fluminense, com características de interesse para o mercado consumidor, visto a fruticultura se mostrar favorável para incremento da economia local em razão das condições edafoclimáticas, da proximidade das instalações portuárias e por representar alternativa viável à monocultura canavieira (Gomes Filho et al., 2010).

2. OBJETIVOS

2.1. GERAL

O presente trabalho visa ao estudo da aplicação de métodos frequentistas e bayesianos na análise e seleção de indivíduos superiores em uma população de irmãos-completos de *Psidium guajava* L.

2.2. ESPECÍFICOS

Comparar a metodologia frequentista (REML/BLUP) e a abordagem bayesiana usando uma *priori* não informativa e uma *priori* informativa, observando o desvio desses métodos em relação à média fenotípica para a seleção de genótipos superiores.

Avaliar diferentes estruturas de matrizes de covariância na modelagem da resposta produtiva de uma população de irmãos-completos de goiabeira.

Estimar o coeficiente de repetibilidade para cada característica e determinar o número mínimo de avaliações necessárias para estimativas que representem os genótipos.

Avaliar a capacidade de predição dos valores genéticos obtidos com a utilização dos modelos bayesianos e dos efeitos dos marcadores SSR para características de produção de uma população de irmãos-completos de goiabeira.

3. CAPÍTULOS

3.1. IMPACTO DA INFERÊNCIA BAYESIANA NA SELEÇÃO DE *Psidium guajava* L.

3.1.1. INTRODUÇÃO

Espécies de plantas perenes como a goiabeira (*Psidium guajava* L.) apresentam características específicas como longo ciclo reprodutivo, alta variação anual em algumas características como produção, diferenças de precocidade e longevidade produtiva (Rodríguez et al., 2010). Isso reduz o poder preditivo dos modelos, o que geralmente significa perdas nos recursos investidos. Do ponto de vista do melhoramento genético e do uso em pomares comerciais, isso gera as seguintes consequências: uso dos mesmos genótipos selecionados por vários anos; necessidade de avaliações repetidas em cada indivíduo ao longo do tempo; e redução na taxa de sobrevivência dos experimentos durante sua vida útil. A última consequência tende a gerar um conjunto de observações desequilibradas que exigem precisão nos métodos de seleção (Rezende, 2002). Assim, o uso de um método de modelagem que produz resultados mais precisos pode, sem dúvida, economizar recursos e, a longo prazo, incrementar o sucesso de experimentos com plantas perenes.

O melhoramento genético de plantas perenes normalmente aplica o procedimento da máxima verossimilhança restrita/melhor preditor linear não viesado (REML/BLUP) para a estimativa de componentes de variância e a previsão de valores genéticos (Rezende, 2002). Junto com a utilização de modelos mistos, essa metodologia tem sido referência para conduzir os programas de melhoramento em plantas e animais (Van Eeuwijk et al., 2019). Essa metodologia, também chamada de frequentista, embora apresente várias propriedades úteis, tem suas limitações, como, por exemplo, o estimador REML fornece apenas intervalos de confiança aproximados (Rezende, 2002).

Algumas limitações podem ser contornadas utilizando a inferência bayesiana, usando uma distribuição *a priori* informativa com modelos mistos. Essa abordagem no melhoramento genético se baseia no conhecimento de uma distribuição *a posteriori*. Nesse processo, a função de probabilidade liga a *priori* (informação anterior do experimento ou da distribuição dos parâmetros de interesse) à distribuição *a posteriori*, que, finalmente, contempla o conhecimento anterior e as informações adicionais obtidas no experimento.

Entre as várias metodologias bayesianas, pode ser aplicado o método de simulação de cadeias de Markov e Monte Carlo para gerar uma cadeia de iterações sucessivas, atualizando as estimativas pela probabilidade a partir de um parâmetro inicial (*priori*). Na distribuição conjunta subsequente, as variações podem ser obtidas, possibilitando a construção de intervalos de confiança mais precisos (definidos como intervalos de probabilidade ou intervalos de credibilidade) e, a partir daí, também a estimativa de parâmetros genéticos (Zhao et al., 2016).

A principal vantagem da abordagem bayesiana em relação à frequentista é a possibilidade de utilização de informação prévia sobre os parâmetros do modelo (Beaumont e Rannala, 2004). Na abordagem frequentista, se você tem informações anteriores sobre o experimento, pode até fazer uma análise conjunta com seu experimento atual, que geralmente é dificultado pela diferença entre tratamentos ou entre o mesmo conjunto de observações incompletas. Mas isso geralmente é colocado como fonte de variação no modelo e não adiciona muita informação. Outra vantagem é que os intervalos de credibilidade são menores do que os intervalos de confiança frequentista se uma *priori* informativa for utilizada. Devido à função de probabilidade, se uma *priori* não informativa for usada, o

desempenho do modelo bayesiano é, pelo menos, equivalente ao frequentista (Sorensen e Gionola, 2007; Junqueira et al., 2016; Sandoval et al., 2017).

Este trabalho tem como objetivo comparar a metodologia frequentista REML/BLUP com a abordagem bayesiana, usando uma *priori* não informativa e uma *priori* adequada, observando o desvio desses métodos em relação à média fenotípica, para a seleção de genótipos superiores em uma população de *Psidium guajava* L.

3.1.2. REVISÃO BIBLIOGRÁFICA

3.1.2.1. Aspectos econômicos da goiabeira

A cultura da goiabeira ocupa importante espaço no agronegócio brasileiro, em razão de características apreciáveis do seu fruto como sabor, aspecto, riqueza em nutrientes e elementos funcionais, podendo ser consumida in natura ou processada, nas mais diversas formas de doces, geleias, compotas, sucos, entre outras (São José et al., 2003). Segundo a FAO (2007), a China, a Índia e o Brasil são os principais produtores de goiaba no mundo e os Estados Unidos, o maior importador de goiaba (pasta, purê e geleias), principalmente do Brasil, República Dominicana, Equador e México. A produção nacional de goiaba, em 2017, atingiu 460.515 mil toneladas, numa área plantada de 20,2 mil hectares, com produtividade média de 22,791 t/ha, gerando um valor de produção estimada em 588,6 milhões de reais (IBGE, 2017).

Os pomares comerciais concentram-se nas regiões Sudeste e Nordeste, sendo os estados de São Paulo, Bahia, Pernambuco e Rio de Janeiro os maiores produtores, com produtividade média podendo alcançar até 40 t/ha em áreas mais tecnificadas (IBGE, 2017). O cultivo da goiabeira desempenha importante papel no desenvolvimento socioeconômico no Estado do Rio de Janeiro, com mais de 300 produtores se dedicando à atividade. Em 2016, o Estado foi o 4º maior produtor nacional, sendo a região Norte Fluminense responsável por, aproximadamente, 39% dessa produção. A região Sudeste, na qual o Estado está inserido, mesmo não sendo a região mais produtora, apresenta maior potencial para a produção visto que, embora o Nordeste tenha acumulado o maior

montante produzido, a produtividade média na região do Sudeste é superior, em torno de 4 toneladas a mais por hectare (IBGE, 2017).

A procura pela fruta in natura e produtos, o aumento de áreas plantadas, o clima favorável e a existência de diversas fábricas de processamento de frutas na região do Norte Fluminense sugerem que a obtenção de novas cultivares com maior qualidade e alta produtividade poderá resultar em ganhos econômicos regionais.

Atualmente o mercado para essa espécie se divide em três sistemas de produção de acordo com o objetivo do produto final: o de goiabas in natura, goiabas para a indústria e goiabas de dupla aptidão, tornando-se uma alternativa muito interessante para os produtores (Lima et al., 2002). É necessária, porém, a adoção de pesquisas em busca do melhoramento da cultura, visando ao aumento da produção e à melhoria da qualidade dos frutos, bem como de características que sejam atraentes ao consumidor e à indústria.

3.1.2.2. Melhoramento da goiaba

Hoje em dia, vem sendo dado maior destaque à importância do consumo de frutos para a alimentação humana como forma alternativa para a obtenção de fontes de nutrientes e compostos funcionais em grande quantidade, especialmente alimentos com características atrativas como grandes concentrações de vitaminas, sais, açúcares, entre outros. Sendo a goiabeira uma espécie que se encaixa nesse grupo de frutos de interesse e elevado valor comercial agregado, é fundamental investir em pesquisas voltadas para o melhoramento vegetal para solucionar alguns problemas que tornam ineficiente o seu cultivo (Evangelista & Vieites, 2006).

Quando se trata do melhoramento de fruteiras, um dos aspectos mais importantes é a qualidade do fruto, uma vez que a demanda do mercado vem buscando por produtos de alta qualidade. Com isso, é de suma importância a observação de variáveis como teor de sólidos solúveis, acidez, teor de vitamina C, entre outros, nos ensaios de melhoramento, para a identificação de genótipos superiores que conciliem essas características com um bom desempenho agrônomico. Os programas buscam também obter novas cultivares com facilidades de propagação vegetativa, visto ser o método mais utilizado para formação de pomares, já que o sistema reprodutivo da goiabeira é intermediário

entre alógamas e autógamias, o que gera uma grande variabilidade dentro de um pomar oriundo de sementes (Medina, 1988).

Para a goiabeira, de maneira geral, o ideal seria, por meio de um programa de melhoramento genético vegetal, uma adaptação ampla a diferentes condições ambientais; aumento da produtividade; resistência a pragas como nematoides, doenças como ferrugem e problemas fisiológicos; ausência de sementes; concentração de características de interesse para a industrialização; resistência mecânica ao transporte e ao manuseio; aumento do tempo de prateleira do fruto; e melhorias no paladar e no valor nutritivo. O melhorista irá definir quais características conciliar de acordo com a exigência do mercado alvo, cabendo a ele definir qual a melhor estratégia a ser traçada e a forma de obtenção das observações (Gomes Filho et al., 2010).

O melhoramento genético da goiabeira assim como em outras culturas teve início desde sua descoberta, pela seleção das plantas com frutos mais atraentes, sem nenhum conhecimento científico (Pereira & Nachtigal, 2002). No Brasil, um dos primeiros trabalhos de melhoramento genético da goiaba foi desenvolvido por Soubiê, na escola Superior de Agricultura Luiz de Queiroz, da Universidade de São Paulo, publicado em 1951. Hoje em dia, encontram-se trabalhos relacionados ao melhoramento de goiabeira feitos pelo IAC - Instituto Agrônomo de Campinas, Esalq - Escola Superior de Agricultura "Luiz de Queiroz", Embrapa - Empresa Brasileira de Pesquisa Agropecuária e UFES - Universidade Federal do Espírito Santo.

3.1.2.3. Melhoramento genético da goiabeira na UENF

O programa de melhoramento genético da goiabeira da Universidade Estadual do Norte Fluminense Darci Ribeiro teve início em 2008, em virtude do potencial para agricultura da região e da baixa disponibilidade de cultivares que atendiam às necessidades dos produtores. Pessanha et al. (2011) iniciaram um pré-melhoramento, tendo estudado 20 acessos de *Psidium guajava* L. em pomar de polinização aberta, nos municípios de São João da Barra e Bom Jesus do Itabapoana – RJ. Com esses acessos, foram desenvolvidos trabalhos de citogenética e diversidade via marcadores RAPD no sentido de selecionar genótipos superiores para posterior cruzamento. As sementes dos frutos desses

cruzamentos, oriundos de genitores com considerável grau de heterozigose, originaram uma população de ampla variabilidade genética.

Tabela 1. Características dos acessos selecionados de goiabeiras para cruzamentos feitos no Sítio Providência, Bom Jesus do Itabapoana/RJ.

Acessos selecionados	Características dos acessos	Localização
1 -UENF 1830	Fruto penca unitária, enxertado (Paluma)	Sítio Providência, Bom Jesus do Itabapoana-RJ
2 -UENF 1831	Precoce, enxertado	Sítio Providência, Bom Jesus do Itabapoana-RJ
3 -UENF 1832	Tardia, Folha Rugosa Estaca e apresenta nível de resistência à ferrugem	Sítio Providência, Bom Jesus do Itabapoana-RJ
4 -UENF 1833	Fruto Grande, piriforme Crescimento horizontal	Sítio Providência, Bom Jesus do Itabapoana-RJ
5 -UENF 1834	Amadurecimento Uniforme (ao mesmo tempo)	Sítio Providência, Bom Jesus do Itabapoana-RJ
6 -UENF 1835	Folha Fina Porte Desuniforme (poda)	Sítio Providência, Bom Jesus do Itabapoana-RJ
7 -UENF 1836	Crescimento horizontal	Sítio Providência, Bom Jesus do Itabapoana-RJ
8-UENF 1837	Cortibel, ES bem tardia	Sítio Providência, Bom Jesus do Itabapoana-RJ
9 -UENF 1839	Goiaba Branca Pé Franco (via semente)	Sítio Providência, Bom Jesus do Itabapoana-RJ

Campos et al. (2013) avaliaram a divergência genética entre genótipos de goiabeira para 26 características físico-químicas e morfoagronômicas em 138 combinações híbridas, obtidas desses cruzamentos biparentais. Os conjuntos de observações obtidos foram analisados pela aplicação das técnicas de Redes Neurais Artificiais e Ward-MLM, tendo sido detectada uma alta variabilidade, feito posteriormente o cruzamento entre os mais promissores (Campos et al., 2016). Esses cruzamentos foram confirmados por meio de marcadores microsatélites, que ratificaram a ampla variabilidade genética. Quintal et al. (2017), por sua vez, trabalharam com as 17 famílias obtidas por Pessanha et al. (2011), que foram avaliadas para características qualitativas e quantitativas, tendo sido identificadas 10 famílias com elevados valores genotípicos.

Posteriormente, Oliveira et al. (2014) avaliaram a diversidade genética dessas 10 famílias mais promissoras em conjunto com 27 acessos de araçás do gênero *Psidium* (resistentes ao fitonematoide *Meloidogyne enterolobii*, que, associado ao fungo *Fusarium solani*, causa a doença declínio da goiabeira) por

meio de marcadores ISSR, tendo sido observada ampla variabilidade genética e ainda identificado um grupo de acessos de araçás mais próximos da goiabeira, o que viabiliza o cruzamento interespecífico.

Paiva et al. (2016) então compararam diferentes métodos de seleção entre e dentro de progênies, utilizando a seleção direta e indireta, o índice de seleção e modelos mistos, para estimar os parâmetros genéticos, tendo selecionado os melhores indivíduos para dar continuidade ao programa.

Dando origem a uma nova vertente do programa de melhoramento de goiabeira, visando a obter híbridos de *Psidium* resistentes ao *M. enterolobii*, Almeida et al. (2017) fizeram 696 cruzamentos, todos eles tendo o *P. cattleyanum* como doador de genes de resistência. Em cada genótipo da geração H₁, foi inoculada uma suspensão com 1.000 espécimes de *M. enterolobii*. Foram avaliadas as características morfológicas, estimada a distância entre os genótipos e selecionados os genótipos resistentes.

Santos (2018) fez a caracterização genética dos 94 indivíduos e seus respectivos genitores, buscando selecionar os genótipos geneticamente mais próximos ao *Psidium guajava* L. para avanço de geração, visando à resistência ao declínio da goiabeira. Atualmente esta população está sendo submetida a uma avaliação de resistência à ferrugem da goiabeira (em andamento).

Como a cultura da goiabeira é conhecida por ter o sistema reprodutivo misto, foi idealizada a obtenção de duas novas populações. A primeira população foi obtida de cruzamentos biparentais e uma segunda população foi obtida da autofecundação dos indivíduos superiores. Bezerra (2020) trabalhou com a primeira, avaliou a divergência genética e estimou os parâmetros genéticos de 11 caracteres relacionados à qualidade do fruto, para promover a seleção de 11 progênies de irmãos-completos de goiabeira. A realização da caracterização genética e da diversidade dessa população está sendo feita com o auxílio de marcadores microssatélites (em andamento).

Simultaneamente, Ambrósio (2020) trabalhou com a população originada pela autofecundação e selecionou os indivíduos superiores via seleção truncada. Para avaliar a ocorrência de depressão endogâmica, está sendo desenvolvido um estudo de depressão endogâmica via atributos de sementes, utilizando famílias S₀, S₁ e S₂ (em andamento).

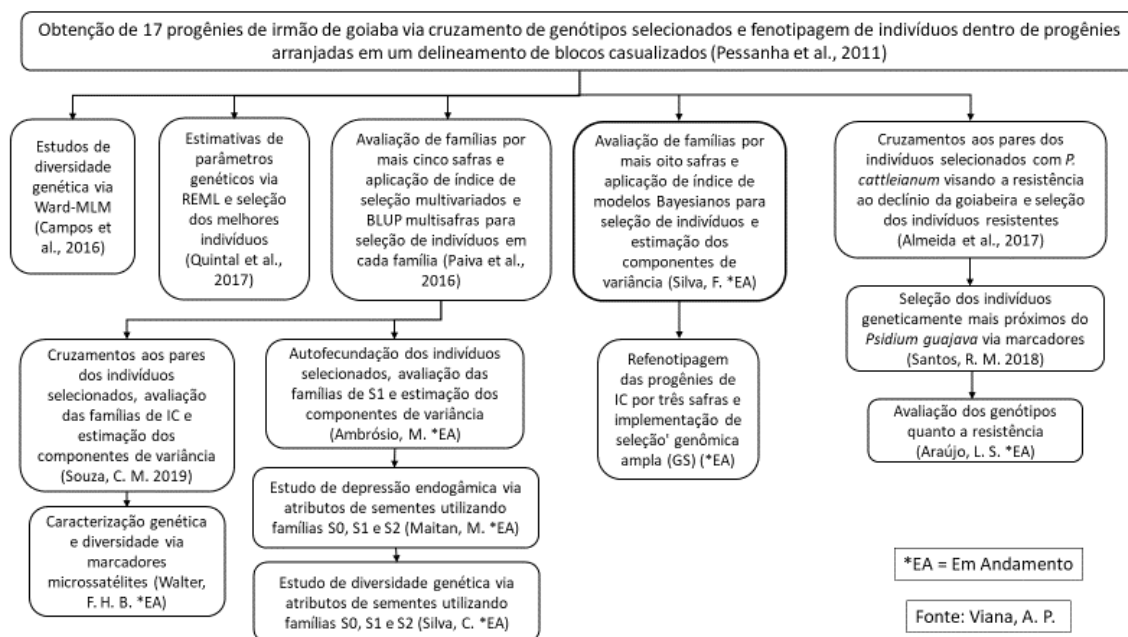


Figura 1. Fluxograma do Programa de melhoramento genético de goiabeira desenvolvido na UENF.

3.1.2.4. Inferência Bayesiana

A inferência bayesiana foi desenvolvida por Thomas Bayes no século XVIII, o qual postulou um teorema conhecido hoje como Teorema de Bayes. Trata-se de um artigo intitulado “An essay towards solving a problem in the doctrine of chances” (Ensaio buscando resolver um problema na doutrina das probabilidades) que o filósofo Richard Price apresentou à Sociedade Real Sociedade dois anos após sua morte. Esse artigo apresenta uma metodologia que por algum tempo ficou esquecida por necessitar de resoluções matemáticas, mas que ressurgiu na década de 60 em alguns trabalhos como o de Jeffreys (1961), que reformulou o teorema, possibilitando sua aplicação nas mais diversas áreas da ciência (Pena, 2006).

Segundo Sorensen e Gianola (2007), os métodos bayesianos vêm sendo aplicados no melhoramento animal desde a década de 80. Estes métodos permitem a análise de grandes conjuntos de observações, não requerem soluções para as equações de modelos mistos e propiciam estimativas diretas e acuradas dos componentes de variância e dos parâmetros genéticos das características de interesse (de Paula, 2006).

O melhoramento de plantas também pode ser conduzido utilizando a abordagem bayesiana. Esta abordagem se baseia no conhecimento da

distribuição *a priori* e possibilita a construção de intervalos de credibilidade mais precisos para as estimativas, podendo gerar uma descrição mais completa sobre a confiabilidade dos parâmetros genéticos quando comparada ao método REML (Gianola e Fernando, 1986; Wang et al., 1994; Wright et al., 2000; Friston et al., 2002; Nogueira et al., 2003; Browne e Draper, 2006).

No teorema de Bayes, supomos que a probabilidade de que observemos simultaneamente um evento θ e um evento y é dada por:

$$P(\theta \cap y) = P(\theta|y).P(y)$$

Também pode ser dada por:

$$P(y \cap \theta) = P(\theta \cap y) = P(y|\theta).P(\theta)$$

Combinando as duas informações, temos:

$$P(\theta|y).P(y) = P(y|\theta).P(\theta)$$

Isolando a probabilidade $P(\theta|y)$:

$$P(\theta|y) = \frac{P(\theta).P(y|\theta)}{P(y)}$$

Em que $P(\theta|y)$ é a probabilidade a posteriori conjunta do evento θ , dado o evento y ; $P(\theta)$ é denominada a probabilidade *a priori* do evento θ ; e $P(y|\theta)$ é a probabilidade condicional do evento y , dado ao evento θ (verossimilhança).

A princípio, o problema da inferência bayesiana consistia em calcular os valores esperados dos parâmetros genéticos seguindo uma função a posteriori. Contudo, segundo Blasco (2001), quando se aplicou o método de Monte-Carlo, baseado em cadeias de Markov (MCMC), para estimar distribuições marginais a posteriori, os procedimentos computacionais foram simplificados e renovados os interesses pelos métodos bayesianos.

Existem vários métodos MCMC, porém, sem dúvida, o mais popular tem sido o algoritmo de amostragem de Gibbs. Como esses algoritmos são processos iterativos, a avaliação da convergência se faz necessária. Entre os métodos mais utilizados para avaliação da convergência das cadeias MCMC, destacam-se os métodos de Heidelberg & Welch (1983), de Geweke (1992) e de Raftery & Lewis (1992).

Entretanto, a importância dessa metodologia se fixa no fato de que é possível conformar regiões de credibilidade (intervalos de confiança no contexto

Frequentista) para as estimativas dos parâmetros (Gianola e Fernando, 1986). Como consequência, a metodologia bayesiana é considerada uma ferramenta importante na avaliação genética, porque leva em consideração a variabilidade existente em todos os parâmetros do modelo (Wright et al., 2000; Nogueira et al., 2003).

Segundo Resende (2002), as seguintes conclusões podem ser relatadas em relação à inferência bayesiana:

1) a análise bayesiana propicia resultados adicionais àqueles obtidos pela abordagem frequentista, destacando-se os intervalos de credibilidade para as estimativas de parâmetros genéticos, valores genéticos e ganhos genéticos;

2) as estimativas dos parâmetros genéticos obtidas do amostrador de Gibbs podem ser mais precisas do que pelo método REML/BLUP;

3) a análise Bayesiana é uma técnica elegante e flexível que permite a simultânea estimação dos parâmetros genéticos, efeitos “sistemáticos” e valores genéticos de maneira precisa, mesmo para pequenas amostras. Entretanto, o autor afirma que, para a obtenção de uma precisão adequada na abordagem bayesiana, é necessário que o pesquisador tenha uma quantidade suficiente de informação experimental.

3.1.3. MATERIAL E MÉTODOS

3.1.3.1. Indivíduos avaliados e condições experimentais

Foram selecionadas para este estudo 17 famílias de irmãos completos, todos pertencentes ao Programa de Melhoramento Genético de Goiabeira da Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), Rio de Janeiro, Brasil. Os genótipos são derivados de cruzamentos entre nove pais contrastantes, escolhidos por estudos de diversidade genética (Pessanha et al., 2011).

O experimento foi conduzido em delineamento de blocos ao acaso, com duas repetições. Cada família foi representada por 24 indivíduos (12 por bloco), com um total inicialmente de 408 indivíduos. O experimento foi conduzido entre 2016 e 2018. O espaçamento foi de 3 por 1,5 m entre linhas e entre plantas,

respectivamente. Todos os tratos culturais foram feitos de acordo com o recomendado para a cultura (Quintal et al., 2017). As colheitas foram feitas em nível individual, tendo a variável produção (kg.plant^{-1}) sido obtida da pesagem de todos os frutos e gerada uma observação por indivíduo. As variáveis massa de frutos (MF em g) e massa de polpa (MP em g) foram constituídas por cinco frutos tomados ao acaso. Alguns genótipos foram perdidos durante o período dos experimentos, resultando em observações desbalanceadas.

3.1.3.2. Modelo estatístico e análises

Primeiro, foi utilizado a metodologia frequentista e, mais tarde, o mesmo modelo com a abordagem bayesiana, usando o modelo misto:

$$y = Xb + Za + Wc + e$$

em que y é o vetor de observação; b é o vetor paramétrico dos efeitos fixos (famílias), associado ao vetor y pela matriz de incidência conhecida X ; a e c são os vetores paramétricos dos efeitos aleatórios (bloco e indivíduo dentro da família, respectivamente), também associados a y pelas matrizes de incidência conhecidas, Z e W , respectivamente; e é o vetor residual, assumindo que a e $c \sim N(0, Gg \text{ e } Ga)$ em que G é a matriz de variância genotípica e aditiva dos efeitos aleatórios e $e \sim N(0, R)$ em que R é a variação residual matriz dos erros aleatórios.

Foi empregado o método de máxima verossimilhança restrita (REML) para obter as melhores estimativas dos componentes de variância, associado a um conjunto de observações não ortogonais e desbalanceados (Thompson, 1969). O método REML/BLUP foi empregado usando o procedimento PROC MIXED no software SAS (SAS, 2010).

