

PROSPECÇÃO DA VARIABILIDADE GENÉTICA E PREDIÇÃO
GENÔMICA VIA MODELOS MISTOS EM POPULAÇÃO DE IRMÃOS
COMPLETOS DE *Psidium guajava* L.

FREDERICO CÉSAR RIBEIRO MARQUES

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY
RIBEIRO

CAMPOS DOS GOYTACAZES – RJ

JUNHO – 2025

PROSPECÇÃO DA VARIABILIDADE GENÉTICA E PREDIÇÃO
GENÔMICA VIA MODELOS MISTOS EM POPULAÇÃO DE IRMÃOS
COMPLETOS DE *Psidium guajava* L.

FREDERICO CÉSAR RIBEIRO MARQUES

“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 Produção Vegetal”

Orientador: Prof. D. Sc., Alexandre Pio Viana

CAMPOS DOS GOYTACAZES – RJ

JUNHO - 2025

PROSPECÇÃO DA VARIABILIDADE GENÉTICA E PREDIÇÃO
GENÔMICA VIA MODELOS MISTOS EM POPULAÇÃO DE IRMÃOS
COMPLETOS DE *Psidium guajava* L.

FREDERICO CÉSAR RIBEIRO MARQUES

“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 Produção Vegetal”

Aprovada em 27 de junho de 2025.

Comissão Examinadora

Documento assinado digitalmente
 **EILEEN AZEVEDO SANTOS**
Data: 24/09/2025 12:39:38-0300
Verifique em <https://validar.iti.gov.br>

Eileen Azevedo Santos (D.Sc., Genética e Melhoramento de Plantas) - UNEMAT

Flavia Alves da Silva (D.Sc., Genética e Melhoramento de Plantas) - UENF

Antonio Teixeira do Amaral Junior (D.Sc., Genética e Melhoramento de Plantas) –
UENF

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

FICHA CATALOGRÁFICA

UENF - Bibliotecas

Elaborada com os dados fornecidos pelo autor.

M357

Marques, Frederico Cesar Ribeiro.

PROSPECÇÃO DA VARIABILIDADE GENÉTICA E PREDIÇÃO GENÔMICA VIA MODELOS MISTOS EM POPULAÇÃO DE IRMÃOS COMPLETOS DE *PSIDIUM GUAJAVA* L. / Frederico Cesar Ribeiro Marques. - Campos dos Goytacazes, RJ, 2025.

80 f. : il.

Inclui bibliografia.

Tese (Doutorado em Produção Vegetal) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Ciências e Tecnologias Agropecuárias, 2025.

Orientador: Alexandre Pio Viana.

1. Ganho de seleção. 2. Melhoramento de frutíferas. 3. Predição de valores genéticos. 4. Rank-Biased Overlap. I. Universidade Estadual do Norte Fluminense Darcy Ribeiro. II. Título.

CDD - 630

Dedico à minha esposa Cibeles, à
minha filha Sofia e aos meus pais,
Antônio e Ilda.

AGRADECIMENTOS

Agradeço primeiramente a Deus, fonte de toda sabedoria, por ter me sustentado em cada etapa desta jornada, iluminando meu caminho mesmo nos momentos mais incertos.

Àqueles que me deram a vida e me ensinaram a caminhar com dignidade, meus pais, Antônio e Ilda Marques, minha eterna gratidão.

À minha esposa Cibele Franco Costa Marques dedica um agradecimento especial e emocionado. Seu apoio firme durante minhas ausências, seu acolhimento nos momentos difíceis e sua confiança inabalável foram meu alicerce.

À minha filha, Sofia Costa Marques, que com sua doçura, alegria e pureza me confortou e renovou minhas esperanças em dias melhores.

Agradeço à minha irmã Taís, ao meu cunhado Tiago e aos meus sobrinhos Igor e Raquel.

À minha sogra, Sandra Yara, minha cunhada Mayara e ao meu sogro, Geraldo Majela (*in memoriam*), cuja presença, mesmo após a partida, permanece como inspiração.

À minha tia Conceição Marques, por sua presença constante em meus desafios intelectuais e nas vitórias mais significativas da minha vida; sua fé em mim nunca passou despercebida. Ao meu tio Geraldo Eustáquio, por cada gesto de orgulho silencioso que me incentivou a seguir firme, mesmo quando as forças pareciam faltar.

Ao Instituto Federal do Espírito Santo (IFES), pela oportunidade de formação proporcionada por meio do programa DINTER.

À Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), ao meu orientador, Professor Dr. Alexandre Pio Viana, meu reconhecimento pelo acolhimento, pela orientação e pelas condições oferecidas para a realização e conclusão deste projeto.

Agradeço à Flávia, ao Caio, ao Natan, à Mariana e à Thaís.

SUMÁRIO

1 INTRODUÇÃO GERAL	1
2 REVISÃO DE LITERATURA	4
2.1 A cultura da goiaba	4
2.2 Estudo da diversidade genética da goiabeira	6
2.3 O melhoramento genético da goiabeira	8
2.4 Marcadores moleculares do tipo SNP	10
2.5 Predição de valores genéticos genômicos	11
2.6 Melhoramento genético da goiabeira na UENF	14
3 OBJETIVOS	17
3.1 Geral	17
3.2 Específicos	17
4 TRABALHOS	18
4.1 Prospecting variability in full-sib families of <i>Psidium. guajava</i> L. through phenotypic information and SNP markers.....	18
INTRODUCTION	19
MATERIAL AND METHODS	21
RESULTS AND DISCUSSION	27
CONCLUSION.....	35
REFERENCES	35

4.2 Modelagem mista via marcadores SNPs na implementação da seleção genômica em uma população de goiabeiras	40
RESUMO	40
INTRODUÇÃO	41
MATERIAL E MÉTODOS	43
RESULTADOS E DISCUSSÃO	48
CONCLUSÕES	57
BIBLIOGRAFIA.....	57
5 CONCLUSÃO	61
REFERÊNCIAS BIBLIOGRÁFICAS	63

RESUMO

MARQUES, Frederico César Ribeiro; Universidade Estadual do Norte Fluminense Darcy Ribeiro; junho de 2025; Prospecção da variabilidade genética e predição genômica via modelos mistos em população de irmãos completos de *Psidium Guajava* L.; Orientador: D. Sc., Alexandre Pio Viana.

A goiabeira (*Psidium guajava* L.) é uma frutífera tropical amplamente cultivada no Brasil e em outros países tropicais, com crescente importância econômica e nutricional. O objetivo desta tese foi avaliar a diversidade genética e a predição de valores genéticos genômicos em uma população composta por 247 indivíduos de 11 famílias de irmãos completos, oriundas de cruzamentos controlados realizados no âmbito do programa de melhoramento da goiabeira da Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF). O primeiro estudo visou caracterizar a diversidade genética da população. Foram avaliados o peso, o comprimento e o diâmetro do fruto, o peso da polpa e da placenta, a espessura do endocarpo, do mesocarpo e da casca da goiaba, teor de sólidos solúveis além da produção total das plantas. A genotipagem foi realizada por meio da tecnologia DArTseq, resultando em 7.874 SNPs após filtragem por qualidade. A divergência fenotípica entre famílias foi estimada por meio da distância generalizada de Mahalanobis dada a sua capacidade de considerar a correlação entre as variáveis e as diferenças nas suas escalas de medida, tornando a comparação entre indivíduos ou tratamentos mais precisa e estatisticamente consistente em contextos multivariados. A contribuição relativa de cada característica quantificada pela estatística de Singh. A distância genética entre os indivíduos foi calculada por meio do complemento aritmético do índice de similaridade não ponderado, refletindo o grau de dissimilaridade genética entre os pares avaliados. A estrutura de agrupamento foi determinada pelos métodos UPGMA e Ward, aplicados separadamente às distâncias obtidas a partir dos dados fenotípicos e moleculares. Com base nos dados fenotípicos, os agrupamentos revelaram a formação de três grupos distintos, enquanto os dados moleculares permitiram a formação de até cinco grupos, indicando maior detalhamento da estrutura genética capturada pelos marcadores