A abordagem bayesiana foi utilizada com o mesmo modelo, aplicando o método Monte Carlo, baseado em cadeias de Markov (MCMC), conforme descrito por Hadfield (Hadfield, 2010), empregando o pacote MCMCglmm :: MCMCglmm no software R (Team 2018). Foi determinada um milhão de iterações (nitt), descartando as primeiras cem mil (burn-in) e completando uma amostragem 1:3 (thin), totalizando uma cadeia com trezentos mil iterações, em que foi obtida a variância componentes (distribuição *a posteriori*). A convergência da cadeia de Markov foi testada pelo critério de Geweke, de acordo com as recomendações de

Cowles e Carlin (Cowles e Carlin, 1996), usando o pacote *coda :: geweke.diag* (Plummer et al., 2006) no software *R* (Team, 2018).

As médias *a posteriori*, intervalos de credibilidade e desvio padrão da amostra MCMC foram obtidos de acordo com o modelo linear misto generalizado:

$$Y_{lik} = \mu_i + b_{ik} + g_{li} + e_{lik}$$

Em que Y_{ijkl} é o l -ésimo = [1, 2, ..., 24] valor fenotípico na i -ésima = [1, 2, ..., 17] família, dentro do k -ésimo = [1, 2] bloco; μ_i é a média geral da i -ésima família; b_{ik} é o efeito da i -ésima família dentro do k -ésimo bloco; g_{li} é o efeito do l -ésimo indivíduo dentro da i -ésima família; e e_{lik} é o termo residual.

A distribuição conjunta das observações (função de probabilidade) foi utilizada sob a abordagem bayesiana: $Y_{lik} | \beta, g, G_0, R_0 \sim N(\begin{bmatrix} x'_{li} \end{bmatrix} \beta + \begin{bmatrix} z'_{li} \end{bmatrix} g, \sigma_e^2 I)$ em que β é o vetor de uma probabilidade *a priori* de efeitos sistemáticos (média geral); $g = \{g_{kl}\} \sim N(0, I \otimes G_0)$ é o vetor de uma probabilidade *a priori* de valores genotípicos, em que I é a matriz de identidade e G_0 é a matriz de variância genotípica; $e = \{e_{lik}\} \sim N(0, I \otimes R_0)$ é o vetor de uma probabilidade anterior de valores residuais com valores idênticos de distribuição independente em que R_0 com x'_i prime e z'_i prime são vetores de incidência relacionados à sistematização dos efeitos do genótipo para o valor fenotípico correspondente; e σ_e^2 é a variação residual considerada homogênea. As informações prévias foram baseadas em metanálises ou nas distribuições *a posteriori* dos parâmetros do ciclo anterior (2011-2015). A distribuição de probabilidade informativa *a priori* para os parâmetros fixos de interesse foi obtida através de $\beta_i \sim N(b_0, V_b)$, em que V_b é uma matriz diagonal da variação *a priori* de β . Para cada G_0 e R_0 , foi assumida uma distribuição de Wishart invertida como *priori* para as matrizes de covariância $G_0 \sim W_1^{-1}(\Sigma_g, n)$ e $R_0 \sim W_1^{-1}(\Sigma_e, n)$, em que Σ_g e Σ_e são matrizes de escala.

A densidade conjunta *a posteriori* de todos os parâmetros, dependentes dos efeitos genotípicos da respectiva matriz, mas que assumem independência *a priori*, é dada por:

$$p(\beta, g, G_0, R_0 | y) \propto p(y | \beta, g, G_0, R_0) p(\beta | b_0, V_b) p(g | I \otimes G_0) p(G_0 | \Sigma_g, n) p(R_0 | \Sigma_e, n)$$

Ao assumir uma *priori* não informativa no modelo, foi considerada *a priori* padrão da função (Hadfield, 2010). Essa *priori* não informativa assume para os efeitos fixos uma matriz de variância $V = I \times 1^{10}$, em que I é uma matriz

identidade de média igual a zero ($\mu = 0$). Para os efeitos aleatórios, assumiu-se variância igual a um ($V = 1$) e parâmetro de grau de confiança centrado em torno de zero ($\nu = 0,002$), sendo estas distribuições equivalentes a distribuições gama inversas (Wishart invertida).

Um esquema de validação cruzada foi utilizado para testar as metodologias. Foram feitas dez folds na validação cruzada. Em cada fold, o conjunto de observações foi dividido em dois subconjuntos. O primeiro foi composto por 90% do conjunto de observações coletados aleatoriamente e utilizado para o treinamento do modelo. O segundo (10% ~ 40 indivíduos) correspondeu aos valores fenotípicos previstos pelo modelo obtido no primeiro fold. Em cada *fold* foi organizado um subconjunto diferente, até que todos os indivíduos avaliados tivessem seus fenótipos previstos.

3.1.4. RESULTADOS E DISCUSSÃO

Primeiramente, foram aplicadas as três metodologias em todo o conjunto de observações, simulando um usuário comum, e observada alguma diferença entre os resultados obtidos. Em seguida, foram estimados os desvios das médias das famílias e da média geral para a principal característica de rendimento (Figura 2). Foi possível observar que a metodologia frequentista apresentou um desvio maior, pois em alguns casos o desvio atinge valores extremos, com erros de aproximadamente 2,4 kg. Vale ressaltar que se esse valor for extrapolado para grandes áreas de pomares a diferença poderá chegar a $\sim 6 \text{ t.ha}^{-1}$. Na abordagem bayesiana de caráter informativo *a priori*, percebe-se que os erros em relação à média eram constantemente menores.

Essas estimativas fazem parte do processo nos modelos mistos aplicados para determinar os componentes de variação, para permitir a adição de informações prévias, melhorando o processo de inferência. Essa análise fornece uma descrição mais precisa das estimativas e previsões do que o método REML (Gianola e Fernando, 1986), com métodos muito menos simples (Efron, 1986), embora a inferência bayesiana tenha objetivos muito semelhantes aos de Fisher, em que o elemento subjetivo é removido da escolha da distribuição *a priori*.

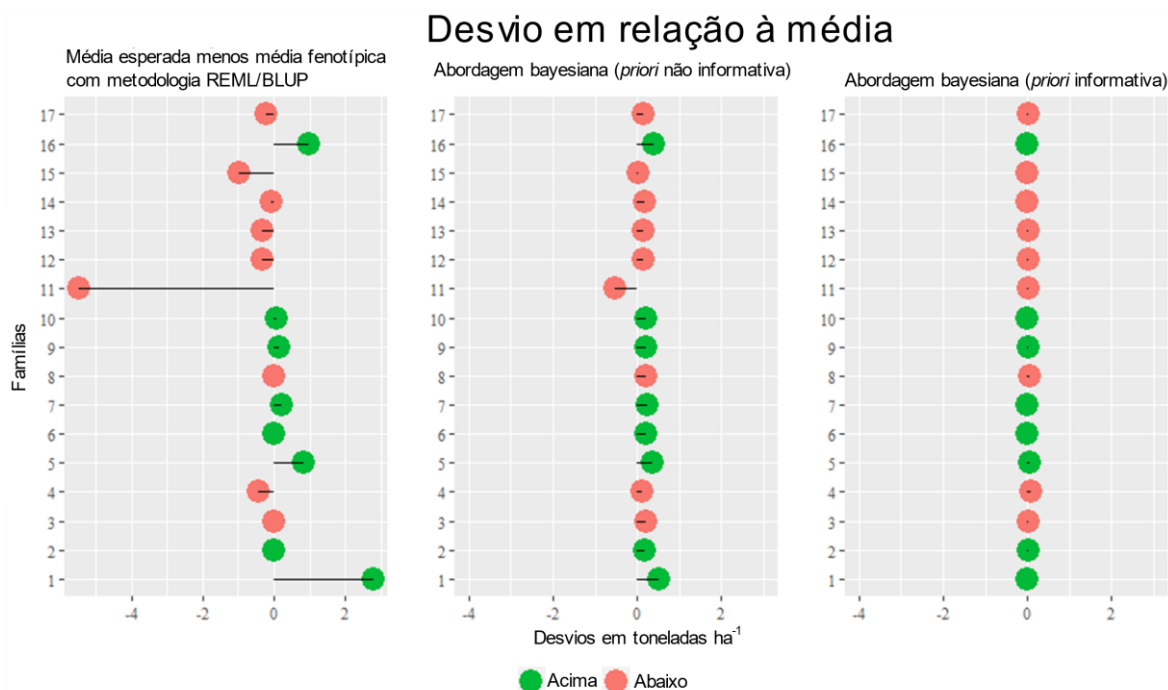


Figura 2. Diferenças entre as estimativas médias obtidas pela metodologia REML/BLUP e inferência bayesiana e os valores médios fenotípicos na variável de produção total (rendimento t.ha⁻¹) em uma população de irmãos completos de goiabeira.

Após observar os desvios, foi utilizada uma validação cruzada para obter medidas de dispersão do ajuste do modelo. Considerou-se como bom ajuste a metodologia que forneceu menor critério de desvio de informação (DIC) e também altos valores para uma probabilidade de ajuste posterior do modelo (W_{prob}) (Tabela 2). Verificou-se o poder preditivo dos modelos pela correlação entre os dados fenotípicos separados para validação e a previsão do modelo obtido pelo treinamento do conjunto de observações em cada fold.

O método bayesiano com prévia apresentou os menores DICs, 4287,9; 17985,8 e 6145,8 para as variáveis massa de fruto, massa de polpa e rendimento, respectivamente, mostrando maiores valores de W_{prob} (probabilidade *a posteriori* do modelo) e correlação. Com os desvios padrão e o delta, é possível notar que entre os folds da validação cruzada houve consistência no ajuste do modelo, com valores menores para inferência bayesiana com *priori* informativa. Assim, sempre que uma porcentagem aleatória dos dados foi usada para testar o modelo, foram obtidos resultados muito próximos, principalmente para a abordagem bayesiana do que para a frequentista.

Tabela 2. Modelos de qualidade de ajuste por validação cruzada (10 folds: 90% de treinamento e 10% de validação), nos mesmos conjuntos de observações amostrais para as metodologias frequentista (REML/BLUP) e bayesiana (com *priori* não informativa e com *priori* não informativa), testados nas variáveis massa de fruto (g), massa de polpa (g) e produção (kg.plant-1) em *P. guajava*.

	Massa de fruto			Massa de polpa			Produção		
	DIC (SD/ Δ)	Wprob	r	DIC (SD/ Δ)	Wprob	r	DIC (SD/ Δ)	Wprob	R
A	14400,8 (94,5/221)	1,60E-25	0,66	18311,4 (1512,3/6049)	1,20E-71	0,31	7195,6 (110,1/322)	1,10E-228	0,70
B	14288,3 (56,1/184)	4,50E-01	0,76	17986,8 (1470,2/4752)	6,20E-01	0,36	6881,1 (238,6/709)	2,10E-160	0,76
C	14287,9 (56,0/183)	5,40E-01	0,81	17985,8 (1470,5/4752)	3,70E-01	0,37	6145,8 (81,4/257)	1,00E+00	082

A = REML / BLUP; B = bayesiano sem *priori*; C = bayesiano com *priori*; DIC = critério de informação de desvio; DP = desvio padrão; Del (delta) = diferença entre o valor mais alto e o mais baixo de DIC; Wprob = probabilidades posteriores modelo; r = correlação entre o Y previsto do modelo (treinamento) e o Y reservado para validação.

Na variável produção, sob um cenário com uma *priori* não informativa, a inferência bayesiana foi pior que a frequentista. Observou-se que uma *priori* ruim prejudicou o modelo, como pode ser observado no DIC, que, embora menor que o frequentista, teve maiores desvios entre os *folds* da validação cruzada, resultado da inconsistência do modelo dependendo das observações. Como os dados de rendimento consistem em uma única observação (produção total), pode-se inferir que a inferência bayesiana contorna bem o problema do pequeno conjunto de observações, desde que seja fornecida uma *priori* adequada (Mostofian e Zuckerman, 2019).

Essa precisão surge porque o método MCMC ainda exibiu grandes variações nas médias das cadeias, portanto, menor significância, já corrigida pela maior consistência da cadeia ao iniciar a partir de uma *priori* informativa (Figura 3). É clara a grande diferença entre as cadeias, usando uma distribuição adequada para *priori* informativa e não informativa. Silva et al. (2019) testaram três distribuições para a *priori*, buscando melhor modelo para variáveis em porcos. Esses autores também mostraram a diferença na precisão que a *priori* adequada fornece. A observação da estabilidade da cadeia também é um critério de controle de qualidade do ajuste do modelo às observações, uma vez que o próprio *burn-in* é uma medida preventiva para descartar o início inconsistente da

cadeia (Carlin e Chib, 1995). Neste trabalho, a importância da *priori* informativa é ainda mais evidente ao observar as cadeias de blocos, plantas e erros (Figuras 2 A e B).

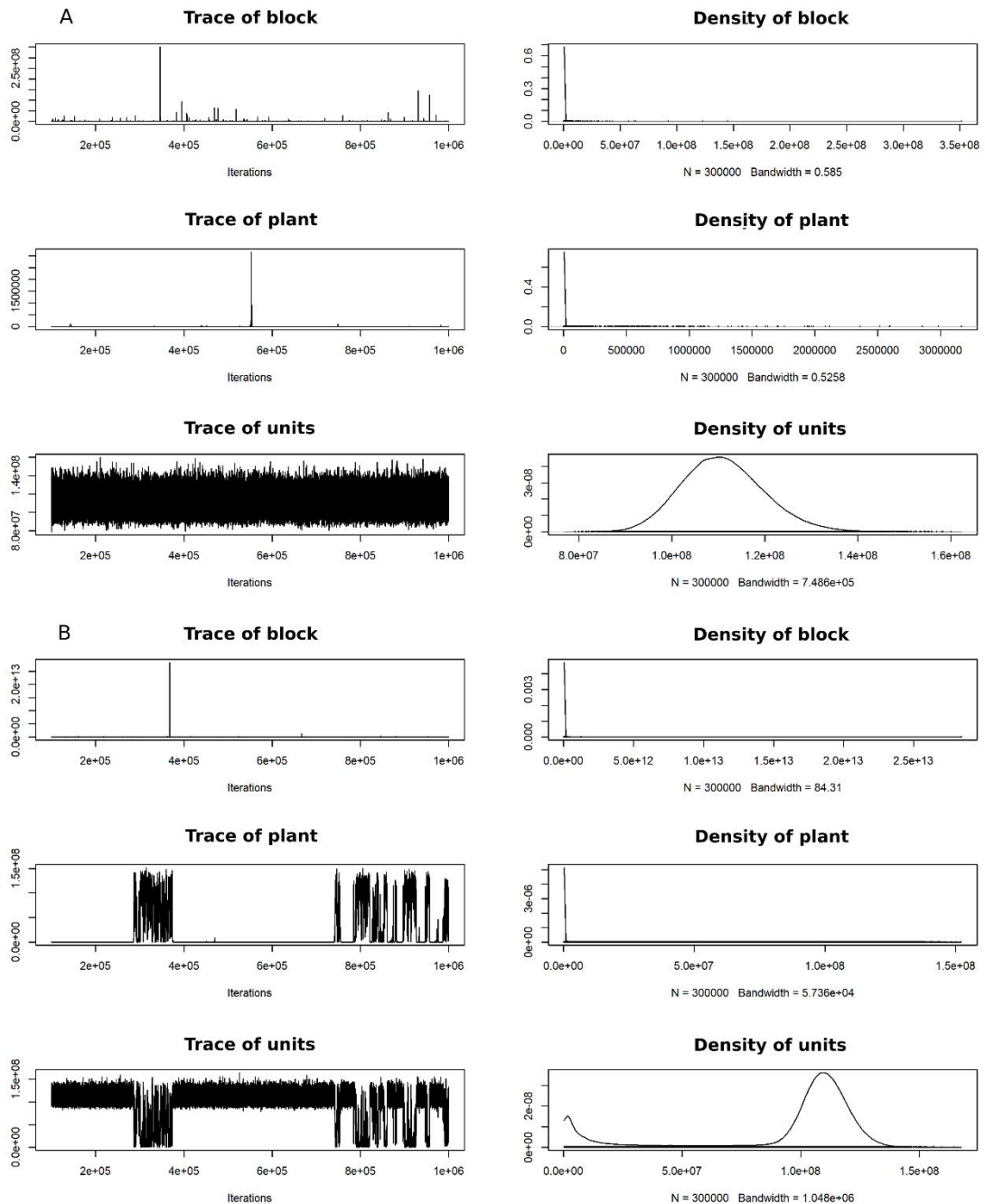


Figura 3. Cadeia da distribuição das estimativas de médias de 300.000 estimativas para a variável PROD nas fontes de variação bloco, plantas e erro do modelo, utilizando uma *priori* informativa (A) utilizar uma *priori* informativa (B) e à direita a função de densidade da distribuição correspondente à cadeia.

Também é importante observar que o critério de iteração de parada no PROC MIXED é quando a diferença entre os parâmetros da distribuição entre uma iteração e outra é menor que $1E-8$ (SAS Institute, 2010). Na abordagem bayesiana, a cadeia de iterações é definida pelo usuário (neste caso, 1 milhão). No início do aquecimento, o método MCMC ainda produz estimativas de médias com variação considerável, que tendem a diminuir com o aumento da cadeia (Hadfield, 2010). Quando o usuário insere *a priori* que representa bem as observações, fornecendo bons parâmetros de distribuição, essa variação entre uma iteração e outra é ainda menor e, juntamente com o tamanho excessivo da cadeia, gera estimativas mais precisas (Resende, 2002). Acredita-se que a *priori* não informativa tenha causado tanta perturbação na cadeia que nem o tamanho excessivo foi capaz de estabilizar os parâmetros e promover boas distribuições *a posteriori*, mas ainda assim apresentaram resultados melhores que a frequentista.

Se for utilizada uma distribuição não informativa para os parâmetros, a inferência bayesiana e o BLUP devem ser equivalentes. Assim, *a priori* altera a distribuição *a posteriori*, de modo que as informações nele contidas não se originam apenas das observações (função de verossimilhança) (Sorensen e Gianola, 2007). Ou seja, ele adiciona mais informações na análise, que não se baseia nas observações. Assim, procedeu-se à seleção dos indivíduos utilizando a abordagem bayesiana com antecedência prévia para obter as médias estimadas e os valores genotípicos preditos. Acredita-se obter valores genotípicos mais precisos porque os métodos bayesianos consideram incertezas nos parâmetros ao longo do processo de inferência. Por outro lado, o BLUP é previsto por estimativas pontuais de componentes de variância, sendo usados valores reais, ignorando a incerteza nos parâmetros de variância (Sorensen, 2009).

A seleção das melhores famílias foi feita para a recombinação e a geração de novas populações. O objetivo é aumentar a média geral da população, tendo sido, para esse fim, selecionadas as nove primeiras famílias, cujas estimativas foram superiores à média geral da população (Figura 4).

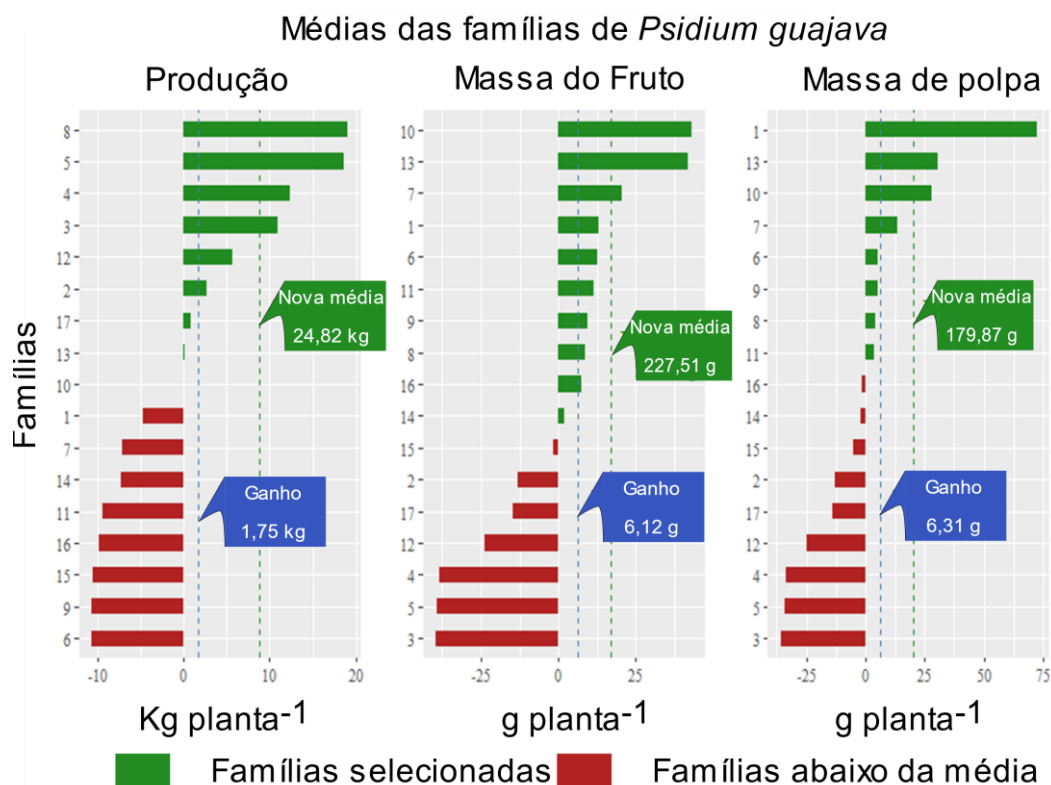


Figura 4. Médias estimadas em uma população de irmãos completos de goiabeiras obtidas por abordagem bayesiana para as variáveis produção total, massa de frutos e massa de polpa.

Os intervalos de credibilidade para esse meio eram geralmente bastante precisos, com um alto grau de confiabilidade. Se observarmos os intervalos de credibilidade para a metodologia bayesiana e os intervalos de confiança para a metodologia REML/BLUP, podemos ver melhores resultados com a inferência bayesiana (Figura 5 e Tabela 3).

Isso ocorre porque o método REML fornece apenas intervalos de confiança aproximados pelo uso de aproximações e suposições de normalidade assintótica. A distribuição e a variância dos estimadores não são conhecidas, assim, as questões relacionadas à efetividade da seleção a ser praticada não podem ser respondidas com rigor. Por outro lado, a análise bayesiana baseia-se no conhecimento da distribuição *posteriori* dos parâmetros e permite a construção de intervalos de confiança exatos (intervalos de probabilidade bayesiana ou intervalos de credibilidade) (Gianola e Rohan, 1986).

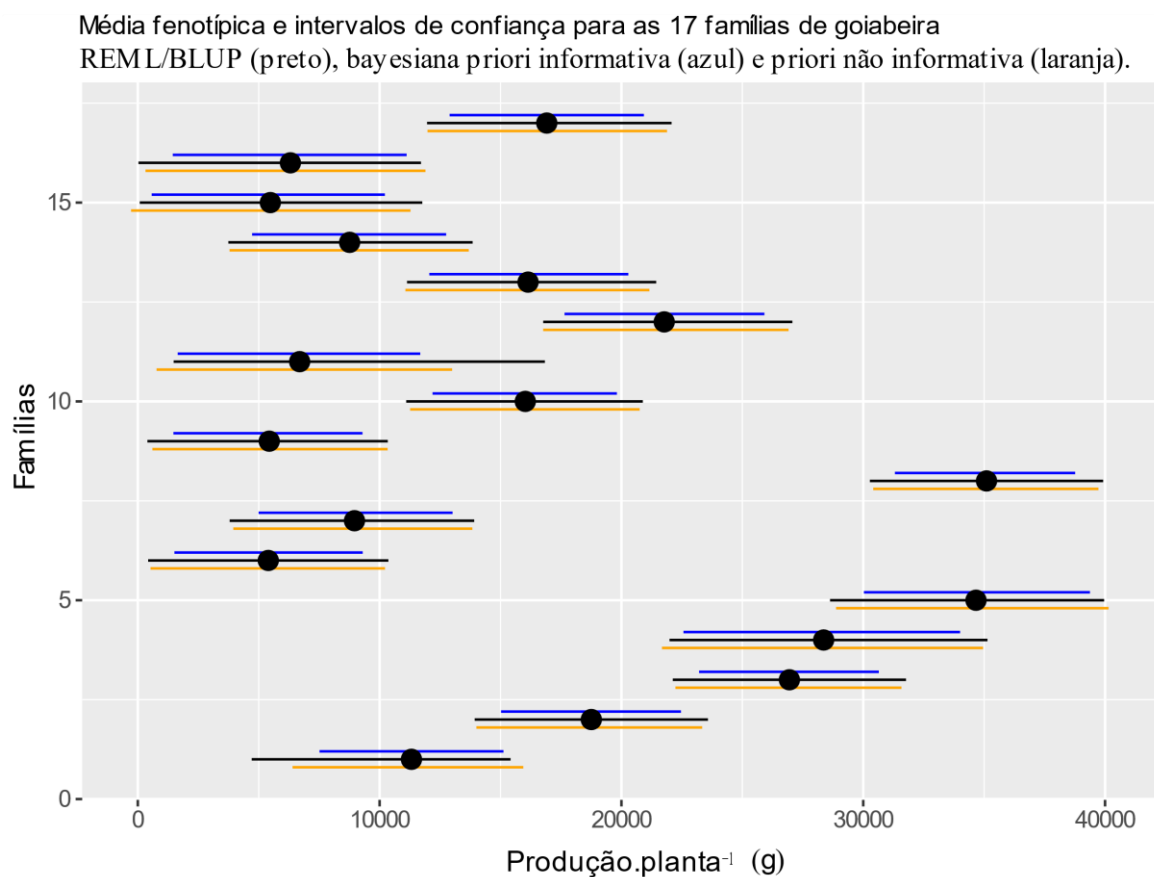


Figura 5. Média fenotípica da característica de produção total para as 17 famílias de goiabeiras e os intervalos de confiança obtidos pela metodologia REML/BLUP e os intervalos de credibilidade obtidos pela abordagem bayesiana com *priori* informativa e com *priori* não informativa.

Tabela 3. Estimativas das médias obtidas pela metodologia frequentista do REML/BLUP e da inferência bayesiana (com *priori* não informativa e *priori* informativa) para as variáveis produção total (kg.plant-1), massa de frutos (g) e massa de polpa (g) em *P. guajava*.

F	PROD		F	MF		F	MP	
	REML/BLUP	ABORDAGEM BAYESIANA		REML/BLUP	ABORDAGEM BAYESIANA		REML/BLUP	ABORDAGEM BAYESIANA
		não informativa informativa			não informativa informativa			não informativa informativa
8	35098,00***	34998,13*** 35078,59***	10	271,80***	250,54** 250,73***	1	230,96***	231,91*** 231,81***
5	34295,00***	34493,25*** 34639,59***	13	248,71***	249,31** 249,24***	10	206,89***	190,29*** 190,21***
4	28563,00***	28303,61** 28329,94***	7	228,47***	227,85** 228,04***	13	190,12***	187,76*** 187,59***
3	26946,00***	26850,92** 26929,71***	1	222,38***	220,30** 220,37***	7	173,08***	173,07*** 172,92***
12	21917,00***	21703,40** 21762,18***	6	220,34***	219,74** 219,86***	6	164,68***	164,83** 164,69***
2	18755,00***	18674,19** 18748,46***	9	217,96***	218,41** 218,86***	9	164,50***	164,68** 164,51***
17	17018,00***	16840,72** 16902,11***	11	217,13***	216,63** 216,83***	11	164,22***	163,99** 163,90***
13	16286,00***	16072,40** 16129,23***	16	216,34***	216,09** 216,16***	8	163,89***	163,69*** 163,36***
10	15992,00***	15924,00** 16023,37***	8	216,19***	214,21** 214,82***	16	158,45***	158,40** 158,20***
1	10064,00***	11079,66** 11312,48***	14	209,23***	209,18** 209,18***	14	157,37***	157,58** 157,45***
11	9158,30**	8848,93* 8964,14***	15	204,26***	206,23** 205,82***	15	154,23***	154,40** 154,50***
7	8855,85***	8682,72* 8756,89***	2	194,24***	194,14** 194,24***	2	146,57***	146,71** 146,57***
14	8795,13***	6924,54* 6685,50*	17	192,38***	192,80** 192,71***	17	145,59***	145,91** 145,75***
15	5922,64*	6139,26* 6314,23*	12	182,88***	183,95** 183,75***	12	134,58***	134,79** 134,65***
16	5869,42*	5475,26 ^{ns} 5480,41*	5	169,18***	169,33** 169,11***	4	125,90***	126,04** 126,12***
6	5395,73**	5339,30* 5428,43*	4	168,07***	167,81** 168,14***	5	125,67***	125,76** 125,51***
9	5368,02**	5305,03* 5394,13*	3	167,89***	167,79** 167,88***	3	123,63***	123,76** 123,65***
\bar{x}	16135,24	15979,73 16051,73		208,67	207,32 207,40		160,61	159,63 159,50

F = famílias (1, ..., 17); ns = não significativo; * = (valor de p <0,05); ** = (valor-p <0,01); *** = (valor-p <0,001) para os intervalos de confiança das médias. A linha de corte no meio da tabela indica os indivíduos que foram selecionados com média acima da média geral para a característica de produção total, considerando a abordagem bayesiana com uma *priori* adequada. Todos os valores na tabela estão em gramas (g).

Outra parte da população foi selecionada para o valor teste de cultivo e uso (VCU) (Tabela 4). Esses indivíduos foram selecionados de acordo com os valores genotípicos previstos e as estimativas de ganho com base na herdabilidade (Tabela 5). As estimativas de herdabilidade mostraram valores dentro da faixa esperada para as características, considerando que elas são controladas por um grande número de genes e altamente influenciadas pelo ambiente (Resende, 2002). A herdabilidade também mostrou alto poder de previsão e menor desvio padrão. Essas medidas são fundamentais para o planejamento do programa de melhoramento, permitindo previsões mais realistas dos próximos passos. Herdabilidade semelhante foi observada em goiabeira (Quintal et al., 2017) ainda mais alta para essas características, mas, como mostrado nos valores de desvio padrão, foram tão altos que se aproximaram das estimativas apresentadas.

Tabela 4. Valores genotípicos e estimativas de ganhos obtidos através da inferência bayesiana para as variáveis rendimento (kg), massa de frutos (g) e massa de polpa (g) em *P. guajava*.

B	F	PL	Produção	EG (kg)	Massa de fruto	EG (g)	Massa de polpa	EG (g)
1	2	8	0,16	44,74	20,48	9,01	0,21	5,4
1	2	11	0,11	2,65	11,98	5,06	0,21	5,27
2	3	9	0,07	14,41	62,24	29,32	0,86	21,38
1	5	7	0,09	18,65	9,4	4,05	0,13	3,38
2	5	11	0,07	14,59	18,78	9,34	0,22	5,31
1	6	3	0,06	28,32	28,01	12,2	0,33	8,16
1	6	5	0,16	26,93	32,01	14,07	0,4	10,03
1	7	8	0,02	12,44	15,97	7,32	0,14	2,9
1	7	11	0,17	30,21	39,81	18,37	0,61	14,37
1	8	1	0,16	17,64	15,04	6,48	0,11	2,77
2	8	2	0,19	46,30	16,66	8,19	0,11	2,82
1	9	1	0,05	14,47	14,86	5,88	0,16	4,11
1	10	6	0,07	3,57	30,34	13,71	0,42	10,01
1	5	2	0,1	22,42	9,55	4,14	0	0,46
2	12	1	0,05	12,63	53,15	24,96	0,5	12,24
2	12	2	0,03	19,90	4,35	2,35	0,15	3,67
2	12	11	0,11	21,88	22,37	10,68	0,39	9,6
1	13	4	0,25	44,37	13,48	5,7	0,08	2,21
2	13	10	0,11	20,88	29,98	14,29	0,37	8,91
1	17	4	0,06	1,55	45,61	20,18	0,23	5,67
2	1	4	0,08	6,85	56,1	24,6	0,12	21,75

Tabela 4 – Cont.

B	F	PL	Produção	EG (kg)	Massa de fruto	EG (g)	Massa de polpa	EG (g)
2	4	6	0	2,66	16,81	6,61	0,3	6,97
1	11	5	0,03	12,23	4,5	3,97	0,05	1,63
2	14	2	0,11	19	47,41	22,17	0,68	16,39
1	16	2	0,08	4,57	20,29	9,68	0,41	10,06
2	2	1	0,24	23,89	3,38	2,06	0,02	0,31
1	3	5	0,23	26,54	14,41	6,2	0,21	4,56
2	3	1	0,15	51,88	8,63	4,47	0,14	3,36
1	5	1	0,13	17,37	3,21	1,19	0,08	2,06
1	8	10	0,22	42,10	12,66	5,38	0,21	5,03

B = bloco; F = família de genótipo; PL = id do genótipo; GE = ganho esperado para a média individual com base na média e herdabilidade de cada família.

Tabela 5. Herdabilidade, previsão de acurácia e desvio padrão para as variáveis massa de fruto (g), massa de polpa (g) e produção (kg.plant⁻¹) em *P. guajava*, obtida via inferência bayesiana.

	Massa de fruto	Massa de polpa	Produção
h²	0,36	0,31	0,20
Desvio padrão	8,20E-03	7,35E-02	9,27E-03
Acurácia preditiva	1,35	0,66	1,83
Média geral	207,40 g	159,50 g	16,05 kg
Média dos selecionados	227,51 g	179,87 g	24,82 kg
Ganho esperado	6,12 g	6,31 g	1,75 kg

h² = herdabilidade em sentido restrito.

Os indivíduos foram selecionados independentemente do objetivo processos industriais, levando em conta a variável produção total ou o consumo in natura, considerando de maior interesse as variáveis massa de frutos e massa de polpa em busca de frutos maiores e mais vigorosos com menos sementes e maior massa de polpa. Como os componentes de variância foram estimados por simulação estocástica (amostragem de Gibbs), acreditamos que os valores genéticos representam melhor o valor real do indivíduo. A ideia por trás desse argumento é a análise exata de amostras de tamanho finito, porque as observações são fixadas na distribuição posterior, em vez de assumir distribuições normais multivariadas. Melhores discussões estatísticas sobre o

BLUP obtidas por inferência bayesiana podem ser encontradas em Henderson (1975), Thompson e Meyer (1986), Gianola et al. (1989) e Resende (2002).

Os programas de melhoramento de plantas perenes têm uma particularidade em comparação com o melhoramento de plantas anuais. Essa diferença é que o período de produção de plantas perenes é muito longo, sendo a quantidade de recursos necessários para melhorar essas espécies muito maior. Assim, para evitar estimativas de componentes de variância com menos precisão, que tornam um programa ainda mais difícil, podemos utilizar a abordagem bayesiana. Outro ponto vantajoso dessa abordagem é a possibilidade de usar informações *a priori* no modelo. Assim, o melhorista pode utilizar melhor as informações disponíveis na literatura, utilizando-as como medidas de distribuição em seu modelo, em vez de apenas comparar seus resultados.

3.1.5. CONCLUSÃO

Em geral, a inferência bayesiana forneceu o melhor ajuste do modelo para esse conjunto de observações, considerando uma população de irmãos completos de *Psidium guajava* L.. Essa abordagem proporcionou um resultado mais completo e confiável, permitindo, assim, uma seleção das melhores famílias para dar continuidade ao programa e dos melhores indivíduos para testar o valor de cultivo e uso. O uso de informações *a priori* é a principal vantagem e, embora seja subjetivo quando a distribuição é informativa, os intervalos de credibilidade são mais estreitos que os intervalos de confiança, sendo essa a principal contribuição para a precisão do modelo, ajudando a contornar problemas de conjuntos de observações pequenos/desequilibrados.

A inferência bayesiana claramente tem vantagens sobre a metodologia frequentista e com os avanços computacionais, essa inferência tende a se tornar popular. Enfatizamos que não dizemos que a abordagem bayesiana seja superior em todos os casos, mas, em razão das vantagens que ela pode oferecer, vale a pena ser testada.

3.2. ESTIMATIVAS DE REPETIBILIDADE EM DADOS LONGITUDINAIS DE GOIABEIRA

3.2.1. INTRODUÇÃO

A presença de ciclos produtivos longos em plantas perenes torna necessária a realização de medidas repetidas ao longo do tempo nos indivíduos. Isso é feito para estimar componentes de variância dos indivíduos da população com maior precisão. Assim, os resultados experimentais representam informações fidedignas do valor real dos indivíduos (Mathew, 2018).

A avaliação genética de indivíduos desta espécie deve então se basear em modelos que consideram um efeito adicional. Esse efeito é denominado efeito de ambiente permanente, quando são feitas medidas repetidas no tempo, e atribuída parte da variância além do erro aleatório (Resende, 2006). O modelo deve considerar também o parâmetro associado à correlação fenotípica entre medidas repetidas em um mesmo indivíduo, que é chamado de repetibilidade (Resende et al., 2017).

O coeficiente de repetibilidade mede a capacidade dos indivíduos em repetir a expressão da característica ao longo de vários ciclos de produção. Do ponto de vista prático, este parâmetro apresenta importância fundamental na predição de valores genotípicos e na inferência sobre a eficiência seletiva de um determinado número de medições. Com ele, é possível identificar quando as

medidas de uma característica se estabilizam, permitindo determinar o número mínimo de medidas que representam uma população. Esse coeficiente de repetibilidade é particularmente útil para acelerar um programa de melhoramento (Cargnin, 2016).

Em culturas perenes, isso significa safras que podem levar um ano ou mais, somado ao tempo até a maturidade da cultura. E a possibilidade de avaliar 5 ou 6 safras ao em vez de 10 ou 12 torna um programa de melhoramento genético mais eficiente, permitindo otimizar diversos recursos como o tempo de avaliação de características de interesse, o tempo até o lançamento de uma nova cultivar e, principalmente, reduzir custos. Permite ainda maximizar o valor da herdabilidade no sentido amplo (Cruz, 2012), o parâmetro de principal interesse do melhorista.

Os modelos lineares mistos podem ser utilizados para descrever dados longitudinais, sendo possível introduzir estruturas de covariâncias associadas aos fatores aleatórios do modelo, assim como também no resíduo (Mohamed, 2019). A escolha das matrizes é feita para representar a forma da dependência existente entre as medidas como a correlação entre diversas colheitas em uma mesma planta.

Entre as estruturas que podem ser utilizadas, o modelo de repetibilidade mais simples assume efeito residual independente e considera correlações ambientais e genéticas constantes entre diferentes registros de medidas repetidas (Henderson, 1984). Embora isso possa não ser uma suposição realista, é frequentemente utilizada. Contudo, ignorar essas correlações quando realmente presentes pode resultar em estimativas tendenciosas dos componentes de variância (Mathew, 2018). Além de dificultar o manejo de um programa de melhoramento, pode gerar significâncias grosseiramente exageradas em testes de médias (Cecon, 2008).

Há diversas estruturas de covariância que tentam aproximar a maneira como se manifesta a dependência das variáveis do modelo. Contudo, deve-se evitar a superparametrização, que pode levar a um sobreajuste do modelo, além de penalizar o critério de AIC, geralmente um valor auxiliar para escolher a qualidade do ajuste. A estrutura de covariância autorregressiva de primeira ordem – AR (1) tem sido indicada para algumas culturas perenes (Barrios, 2016; Quintal et al., 2017; Owusu-Ansah, 2018). A estrutura AR (1) tem apenas mais um

parâmetro que o modelo de repetibilidade simples e permite covariâncias não constantes (Wade, 1993; Wolfinger, 1993).

Porém, nenhuma estrutura ou padrão predeterminado é assumido para medidas longitudinais em plantas perenes, bem como para goiabeiras. Nesse contexto, o objetivo deste trabalho foi analisar seis estruturas de covariância, buscando a que melhor descreve a correlação entre as medidas repetidas no tempo em características de irmãos completos de goiabeira. Por fim, foi estimado o coeficiente de repetibilidade para cada característica e determinado o número mínimo de avaliações necessárias para estimativas que representem os genótipos.

3.2.2. REVISÃO

3.2.2.1. Análises de dados longitudinais

O uso de medidas longitudinais é empregado em diversas áreas, com a finalidade de avaliar uma mesma característica no indivíduo ao longo do tempo. No melhoramento de fruteiras perenes, há necessidade de avaliar informações referentes a diferentes tratamentos, aplicados num mesmo indivíduo, de forma repetida, em momentos distintos, considerando que os indivíduos constituirão o pomar por um longo período.

A principal vantagem é a redução do erro experimental, além de gerar muitas unidades de observação por meio de uma única unidade experimental. O estudo de medidas longitudinais consiste em delineamentos experimentais, completamente aleatorizados, em que os tratamentos são alocados às unidades experimentais e as observações, coletadas mais de uma vez em cada unidade. Dessa forma, é possível estudar pelo menos dois fatores: tratamentos e tempo. O tratamento é o fator entre indivíduos e o tempo, o fator intraindivíduos (Cestari et al., 2012).

As principais desvantagens de estudos com medidas longitudinais são duas. A primeira é que a análise é mais complicada por requerer que as correlações das medidas realizadas no mesmo indivíduo sejam modeladas. A segunda desvantagem está relacionada à presença de observações incompletas

ou omissas, o que, na prática, ocorre com mais frequência do que em outros tipos de observações (Diggle et al., 2002).

Para o emprego de medidas longitudinais em análises de observações, é importante considerar a existência de correlação entre as medidas obtidas numa mesma unidade experimental. As medidas longitudinais são obtidas nos mesmos indivíduos, em tempos diferentes e, possivelmente, sob diferentes condições experimentais. É comum que, para uma mesma unidade experimental, as observações sejam mais fortemente correlacionadas do que as respostas de tempos mais distantes, sendo variável a estrutura da matriz (Malheiros, 2004).

Tal correlação pode ser modelada através de uma estrutura de covariância das observações, sendo que, para outros tipos de observações, é usual assumir que os erros sejam independentes. Modelar a estrutura de covariância adequada da variância residual ao longo do experimento é fundamental para a obtenção de maior acurácia na estimativa dos parâmetros (Toral et al., 2009).

Diversos programas podem ser utilizados para fazer as análises em observações longitudinais, e o procedimento MIXED, do programa SAS®, tem sido bastante utilizado. Esse procedimento possibilita selecionar entre as diferentes estruturas de covariância residuais aquela mais adequada para representar a variação de medidas repetidas dentro de indivíduos. Para isto, utiliza-se um modelo misto considerando a variação entre as unidades experimentais, decorrentes das diferenças entre os níveis dos tratamentos, além da variação dentro das unidades experimentais, decorrentes dos diferentes tempos.

3.2.2.2. Estruturas de matriz de covariância

O modelo de matriz de covariância depende da forma como as observações foram obtidas e da origem genética do material observado. A matriz de covariâncias deve ser suficientemente flexível para incluir no mínimo três fontes diferentes de variação aleatória: a) a variação devida a efeitos aleatórios, quando as observações coletadas advêm de uma amostra aleatória da população em estudo; b) a variação que pode ser explicada por correlação serial, em que se esperam observações próximas mais fortemente correlacionadas que as observações mais distantes; e c) a variação devida ao erro de medição (Diggle et al., 2002).

Atualmente há cerca de 40 tipos de estruturas de covariâncias residuais, o que possibilita ao pesquisador selecionar aquela mais adequada para representar a variação entre medidas longitudinais em um mesmo indivíduo. Das possíveis estruturas de covariância disponíveis na literatura, podemos citar as mais empregadas:

AR – autorregressiva de primeira ordem: com observações de séries temporais igualmente espaçadas e correlações diminuindo exponencialmente, ou seja, a covariância entre duas observações decresce a medida que aumenta o intervalo de tempo entre elas (Freitas et al., 2005);

VC – componentes de variância: com variâncias iguais e observações independentes;

CS – simetria composta: em que as variâncias são constantes entre quaisquer observações de uma mesma unidade em razão de erros independentes;

ARH – autorregressiva de primeira ordem heterogênea: tem variância e covariância desiguais (Littell et al., 1996);

CSH – simetria composta com variância heterogênea: tem parâmetros de variâncias diferentes para cada elemento da diagonal principal e raiz quadrada desses parâmetros nos elementos fora da diagonal principal;

UN – não estruturada: em que todas as variâncias e as covariâncias podem ser desiguais. Especifica uma matriz completamente geral, parametrizada diretamente em termos de variâncias e covariâncias. As variâncias são restritas a valores não negativos e as covariâncias não têm restrições (Singer & Andrade, 2000).

No melhoramento de plantas, essa modelagem dos fatores aleatórios não é comum. Em plantas perenes, a abordagem vem ganhando espaço em relação aos métodos tradicionais em várias culturas. Para selecionar um modelo, há que ter em mente que não existe um modelo verdadeiro. Na verdade, há apenas modelos aproximados da realidade, desse modo, é necessário selecionar o “melhor modelo” entre aqueles que foram ajustados para explicar de maneira mais fidedigna o fenômeno em estudo, sendo de suma importância selecionar um modelo com base em princípios científicos. Para isso, utilizamos uma análise comparativa tendo em vista critérios de informação (Burnham & Anderson, 2004).

3.2.2.3. Estimativa de repetibilidade

Durante o processo de melhoramento visando à seleção de indivíduos para escolha de parentais ou ao lançamento de uma nova cultivar, faz-se necessária a certeza da superioridade genética dos indivíduos (Neves et al., 2010). Como o desempenho das plantas perenes pode variar no passar dos anos, essa característica dificulta a seleção de genótipos promissores. Diante disso, a realização de medidas repetidas nos mesmos indivíduos ao longo do tempo tem se mostrado um procedimento interessante, pois esse parâmetro mede a capacidade de o indivíduo repetir a expressão do caráter no decorrer de sua vida (Cruz et al., 2004).

No melhoramento de plantas perenes, a seleção de indivíduos superiores tem sido feita por avaliações sucessivas em intervalos de tempo previamente estabelecidos. Entretanto, esse processo envolve muitas etapas e avaliação de muitas características, demandando, assim, bastante tempo e recursos (Oliveira et al., 2011).

Uma maneira de reduzir os recursos gastos seria avaliando um número de repetições mínimo necessário. Assim, determinar um número reduzido sem causar erros de estimação de parâmetros e falhas na identificação dos genótipos superiores torna-se de grande valia para o avanço e o sucesso de programas de melhoramentos que necessitam desse procedimento. Diante disso, uma excelente alternativa para contornar esses desafios é a utilização do seu coeficiente de repetibilidade (Cardoso, 2006). Por meio dessa estimativa, é possível determinar o número mínimo de medições que deverão ser feitas para que a avaliação fenotípica apresente alta eficiência (Cruz et al., 2012).

O coeficiente de repetibilidade fornece ainda o valor máximo que a herdabilidade pode atingir, expressa a proporção da variância total de um caráter explicada pelos efeitos genéticos dos indivíduos. Assim sendo, a repetibilidade é fundamental para orientar os trabalhos de melhoramento (Cruz et al., 2012).

3.2.3. MATERIAL E MÉTODOS

As famílias de goiabeira (*Psidium guajava* L.) utilizadas foram obtidas por cruzamentos estabelecidos considerando a diversidade genética (Pessanha, 2011) (Tabela 1). A população faz parte do programa de melhoramento genético da goiabeira da Universidade Estadual do Norte Fluminense Darcy Ribeiro. As colheitas foram iniciadas após o fim do período juvenil das plantas e seguiram os tratos culturais cíclicos, compreendendo o período intermitente vegetativo, a poda de produção, a adubação e a produção (Miranda, 2011).

Tabela 1. Famílias de *Psidium guajava* L. e seus pais (acessos).

♀ acesso	♂ acesso	Família
UENF 1834	UENF 1833	1
UENF 1831	UENF 1830	2
UENF 1831	UENF 1832	3
UENF 1831	UENF 1837	4
UENF 1831	UENF 3839	5
UENF 1833	UENF 1832	6
UENF 1834	UENF 1839	7
UENF 1835	UENF 1834	8
UENF 1834	UENF 1836	9
UENF 1836	UENF 1835	10
UENF 1833	UENF 1836	11
UENF 1831	UENF 1835	12
UENF 1833	UENF 1835	13
UENF 1832	UENF 1833	14
UENF 1834	UENF 1837	15
UENF 1834	UENF 1831	16
UENF 1832	UENF 1835	17

Foi utilizado o delineamento de blocos casualizados, com 2 repetições, 17 famílias segregantes e 12 plantas por família, avaliadas por um período de nove colheitas de avaliação. Foram avaliadas três características a nível de indivíduo: massa de fruto em g (MF) – peso de cinco frutos, com auxílio de uma balança semianalítica; número de frutos total (NF) – contagem de todos os frutos colhidos de um mesmo indivíduo; e produção em g (PROD) – pesagem de todos os frutos colhidos de um mesmo indivíduo, com o auxílio de uma balança digital.

Foi seguido o procedimento para análise de modelo misto proposto por Littell (2000). Primeiramente foram escolhidas as estruturas de covariância, posteriormente, especificados os efeitos fixos, seguido de escolha/estimativa da estrutura de covariância. Na sequência, foram avaliados os efeitos do tratamento e do tempo usando mínimos quadrados generalizados com a covariância estimada e, no final, inferência estatística com base nos resultados.

O modelo foi ajustado para cada estrutura de covariância por vez com o auxílio do REPEATED no procedimento PROC MIXED, usando o software SAS (Littell, 2006). Foi utilizada a máxima verossimilhança restrita como estimador (Patterson, 1971). Foi utilizado o modelo:

$$(1) \quad Y_{ijklm} = \mu + P_i + F_{ij} + B_{ijk} + M_l + e_{ijkl}$$

em que:

Y_{ijkl} denota a medida na l^{th} safra no k^{th} bloco na j^{th} família do i^{th} planta
 $\mu + P_i + F_{ij} + B_{ijk} + M_l$ é a média da planta i dentro da família j do bloco k na safra l , contendo os efeitos de família, bloco e safra respectivamente.

e_{ijkl} é o erro aleatório associado à medida na safra l na i^{th} planta associada a j^{th} família do k^{th} bloco, com $\sim \text{NID}(0, R)$.

A característica distintiva de um modelo de medidas repetidas é a estrutura de variância e covariância do erro e_{ijkl} . Embora as plantas tenham sido atribuídas aleatoriamente às famílias, que foram atribuídas aleatoriamente aos blocos, os níveis do fator de medidas repetidas, nesse caso, o tempo, não são atribuídos aleatoriamente às unidades dentro das plantas. Portanto, os erros aleatórios e_{ijkl} para a mesma planta não são independentes. Em vez disso, assumiu-se que erros para diferentes plantas sejam independentes, dando

$$\text{Cov}[e_{ijkl}, e_{i'j'k'l}] = 0$$

se $i \neq i'$, $j \neq j'$ or $k \neq k'$.