SNPs. A análise de variância molecular (AMOVA de Excoffier) indicou que as famílias não são diferenciáveis pelos dados fenotípicos ou moleculares apresentados, entretanto, existe variabilidade significativa entre os grupos formados pelas metodologias de agrupamento utilizadas. O segundo estudo teve como objetivo comparar o modelo misto BLUP tradicional (baseado em matriz de parentesco médio), o RRBLUP e o GBLUP (ambos baseados em informações moleculares) na predição de valores genéticos para as mesmas características. Foram estimados os parâmetros de herdabilidade, viés de predição, erro quadrático médio (MSE), capacidade preditiva e a similaridade entre os rankings dos modelos por meio do índice *Rank-Biased Overlap* (RBO). O modelo BLUP tradicional apresentou os melhores resultados em todas as características avaliadas, com maiores capacidades preditivas, menores MSEs e viés mais próximo de zero. Os modelos RRBLUP e GBLUP não foram ajustados para espessura do mesocarpo, massa da polpa e produção total, devido às estimativas de herdabilidade próximas de zero. Por outro lado, RRBLUP apresentou as maiores herdabilidades para características como massa, comprimento, diâmetro do fruto, espessura do endocarpo e massa da placenta. A semelhança entre os rankings gerados por RRBLUP e GBLUP foi alta para algumas variáveis (RBO > 0,9), enquanto a concordância entre BLUP e os modelos genômicos foi baixa para quase todas as características. Os resultados demonstram que, para a população estudada onde verificou-se reduzida variabilidade genética, o uso de modelos tradicionais baseados na matriz de parentesco médio foi mais eficiente do que modelos genômicos, ressaltando a importância de considerar a estrutura da população na escolha da metodologia preditiva.

PALAVRAS-CHAVE: Análise multivariada; Marcadores moleculares; Melhoramento de fruteiras tropicais; Modelagem mista; Polimorfismo de nucleotídeo único.

ABSTRACT

MARQUES, Frederico César Ribeiro; Universidade Estadual do Norte Fluminense Darcy Ribeiro; June 2025; Prospecting genetic variability and genomic prediction via mixed models in a full-sib population of *Psidium Guajava* L.; Advisor: Ph.D., Alexandre Pio Viana.

The guava tree (*Psidium guajava* L.) is a tropical fruit crop widely cultivated in Brazil and other tropical countries, with growing economic and nutritional importance. The objective of this thesis was to evaluate the genetic diversity and the prediction of genomic breeding values in a population composed of 247 individuals from 11 full-sib families, derived from controlled crosses carried out within the guava breeding program of the Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF). The first study aimed to characterize the genetic diversity of the population. Fruit weight, length, and diameter, pulp and placenta weight, endocarp, mesocarp, and peel thickness, soluble solids content, and total plant yield were evaluated. Genotyping was performed using DArTseq technology, resulting in 7,874 SNPs after quality filtering. Phenotypic divergence among families was estimated using the generalized Mahalanobis distance due to its ability to account for correlations among variables and differences in measurement scales, making comparisons between individuals or treatments more accurate and statistically consistent in multivariate contexts. The relative contribution of each trait was quantified by Singh's statistic. Genetic distance between individuals was calculated using the arithmetic complement of the unweighted similarity index, reflecting the degree of genetic dissimilarity between the evaluated pairs. The clustering structure was determined by the UPGMA and Ward methods, applied separately to distances obtained from phenotypic and molecular data. Based on phenotypic data, the clusters revealed the formation of three distinct groups, whereas molecular data allowed the formation of up to five groups, indicating a greater level of detail in the genetic structure captured

by SNP markers. The molecular analysis of variance (Excoffier's AMOVA) indicated that families are not differentiable by the phenotypic or molecular data presented; however, there is significant variability among the groups formed by the clustering methodologies used. The second study aimed to compare the traditional BLUP mixed model (based on an average pedigree matrix), RRBLUP, and GBLUP (both based on molecular information) in predicting breeding values for the same traits. Heritability parameters, prediction bias, mean squared error (MSE), predictive ability, and similarity between the model rankings were estimated using the Rank-Biased Overlap (RBO) index. The traditional BLUP model showed the best results for all traits evaluated, with higher predictive abilities, lower MSEs, and bias closer to zero. The RRBLUP and GBLUP models were not fitted for mesocarp thickness, pulp weight, and total yield due to heritability estimates close to zero. On the other hand, RRBLUP showed the highest heritability estimates for traits such as fruit weight, length, diameter, endocarp thickness, and placenta weight. Similarity between the rankings generated by RRBLUP and GBLUP was high for some variables (RBO > 0.9), whereas agreement between BLUP and the genomic models was low for almost all traits. The results demonstrate that, for the studied population where reduced genetic variability was observed the use of traditional models based on the average pedigree matrix was more efficient than genomic models, highlighting the importance of considering population structure when choosing the predictive methodology.

KEYWORDS: Mixed modeling; Molecular markers; Multivariate analysis; Single nucleotide polymorphism; Tropical fruit breeding.