Além disso, como as medições na mesma planta duram um período, elas podem ter variações diferentes, e as correlações entre pares de medições podem depender da duração do intervalo de tempo entre as medições. Portanto, em geral, foi assumido:

$$\text{Var}[e_{ijkl}] = \sigma_i^2$$

e

$$\text{Cov}[e_{ijkl}, e_{ij'kl}] = \sigma_{ii'}$$

Foi permitido que a variação de e_{ijkl} dependa do tempo de medição/colheita l , e a covariância entre os erros em duas colheitas, l e l' , para a mesma planta, dependa das colheitas. O modelo para a covariância pode ser expresso de acordo com alguma estrutura que envolve menos parâmetros em R . Foram avaliadas as seguintes estruturas de (co)variância dos erros:

Simetria Composta (CS), caracterizada pela igualdade de variâncias e covariâncias:

$$\sigma^2 \begin{bmatrix} 1.0 & \rho & \dots & \rho \\ & 1.0 & \dots & \rho \\ & & \ddots & \rho \\ & & & 1.0 \end{bmatrix} = \begin{bmatrix} \sigma_l^2 + \sigma_e^2 & \sigma_l^2 & \dots & \sigma_{lth}^2 \\ & \sigma_l^2 + \sigma_e^2 & \dots & \sigma_{lth}^2 \\ & & \ddots & \sigma_{lth}^2 \\ & & & \sigma_{lth}^2 + \sigma_e^2 \end{bmatrix}$$

Autorregressiva de primeira ordem (AR), caracterizada pela igualdade de variâncias, sendo que a covariância decresce à medida que as distâncias entre as safras aumentam:

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & 1 & \rho & \rho^2 \\ & & 1 & \rho \\ & & & 1 \end{bmatrix}$$

Componente de Variância (CV), com variância homogênea das safras e covariância ausente:

$$\begin{bmatrix} \sigma_l^2 & 0 & 0 & 0 \\ & \sigma_l^2 & 0 & 0 \\ & & \sigma_l^2 & 0 \\ & & & \sigma_l^2 \end{bmatrix}$$

Autorregressiva de primeira ordem heterogênea (ARH), variâncias heterogêneas, sendo a covariância entre duas medidas adjacentes igual à correlação (r) e a covariância entre duas medidas não adjacentes é a correlação elevada ao número de medidas entre as duas -1:

$$\begin{bmatrix} \sigma_1^2 & \sigma_2 \sigma_1 \rho & \sigma_3 \sigma_1 \rho^2 & \sigma_4 \sigma_1 \rho^3 \\ & \sigma_2^2 & \sigma_3 \sigma_2 \rho & \sigma_4 \sigma_2 \rho^2 \\ & & \sigma_3^2 & \sigma_4 \sigma_3 \rho \\ & & & \sigma_4^2 \end{bmatrix}$$

Simetria Composta com variância heterogenia (CSH), caracterizada pela desigualdade das variâncias:

$$\begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ & & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ & & & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

Não Estruturada (UN), usando variâncias diferentes para cada um das i^{th} safras e covariâncias diferentes entre medidas feitas em safras distintas:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ & & \sigma_3^2 & \sigma_{34} \\ & & & \sigma_4^2 \end{bmatrix}$$

Foram obtidas para cada modelo duas medidas de ajuste do modelo. A primeira foi o Critério de Informação de Akaike (AIC) (Akaike, 1974):

$$AIC = -2\log L(\hat{\theta}) + 2d$$

em que d representa o número total de parâmetros de efeito fixo e componentes de variação estimados no modelo. A segunda medida de ajuste do modelo foi o Critério de Informação Bayesiano (BIC) (Schwarz, 1978):

$$BIC = -2\log f(x_n|\theta) + p \log n$$

em que:

$f(x_n|\theta)$ é o modelo escolhido;

p é o número de parâmetros a serem estimados; e

n é o número de observações da amostra.

A função da acurácia para os efeitos fenotípicos permanentes foi obtida por:

$$r_{fp}^2 = \frac{mp}{1 + (m-1)p}$$

em que:

r_{fp}^2 = acurácia fenotípica permanente;

m = número de medições por indivíduos; e

p = coeficiente de repetibilidade.

A eficiência em relação ao uso de apenas uma colheita foi calculada conforme descrito por Viana e Resende (2014):

$$r_{ia}^2 = \left\{ \frac{m}{1 + (m-1)p} \right\}^{1/2}$$

em que:

r_{ia}^2 = eficiência em relação ao uso de apenas uma colheita;

m = número de medições por indivíduos; e

p = coeficiente de repetibilidade.

O coeficiente de determinação, que representa a certeza da predição do valor real do indivíduo para as variáveis considerando o número de medições realizadas, foi calculado pela equação (Cruz, 2012):

$$R^2 = \frac{m \cdot p}{1 + p \cdot (m - 1)}$$

em que:

R^2 = coeficiente de determinação para o número de repetições realizadas;

m = número de medições por indivíduos; e

p = coeficiente de repetibilidade.

A estimativa do número de medições (n_0) necessárias para prever o valor real dos indivíduos com o valor de determinação genotípica (R^2) desejado foi determinada de acordo com a equação (Cruz, 2012):

$$n_0 = \frac{R^2(1 - r)}{\{(1 - R^2)r\}}$$

em que:

R^2 = coeficiente de determinação para o número de repetições realizadas;

e

r = número de medições.

3.2.4. RESULTADOS E DISCUSSÃO

O termo medidas repetidas refere-se a conjuntos de observações com várias medidas de uma variável de resposta na mesma unidade experimental. Na maioria das aplicações, as várias medições são feitas durante um período de tempo. De modo geral, quaisquer dados medidos repetidamente ao longo do tempo ou no espaço são dados de medidas repetidos. O que diferencia a análise de observações de medidas repetidas é a estrutura de covariância dos dados observados. Nos experimentos de blocos aleatórios, os tratamentos são aleatórios para unidades dentro de um bloco. Isso torna todas as observações dentro de um determinado bloco igualmente correlacionadas. Porém, em experimentos de medidas repetidas, duas medições feitas em pontos de tempo

adjacentes são tipicamente mais altamente correlacionadas do que duas medições feitas com vários pontos de tempo.

O ponto chave nesses modelos é uma especificação correta da estrutura de covariância para obter estimativas eficientes na análise de medidas repetidas. Nesse trabalho, foram testadas seis estruturas de covariâncias. A escolha do melhor modelo foi feita com base no critério de informação de Akaike (AIC) e Schwarz's Bayesian (BIC) para variáveis de desempenho agrônômico de irmãos completos de *Psidium guajava* L. (Tabela 2).

Tabela 2. Valores para os critérios de informação de AIC (Critério de Informação de Akaike) e BIC (Schwarz's Bayesian) para o ajuste de modelos, considerando seis diferentes estruturas de covariância.

Variáveis	Critérios	Estruturas					
		AR	CS	CV	ARH	CSH	UN
MF	AIC	47052,8	48623,8	48864,8	*	*	*
	BIC	47064,8	48635,9	48753,9	*	*	*
NF	AIC	32715,9	35501,4	36321,5	*	*	*
	BIC	32727,9	35513,5	36319,4	*	*	*
PT	AIC	64176,7	66832,5	67502,2	*	*	*
	BIC	64188,7	66844,5	67485,9	*	*	*

MF – massa de fruto; NF – número de frutos e PT – produção total por planta. Autorregressiva de primeira ordem (AR), simetria composta (SC), componentes de variância (CV), autorregressiva heterogênea (ARH), simetria composta heterogênea (CSH) e não estruturada (UN). * não convergiram.

Para os modelos que consideraram as estruturas de covariância ARH, CSH e UN não houve convergência do processo iterativo. Os motivos pelos quais isso pode ter ocorrido é que em cada iteração foi calculada a variação residual após a resolução das equações do modelo misto e obtido o -2 Log Res Like. Se a diferença entre -2 Res Log Like de cada iteração for menor que 1E-8, o modelo convergirá. Se a resposta do modelo continuar a variar entre as iterações, ela poderá não convergir (Littell, 2006). Esse resultado indica que essas estruturas podem não ser adequadas para a modelagem dos resíduos do conjunto utilizado

neste trabalho com o objetivo de obter estimativas de repetibilidade com maior precisão.

Outra razão para a não conversão do modelo é que a matriz de covariância pode ser definida como não positiva, ou seja, um prerequisite. Uma alternativa para lidar com problemas de convergência pode ser o uso de modelos não lineares, modelos mistos generalizados e métodos bayesianos. No software SAS, é possível usar uma pontuação de Fisher associada ao método de estimativa até um número de iteração predeterminado (Littell, 2006). Contudo, foi escolhido não forçar a convergência, porque os modelos que não convergiram continuam muitos parâmetros na matriz de covariância e essa hiperparametrização não é desejada.

Os critérios de AIC e BIC permitem a comparação de modelos não aninhados e penalizam modelos mais parametrizados, sendo o BIC o mais rigoroso, ou seja, é um critério que favorece modelos mais parcimoniosos (Wolfinger, 1993). Pode ser que, se os modelos ARH, CSH e UN tivessem convergido, ambos os critérios penalizariam e não seriam escolhidos.

A estrutura de covariância não estruturada é a mais complexa porque está estimando correlações únicas para cada par de pontos no tempo, sendo um modelo hiperparametrizado. Não é incomum a impossibilidade de utilização dessa estrutura, que faz dele um modelo incomum para culturas perenes. Por outro lado, a estrutura de simetria composta também é uma matriz de covariância muito parametrizada, porém ela é usada em alguns trabalhos, onde também são citados AR e CS (Quintal et al., 2017; Maia et al., 2013; Santos et al., 2016).

Quando se negligencia uma possível existência de correlação linear entre as medidas da unidade experimental, possivelmente estará sendo inflacionado o componente de variância residual, já que tudo o que não está no modelo vai para o resíduo (Islam e Chowdhury, 2017). Isso é observado quando são usados os componentes de variância simples, que consideram uma covariância zero entre as medidas, não representando a relação entre as medidas. Isso pode ser observado nos trabalhos de Cargnin (2016), Soares et al. (2017), Woyann et al. (2018) e Bergo et al. (2019).

A estrutura de variância utilizada nos modelos de repetibilidade simples (CV) apresentou os maiores valores de AIC e BIC, sendo a menos verossímil, como já esperado. Isso se deve provavelmente porque essa estrutura de covariância pressupõe a falta de correlação entre as medidas, desta forma, a

suposição de independência não pode ser admitida como regra para sustentar o modelo de análise de variância clássica para esse estudo (Freitas, 2008; Silva et al., 2015).

A estrutura autorregressiva apresentou o menor valor em ambos os critérios de seleção em todas as características. Nessa estrutura, as covariâncias entre observações no mesmo indivíduo diminuem ao longo do tempo, sendo a mais adequada para representar a correlação existente entre as medidas. Quintal et al. (2017), trabalhando com três safras, concluíram que as estruturas mais adequadas para modelar variáveis de produção na cultura da goiabeira foram AR e CS, respectivamente. Resultados semelhantes apontaram que a modelagem espacial dos erros em clones de laranjeira pera pode ser feita utilizando modelo autorregressivo de primeira ordem. Esta estrutura de covariância permitiu uma melhor qualidade de ajuste entre os modelos avaliados (Maia et al., 2012).

Foram estimados os parâmetros da matriz de covariância ($\hat{\rho}$) para a estrutura que apresentou os menores valores de AIC e BIC, a autorregressiva. Os valores estimados foram 0,57; 0,87 e 0,87 para as variáveis massa de fruto, número de frutos e produção total por planta, respectivamente (Tabela 3).

Tabela 3. Estimativas dos parâmetros de (co)variância para as variáveis massa de fruto (MF), número de frutos (NF) e produção total por planta (PT) com uso da estrutura de covariância autorregressiva (AR).

Estimativa	MF	NF	PT
rho (ρ)	0,57	0,87	0,87
Indivíduo	269,93	1867,8	42340327
Residual	621,1	5263,89	11491000

O parâmetro ρ estimado para a estrutura de covariância AR vai aproximando de zero a correlação conforme se passam as safras. Por exemplo, para produção, que é a principal variável de interesse na cultura, quando se plota o reposta dessa variável ao longo do tempo, é possível perceber que cada safra sofre influência das condições climáticas naquele tempo. Assim, conforme avança o tempo, apesar de serem medidas tomadas em uma mesma localização, o clima de uma medida pouco influencia em outra medida tomada muito tempo depois.

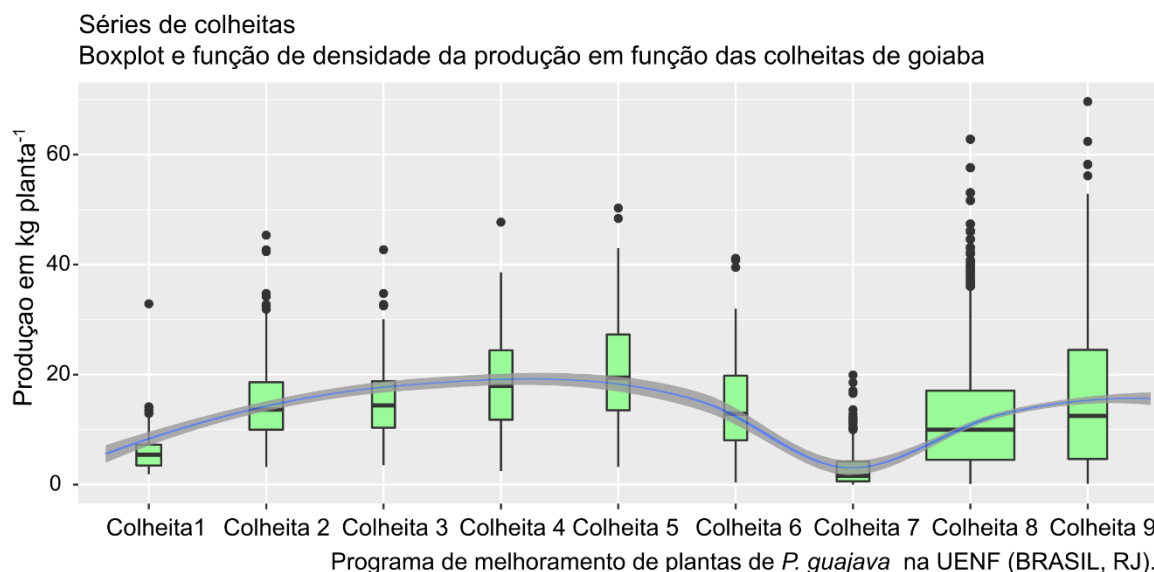


Figura 1. Perfil de produtividade de *Psidium guajava* L. em nove colheitas. A linha azul é a função de densidade de 17 famílias de goiabas ao longo do tempo. Rodear a linha azul é o erro padrão. O boxplot verde é o quantil e a mediana de 17 famílias de goiabas em cada vez.

Uma vez selecionado o melhor modelo, estimou-se o coeficiente de repetibilidade (Tabela 4). Este coeficiente foi estimado para ambas as variáveis massa de frutos e produção total de plantas, que apresentaram estimativas de repetibilidade de 0,24 e 0,54, respectivamente. Essa variação dos coeficientes de repetibilidade pode estar relacionada com a natureza das características, com as propriedades genéticas da população e se os indivíduos avaliados estão ou não estabilizados (Cruz, 2012).

As estimativas do coeficiente de repetibilidade é considerada alta quando as estimativas forem iguais ou superiores a 0,6, mediana quando apresentarem valores entre 0,6 e 0,3 e baixa quando os valores estiverem abaixo de 0,3 (Resende et al., 2006). Variáveis com coeficiente de repetibilidade acima de 0,5 com coeficiente de determinação acima de 80% mostram confiabilidade do valor fenotípico em prever o valor real dos indivíduos (Bergo, 2013).

Tabela 4. Valores de acurácia (A), eficiência (E) e coeficiente de repetibilidade (R^2) para as variáveis massa de fruto (MF), número de frutos (NF) e produção total por planta (PT) em uma população de goiabeira.

Repetibilidade	MF			NF			PT		
	0,24			0,54			0,53		
Safras	R^2	A	E	R^2	A	E	R^2	A	E
1	0,24	0,49	1	0,54	0,74	1	0,53	0,73	1
2	0,39	0,62	1,26	0,70	0,84	1,13	0,69	0,83	1,14
3	0,49	0,70	1,41	0,78	0,88	1,19	0,77	0,88	1,20
4	0,56	0,75	1,51	0,83	0,91	1,22	0,82	0,90	1,23
5	0,61	0,78	1,58	0,85	0,92	1,25	0,85	0,92	1,26
6	0,66	0,81	1,64	0,87	0,93	1,26	0,87	0,93	1,27
7	0,69	0,83	1,68	0,89	0,94	1,27	0,88	0,94	1,28
8	0,72	0,84	1,71	0,90	0,95	1,28	0,90	0,94	1,29
9	0,74	0,86	1,74	0,91	0,95	1,29	0,91	0,95	1,30

As variáveis número de frutos e produção total por planta apresentaram estimativas superiores a 0,5, valores considerados moderados. Lira Junior (2010), testando diferentes métodos para estimar o coeficiente de repetibilidade e trabalhando com as mesmas variáveis em pitangueira, obteve valores semelhantes e concluiu que os coeficientes estimados indicaram que a variância ambiental pouco influenciou nessas variáveis de uma safra para outra.

Porém, é muito importante ressaltar que há um trato cultural que muito influencia essas variáveis, a poda. Na goiabeira, comumente é feita uma poda de produção e posterior desbrota, sendo mantida uma quantidade de ramificações que darão origem aos botões florais. Se na desbrota forem mantidas muitas ramificações, o número de frutos aumenta, mas, conseqüentemente, a massa dos frutos diminui, pois há uma distribuição dos recursos da planta. Assim, o “ambiente desbrota” exercido pelo melhorista influencia nas variáveis, mas deve ser constante ao longo do tempo para que não influencie parâmetros como herdabilidade.

Quanto ao número estimado de medições, podemos estabelecer quatro colheitas como sendo o número mínimo de observações necessárias no mesmo indivíduo para as variáveis número de frutos e produção total por planta. Essas medições podem gerar observações confiáveis que permitem a seleção de indivíduos, com confiabilidade superior a 90% e com custo e mão de obra mínimos. Quintal et al. (2017), trabalhando com três colheitas de goiabeira, tentaram estimar o número de safras necessárias com predições a partir da

terceira. Os autores concluíram que seria necessário maior número de colheitas, cinco, para atingir certo grau de precisão. Contudo, neste trabalho, em que as medições foram de fato feitas, não previstas, foram indicadas quatro colheitas.

A repetibilidade ao nível de média das quatro safras (coeficiente de determinação) corresponde a 0,83 (NF) e 0,82 (PT). Coeficientes de determinação superiores a 0,8 mostram confiabilidade do valor fenotípico em prever o valor real dessa população. Estimativas semelhantes foram apresentadas por Costa (2003) em um estudo de repetibilidade com variáveis de produção de manga.

A acurácia individual foi de 0,90 para as variáveis NF e PT. A acurácia seletiva resulta da estimativa de herdabilidade, da repetibilidade da variável e das metodologias aplicadas na predição dos valores genéticos. Como esta medida está associada à correlação entre valores genéticos preditos e valores genéticos verdadeiros dos indivíduos, quanto maior a acurácia na avaliação de um indivíduo, maior é a confiabilidade na avaliação do indivíduo (Viana, 2014).

A eficiência do uso de cinco safras em relação ao uso de apenas uma é de 1,22 e 1,23 para as variáveis NF e PT, ou seja, com o uso de quatro safras obtivemos, em média, aumento superior a 20% de eficiência em relação a somente uma safra. A partir da quarta colheita, o aumento no número de safras apresentou pequeno ganho na eficiência, sendo inviável este aumento.

A característica massa de fruto apresentou estimativa de repetibilidade de 0,24 mostrando baixa regularidade na repetição do caráter de um ciclo para o outro. Para estabelecer um número mínimo de medições necessárias para prever o valor real dos indivíduos, com base no coeficiente de determinação de 0,85, considerado valor confiável para característica com baixa herdabilidade (Bergo, 2013) para esta variável, foi utilizada a equação:

$$n_0 = \frac{R^2(1-r)}{(1-R^2)r} = \frac{0,85(1-0,24)}{(1-0,85)0,24} = 17,94$$

Assim, para alcançar um coeficiente de determinação de 85%, é necessário fazer aproximadamente 18 medições por indivíduo. Desta forma, para obter ganhos em relação a esta variável, é necessária a aplicação de métodos de melhoramento que tenham um bom controle parental dos indivíduos e fazer uma seleção indireta, por meio do estudo de correlações com características de melhor controle genético (Lédo, 2008).

Esse baixo valor do coeficiente de repetibilidade para a variável MF pode ter ocorrido em função da diferença genética entre os genótipos analisados no experimento, do controle experimental e das variações ambientais ao longo tempo de exposição das plantas ao ambiente (perenidade). Contudo, a variável MF é uma variável relevante no melhoramento das fruteiras, mesmo sendo uma característica de baixa repetibilidade.

Nesse caso, a seleção indireta por meio do estudo de correlações pode ser uma boa estratégia. Por exemplo, massa de polpa, diâmetro do fruto e comprimento do fruto apresentaram correlação de 0,95; 0,9 e 0,78 com a massa de fruto (Silva et al., 2020), podendo ser usadas para selecionar indiretamente essa variável.

Em razão do baixo valor de repetibilidade dessa variável, a realização de 18 colheitas é inviável num programa de melhoramento genético. Resultados semelhantes mostram não ser viável o aumento do número de medições para alcançar níveis de precisão superiores, sendo necessário buscar um método de determinação da característica que seja mais estável, o que foi obtido para a cultura da acerola (Lopes, 2001), açaí (Oliveira, 2001) e mangueira (Costa, 2003).

Assim, a principal característica de interesse na cultura da goiabeira, que é a produção, pode ser avaliada com poucas safras e ainda assim obter observações confiáveis para predições nos indivíduos. Já para outras características como massa do fruto, que é uma característica comumente buscada para cultivares destinadas a frutos de mesa, as goiabeiras podem ser selecionadas com o auxílio de seleção indireta por outras variáveis correlatas.

3.2.5. CONCLUSÃO

A estrutura autorregressiva apresentou os melhores resultados para todas as variáveis, podendo ser indicada para modelar esse tipo de experimento.

Quatro medições podem ser utilizadas para estimar um valor próximo do real dos indivíduos, com acurácia superior a 0,90, para as variáveis NF e PT.

A variável MF apresentou valor do coeficiente de repetibilidade baixo, indicando necessidade de mais observações para obter uma alta precisão, inviabilizando se basear somente nessa variável.

3.3. REGRESSÃO DE CUMEEIRA BAYESIANA APRESENTA MELHOR AJUSTE PARA MARCADORES SSR EM *Psidium Guajava* L. ENTRE MODELOS BAYESIANOS

3.3.1. INTRODUÇÃO

As fruteiras tropicais têm grande valor comercial mundial, pois, além de amplamente consumidas nos países que as produzem, são muito apreciadas e com grande valor agregado ao redor do mundo (FAO, 2016). Uma dessas fruteiras perenes é a goiabeira (*Psidium guajava* L.), que está ganhando espaço no mercado em razão dos métodos de seleção cada vez mais eficientes no melhoramento da espécie.

Um desses métodos é a seleção de indivíduos superiores com o auxílio dos marcadores moleculares, como a seleção genômica. Esse método caracteriza a associação ideal entre o melhoramento convencional baseado em observações fenotípicas e as técnicas modernas moleculares disponíveis atualmente. Sua utilização é de grande impacto dentro dos programas de melhoramento genético (Leon et al., 2016). Isso por que seu objetivo principal é obter estimativas mais acuradas e precisas, o que possibilita um melhor planejamento.

O melhorista tem disponível diversos modelos estatísticos para associar as marcas com os fenótipos, o que torna um desafio escolher um modelo adequado para a resposta da espécie e das marcas. Recentemente, entre esses modelos, as abordagens bayesianas têm ganhado muito destaque com o advento do poder

computacional. Usando uma abordagem bayesiana, os efeitos dos marcadores podem ser estimados em conjunto para prever os valores genômicos para uma característica quantitativa sem fazer uma seleção prévia no painel dos marcadores (Mutshinda e Sillanpaa, 2010). Essa seleção genômica bayesiana tem como principais vantagens a inclusão de uma informação a priori no modelo, além de gerar intervalos de credibilidade mais precisos (Alencar, 2019).

A precisão varia entre os modelos de seleção genômica, de acordo com suas suposições e tratamentos dos efeitos dos marcadores. Por exemplo, foi identificado que os modelos bayesianos (*Bayesian LASSO - BL*) e modelos de regressão de cume (BRR) apresentaram desempenho superior para características controladas por efeitos genéticos aditivos (Desta e Ortiz, 2011).

Entre as abordagens bayesianas disponíveis para a estimativa de valores genômicos no melhoramento, pode-se citar a LASSO Bayesiano – BL, que combina tanto métodos de seleção como de contração de variável, sendo vantajoso em relação ao método frequentistas, que não usa seleção de variáveis. Tem uma *priori* exponencial na variância dos marcadores, resultando em uma distribuição exponencial dupla. A distribuição exponencial dupla tem uma densidade de massa alta em zero e caudas da *priori* mais pesadas em comparação com uma distribuição gaussiana (de Los Campos et al., 2009; Li e Sillanpaa 2012). A regressão de cume Bayesiana -BRR induz encolhimento homogêneo de todos os efeitos do marcador para zero e produz uma distribuição gaussiana dos efeitos do marcador (de Los Campos et al., 2013).

Bayes A utiliza uma qui-quadrada inversa (x^2) na variância dos marcadores, produzindo uma distribuição t escalada para efeitos dos marcadores. Similarmente ao BL e ao contrário do BRR, encolhe os marcadores com pequenos efeitos para valores próximos a zero e os marcadores com maiores efeitos são mantidos. A distribuição final das marcas apresenta um pico mais alto de densidade de massa próximo a zero em comparação com a distribuição dupla exponencial (Meuwissen et al., 2001; de Los Campos et al., 2009). Bayes B é semelhante e também usa um x^2 inversa, mas utiliza métodos de contração e seleção de variável. E quando o parâmetro da priori $\pi = 0$, é semelhante a Bayes A (Meuwissen et al., 2001; Heffner et al., 2011). Bayes C também aplica os métodos de contração e seleção de variável e gera uma distribuição gaussiana nos efeitos dos marcadores. Bayes B e Bayes C consistem em densidade

próxima a zero na distribuição quando usam uma priori baixa (de Los Campos et al., 2009; Habier et al., 2011; de Los Campos et al., 2012).

O melhorista descobrir qual modelo se ajusta melhor ao seu objeto de estudo é de suma importância para o planejamento do programa de melhoramento. Para a goiabeira, ainda não há um estudo buscando qual modelo se aplica melhor para a associação de marcas, embora já tenham sido aplicados também primers para sequências simples repetidas (SSR), como observado em Dinesh et al. (2016).

O objetivo deste estudo foi aplicar e comparar o ajuste de diferentes modelos bayesianos BRR, Bayes A, Bayes B, Bayes B (fixando o valor de muito baixo para $\pi = 10^{-5}$) e BayesC e Lasso Bayesiano (LASSO) para predições dos valores genéticos genômicos em características produtivas e de qualidade de uma população de goiabeira.

3.3.2. REVISÃO BIBLIOGRÁFICA

3.3.2.1. Marcadores moleculares no melhoramento de plantas

Ainda que a eficiência das técnicas clássicas de melhoramento seja comprovada, refletida pelos contínuos ganhos genéticos em produtividade obtidos ao longo de vários anos para várias culturas, novas alternativas podem ser utilizadas para aumentar a eficiência dos programas de melhoramento (Pereira et al., 2008).

Desejando aumentar a eficiência no processo de seleção, uma das técnicas que vem sendo empregada e que supre muitas das limitações existentes no melhoramento convencional é a seleção assistida por marcadores (Miedaner & Korzun, 2012). Técnica esta caracterizada pela utilização de marcadores moleculares no processo de seleção, de modo a auxiliar as avaliações fenotípicas (He et al., 2014).

Podendo ser empregada em qualquer fase de desenvolvimento da cultura, e por não sofrer influência do ambiente, a seleção assistida pode auxiliar na redução de tempo para a obtenção de indivíduos geneticamente superiores. Em

contrapartida, para sua implantação, faz-se necessário um alto investimento (Collard & Mackill, 2008).

Trata-se de uma técnica útil em programas de melhoramento em quatro principais situações: quando a seleção fenotípica é inviável em função do custo e do tempo demandado ou em razão da herança complexa da característica (Rex e Yu, 2007); quando a seleção é dependente de condições ambientais ou estágios específicos de desenvolvimento para a expressão da característica (Rosa et al., 2007); para acelerar a recuperação do genoma recorrente ou manter os alelos recessivos em retrocruzamentos (Fonseca et al., 2009); e para a piramidação de características monogênicas múltiplas ou QTLs de características de herança complexa (Schuster, 2011).

Nas últimas décadas, vários tipos de marcadores como os RFLPs (Restriction Fragment Length Polymorphism), RAPDs (Random Amplified Polymorphic DNA), AFLPs (Amplified Fragment Length Polymorphism), SSRs (Simple Sequence Repeats), entre outros, foram desenvolvidos e empregados na seleção assistida por marcadores, cabendo ao melhorista definir qual técnica será adotada.

Para o emprego de técnicas moleculares, são necessárias infraestrutura apropriada e qualificação técnica para conduzir as diferentes fases da metodologia, que variam de acordo com o tipo de marcador. De modo geral, são necessários equipamentos para a extração de DNA, amplificação via reação em cadeia da polimerase (PCR), separação por eletroforese, fotodocumentação e análise estatística dos marcadores gerados.

3.3.2.2. Marcadores moleculares na goiabeira

O processo de melhoramento genético convencional da goiabeira é bastante longo, tendo por base o rearranjo aleatório de genes entre duas plantas parentais. O emprego de técnicas moleculares tem apresentado benefícios importantes para o estudo das espécies na família, com aplicações em estudos de diversidade genética, filogenia, mapeamento genético, seleção assistida por marcadores moleculares, conservação e taxonomia. Permitindo a redução do tempo para o alcance de resultados efetivos, exclui os efeitos ambientais e fornece informações genéticas (Kalia et al., 2011).

Na cultura da goiabeira, os principais estudos via marcadores são RAPD (Dahiya et al., 2002; Prakash et al., 2002; Rueda et al., 2006; Chen et al., 2007; Gomes Filho et al., 2010; Pessanha et al., 2011), AFLP (Rodriguez et al., 2004; Hernandez-Delgado et al., 2007; Rodriguez et al., 2007; Teyer et al., 2010), SSR (Gupta & Varshney, 2000; Risteruci et al., 2005; Rodriguez et al., 2007; Valdés Infante et al., 2007; Herrero et al., 2010; Kidaha et al., 2014; Nogueira et al., 2015) e ISTR (Medina et al., 2010).

A associação de informação molecular com informações fenotípicas é uma das principais estratégias para reduzir o tempo de obtenção de resultados efetivos. O primeiro passo para implementar essa associação é a construção de um mapa de ligação. Na cultura da goiaba, o primeiro mapa de ligação baseado em marcadores foi feito por Valdés Infante et al. (2003), tendo sido encontradas 167 marcas, formando 11 grupos de ligação, a partir daí, foram desenvolvidos outros trabalhos nesse sentido (Padmakar et al., 2015; Padmakar et al., 2016). Cada vez mais, os programas de melhoramento da goiabeira vêm aplicando ferramentas moleculares com o objetivo de aumentar o ganho genético e poupar tempo e mão de obra.

3.3.2.3. Alfabeto bayesianos

A abordagem bayesiana também pode ser aplicada na predição dos valores genéticos dos indivíduos e efeitos de marcadores, sendo as estimativas obtidas considerando que a variância pode ser diferente entre loci, em que a variância explicada pelo i -ésimo locus é obtida por uma distribuição (Meuwissen et al., 2001).

Essa abordagem envolve métodos estatísticos que atribuem probabilidades aos parâmetros, permitindo incorporar informações em relação à quantidade de interesse desconhecida por meio de $p(\theta)$, sendo θ o parâmetro desconhecido. A atualização da incerteza é feita utilizando o Teorema de Bayes, pela combinação da priori ($p(\theta)$) e verossimilhança ($L(x|\theta)$), levando-nos à distribuição *a posteriori* de θ , dada por:

$$p(\theta|x) \propto L(x|\theta)p(\theta)$$

Assim, utilizando a metodologia bayesiana, é possível obter um resumo da amostra estudada, o qual é dado pela verossimilhança, e podemos incorporar

informações pela distribuição *a priori* (Bolstad, 2007). Nesse contexto, o modelo estatístico é:

$$y = \mu \mathbf{1}_n + Z\alpha + e$$

Em que y é o vetor das observações; μ é a média geral; $\mathbf{1}_n$ é um vetor com valores um, de tamanho n ; Z é a matriz de genótipos (com valores 0, 1 e 2), sendo 0 e 2 para os genótipos homozigotos e 1 para heterozigoto; α é o vetor aleatório dos efeitos dos marcadores; e é o vetor de efeitos residuais.

Na estimação bayesiana, para cada parâmetro devem ser atribuídas distribuições *a priori*. Para β , é atribuída uma distribuição não informativa, seguindo a distribuição normal $N \sim (0, |\sigma_e^2|)$, em que σ_e^2 é a variância do resíduo considerada desconhecida, com uma distribuição *a priori* qui-quadrado escalonada invertida. A diferença entre os modelos bayesianos está na distribuição *a priori* marginal, que é assumida para o vetor de parâmetros α .

No modelo RR-BLUP (“Ridge Regression”), uma distribuição normal $N \sim (0, \sigma^2 \alpha_i)$ é atribuída como *priori* para α , e para a variância $\sigma_{\alpha_i}^2$, é usada uma distribuição qui-quadrada escalonada invertida.

No modelo BayesA, a *priori* para α é uma distribuição univariada t , com média zero. Neste modelo, é assumida uma variância diferente para cada marcador (Meuwissen et al., 2001).

No modelo BayesB, a *priori* assumida para os marcadores é uma mistura de distribuições, tendo um ponto de massa em zero com probabilidade π e uma distribuição univariada t com probabilidade $1 - \pi$ e média zero. Assim BayesA é um caso especial de BayesB, quando $\pi = 0$ (Habier et al., 2011).

No modelo BayesC, a *priori* assumida para os efeitos dos marcadores é uma mistura de distribuições, tendo um ponto de massa em zero com probabilidade π e uma distribuição normal univariada $N \sim (0, \sigma_{\alpha}^2)$ com probabilidade $1 - \pi$, a variância σ_{α}^2 segue uma distribuição qui-quadrada escalonada invertida. RR-BLUP é um caso especial de BayesC quando $\pi = 0$ (Habier et al., 2011). Em BayesB e BayesC, π pode ser tratado como desconhecido e ser inferido com base nas observações, atribuindo uma distribuição beta *a priori* para π .

No LASSO Bayesiano, a distribuição *a priori* atribuída para α é a dupla exponencial, que tem maior concentração de massa em zero, quando comparada com as distribuições normal e t . Este modelo usa o parâmetro λ , que controla o

nível do “shrinkage” aplicado aos efeitos dos m. O valor de λ pode ser definido por validação cruzada ou inferido com base nas observações. Para este último caso, uma distribuição beta ou gama é atribuída como *priori* ao parâmetro λ (Park & Casella, 2008; De los Campos et al., 2009).

3.3.3. MATERIAL E MÉTODOS

3.3.3.1. Material genético

As observações utilizadas neste estudo foram obtidas num experimento conduzido no Programa de Melhoramento Genético da Goiabeira da UENF, na área experimental situada na Escola Técnica e Agrícola Antônio Sarlo, Campos dos Goytacazes, Rio de Janeiro. O experimento de campo foi conduzido em delineamento de blocos completos, com duas repetições e 12 plantas por parcela.

Foram avaliadas 17 famílias segregando goiabeiras: 1) UENF 1834 × UENF 1833; 2) UENF 1831 × UENF 1830; 3) UENF 1831 × UENF 1832; 4) UENF 1831 × UENF 1837; 5) UENF 1831 × UENF 3839; 6) UENF 1833 × UENF 1832; 7) UENF 1834 × UENF 1839; 8) UENF 1835 × UENF 1834; 9) UENF 1834 × UENF 1836; 10) UENF 1836 × UENF 1835; 11) UENF 1833 × UENF 1836; 12) UENF 1831 × UENF 1835; 13) UENF 1833 × UENF 1835; 14) UENF 1832 × UENF 1833; 15) UENF 1834 × UENF 1837; 16) UENF 1834 × UENF 1831; e 17) UENF 1832 × UENF 1835. Os cruzamentos entre os acessos foram estabelecidos tendo em conta informação de diversidade genética obtida por Pessanha et al. (2011).

Para cada indivíduo, foram mensuradas cinco variáveis explicativas: massa do fruto (MF) – mensurada com o auxílio de uma balança semianalítica, expressa em gramas; massa da polpa (MP) – mensurada com o auxílio de uma balança semianalítica, expressa em gramas; teor de sólidos solúveis (°BRIX) – obtido com o auxílio de um refratômetro digital; número de frutos total (NF) – feita a contagem de todos os frutos colhidos em um mesmo indivíduo; e produção total por planta (PROD) – obtida pela pesagem de todos os frutos colhidos em um mesmo indivíduo, com o auxílio de uma balança digital, expressa em gramas. Foram

obtidas cinco observações de todas as variáveis, exceto para NF e PROD, para as quais foi realizada uma observação por indivíduo.

3.3.3.2. Extração e quantificação do DNA

A extração do DNA foi feita das folhas jovens coletadas individualmente de cada planta, utilizando o método CTAB padrão, com modificações (Doyle & Doyle, 1990). Em seguida, o DNA foi quantificado por análise em gel de agarose a 1 % com tampão TAE 1X (Tris, Acetato de Sódio, EDTA, pH 8,0), utilizando o marcador Lambda (λ) de 100 pb ($100 \text{ ng} \cdot \mu\text{L}^{-1}$) (Invitrogen, USA), por meio de comparação das bandas. Para esse procedimento, as amostras foram coradas utilizando a mistura de Gel RedTM e Blue Juice (1:1) e a imagem capturada pelo sistema de fotodocumentação MiniBis Pro (Bio-Imaging Systems). Posteriormente, as amostras de DNA foram diluídas para a concentração de trabalho de $10 \text{ ng} \cdot \mu\text{L}^{-1}$.

3.3.3.3. Triagem dos Primers

Foram testados 200 marcadores SSR (GuavaMap, 2008) em cinco indivíduos de goiabeira visando a identificar os locos polimórficos. Após rastreio, um conjunto de 44 iniciadores polimórficos foram selecionados para as reações de ampliação em uma população de 96 indivíduos na Tabela 1, em anexo.

3.3.3.4. Reação em cadeia da polimerase (PCR)

As reações de PCR foram feitas em termocicladores da Applied Biosystems/Veriti 96 well, em um programa de 38 ciclos, obedecendo às seguintes temperaturas e tempo: 94 °C durante um minuto (desnaturação inicial), 94 °C por dois minutos (desnaturação cíclica), temperatura específica de cada iniciador, em °C, por um minuto (anelamento), 72 °C por três minutos (extensão cíclica), 72 °C por 10 minutos (extensão final) e 4 °C. O volume final foi de 13 μL de cada amostra, sendo 2 μL de DNA ($10 \text{ ng} \cdot \mu\text{L}^{-1}$), 1.50 μL de Tampão 10X (NH_4SO_4), 1.5 μL de MgCl_2 (25 mM), 1.5 μL de dNTPs (2 mM), 1 μL de iniciador (R+F) (5 μM) e 0.12 μL de Taq-DNA polimerase (5 U/ μL) (Invitrogen, Carlsbad, Califórnia, EUA). Os produtos de amplificação foram separados em gel de agarose Metaphor 4%, corado com GelRedTM e visualizado pelo sistema de fotodocumentação MiniBis Pro (Bio-Imaging Systems).

3.3.3.5. Análise estatísticas

Os fenótipos de cada indivíduo foram utilizados como variáveis resposta nas predições genômicas com o uso dos seguintes modelos: Regressão de cunha Bayesiana (BRR – Bayesian Ridge regression), Bayes A, Bayes B, Bayes B (fixando o valor de muito baixo de $\pi = 10^{-5}$), Bayes C e Lasso Bayesiano (Bayesian Lasso - BL, assumindo distribuição marginal a priori exponencial dupla para os efeitos dos marcadores). O modelo geral para as predições genômicas pode ser descrito na forma matricial como:

$$y = \mu + Wg + e$$

em que y é o vetor das observações de cada característica; μ é a média geral; g é o vetor com os efeitos dos marcadores, cujas pressuposições dependem do modelo utilizado; W é a matriz dos genótipos (codificados como 0, 1 e 2) de cada planta para cada marcador; e e é o vetor dos resíduos.

Uma descrição completa sobre o cálculo da herdabilidade e as especificações das distribuições de probabilidade dos efeitos do modelo geral acima, para o uso dos métodos bayesianos, pode ser encontrada em Pérez & de los Campos (2014). Todas as análises bayesianas foram feitas no pacote BGLR do *software* R (R Core Team, 2018), com a função BGLR, ajustada para 1E6 iterações com os primeiros 2E5 ciclos descartados como *burn in* e *thin* assumindo o valor 4.

Os modelos foram comparados com base no Deviance Information Criterion (DIC) proposto por Spiegelhalter et al. (2002). O DIC pode ser descrito como $DIC = D(\hat{\theta}) + 2p_D$, sendo que o primeiro termo é uma medida bayesiana de ajuste do modelo ($D(\hat{\theta})$), que é definido como a média *a posteriori* da *deviance*, e o segundo componente (p_D) mede a complexidade do modelo por meio do número efetivo de parâmetros. Para facilitar a interpretação dos valores DIC em termos da superioridade de um modelo sobre o outro, as probabilidades *a posteriori* dos modelos foram calculadas utilizando a aproximação apresentada por Wilberg e Bence (2008), dada por:

$$p(M_t | l) = \frac{\exp(-\Delta_t/2)}{\sum_{t=1}^6 \exp(-\Delta_t/2)}, t = 1,2,3,4,5,6$$

em que $p(M_t | l)$ é a probabilidade *a posteriori* do modelo t e Δ_t é a diferença entre o DIC do modelo t e o modelo com menor DIC.

3.3.4. RESULTADOS

Foram aplicados seis modelos bayesianos para detectar o efeito dos marcadores juntamente com observações fenotípicas de uma população de goiabeira. No processo de modelagem, foi utilizada uma validação cruzada com oito folds, para obter alguns parâmetros de ajuste dos modelos em todas as folds (Tabela 1). Entre os modelos utilizados, o modelo Regressão de Cumeeira Bayesiana - BRR apresentou valor médio menor para um valor de ajuste comparativo ($<DIC$ - *Deviance Information Criterion*), de acordo como os parâmetros utilizados para a variável teor de sólidos solúveis (TSS).

O DIC é particularmente útil em problemas de seleção de modelos bayesianos, tendo as distribuições posteriores dos modelos sido obtidas pela simulação Monte Carlo da cadeia de Markov (MCMC). DIC é uma aproximação assintótica à medida que o tamanho da amostra se torna grande como o AIC. Só é válido quando a distribuição *posterior* é aproximadamente multivariada normal. Assim, foram verificadas as convergências das cadeias e a distribuição *a posteriori* para todas as variáveis no modelo BRR, Figura 1, para verificar se o DIC poderia ser utilizado com critério de seleção de modelos.

Foi observada uma boa estabilização das cadeias de Markov e Monte Carlo com os valores médios correspondendo às médias fenotípicas observadas. As curvas de densidade *a posteriori* das cadeias apresentaram distribuição normal para todas as variáveis. Sendo assim, é possível utilizar com segurança os valores de DIC para selecionar os modelos.

Foram obtidos também os desvios (Δ) dos critérios de informação para cada variável em relação ao menor valor, assumido como o modelo que apresentou o melhor ajuste as observações. Considerando estes parâmetros, foram obtidos também auxiliares na classificação dos modelos como valores de probabilidade de ajuste *a posteriori* do modelo (W_{prob}) e a razão de evidência (ER) para os modelos. Todos os parâmetros de ajuste do modelo BRR foram superiores aos demais modelos bayesianos utilizados para a variável TSS.

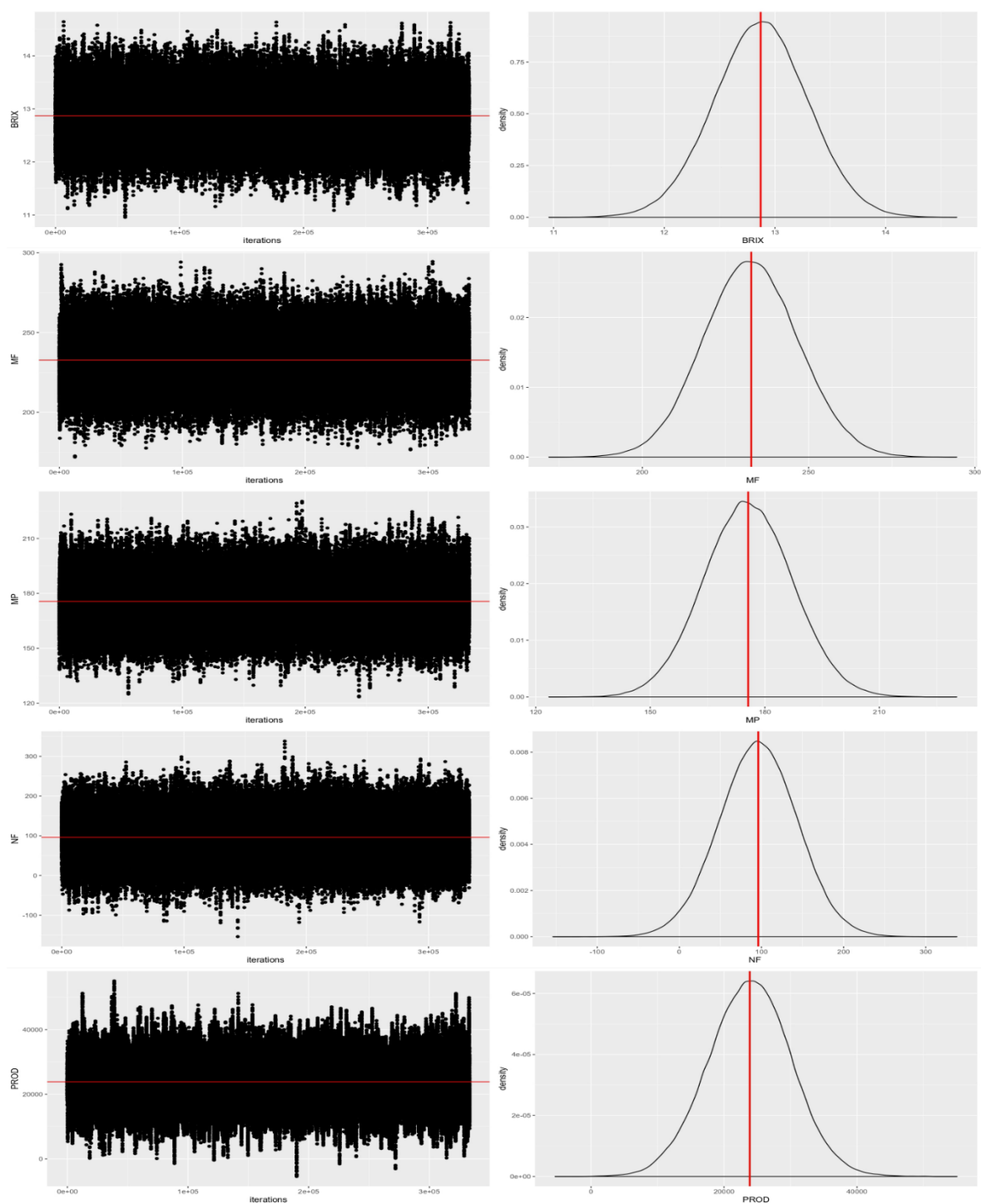


Figura 1. Cadeias de Markov e Monte Carlo com os valores médios (linha vermelha) e curva de distribuição para cinco variáveis observadas em goiabeira, geradas para relacionar marcas SSR a observações fenotípicas.

Além dos valores de ajuste, consideramos para a escolha do modelo a capacidade do modelo em prever os valores fenotípicos de uma subamostra com indivíduos aleatórios, em cada fold da validação cruzada. Os valores médios da

correlação dos preditos e os fenotípicos observados (r), juntamente com um valor de probabilidade de r ter ausência de correlação linear (Tabela 1).

Para a variável TSS, o modelo BRR também apresentou o maior valor de r com a menor probabilidade, sendo uma correlação consistente entre as subamostras. Os demais modelos apresentaram um desempenho muito semelhante, com exceção do modelo BayesL, em que foi observado um valor discrepante de DIC, e do modelo Bayes B2, que, apesar de mostrar um bom ajuste com DIC semelhante, apresentou uma capacidade preditiva baixa, com $r = 0,35$ em relação ao BRR, com $r = 0,65$.

Resultado semelhante nos critérios de ajuste e predição do modelo foi observado para as demais variáveis, como número de frutos por planta (NF). Em NF foram observados valores de ajuste dos modelos muito semelhantes, próximos a 980,19 (DIC), sendo o modelo BRR escolhido pela melhor capacidade preditiva, com $r = 0,82$ para NR, 0,65 para MF, 0,64 para MP e 0,84 para PROD.

Com os critérios de ajuste do modelo muito próximos entre os modelos utilizados e grandes diferenças entre o poder preditivo de cada modelo dentro das variáveis, foi possível observar que escolher um valor de π para o modelo Bayes B causou um *overfit* do modelo, sendo observado que o poder preditivo desse modelo BayesB na maioria apresentou os piores resultados (r).

Tabela 1. Qualidade de ajuste de seis modelos bayesianos BL, BRR, Bayes A, Bayes C, Bayes B e Bayes B com $\pi = 1e-5$ (Bayes B2) para associar marcadores SSR e observações fenotípicas em *P. guajava* nas variáveis teor de sólidos solúveis, massa de frutos, massa de polpa, número de frutos e produtividade. Os valores de viés foram obtidos por validação cruzada com 8 folds (88% das observações para treinamento e 12% para validação), nos mesmos conjuntos de amostras para cada modelo.

	DIC	Δ	Wprob	ER	R	p-value
Teor de sólidos solúveis						
BRR	1348,95	0,00E+00	5,68E-01	1,00E+00	0,65	1,76E-11
BayesA	1352,36	3,41E+00	1,03E-01	5,50E+00	0,65	1,82E-11
BayesL	1455,71	1,07E+02	3,73E-24	1,52E+23	0,65	3,34E-11
BayesC	1352,66	3,71E+00	8,90E-02	6,38E+00	0,65	2,50E-11
BayesB	1353,17	4,22E+00	6,87E-02	8,26E+00	0,65	3,29E-11
BayesB2	1351,35	2,40E+00	1,71E-01	3,31E+00	0,35	3,00E-10
Massa de frutos						
BRR	4330,52	7,03E-01	2,58E-01	1,42E+00	0,65	5,53E-12
BayesA	4332,11	2,29E+00	1,16E-01	3,15E+00	0,65	8,88E-12
BayesL	4377,44	4,76E+01	1,67E-11	2,20E+10	0,64	1,20E-11
BayesC	4331,05	1,23E+00	1,98E-01	1,85E+00	0,65	7,59E-12
BayesB	4333,38	3,57E+00	6,16E-02	5,95E+00	0,65	1,04E-11
BayesB2	4329,82	0,00E+00	3,66E-01	1,00E+00	0,52	2,46E-12
Massa de polpa						
BRR	4179,22	8,37E-01	2,64E-01	1,52E+00	0,64	3,04E-11
BayesA	4181,10	2,71E+00	1,03E-01	3,89E+00	0,64	5,58E-11
BayesL	4224,53	4,61E+01	3,83E-11	1,05E+10	0,62	7,56E-11
BayesC	4180,16	1,78E+00	1,65E-01	2,43E+00	0,63	4,36E-11
BayesB	4181,99	3,61E+00	6,60E-02	6,08E+00	0,63	6,04E-11
BayesB2	4178,39	0,00E+00	4,01E-01	1,00E+00	0,50	1,28E-11

Tabela 1 – Cont.

	DIC	Δ	Wprob	ER	R	<i>p-value</i>
Número de frutos						
BRR	980,19	6,57E-01	1,71E-01	1,39E+00	0,82	4,36E-13
BayesA	981,18	9,87E-01	1,45E-01	1,64E+00	0,79	8,92E-13
BayesL	980,75	5,58E-01	1,79E-01	1,32E+00	0,76	4,31E-13
BayesC	981,36	1,17E+00	1,32E-01	1,79E+00	0,80	2,88E-13
BayesB	980,15	0,00E+00	2,37E-01	1,00E+00	0,79	2,39E-13
BayesB2	981,31	1,12E+00	1,35E-01	1,75E+00	0,73	8,51E-13
Produção						
BRR	1825,73	8,71E-01	1,99E-01	1,55E+00	0,84	2,02E-13
BayesA	1826,51	1,64E+00	1,36E-01	2,27E+00	0,82	5,34E-13
BayesL	1828,81	3,95E+00	4,28E-02	7,20E+00	0,77	4,08E-13
BayesC	1825,92	1,06E+00	1,81E-01	1,70E+00	0,82	1,47E-13
BayesB	1826,55	1,69E+00	1,33E-01	2,33E+00	0,81	3,53E-13
BayesB2	1824,86	0,00E+00	3,08E-01	1,00E+00	0,74	1,10E-12

DIC = critério de desvio de informação; Del (delta) = diferença entre o valor mais alto e o mais baixo de DIC; Wprob = modelo de probabilidades posteriores; ER = razão de evidência; Erro = erro atribuído ao Wprob; r = correlação entre o previsto pelo modelo e as observações da validação reservados; Valor-p = significância para correlação.

Vale ressaltar que os modelos bayesianos levam um tempo considerável para ser executados. Mesmo com o avanço do poder computacional, a resolução de modelos mais complexos demanda um período maior de processamento. Essa é uma informação amplamente conhecida, mas pouco mensurada, que deve ser levada em consideração na escolha do modelo. Neste trabalho, foi mensurado o tempo investido para resolver os modelos bayesianos, repetindo 10 vezes em loop cada cadeia (Tabela 2).

Tabela 2. Estimativas de médias de tempo para resolução de diferentes modelos bayesianos com 10^6 iterações, burn-in de 10^4 e amostragem da cadeia igual a 3. Os cálculos foram feitos no software R (versão 3.5.1) com o pacote BGLR (versão 1.0.8).

Modelo	Tempo para processamento*
Regressão de cumeira Bayesiana - BRR	15,3 horas (+- 0,15)
Bayes A	15,35 horas (+- 0,10)
Bayes B	15,32 horas (+- 0,13)
Bayes B $\pi = 10^{-5}$	15,3 horas (+- 0,10)
Bayes C	15,5 horas (+- 0,2)
Lasso Bayesiano	14,95 horas (+- 0,5)

*Foi utilizado um núcleo de um processador Intel I7-7500U 2,7 GHz.

Foram estimados ainda os valores de herdabilidade no sentido restrito para as variáveis observadas em goiabeira com o modelo que apresentou o melhor desempenho (Tabela 3). Os extremos dos valores foram para teor de sólidos solúveis (TSS = 0,32), sendo o maior valor de herdabilidade observado, e o número de frutos por planta (NF = 0,07) foi a menor herdabilidade observada. Em geral, os valores de herdabilidade foram baixos, mas acompanhados das medidas de desvio e acurácia podem fornecer estimativas mais precisas para o avanço de gerações no programa de melhoramento. Os valores da medida de desvio da herdabilidade foram baixos. Isso indica valores mais precisos para a herdabilidade em contraponto a estimativas de herdabilidade obtidas pontualmente como é feita comumente. As estimativas de erro para a herdabilidade foram obtidas com a estimativa da herdabilidade em cada iteração da validação cruzada, sendo assim estimada em diversos subconjuntos que representam a população.

Já na acurácia preditiva, foram observados valores altos, sendo um bom indicativo de que a herdabilidade estimada representa muito bem a população, em especial o valor da acurácia preditiva para a variável MP, estimada em 0,9708. Contudo, quando se observa a acurácia preditiva da herdabilidade para a variável de principal interesse PROD, foi obtido um valor próximo a 0,51, que é baixo. Foram observados resultados muito semelhantes para massa do fruto e massa de polpa. Os valores de herdabilidade foram 0,1581 e 0,1478 para MF e MP, respectivamente. Os desvios padrão das herdabilidades também ficaram

próximos e baixos. A acurácia preditiva foi melhor em MP em relação a MF, indicando que o volume no interior dos frutos depende menos do tamanho do fruto em si, sendo mais aleatório ou influenciado por outro fator não observado nesse experimento.

Tabela 3. Acurácia preditiva e desvio padrão da herdabilidade para teor de sólidos solúveis (TSS), massa de frutos (MF), massa de polpa (MP), número de frutos por planta (NF) e produção por planta (PROD) observados em goiaba (*Psidium guajava* L.), estimados, utilizando um modelo com marcadores SSR e Regressão de Cumieira Bayesiana - BRR.

	H2	Desvio padrão	Acurácia preditiva
TSS	0,3261	0,0706	0,6302
MF	0,1581	0,0215	0,8779
MP	0,1478	0,0310	0,9708
NF	0,0732	0,0146	0,6939
PROD	0,1058	0,0154	0,5095

Com a matriz de marcas dos indivíduos e o peso que cada marcador recebeu no modelo de regressão de cumieira bayesiana, foram estimados os valores genéticos dos indivíduos para as características e obtida a matriz de correlação genética entre as características (Figura 3). Foi observada forte correlação linear entre as variáveis MP e MF no valor de 0,9393 e também entre as variáveis NF e PROD no valor de 0,9641. Foi observada também correlação entre o teor de sólidos solúveis e as duas variáveis do fruto, com valor de 0,3060 entre TSS e MF e de 0,3705 entre TSS e MP. Foi observado também que TSS apresentou correlação negativa com as variáveis de PROD e NF, mas que não foi significativa.

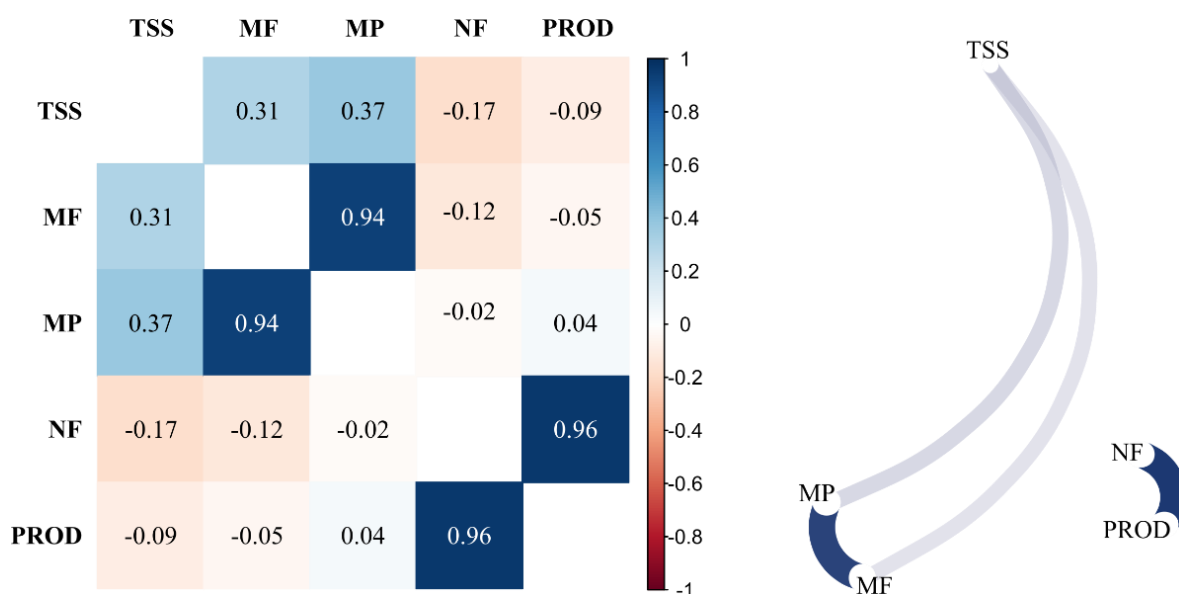


Figura 2. Correlação genética entre teor de sólidos solúveis (TSS), massa de frutos (MF), massa de polpa (MP), número de frutos por planta (NF) e produção por planta (PROD) observados em goiaba (*Psidium guajava* L.), estimados, utilizando um modelo com marcadores SSR e Regressão de Cumeeira Bayesiana - BRR.

Tabela 4. Lista de indivíduos selecionados que apresentaram valores genéticos positivos para as variáveis teor de sólidos solúveis (TSS), massa de frutos (MF), massa de polpa (MP), número de frutos por planta (NF) e produção por planta (PROD).

Indivíduo	TSS	MF	MP	NF	PROD
B1F15P12	0,76	10,68	8,45	24,57	4507,96
B2F8P4	0,19	6,54	4,12	13,11	3115,68
B2F1P4	0,64	4,99	12,08	14,53	2958,55
B1F15P10	2,51	18,14	16,13	1,25	2848,35
B1F2P8	1,21	33,01	29,67	13,28	2268,82
B2F2P10	0,86	12,18	12,33	16,93	2206,82
B2F3P11	0,89	19,74	19,79	9,29	1421,55
B1F8P1	1,04	18,97	15,93	7,67	1264,91
B2F12P9	0,96	6,90	5,19	9,00	1245,41
B2F17P2	0,27	7,61	4,86	6,15	622,86

Os indivíduos foram classificados em ordem decrescente, considerando a variável PROD.

Os indivíduos contidos na Tabela 4 foram selecionados por apresentarem valores positivos em todas as características, contudo, é possível utilizar um índice de seleção caso o objetivo seja selecionar novos indivíduos para compor uma nova população dentro de um programa de melhoramento genético. As famílias 8, 10 e 17 foram as que continham mais indivíduos no ordenamento dos valores genéticos para a PROD. Sendo assim, essas famílias apresentam baixa variabilidade entre si, mas com grande capacidade produtiva, sendo recomendadas para seleção e continuidade para experimento de valor de cultivo e uso. A Tabela 3, em anexo, apresenta uma lista completa com os indivíduos e os valores genéticos de todas as características.

3.3.5. DISCUSSÃO

Usando uma abordagem bayesiana, os efeitos de marcadores podem ser estimados em conjunto para prever os valores genômicos para uma característica quantitativa sem fazer a seleção do marcador. Essa abordagem é chamada seleção genômica. Diversos métodos penalizados e de estimação de contração bayesiana estão disponíveis, como, por exemplo, os homólogos bayesianos da regressão de cumeeira (Ridge Regression - RR) (Hoerl e Kennard 1970), operador de seleção e contração menos absoluto (Least Absolute Shrinkage and Selection Operator - LASSO) (Tibshirani 1996) e modelos como BayesA e BayesB e suas extensões (Meuwissen et al., 2001). Esses modelos são frequentemente testados para diversas culturas de interesse, contudo, para a goiabeira, essa informação ainda é escassa. Aqui é discutido o desempenho de seis modelos bayesianos para ajuste de marcadores SSR na goiabeira e estimados parâmetros de interesse para o melhorista em um programa de melhoramento.

Embora existam diferenças entre os métodos, em suposições *a priori* sobre os efeitos dos marcadores, foi observado que os parâmetros de ajuste dos modelos foram semelhantes. Nenhuma diferença evidente foi detectada para qualquer uma das características, principalmente para DIC. Assim, os modelos foram escolhidos considerando não somente os parâmetros de ajuste, como

também sua capacidade preditiva e como se comportam em relação aos marcadores para gerar o modelo regressivo.

O BayesL produz um encolhimento mais forte dos coeficientes de regressão próximos de zero e menos encolhimento daqueles com grandes valores absolutos, levando a um modelo mais escasso, enquanto o BRR encolhe mais fortemente os coeficientes de regressão com um valor alto (Heslot, 2012). Assim, foi observado que BayesL apresentou desempenho mediano, possivelmente porque a quantidade de marcas significativas com grandes efeitos no modelo estava muito escassa para explicar as características quantitativas avaliadas. Intuitivamente o inverso ocorreu com o modelo BRR, que considerou mais os efeitos das marcas, gerando um modelo com mais marcas para explicar características controladas por diversos genes. Isso significa que a distribuição das marcas foi, em média, um pouco menos de picos para a grade de pesquisa de efeitos no modelo BRR.

Estudos que buscam os melhores modelos para diferentes espécies são importantes para direcionar programas de melhoramento. Por exemplo, para outra planta perene, maracujá, *Passiflora edulis*, foi observado que o modelo BayesC foi o melhor para diversas variáveis avaliadas nessa espécie (Viana et al., 2017). Esse modelo assume variância comum para todos os efeitos de marcadores como também pressupõe que alguns marcadores não tenham efeito (Habier et al. 2011). Assim, genes com a mesma frequência alélica provavelmente explicam a mesma porção de variação genética, sugerindo que as variáveis são controladas por diversos genes com poucos efeitos sobre as variáveis quantitativas observadas também nesse trabalho. Nos resultados, foi possível observar que esse modelo também apresentou um desempenho satisfatório para variáveis em goiabeira, podendo ser escolhido como um modelo alternativo.

Resultados semelhantes entre métodos bayesianos como BayesA e BayesB e outros deles derivados também foram observados por Habier et al. (2011), que trabalharam com a aplicação de métodos bayesianos em dados reais de touros. Esse resultado semelhante já era esperado, visto que os modelos apresentam poucas variações entre si. Por exemplo, BayesB e BayesA são mais tolerantes quanto à suposição de uma variância comum entre os efeitos dos marcadores. *A priori* assumida nesses modelos para o efeito de um j -ésimo

marcador é uma distribuição conjunta com uma probabilidade π para o beta referente à marca igual a zero.

Quando o modelo BayesB foi proposto, π foi sugerido com um valor próximo a 0,95 (Meuwissen et al., 2001). Contudo, com poucas marcas é possível escolher valores mais baixos para π , em que BayesB com π reduzido a zero equivale ao modelo BayesA. Como foi observado um possível overfitting do modelo BayesB quando utilizamos um valor de $\pi = 1e^{-5}$, foi forçado que as marcas tinham grande probabilidade de ter efeito sobre a variável de interesse. Assim, era obtido um modelo em que os betas referentes às marcas se ajustavam muito bem para explicar a subamostra em cada folds da validação cruzada, mas falhava para prever a amostra de validação como observado para a maioria das características (Tabela 1).

Se tivessem sido utilizados somente os parâmetros de ajuste do modelo como o DIC, que é amplamente utilizado, talvez não teria sido observado que o poder preditivo do modelo BayesB tivesse apresentado o pior desempenho. Isso evidencia a importância da validação cruzada. A validação cruzada foi usada para avaliar como os resultados de um modelo estatístico se assemelham a outro conjunto de observações. Por exemplo, como um modelo ajustado preverá observações que não foram usados para ajustar o modelo. Prever o desempenho de genótipos com fenótipos ainda a serem observados, como, por exemplo, linhas ou linhas recém-desenvolvidas, que foram avaliadas em poucos ambientes, é essencial no melhoramento de plantas. Portanto, a validação cruzada parece ser uma maneira natural de avaliar o desempenho do modelo sob a perspectiva do melhorista (Crossa et al., 2011).

Estudos de simulação mostraram que a seleção genômica usando marcadores isoladamente pode ajustar o modelo com uma precisão de até 85% (Meuwissen et al., 2001). A precisão de 85% é a correlação entre os verdadeiros valores genéticos e os valores previstos de indivíduos na próxima geração. Valores genéticos verdadeiros são conhecidos apenas em estudos de simulação. Na análise de observações reais, a previsibilidade de um modelo deve ser extraída de um estudo de validação cruzada. A previsibilidade obtida da validação cruzada e a qualidade do ajuste do modelo não necessariamente concordam entre si. Começando com um pequeno número de marcadores, ambos podem aumentar à medida que o número de marcadores aumenta. Aumentar ainda mais

o número de marcadores pode continuar a melhorar a qualidade do ajuste do modelo, mas a previsibilidade pode cair (Che and Xu, 2010).

O coeficiente de herdabilidade influencia na predição de valores genéticos genômicos, na capacidade preditiva e nas análises de associação em todo o genoma. Assim, quanto maior a herdabilidade de um fenótipo, há melhorias na identificação de indivíduos a serem utilizados como pais nas próximas gerações, além de favorecer a identificação de regiões associadas a uma característica de interesse (Viana et al., 2017).

A herdabilidade da característica TSS foi a maior observada, e o valor confirma, dentro do intervalo, um trabalho que avaliou uma grande população de goiabeiras na Índia (Dinesh et al., 2017). Os autores também detectam correlação entre essa variável e massa do fruto, possibilitando uma seleção indireta. É sugerido ainda que pode haver um possível efeito não aditivo nos genes controlando essa característica, pois observaram uma variância fenotípica maior que a variância genotípica. Nosso modelo apresentou baixa acurácia preditiva para a herdabilidade apesar do maior valor, sendo que em aproximadamente em 40% da subamostra de validação o modelo predisse enviesadamente, confirmando também a ideia da ação de genes com efeitos não aditivos para esta característica.

Para as demais características, os valores de herdabilidade foram baixos como esperado. Os valores geralmente relatados para características como massa de frutos, massa de polpa, número de frutos e produção geralmente ficam próximos de 0,60 (Bihari, 2011). Nossas estimativas possivelmente são menores por se tratar de estimativas oriundas de um modelo que considera o efeito de marcas, e as estimativas comuns são obtidas de observações fenotípicas que têm muito mais fontes de variação, muitas vezes não consideradas. Apesar de baixas, as herdabilidades mostraram na validação cruzada uma boa acurácia preditiva, chegando a 0,97 para a massa de polpa.

A variável massa de polpa está fortemente correlacionada com a variável massa do fruto, o que, do ponto de vista da fisiologia da planta, já era esperado. MF e MP são variáveis obtidas de maneiras semelhantes: em uma se mensura a massa do fruto inteiro e na outra, é removida a placenta, que contém as sementes, parte que não interessa no processamento do fruto. Ambas as variáveis apresentaram herdabilidades semelhantes de 0,14 e 0,15, que foram

superiores às variáveis de interesse referentes à produção (NF e PROD). Essas duas variáveis são geralmente observadas em colinearidade, e essa colinearidade entre essas duas variáveis é particularmente interessante para estudos de correlações entre variáveis em goiabeira, podendo envolver modelagens de equações estruturais como a análise de trilha, que busca variáveis que podem ser selecionadas indiretamente.

Como a herdabilidade é muito semelhante nas duas variáveis e a correlação genética entre as duas também é alta, um programa pode direcionar a seleção indivíduos com mais massa de polpa, selecionando indiretamente pela massa do fruto. Nas etapas de seleção, há uma grande diferença de tempo e recurso gasto entre somente mensurar a massa de um conjunto de frutos contra abrir um fruto remover a placenta e mensurar a massa da polpa. Assim, a seleção da massa de polpa, que é o produto de interesse para a indústria extrair e processar os subprodutos derivados da goiabeira, pode ser mais eficiente.

Já nas variáveis número de frutos e produção, a herdabilidade foi muito baixa, juntamente com as estimativas de acurácia preditiva. Como essas também são variáveis quantitativas, geralmente controladas por um grande número de genes, era esperada uma herdabilidade baixa. Mas provavelmente nosso modelo não foi capaz de capturar todos os efeitos para essas características, visto que o modelo, apesar de predizer com mais de 50% a subamostra de validação, tem desempenho não satisfatório.

Vale ressaltar que essas características avaliadas sofrem grande influência do ambiente e, principalmente, do manejo (Thaipong e Boonprakob, 2005). Por exemplo, um trato cultural comum em goiabeiras é a poda de formação e posterior desbaste dos novos brotos que surgem após a poda. Isso serve para controlar tanto a altura da planta para facilitar a colheita, quanto o número e o tamanho de frutos. Assim, as inflorescências que originaram os frutos surgem em gemas nas axilas dos novos brotos. Se forem mantidos muitos brotos após a poda, o número de frutos tende a aumentar, mas a massa dos frutos é menor em razão da maior distribuição dos recursos disponíveis da planta mãe. Isso nos leva a procurar uma correlação entre número de frutos/produção, como, por exemplo, a massa do fruto, que não foi observada aqui, ou pelo menos é uma correlação não linear, já que as correlações entre NF e MF são próximas de zero (Figura 2).

Foram observados valores genéticos discrepantes entre os indivíduos selecionados, e uma possível explicação para esse fato é que a população apresenta alta variabilidade genética, implicando diferenças entre os valores genéticos dos indivíduos, deixando-as mais pronunciadas, facilitando uma classificação mais acurada dos indivíduos.

3.3.6. CONCLUSÃO

O modelo de regressão de cumeira bayesiana apresentou os melhores resultados e foi escolhido para predizer os valores genéticos dos indivíduos nas variáveis teor de sólidos solúveis, massa de frutos, massa de polpa, número de frutos e produção. Os valores de herdabilidade apresentaram boa acurácia preditiva. Correlações genéticas foram obtidas para verificar a relação entre as variáveis e o modelo, evidenciando fortes correlações entre algumas variáveis, permitindo seleção indireta.

REFERÊNCIAS BIBLIOGRÁFICAS

- Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control.*, Boston, v.19, n.6, p.716-723.
- Almeida, O. F.; Viana, A. P.; Gomes, V. M.; De Souza, Ricardo Moreira; Santos, P. R.; Amaral Junior, A. T.; Rodrigues, Daniele Lima; Rodrigues, D. L.(2017) Interspecific cross of *Psidium* species: crossability and genetic diversity of hybrids resistant to *Meloidogyne enterolobii*. *Crop Breeding and Applied Biotechnology*.
- Alencar. X. (2019) Efficient estimation of marker effects in plant breeding. *G3: Genes, Genomes, Genetics*, v. 9, n. 11, p. 3855-3866.
- Amaral Júnior, A. T.; Viana, A. P.; Gonçalves, L. S. A.; Barbosa, C. D. (2010) Procedimentos Multivariados em Recursos genéticos vegetais. **In:** Pereira, T.N.S. Germoplasma: Conservação, Manejo e Uso no Melhoramento de Plantas. Viçosa, MG: Arca. p.205- 254.
- Ambrósio, M.; Viana, A.P.; Ribeiro, R. M.; Preisigke, S. C.; Cavalcante, N. R.; Alves, F.; Torres, G. X.; Bezerra, C. M. (2020) Genotypic Superiority of *Psidium guajava* L. S1 Families Using Mixed Modeling for Truncated and Simultaneous Selection. *Scientia Agricola*.

- Barros, M. A.; Rocha, M. De M.; Gomes, R. L. F.; Silva, K. J. D.; Neves, A. C. (2013) Adaptabilidade e estabilidade produtiva de feijão-caupi de porte semiprostrado. *Pesquisa Agropecuária Brasileira*, v.48, p.403-410.
- Beaumont, M. A. and B. Rannala. (2004) "The Bayesian revolution in genetics." *Nature Reviews Genetics* **5**(4): 251.
- Bergo, C. L.; Miqueloni, D. P.; Pedrozo, C. A.; Resende, M. D. V. (2019) Parâmetros genéticos para caracteres de produção de palmito em progênies de meios-irmãos de pupunheira. *Cienc. Rural* vol.49 no.7 Santa Maria. DOI: <http://dx.doi.org/10.1590/0103-8478cr20180638>.
- Bergo, C. L.; Negreiros, J. R. S.; Miqueloni, D. P.; Lunz, A. M. P. (2013) Estimativas de repetibilidade de caracteres de produção em pupunheiras para palmito da raça putumayo. *Revista Brasileira de Fruticultura*, v. 35, p. 829-836.
- Bezerra, C. M.; Ribeiro, R. M.; Viana, A. P.; Cavalcante, N. R.; Silva, A. F.; Ambrósio, M. (2020) Guava breeding via full-sib family selection: conducting selection cycle and divergence between parents and families. *Crop Breeding and Applied Biotechnology*.
- Blasco, A. (2001) The Bayesian controversy in animal breeding. *Journal of Animal Science*, 79:2023–2046.
- Bihari, M. A. N.; Suryanarayan. (2011) Genetic diversity, heritability, genetic advance and correlation coefficient in guava (*Psidium guajava* L.). *Indian Journal of Agricultural Sciences*, v. 81, n. 2, p. 107-110.
- Bolstad. W. M. (2007) Introduction to Bayesian statistics 2nded Hoboken John Wiley. '007. 463p.
- Boti, J. B.; Maladon, F. Z.; Oliveira, B. R.; Haddade, I. R. (2016) Insetos provocadores de danos em folhas, flores e frutos da goiabeira (*Psidium guajava* L., *Myrtaceae*) nos pomares conduzidos em sistema de cultivo convencional e orgânico, no município de Santa Teresa-ES. *Natureza on line*. v.14, n.1, p.040-044.

- Browne, W. J.; Draper, D. (2006) A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*. v.1, n. 3, p. 473-514.
- Burla, R. S. ; Souza, R.M ; Gomes, V. M. ; Correa, F. M. (2010) Comparação entre níveis de inóculo, época de avaliação e variáveis para seleção de *Psidium* spp. visando à resistência a *Meloidogyne mayaguensis*. *Nematologia Brasileira*, v. 34, p. 82-90.
- Burnham, K. P.; Anderson, D. R. (2004) Multimodel inference: understanding aic and bic in model selection. *Sociological Methods and Research*. Beverly Hills, v.33, n.2, p.261-304.
- Campos, B. M.; Viana, A.; P.; Quintal, S. S. R.; Barbosa, C. D.; Daher, R F. (2016) heterotic group formation in *Psidium guajava* L. By artificial neural network and discriminant analysis. *Revista Brasileira de Fruticultura*, v. 38, p. 151-157.
- Campos, B. M.; Viana, A. P.; Quintal, S. S. R.; Gonçalves, L. S. A.; Pessanha, P. G. O. (2013) Quantificação de divergência genética entre acessos de goiabeira por meio da estratégia Ward-MLM. *Revista Brasileira de Fruticultura (Impresso)*, v. 35, p. 87-94.
- Cardoso, A. I. I. (2006) Número mínimo de colheitas em pepino híbrido estimado por meio do coeficiente de repetibilidade. *Bragantia*. 65:591-595.
- Carlin, B. P.; Chib, S. (1995). "Bayesian model choice via Markov chain Monte Carlo methods." *Journal of the Royal Statistical Society* 57(3): 473-484.
- Cargnim, A. (2016) Repetibilidade e número de colheita de características para seleção de clones de variedades viníferas. *Ciência Rural*, v.46, n.2, p.221-226.
- Cecon, P. R.; Silva, F. F.; Ferreira, A.; Ferrão, R. G.; Carneiro, A. P. S.; Detmann, E.; Faria, P. N.; Morais, T. S. S. (2008) Análise de medidas repetidas na avaliação de clones de café 'Conilon'. *Pesq. agropec. bras.* vol.43 no.9 Brasília Sept. DOI: <http://dx.doi.org/10.1590/S0100-204X2008000900011>.

- Cestari, A. A.; Costa, S. C.; Minho, A. P. (2012) Análise de dados longitudinais em experimentação animal. *Semina. Ciências Agrárias (Online)*, v. 33, p. 1565-1579. DOI: 10.5433/1679-0359.2012v33n4p1565.
- Chen, T.W.; Chaing, C.; Wang, C.Y.; Shyu, Y.T. (2007) Molecular Identification and Analysis of *Psidium guajava* L. from Indigenous Tribes of Taiwan. *Journal of Food and Drug Analysis*, v. 15, n. 1, p. 82-88.
- Che X.; Xu S. (2010) Significance test and genome selection in Bayesian shrinkage analysis. *International Journal of Plant Genomics* 2010: 11 pages. doi: 10.1155/2010/893206.
- Collard, B. C. Y.; Mackill, D. J. (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B*, v. 363, p. 557–572.
- Costa, J. G. (2003) Estimativas de repetibilidade de alguns caracteres de produção em mangueira. *Cienc. Rural* vol.33 no.2 Santa Maria Apr. <http://dx.doi.org/10.1590/S0103-84782003000200013>.
- Cowles, M. K.; Carlin, B. P. (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* 91(434): 883-904.
- Crossa, J.; Pérez, P.; de los Campos, G.; Mahuku, G.; Dreisigacker, S.; Magorokosho, C. (2011) Genomic Selection and Prediction in Plant Breeding, *Journal of Crop Improvement*, 25:3, 239-261, DOI: 10.1080/15427528.2011.558767.
- Cruz, C. D.; Regazzi, A. J.; Carneiro, P. C. S. (2012) Modelos biométricos aplicados ao melhoramento genético. Viçosa, MG: UFV, v. 1, 514p.
- Cruz, C. D., Regazzi, A. J., Carneiro, P. C. S. (2012) Modelos biométricos aplicados ao melhoramento genético Volume1: Ed. UFV. 514p.
- Cruz, C.D., Carneiro, P.C.S., Regazzi, A.J. (2004) Modelos biométricos aplicados ao melhoramento genético, 3, ed, Viçosa: UFV, 480p.

- Dahiya, K. K.; Sunil, A.; Karihaloo, J. L. (2002) DNA fingerprinting of guava (*Psidium guajava* L.) cultivars using RAPD markers. Indian Journal of Plant Genetic Resources, v. 15, n. 2, p 112-115.
- De Los Campos, G.; Naya, H.; Gianola, D.; Grossa, J.; Legarra, Manfredi, E.; Weigel, K.; Cotes, J. M. (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics, vol. 182 no. 1 375-385; <https://doi.org/10.1534/genetics.109.101501>.
- De los Campos, G.; Hickey, J. M.; Pong-Wong, R.; Daetwyler, H. D.; Calus, M. P. L. (2012). Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. Genetics, 193(2), 327–345. doi:10.1534/genetics.112.143313
- De Los Campos, G.; Vazquez, A. I.; Rohan, F.; Klimentidis, Y. C.; Sorensen, D. (2013) Prediction of complex human traits using the genomic best linear unbiased predictor. PLoS genetics, v. 9, n. 7, p. e1003608, doi.org/10.1371/journal.pgen.1003608
- Desta, Z. A.; Ortiz, R. (2014) Genomic selection: genome-wide prediction in plant improvement. Trends in plant science, v. 19, n. 9, p. 592-601.
- Doyle, J.J.; Doyle, J.L. (1990) Isolation of plant DNA from fresh tissue. Focus. 12:13- 15.
- Diggle, P.; Liang, K.; Zeger, S. (2002). Analysis of Longitudinal Data. Oxford University Press, Oxford.
- Dinesh, M. R.; Bharathi, K.; Vasugi, C.; Gowda, V. G. L.; Ravishankar, K.; Nischita, P. (2017) Inheritance studies and validation of hybridity in guava (*Psidium guajava* L.). 87. 42-45.
- Eberhart. S. A.; Russell. W. A. (1966) Stability parameters for comparing varieties. Crop Science, v.6, p.36-40.
- Efron, B. (1986). "Why isn't everyone a Bayesian?" The American Statistician 40(1): 1-5.

- Evangelista, R. M.; Vieites, R. L. (2006) Avaliação da qualidade de polpa de goiaba congelada, comercializada na cidade de São Paulo. *Segurança Alimentar e Nutricional*, v. 13, n. 2, p. 76-81.
- Falconer, D. S.; Makcay, T. F. C. (1996) *Introduction to quantitative genetics*. 4 ed. Essex: Longman, 464p.
- FAO. FAOSTAT. Food and Agriculture Organization of the United Nations. *Estatistical Databases Agriculture*. Disponível em: <<http://faostat.fao.org/>>. Acesso em:28/12/2019.
- Fonseca, K. G.; Faleiro, F. G.; Peixoto, J. R.; Junqueira, N. T. V.; Silva, M. S.; Bellon, G.; Junqueira, K. P.; Vaz, C. F. (2009) Análise de recuperação do genoma recorrente em maracujazeiro-azedo por meio de marcadores RAPD. *Revista brasileira de fruticultura*, Jaboticabal, v.31, n.1, p.145-153, doi: 10.1590/S0100-29452009000100021.
- Freitas, A. R. De; Presotti, C. V.; Toral, F. L. B. (2005) Alternativas de Análises em Dados de Medidas Repetidas de Bovinos de Corte. *R. Bras. Zootec.*, v.34, n.6, p 2233-2244.
- Freitas, E.G.; Barbin, D.; Barbosa, G. V. S.; Carneiro, M. S.; Bassinello, A. T. (2008) Modelo univariado aplicado a dados longitudinais de cana-de-açúcar. *Revista Brasileira de Biometria*, v.26, p.93-16.
- Freire Filho, F. R; Ribeiro, V. Q.; Rocha, M. De M.; Lopes, Â. C. A. (2002) Adaptabilidade e estabilidade da produtividade de grãos de linhagens de caupi de porte enramador. *Revista Ceres*, v.49, p.383-393.
- Friston, K. J.; Glaser, D. E.; Henson, R. N. A.; Kiebel, S.; Phillips, C.; Ashburner, J. (2002) Classical and Bayesian inference in neuroimaging: Applications. *NeuroImage* 16: p. 484-512.
- Geweke, J. (1992) Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *Bayesian Statistics 4* (Eds.) Oxford University Press, 169-193.

- Gianola, D.; Rohan, L. F.; Sotan, I.; Foulley, J. L. (1989). "Likelihood estimation of quantitative genetic parameters when selection occurs: models and problems." *Genome* 31(2): 768-777.
- Gianola, D.; Rohan, L. F. (1986) Bayesian methods in animal breeding theory. *J Anim Sci* 63, 217-244.
- Gomes Filho, A.; Oliveira, J.G.; Viana, A.P.; Siqueira, A.P.O.; Oliveira, M.G.; Pereira, M.G. (2010) Marcadores moleculares RAPD e descritores morfológicos na avaliação da diversidade genética de goiabeiras (*Psidium guajava* L.) *Acta Scientiarum Agronomy*, v. 32, p. 627-633.
- Gomes, S. M; Somavilla, N. S. D.; Gomes-Bezerra, K. M.; Miranda, S. C.; De-Carvalho, P. S.; Graciano-Ribeiro, D. (2009) Anatomia Foliar de Espécies de Myrtaceae: contribuição à Taxonomia e Filogenia. *Acta Botânica Brasílica* (23)223-238.
- González Barrios, P.; Speranza, P.; Glison, N.; Piccardi, M.; Gutiérrez, L. (2016) Analysis of flowering dynamics heritability in the perennial warm-season grass *Paspalum dilatatum*. *Grass and forage science*, v. 71, n. 1, p. 123-131. DOI: <https://doi.org/10.1111/gfs.12159>.
- Guavamap, Generation and screening of microsatellite markers (SSRs) in Guave. 2008. Disponível em: <<http://www.neiker.net/neiker/guavamap/for1-6a.htm>>.
- Gupta, P. K.; Varshney, R. K. (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica*, v. 113, p 163-185.
- Habier, D.; Fernando, R. L.; Kizilkaya, K.; Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*, 12(1). doi:10.1186/1471-2105-12-186.
- Hadfield, J. D. (2010). "MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package." *Journal of Statistical Software* 33(2): p1-22.

- He, J.; Zhao, X.; Laroche, A.; Lu, Z.; Liu, H.; Li, Z. (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*, v. 5, p. 1-8.
- Heffner, E. L.; Jannink, J. L.; Sorrells, M. E. (2011). Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. *The Plant Genome Journal*, 4(1), 65. doi:10.3835/plantgenome.2010.12.0029.
- Henderson, C. R. (1975). "Best linear unbiased estimation and prediction under a selection model." *Biometrics* 31(2): 423-447.
- Henderson, C. R. (1984) Aplicações de modelos lineares na criação de animais. Guelph: Universidade de Guelph, p 462.
- Heidelberger, P.; Welch, P. D. (1983) Simulation run length control in the presence of an initial transient. *Oper. Res.*, 31:1109–1144.
- Hernández-Delgado, S.; Padilla-Ramírez, J. S.; Nava-Cedillo, A.; Mayek-Pérez, N. (2007) Morphological and genetic diversity of Mexican guava germplasm. *Plant Genetic Resources: Characterization and Utilization*, v. 5, n. 3, p. 131-141.
- Herrero, J. V. I.; Medina, N. N. R; Alor, M. B.; Garcia, M. M. O.; Moreno, A. Q.; Teyer, L. F. S.; Risterucci, A. M.; Rohde, W. (2010) Microsatélites desarrollados en guayabo (*Psidium guajava* L.) y su utilidad para evaluar diversidad en la familia Myrtaceae. *Revista Colombiana de Biotecnología*, v. XII, n. 1, p. 64-76.
- Heslot, N.; Yang, H.P.; Sorrells, M. E.; Jannink, J.L. (2012) Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Science*, 52(1), 146. doi:10.2135/cropsci2011.09.0297
- Hoerl, A. E.; Kennard, R. W. (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12:55–67.
- IBGE. (2017) Instituto Brasileiro de Geografia e Estatística. Sistema IBGE de Recuperação Automática – SIDRA. Disponível em: <http://www.sidra.ibge.gov.br/bda/tabela/protabl.asp?c=1613&z=t&o=11&i=>.

- Islam, M.A.; Chowdhury, R.I. (2017) Analysis of Repeated Measures Data. Springer Nature: Singapore. p 257.
- Jeffreys, H. (1961) Theory of Probability. Oxford: Clarendon Press.
- Junqueira, V. S.; Peixoto, L. A.; Laviola, B. G.; Bhering, L. L.; Mendonça, S. Costa T. S. A.; Antoniassi, R. (2016). "Bayesian multi-trait analysis reveals a useful tool to increase oil concentration and to decrease toxicity in *Jatropha curcas* L." PloS One 11(8): e0157038.
- Lédo, F. J. S.; Pereira, A. V.; Sobrinho, F. S.; Auad, A. M.; Jank, L.; Oliveira, J. S. (2008) Estimativas de repetibilidade para caracteres forrageiros em *Panicum maximum*. Ciência e Agrotecnologia, v.32, n.4, p.1299-1303.
- Leon, N.; Jannink, J. L.; Edwards, J. W.; Kaeppler, S. M. (2016) Introdução a uma edição especial sobre genótipo por interação com o ambiente. Crop Science 56 , 2081 . doi: 10.2135 / culturasci2016.07.0002in.
- Lessa, L. S.; Ledo, C. A. S.; Amorin, E. P.; Silva, S. O. (2014) Estimativa de repetibilidade de híbrido diploides (AA) de bananeira. Pesquisa Agropecuária Brasileira, v.49, p.109-117.
- Li, Z.; Sillanpää, M. J. (2012) Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. Theoretical and applied genetics, v. 125, n. 3, p. 419-435.
- Lima, M. A. C. D.; Assis, J. S. D.; Gonzaga Neto, L. U. I. Z. (2002) Caracterização dos frutos de goiabeira e seleção de cultivares na região do submédio São Francisco. Revista Brasileira de Fruticultura.
- Littell, R. C.; Milliken, G. A.; Stroup, W. W.; Wolfinger, R. D. (1996). SAS System for Mixed Models. Cary: SAS Institute.
- Littell, R. C.; Milliken, G. A.; Stroup, W. W.; Wolfinger, R. D.; Schabenberger, O. (2006) SAS for mixedmodels. SAS publishing. p828.

- Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., Dhawan, A. K. (2011) Microsatellite markers: an overview of the recent progress in plants. *Euphytica*, 177: 309–334.
- Kidaha, L. M., Alakonya, A. E., Nyende, A. B. (2014) Molecular Characterization of Guava Landraces in Kenya (Western and South Coast). *Journal of Biology, Agriculture and Healthcare*, 4(15): 81-87.
- Lopes, R.; Bruckner, C. H.; Cruz, C. D.; Lopes, M. T. G.; Freitas, G. B. (2001) Repetibilidade de características do fruto de aceroleira. *Pesquisa Agropecuária Brasileira*, Brasília, v.36, n.3, p.507-513.
- Maia, M. C. C.; Oliveira, L. C.; Vasconcelos, L. F. L.; Lima Neto, F. P.; Yokomizo, G. K.; Araújo, L. B. (2017) Repetibilidade de características quantitativas de frutos em seleções elite de manga rosa. *Revista Agro@ambiente On-line*, v. 11, n. 1, p. 56-62.
- Maia, E.; Siqueira, D. L.; Carvalho, S. A.; Peternelli, L. A.; Latado, R. R. Aplicação da análise espacial na avaliação de experimentos de seleção de clones de laranjeira Pêra. *Cienc. Rural* vol.43 no.1 Santa Maria. 2013.DOI:<http://dx.doi.org/10.1590/S0103-84782012005000134>.
- Malheiros, E. B. (2004) Accuracy of univariate F tests used in repeated measures in time when the sphericity condition is not available. *Rer. Mat. Estat.*, São Paulo, v.22, n.2, p.23-29.
- Malheiros, E. B. (2004) Precisão de teste F univariados usados em experimentos com medidas repetidas no tempo, quando a condição de esfericidade da matriz de covariâncias não é verificada. *Rev. Mat. Estat.*, São Paulo, v.22, n.2, p.23-29.
- Manfio, C. E.; Motoikei, S. Y.; Santos, C. E. M.; Pimentel, L. D.; Queiroz, V.; Sato, A. Y. (2011) Repetibilidade em características biométricas do fruto de macaúba. *Ciência Rural*, v. 41, n. 1, p.70-76.
- Mathew, B.; Leon, J.; Sillanpaa, M. J. (2018) Impacto das estruturas de covariância residual na capacidade de previsão genômica em ensaios multi-ambiente. *PLoS ONE*, vol. 13, n. 7.

- Medina, N. N. R.; Herrero, L.; Lanes, M A.K.; Sánchez, A. O.; Rohde, W. (2010) Estudio de diversidad de accesiones de guayabo (*Psidium guajava* L) mediante el marcador molecular ISTR. Revista Colombiana de Biotecnología v. 12, n. 1, p. 113-123.
- Medina, J. C.; Castro, J. V.; Sigrist, J. M. M.; Martín, Z. J.; Kato, K.; Maia, M. L.; Lopez García, A. E. B.; Leite, R. S. S. F. (2014) Goiaba: cultura, matéria-prima, processamento e aspectos econômicos. 2TM.ed. revisada. e ampliada. Campinas: Instituto de Tecnologia de Alimentos, ITAL, Frutas Tropicais, 6, p.01-119.
- Mehmood, A., Jaskani, M.J., Khan, I.A., Ahmad, S., Ahmad, R., Luo, S. and Ahmad, N.M., (2014) Genetic diversity of Pakistani guava (*Psidium guajava* L.) germplasm and its implications for conservation and breeding. Scientia Horticulturae, 172, pp.221-232.
- Meuwissen, T. H.; Hayes, B. J.; Goddard, M. E. (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.
- Miedaner, T.; Korzun, V. (2012) Marker-Assisted Selection for Disease Resistance in Wheat and Barley Breeding. Phytopathology, v. 102, p. 560-566.
- Miranda, G.B.; Souza, R. M.; Viana, A. P. (2011) Assessment of methods and criteria for screening *Psidium* spp. for resistance to *Meloidogyne enterolobii*. Nematologia Brasileira, v. 34, p. 211-219.
- Mohammad, S.; Fikret, I. (2019) Genetic parameter estimates and GxE interaction in a large-cloned population of *Pinus taeda* L. Tree Genetics & Genomes, Vol.15(3), p 1-13.
- Mostofian, B.; D. M. Zuckerman (2019). "Statistical uncertainty analysis for small-sample, high log-variance data: Cautions for bootstrapping and Bayesian bootstrapping." Journal of chemical theory and computation 15(6): 3499-3509.
- Mutshinda, C. M.; Sillanpää, M. J. (2010) Extended Bayesian LASSO for multiple quantitative trait loci mapping and unobserved phenotype prediction. Genetics, v. 186, n. 3, p. 1067-1075.

- Negreiros, J. R. S.; Andrade Neto, R. C.; Miqueloni, D. P.; Lessa, L. S. Estimativa de repetibilidade para caracteres de qualidade de frutos de laranja-doce. *Pesquisa Agropecuária Brasileira*, v.49, n.1, p.40-48, 2014.
- Neves, L. G.; Bruckner, C. H.; Cruz, C. D.; Barelli, M. A. A. (2010) Avaliação da repetibilidade no melhoramento de famílias de maracujazeiro. *Ceres*, 57:480-485.
- Nogueira, D. A.; Sáfadi, T.; Bearzoti, E.; Filho, J. S. D. S. B. (2003) Análise clássica e bayesiana de um modelo misto aplicado ao melhoramento animal: uma ilustração. *Cienc. Agrotec.*, Lavras, p. 1614-1624.
- Nogueira, A. M.; Ferreira, A.; Ferreira, M. F. S. (2015) Transferability of Microsatellites from *Psidium guajava* L. to *Eugenia*, *Myrciaria*, *Campomanesia*, and *Syzygium* Species (Myrtaceae). *Plant Molecular Biology Reporter*, 34(1): 249-246.
- Oliveira, I. P., Oliveira, L. C., de Moura, C. S. F. T., de Lima Júnior, A. F., & da Rosa, S. R. A. (2012) Cultivo da goiabeira: do plantio ao manejo. *Revista Eletrônica Faculdade Montes Belos*, 5(4).
- Oliveira, M. S. P.; Fernandes, G. L. C. (2001) Repetibilidade de caracteres do cacho de açaizeiro nas condições de Belém-PA. *Rev. Bras. Frutic.*, Jaboticabal - SP, v. 23, n. 3, p. 613-616.
- Oliveira, N. N. S.; Viana, A. P.; Quintal, S. S. R.; Paiva, C. L.; Marinho, C. S. (2014) Análise de distância genética entre acessos do gênero *Psidium* via marcadores ISSR. *Revista Brasileira de Fruticultura (Impresso)*, v. 36, p. 917-923.
- Oliveira, T. N.; Santos, M. V. F.; Lira, M. A.; Mello, A. C. L.; Lira Júnior, M. A.; Silva, N. G. M. (2011) Estimativa de repetibilidade para mancha ocular em clones de *Pennisetum*. *Arquivos de Zootecnia*, 60 (231):797- 808.
- Owusu-Ansah, F.; Curnow, R. N. (2018) Planning clonal selection programmes for perennial crops. *Experimental Agriculture*, v. 54, n. 6, p. 957-965.

- Padmakar, B.; Sailaja, D.; Aswath, C. (2015) Molecular exploration of guava (*Psidium guajava* L.) genome using SSR and RAPD markers: a step towards establishing linkage map. *Journal of Horticultural Science*, v. 10, n. 2, p. 130-135.
- Padmakar, B.; Kanupriya, C.; Latha, P. M.; Vasugi, C.; Dinesh, M. R.; Sailaja, D.; Aswath, C. (2016) Enrichment of Genetic Linkage Maps and Mapping QTLs Specific to Seed Strength-Hardness/Softness-In Guava (*Psidium guajava* L.). *Journal of Horticultural Science*, 11(1), 13-20.
- Paiva, C. L.; Viana, A. P.; Santos, Eileen A.; Freitas, J. C. O.; Amaral Júnior, A. T. (2016) Genetic gain estimated by different selection criteria in guava progenies. *Bragantia* (São Paulo, SP. Impresso), v. 75, p. 418-427.
- Park, T.; Casella, G. (2008) The Bayesian LASSO. *Journal of the American Statistical Association*, New York, v. 103, n. 482, p. 681-686.
- Pedrozo, C. A.; Tononi, H.; Resende, M. D. V.; Jordão, S. M. F. (2015) Repeatability of fruits and seeds production and selection of brazil nut genotypes in native populations in Roraima. *Revista Árvore*, v.39, n.5, p.863-871.
- Pena, S. D. (2006) Tomas Bayes “O cara”. *Ciência Hoje*. 28. N° 228.
- Pereira, F.M. (1995) *Cultura da goiabeira*. Jaboticabal: Universidade Estadual Paulista. UNESP, FUNEP, 47p.
- Pereira, F. M.; Natchtigal, J. C. (2002) Goiabeira. In: *Melhoramento de Fruteiras Tropicais*, ed. Viçosa, UFV. p267-289.
- Pereira, H. S.; Santos, J. B.; Souza, T. P.; Lima, I. A. (2008) Seleção fenotípica e assistida por marcadores moleculares de famílias de feijoeiro-comum com alta produtividade. *Pesquisa Agropecuária Brasileira*, Brasília, v. 43, p. 1551-1558.
- Pérez, P.; de los Campos, G. (2014) Genome-wide regression & prediction with the BGLR statistical package. *Genetics*, p. genetics. 114.164442

- Pessanha, P. G. D. O.; Viana, A. P.; Amaral Júnior, A. T.; Souza, R. M.; Teixeira M. C.; Pereira, M. G. (2011) Avaliação da Diversidade Genética em Acessos de *Psidium* spp. via marcadores RAPD. Revista Brasileira de Fruticultura 33(1): 129-136.
- Plummer, M.; Best, N.; Cowles, K.; Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. R news 6(1): 7-11.
- Prakash, D. P.; Narayanaswamy, P.; Suresh, N.; Sondur, S. N. (2002) Analysis of molecular diversity in guava using RAPD markers. Journal of Horticultural Science and Biotechnology, v. 77, n 3, p. 287-293.
- Quintal, S. S. R.; Viana, A. P.; Campos, B.; Vivas, M.; Amaral Júnior, A. T. (2017) Selection via mixed models in segregating guava families based on yield and quality traits. Revista Brasileira de Fruticultura 39(2): e-866.
- R Core Team (2018). R: A Language and Environment for Statistical Computing. Viena, Austria, ISBN 3-900051-07-0: R Foundation for Statistical Computing.
- Raftery, A. E.; Lewis, S. (1992). How many iterations in the Gibbs sampler. Bayesian statistics, 4(2), 763-773.
- Resende, M. D. V. (2002). Genética biométrica e estatística no melhoramento de plantas perenes, Embrapa Florestas.
- Resende, R. M. S.; Simões, R. M.; Fernandes, J. S. C. (2000) Blup individual com medidas repetidas aplicado ao melhoramento de espécies perenes. Ciência Agrotécnica, Lavras. v.24, p.986-997.
- Resende, M. D.; Thompson, R.; Welhan. S. U. E. (2006) Multivariate spatial statistical analysis of longitudinal data in perennial crops. Mat. Estat., São Paulo, v.24, n.1, p.147-169.
- Resende, R. T.; Resende, M. D. V.; Silva, F. E. F.; Takahashi, E. K. (2017) Predictive accuracy of *Eucalyptus* spp. Clonal trials using additive kinship effects and crossvalidation Scientia Forestalis/Forest Sciences, Vol.45(113), p 39-47.

- Rex, B.; Yu, J. (2007) Prospects for genomewidw selection for quantitative traits in maize. *Crop Science*. Vol. 47, no. 3, p. 1082-1090, doi: 10.2135/cropsci2006.11.0690.
- Risterucci, A. M.; Duval, M. F.; Rohde, W.; Billotte, N. (2005) Isolation and characterization of microsatellite loci from *Psidium guajava* L. *Molecular Ecology Notes*, n. 5, p. 745-748.
- Rodríguez, N. N.; Valdès-Infante, J.; Becker, D.; Velásques, B.; Coto, O.; Rohde, W.; Ritter, E. (2004) Caratterizzazione morfologica, agronômica e molecolare delle accessioni cubane di guaiava (*Psidium guaiava* L). *Journal of Genetics and Breeding, Italy*, v. 58, n. 1, p. 79-90.
- Rodríguez, N.; Valdés-Infante, J.; Becker, D.; Velázquez, B.; González, G.; Sourd, D.; Rodríguez, J.; Billotte, N.; Risterucci, A.M.; Ritter, E.; Rohde, W. (2007) Characterization of guava accessions by SSR markers, extension of the molecular linkage map, and mapping of QTLs for vegetative and reproductive characters. *Acta Horticulturae*, v. 735, p. 201-216.
- Rodríguez, N. N.; J. Valdés, J. A.; Rodríguez, J. B.; Velásquez, D.; Rivero, F.; Martínez, G.; González, D. G.; Sourd, L.; González J. C. (2010). Genetic resources and breeding of guava (*Psidium guajava* L.) in Cuba. *Biotecnología Aplicada* 27(3): 238-240.
- Rocha, M. De M.; Freire Filho, F. R.; Ribeiro, V. Q.; Carvalho, H. W. L. De; Belarmeno Filho, J.; Raposo, J. A. A.; Alcântara, J. Dos R; Ramos, S. R. R.; Machado. C. De F. (2007) Adaptabilidade e estabilidade produtiva de genótipos de feijão-caupi de porte semiereto na Região Nordeste do Brasil. *Pesquisa Agropecuária Brasileira*, v.42, p. 1283-1289.
- Rosa, G. J. M.; Rocha, L. B.; Furlan, L. R. (2007) Estudos de expressão gênica utilizando microarrays: delineamento, análise e aplicações na pesquisa zootécnica. *R. Bras. Zootec.* vol.36, doi.org/10.1590/S1516-35982007001000018.
- Rueda, A., Palacio, J. D., Munoz, J. E., Saavedra, R., Baravo, E. (2006) Caracterizacion molecular dei banco de germoplasma de guayaba *Psidium*

spp Del centro de investigacion corpoica-palmira. Fitotecnia Colombiana, v. 6, n.2, p. 26-32.

Sandoval, V. J. C.; Silva, F. F.; Resende, M. D. V.; Macedo, L. R.; Cecon, P. R. (2017). Bayesian random regression for genetic evaluation of South American Leaf Blight in rubber trees. *Revista Ciência Agronômica* 48(1): p151-156.

Santos, R. R.; Martins, F. P.; Ribeiro, I. J. A.; Nascimento, L. M.; Igue, T. (1998) Avaliação de variedades de goiabeira em Monte Alegre do Sul- SP. *Bragantia*. v. 57, n. 1, p. 117-126.

Santos, N. P. S.; Guimarães, F. F.; Sarmiento, J. L. R.; Sousa Júnior, A.; Rego Neto, A. A.; Sena, L. S.; Santos, G. V. (2016) Estrutura de covariância para características de carcaça e tamanho corporal com medidas repetidas em ovinos de diferentes grupos genéticos. *Rev. bras. saúde prod. anim.* vol.17 no. 4. <http://dx.doi.org/10.1590/s1519-99402016000400009>.

Santos, R. M. (2018) Caracterização e estruturação genética de populações segregantes de *Psidium* spp. resistentes ao *Meloidogyne enterolobii* via marcadores microssatélites. Tese (Mestre em Genética e Melhoramento de Plantas) – Campos dos Goytacazes – RJ, Universidade Estadual do Norte Fluminense – UENF, 51p

SAS Institute (2010). SAS/STAT 9.22 User's Guide: The PROC MIX Procedure first.

Schuster, I. (2011) Marker-assisted selection for quantitative traits. *Crop Breeding and Applied Biotechnology*, v. 1, p. 50-55.

Schwarz, G. (1978) Estimating the dimensional of a model. *Annals of Statistics*, Hayward, v.6, n.2, p.461-464.

Silva, C. A.; Detoni, J. L.; Costa, P. R.; Schmilth, O.; Alexandre, R. S.; Schildt, E. R. (2015) Estimativa de repetibilidade em características de cajá-mirim no Norte do Espírito Santo. *Agro@mbiente On-line*, v. 9, p. 284-291.

Silva, H. T.; Marques, D. B.; Lopes, D. P. S.; Paiva, J. T.; Veroneze, R.; Camargo, E. G.; Figueiredo E. A. P.; Silva, F. F. (2019) Alternative count Bayesian

models for genetic evaluation of litter traits in pigs. *Livestock Science* 225: 140-143.

Singer, J. M.; Andrade, D. F. (2000) Analysis of Longitudinal Data. *Handbook of statistics*, 18, 115-160.

Soares, B. C.; Santos, P. H. R.; Silva, F. H. L.; Esteves, E. A.; Titon, M.; Fernandes, J. S. C. (2017) Repetibilidade de características físicas e químicas em frutos de pequi. *Rev. Bras. Frutic.* vol.39 n. 2. DOI: <http://dx.doi.org/10.1590/0100-29452017181>

Sorensen, D. (2009) Developments in statistical analysis in quantitative genetics. *Genetica* 136(2): 319-332.

Sorensen, D.; Gianola, D. (2007) Likelihood, Bayesian, and MCMC methods in quantitative genetics, Springer Science & Business Media.

Soubihe Sobrinho, J. (1951) Estudos básicos para o melhoramento da goiabeira (*Psidium guajava* L.). São Paulo: ESALQ. 166p. Tese de Doutorado.

Spiegelhalter, D. J.; Best, N. G.; Carlin, B. P.; van der Linde, A. (2002) Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B* 64, 583–639.

Teodoro, P. E.; Nascimento, M.; Torres, F. E.; Barroso, L. M. A.; Sagrilo, E. (2015) Perspectiva bayesiana na seleção de genótipos de feijão-caupi em ensaios de valor de cultivo e uso. *Pesquisa Agropecuária Brasileira*, v.50, n.10, p.878-885.

Teyer, L. F. S.; Morales, A. B.; Keb, L.; Barredo, F.; Moreno, A. Q.; Sancliez, A. O.; Ramirez, J. S. P. (2010) Assessment of genetic diversity of mexican guava germplasm using DNA molecular markers. *Acta Horticulturae*, v. 849, p. 133-138.

Thaipong, K.; Boonprakob, U. (2005) Genetic and environmental variance components in guava fruit qualities. *Scientia Horticulturae*, v. 104, n. 1, p. 37-47.

- Thompson, R. (1969). Iterative estimation of variance components for non-orthogonal data. *Biometrics* 25: 767-773.
- Thompson, R.; Meyer, K. (1986). A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livestock Production Science* 15(4): 299-313.
- Tibshirani, R. (1996) Regression shrinkage and selection via the LASSO. *J. Royal. Statist. Soc. B.* 58:267–288.
- Toral, F. L. B.; Alencar, M. M.; Freitas, A. R. (2009) Estruturas de variância residual para estimação de funções de covariância para o peso de bovinos da raça Canchim. *R. Bras. Zootec.*, v.38, n.11, p.2152-2160.
- Yates. F.; Cochran, W.G. (1938) The analysis of group of experiments. *The Journal of Agricultural Science.* v.28, p.556-580.
- Wang, C. S.; Rutledge, J. J.; Gianola, D. (1994) Bayesian analysis of mixed linear models via Gibbs sampling with na application to litter size in Iberian pigs. *Genetics Selection Evolution.* v.26, p. 91.
- Wilberg, M. J.; Bence, J. R. (2008). Performance of deviance information criterion model selection in statistical catch-at-age analysis. *Fisheries Research* 93, 212–221.
- Wright, D. R.; Stern, H. S.; Berger, J. (2000) Comparing traditional and Bayesian analyses of selection experiments in animal breeding. *J. Agric. Biol. Environ. Stat., Alexandri*, v. 5, n. 2, p. 240-256.
- Wolfinger, R. D. (1993) Covariance structure in general mixed models. *Communications in Statistics*, v.22B, p.1079-1106.
- Woyann, L. G.; Milioli, A. S.; Bozi, A. H.; Dalló, S. C.; Matei, G.; Storck, L.; Benin, G. (2018) Repeatability of associations between analytical methods of adaptability, stability, and productivity in soybean. *Pesq. agropec. bras.*, Brasília, v.53, n.1, p.63-73. DOI: 10.1590/S0100-204X2018000100007.

- Valdés-Infante, J.; Rodríguez, N. N.; Becker, D.; Velázquez, B.; Sourd, D.; Espinosa, G.; Rohde, W. (2007) Microsatellite characterization of guava (*Psidium guajava* L.) germoplasm collection in Cuba. *Cultivos Tropicales*, v. 28, n. 3, p. 61-67.
- Van Eeuwijk, F. A.; Bustos-Korts, D.; Millet, E. J.; Boer, M. P.; Kruijer, W.; Thompson, A.; Malosetti, M.; Iwata, H.; Quiroz, R.; Kuppe, C. (2019) Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Science* 282: 23-39.
- Viana, A. P.; Resende, M. D. V. (2014) *Genética quantitativa no melhoramento de fruteiras*. 1. ed. Rio de Janeiro, RJ: Editora Interciência, p282.
- Viana, A. P.; Silva, F. H. L.; Gloria, L. S.; Ribeiro, R. M.; Krause, W.; Boechat, M. S. B. (2017) Implementing genomic selection in sour passion fruit population. *Euphytica*, v. 213, n.10, p.228.doi.org/10.1007/s10681-017-2020-3.
- Zhao, T.; Wang, Z.; Cumberworth, A.; Gsponer, J.; Freitas, N.; Bouchard-Côté A. (2016). Bayesian analysis of continuous time Markov chains with application to phylogenetic modelling. *Bayesian Analysis* 11(4): 1203-1237.

ANEXO

Tabela 1. Sequência dos 44 pares de iniciadores microssatélites utilizados na análise de 96 indivíduos de uma população de *Psidium guajava* L.

Locus	Sequências (5'3')	TA	NA	Referência
mPgCIR027	F: AGCACTTAGGGACAAATTCA R: CTCACTCTCCTCCATTCAAG	54°C	2	GuavaMap 2008
mPgCIR029	qF: CTCGCTTCAATCTCCATCTA R: AGCGACACAGACTCTTCATT	54°C	2	GuavaMap 2008
mPgCIR030	F: CTCAAAGCACTATCATGTCTG R: CCTTGTGGGTTCTCTTTTG	54°C	2	GuavaMap 2008
mPgCIR035	F: TTGACCTGGCATTAAACAGA R: CATTGGGAAAGGGAAGAA	54°C	2	GuavaMap 2008
mPgCIR037	F: GCGACGTTGTTGACTGAT R: AGTGCGATAAAGACATTAC	54°C	2	GuavaMap 2008
mPgCIR040	F: TGAATCTCCAGTGTCTTATCG R: TGATTTCAACTGCGTATGTC	53°C	2	GuavaMap 2008
mPgCIR042	F: CTCACCCAAAATCTACACAAG R: AAGGGACTGGACGATGTT	52°C	2	GuavaMap 2008
mPgCIR094	F: CAACCTTCCCCTGATTATT R: CTAGCTTCTTCAGTGGGAAC	53°C	2	Risterucci et al. 2005
mPgCIR096	F: ACGCTGCAAACGATACTAAT R: AACTCACACGAGCACAGAG	54°C	2	GuavaMap 2008
mPgCIR099	F: TCAAAGTCCAAAACATCATGC R: GGGATGGAGTAAAGATGAAA	50°C	2	GuavaMap 2008
mPgCIR102	F: AATTGGTGTAGCATCTGGA R: GCCTACCATGAACAGAGAAA	53°C	2	GuavaMap 2008
mPgCIR106	F: GACTCGACAGGAAGGTCTC R: CAGCTTTGTATATCGCAAGA	54°C	2	GuavaMap 2008
mPgCIR111	F: CAACCTCGTTTGAGTCTTCT R: AACATCATTGGGACCATTCC	54°C	3	GuavaMap 2008
mPgCIR131	F: GAGGTTGAGAGTTCAAGGAT R: GGTTTGCTCTGAAATCACTC	53°C	3	GuavaMap 2008
mPgCIR135	F: CAGATAGCAAACACTGCCTCT R: ATATCCCTCTCGCCTTCTT	54°C	2	GuavaMap 2008
mPgCIR138	F: GGTGAGACCACTGAGTTCC R: GAAAGACCAATGAAATCGAC	54°C	2	GuavaMap 2008
mPgCIR140	F: GTGGTGAAGGAGTAAAGCTG R: GCAGTATAAAGCAACAGATGG	54°C	3	GuavaMap 2008
mPgCIR143	F: TCTGGATTTTCTCCAATGC R: GTATGCACCACCATCTGC	54°C	2	GuavaMap 2008
mPgCIR147	F: ACTGACATCTCTGACCATAGC R: GATTGCCATAGGAACTGAAA	54°C	3	GuavaMap 2008
mPgCIR149	F: CTTTCGTGGAAGAGGATGAC R: AATATAAACATCGCCACAGG	54°C	2	GuavaMap 2008
mPgCIR153	F: GCCTCTGGTAAATCTGTTGA R: ACATACGGATCAAGTCCAAA	54°C	2	GuavaMap 2008
mPgCIR160	F: TGGCTATAAGAATGGGAGAT R: GACGAGCTTAGCCTCTGAAT	54°C	2	GuavaMap 2008
mPgCIR161	F: TCTCAAGGACCAACAAGAAG R: AGGACTTAGCTTGGGTTTTTC	54°C	2	GuavaMap 2008
mPgCIR163	F: TCTTTGCACATCAAACCTCG R: CATGGTATCAATAGGTCAAGC	54°C	2	GuavaMap 2008
mPgCIR164	F: TCCCGTAAGTTGTTCTGTTT R: CGTAAGAGATCGTGAAGGAG	54°C	2	GuavaMap 2008
mPgCIR166	F: CTTTCCCATCAAACGTAAG R: CCAATTCATGCACTTAGACA	50°C	2	GuavaMap 2008
mPgCIR169	F: TTCAGGCAGATCGTGTTACT R: GTGCCTAACCTACACCCTAA	54°C	2	GuavaMap 2008

Tabela 1 – Cont.

Locus	Sequências (5'3')	TA	NA	Referência
mPgCIR180	F: CATGGATTCAACTCTTGTCG R: CTACATTGGAAGCAGAATGG	53°C	2	GuavaMap 2008
mPgCIR181	F: AGACACTACCTCACCGCTAT R: ATAGAGAAGCCGAAGGAAAC	54°C	2	GuavaMap 2008
mPgCIR184	F: AAGCTACAATCGACGAAAAC R: CACTATTAGCGAACCTGCAT	54°C	2	GuavaMap 2008
mPgCIR187	F: AAGCCTAATAATGCGGAAGT R: TCTTTCCAGGATGGAGTA	52°C	2	GuavaMap 2008
mPgCIR188	F: TGGATGAATCAGGAGGATTA R: TTGTGGGGAAGAACTACTG	51°C	2	GuavaMap 2008
mPgCIR191	F: GACCCTCCACTTATATTTTG R: AAGCTGACATAACAGTCGAA	53°C	2	GuavaMap 2008
mPgCIR193	F: GAACGTGGGTTACATACCAT R: ATCACCGTCCTCCTAAATCT	52°C	2	GuavaMap 2008
mPgCIR200	F: CCTTGCTTTGGTGAGGTC R: GCTAATTCAGTCCTTCCAAC	54°C	2	GuavaMap 2008
mPgCIR202	F: CCATTAGAGCCGACAAAA R: GACGAGAAACCCTAAACGTA	54°C	2	GuavaMap 2008
mPgCIR204	F: GTCGGATCATGGAGATCA R: GCGGCTAAAGAAATCTGG	53°C	2	GuavaMap 2008
mPgCIR207	F: CAAGATTTGCCTCAAGAAAC R: AACTAAATAGCCTGCTGGTG	53°C	2	GuavaMap 2008
mPgCIR209	F: CTAAAGCCACATCCAGCA R: CTAACATTTGCCTTCTACAGC	54°C	2	GuavaMap 2008
mPgCIR210	F: CTACGAGGGTGTAAACGAAAA R: CTACAACAGCCAACGTGAG	53°C	3	GuavaMap 2008
mPgCIR221	F: CTAAGCCTGAAGTCCCAAAT R: CCTCTTCTAAAGGCAACGAC	54°C	3	GuavaMap 2008
mPgCIR227	F: GGGATGCTCAAACTGTAAG R: CCTGTTACATTGACGAATCA	53°C	2	GuavaMap 2008
mPgCIR230	F: CACATTTGCTCCTGATTTTC R: GCTCTTCAACGACCATCTT	53°C	2	GuavaMap 2008
mPgCIR249	F: TTTGTCTGGTCGTCTAGTT R: CTTCAGTCCATCAGCAAAAT	52°C	2	GuavaMap 2008

TA= temperatura de anelamento, NA= número de alelos por loco.

Tabela 2. Valores genômicos preditos para todos os indivíduos para as variáveis teor de sólidos solúveis (TSS), massa de frutos (MF), massa de polpa (MP), número de frutos por planta (NF) e produção por planta (PROD).

Indivíduos	BRIX	MF	MP	NF	PROD
B2F2P7	-1,26	-9,43	-2,46	88,57	16936,59
B1F6P5	-0,45	25,06	26,52	36,83	7622,50
B2F17P6	-1,99	-14,71	-14,65	36,11	7453,87
B2F17P5	-0,04	-0,64	1,52	29,48	4906,31
B1F15P12	0,76	10,68	8,45	24,57	4507,96
B2F6P11	-0,15	11,59	11,37	23,72	4149,27
B1F10P11	-0,55	-2,55	-3,82	17,23	3857,22
B1F10P5	0,92	-6,00	-2,98	17,02	3715,07
B2F17P4	-0,22	-5,30	-3,87	21,84	3675,71
B2F8P4	0,19	6,54	4,12	13,11	3115,68
B1F10P8	-0,57	-6,15	-5,92	12,63	3049,19
B1F11P11	-0,27	-22,29	-19,53	21,88	3004,88
B2F1P4	0,64	4,99	12,08	14,53	2958,55
B2F1P5	-1,04	-29,24	-23,30	23,74	2881,99
B1F15P10	2,51	18,14	16,13	1,25	2848,35
B2F1P12	0,34	-12,94	-10,82	20,13	2679,10
B2F8P6	-0,45	32,36	26,36	11,21	2510,58
B1F2P8	1,21	33,01	29,67	13,28	2268,82
B2F2P10	0,86	12,18	12,33	16,93	2206,82
B2F15P2	0,60	-27,09	-25,10	8,39	2145,49
B2F10P3	-0,13	27,87	24,14	3,46	1953,60
B1F13P1	0,22	-17,26	-12,36	0,35	1599,54
B2F3P11	0,89	19,74	19,79	9,29	1421,55
B1F8P3	-0,45	-18,94	-19,16	10,39	1322,52
B1F8P12	-0,83	-2,97	-3,35	11,14	1290,42
B1F8P1	1,04	18,97	15,93	7,67	1264,91
B2F12P9	0,96	6,90	5,19	9,00	1245,41
B2F10P8	0,45	-1,37	-2,33	7,81	1211,01
B2F10P1	-0,37	-11,44	-9,95	-0,71	1053,72
B2F2P2	-0,93	-27,53	-25,76	8,80	965,44
B1F6P10	-0,27	-27,94	-24,25	-5,57	959,25
B1F13P4	0,77	26,74	25,09	-7,57	953,69
B1F8P2	-0,87	-6,62	-8,79	10,00	892,75
B2F12P10	0,64	-24,72	-19,60	5,73	834,45
B2F1P8	-0,70	-32,35	-29,06	5,57	802,53
B1F10P12	0,20	-4,74	-4,73	4,07	786,74
B2F17P7	0,47	18,03	15,39	-2,37	698,88
B2F17P2	0,27	7,61	4,86	6,15	622,86
B1F2P4	-1,09	-28,14	-27,96	6,40	552,34
B2F10P5	-0,37	4,71	2,18	9,97	519,98
B2F12P1	0,42	0,55	-1,22	0,76	487,86
B2F8P8	-0,32	-5,87	-5,97	5,62	448,45

Tabela 2 – Cont.

Indivíduos	BRIX	MF	MP	NF	PROD
B2F2P8	-1,04	14,73	14,74	-0,66	298,24
B2F10P7	-0,59	15,68	13,82	1,24	258,26
B1F10P7	0,16	29,27	25,38	-2,99	128,80
B2F15P11	1,32	-1,42	-3,75	-4,00	111,05
B1F2P5	0,00	2,23	2,77	-2,29	69,95
B1F10P1	-0,33	9,37	8,03	1,49	-53,06
B2F1P9	0,19	-31,34	-25,20	4,88	-96,14
B2F2P9	-0,78	-14,55	-14,55	1,03	-167,01
B2F8P12	-0,85	-5,52	-6,98	-4,07	-206,77
B2F17P8	-0,15	10,06	8,55	0,73	-288,78
B2F10P6	-0,97	-11,90	-11,08	-5,60	-350,53
B1F12P5	0,10	3,13	3,78	-3,32	-414,67
B2F12P12	0,92	8,02	10,51	1,86	-415,49
B2F6P10	0,94	-24,20	-16,95	-6,65	-471,75
B1F10P2	0,40	28,49	23,78	-9,03	-524,53
B1F8P10	-0,32	8,06	7,18	-1,98	-546,28
B2F2P5	-0,23	8,25	4,87	-4,35	-613,17
B2F11P10	-0,38	-12,29	-12,79	-0,98	-803,87
B2F8P7	-0,09	3,75	5,70	-1,69	-805,72
B2F12P11	0,30	12,43	11,52	-9,83	-817,18
B1F15P11	1,40	10,43	10,25	-3,18	-859,82
B1F6P11	-0,13	-16,41	-15,25	-3,11	-880,87
B2F6P3	0,30	-2,14	-4,09	-3,06	-968,72
B2F1P3	-0,45	-28,13	-23,80	-1,62	-1036,02
B2F6P6	0,78	-23,01	-19,05	-0,02	-1094,85
B2F6P2	0,48	-0,47	1,92	-6,80	-1419,28
B2F13P8	-0,25	17,34	16,14	-8,43	-1448,17
B1F12P6	0,51	-4,83	-3,68	-8,39	-1604,44
B2F1P11	0,13	5,74	6,60	-2,41	-1641,97
B2F8P5	-0,47	-6,37	-6,15	-6,44	-1748,99
B2F8P10	-0,17	3,32	0,59	-13,10	-1871,62
B2F12P8	-0,80	-6,13	-6,45	-8,89	-1958,88
B2F6P4	-0,18	-0,64	5,19	-4,88	-2360,19
B2F8P1	-0,61	-0,68	-4,93	-12,88	-2563,43
B2F8P9	-0,19	3,68	2,98	-10,49	-2623,56
B2F1P10	0,96	-7,23	-10,37	-14,75	-2720,17
B1F2P9	1,84	41,77	38,52	-25,45	-2756,83
B2F2P12	-0,13	6,43	1,27	-23,06	-2902,44
B2F8P3	-0,88	13,15	8,95	-6,36	-2946,93
B2F6P12	-0,81	-0,64	-3,69	-17,22	-2998,35
B2F13P2	0,47	1,95	3,49	-15,43	-3006,53
B1F6P8	-0,03	-12,70	-10,69	-14,38	-3060,24
B1F2P10	0,20	-9,76	-8,56	-23,15	-3490,28
B1F12P4	-0,19	-6,74	-8,37	-23,59	-3658,66

Tabela 2 – Cont.

Indivíduos	BRIX	MF	MP	NF	PROD
B1F12P3	0,24	-16,73	-15,45	-22,82	-3976,89
B1F1P11	-0,33	-8,54	-6,83	-20,95	-4278,83
B2F13P3	0,13	11,90	6,61	-28,51	-4390,70
B2F8P11	-0,39	25,82	19,83	-22,85	-4591,40
B1F10P10	-0,45	13,49	11,78	-23,46	-4760,94
B2F10P4	-1,14	10,42	6,88	-21,32	-4801,80
B1F1P7	0,07	-12,10	-9,07	-24,56	-5198,11
B1F7P9	0,41	7,15	8,76	-33,14	-5353,08
B1F1P6	-0,42	-2,45	0,51	-26,15	-6779,38
B1F1P2	0,48	1,40	2,65	-31,33	-6920,83