1 INTRODUÇÃO GERAL

A goiabeira (*Psidium guajava* L.), frutífera perene e tropical originária entre o México e o noroeste da Argentina (Diaz-Garcia et al., 2023), teve sua domesticação iniciada na Amazônia brasileira, expandindo-se pelos Andes, América Central e México, alcançando diversas regiões tropicais graças à ação de exploradores europeus e à sua adaptabilidade às diferentes condições edafoclimáticas (Kumari et al., 2018; Angulo-López et al., 2021; Arévalo-Marín et al., 2024). Rica em fibras, ácido ascórbico e compostos fenólicos, destaca-se pelo valor nutritivo e versatilidade, sendo consumida *in natura* ou processada em sucos, doces, polpas e purês, enquanto folhas e outros órgãos são empregados na medicina tradicional por suas propriedades antimicrobianas, hepatoprotetoras, antidiabéticas e anti-inflamatórias (Menezes et al., 2021; Chu et al., 2022).

No Brasil, a goiabeira se destaca com sua presença expressiva nas regiões Sudeste e Nordeste, concentrando a maior parte da produção nacional. Segundo os dados do Instituto Brasileiro de Geografia e Estatística (IBGE), referentes ao ano de 2023, a produção nacional de goiaba alcançou, aproximadamente, 582.832 toneladas, com valor de produção estimado em 1,306 bilhão de reais (IBGE, 2023). Esses números refletem não apenas a importância econômica crescente da cultura no país, mas também o avanço de tecnologias e práticas agrícolas que contribuem para o aumento da produtividade.

A melhoria das qualidades agronômicas da goiaba, como doçura, cor, consistência da polpa, aparência, espessura da casca e valor nutricional, é o

principal objetivo de diversos programas de melhoramento genético ao redor do mundo (Maan et al., 2023). No Brasil, a Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF) desenvolve, desde 2007, um programa de melhoramento genético da goiabeira com o intuito de obter variedades superiores e mais resistentes (Pessanha et al., 2011).

Dentre as diversas etapas de um programa de melhoramento genético, o estudo da diversidade genética e a seleção de indivíduos com maior potencial de transmitir genes favoráveis às características de interesse assumem papel fundamental. Quando se utiliza exclusivamente informações fenotípicas, a caracterização da diversidade genética por meio de análises multivariadas, bem como as estimativas de BLUPs (*Best Linear Unbiased Predictions*) obtidas a partir de matrizes de parentesco médio, configuram metodologias amplamente empregadas, com resultados consistentes (Maan et al., 2023).

O desenvolvimento de marcadores moleculares representou um avanço significativo no melhoramento genético tradicional, ao aumentar a acurácia dos modelos preditivos e possibilitar maior precocidade nas seleções. Dentre os marcadores moleculares, os SNPs (*Single Nucleotide Polymorphisms*) destacam-se por sua alta resolução, codominância, densidade genômica e compatibilidade com plataformas de sequenciamento de nova geração (NGS), como a tecnologia DArTseq (Grossi et al., 2021; Koura et al., 2024). Essas características permitem sua aplicação em estudos de diversidade, estrutura populacional, associação genômica ampla (GWAS) e seleção genômica (GW).

A seleção genômica (GW) emprega dados genotípicos de milhares de SNPs para prever o valor genético dos indivíduos. Para tanto, utilizam-se métodos estatísticos como o RRBLUP (*Ridge Regression BLUP*) e o GBLUP (*Genomic BLUP*), ambos com eficácia comprovada na seleção precoce de genótipos superiores (Das et al., 2022).

O estudo da diversidade genética é fundamental para apoiar programas de melhoramento e estratégias de conservação. Nesse contexto, o desenvolvimento e a aplicação de marcadores, tanto de baixa densidade como é caso dos marcadores RAPD ou SSR, quanto de alta densidade, como os SNPs, representam um avanço

significativo, permitindo análises mais precisas da estrutura populacional, da variação genética e da identificação de regiões genômicas associadas a caracteres de interesse agrônomo (Diaz-Garcia et al., 2023).

Apesar dos avanços, ainda são escassos os estudos no Brasil que avaliem a diversidade genética da goiabeira com base em dados moleculares de alta resolução, especialmente em populações com alto grau de parentesco, como as de irmãos completos (Diaz-Garcia et al., 2023). Neste trabalho, foram analisadas 11 famílias de irmãos completos de goiabeira, oriundas de ciclos sucessivos de avaliação e seleção em populações constituídas a partir de acessos coletados em pomares comerciais. A investigação da diversidade genética teve como objetivo identificar os efeitos dos ciclos de seleção e dos cruzamentos sobre a variabilidade genética remanescente. Complementarmente, a utilização de modelos de seleção baseados em matrizes de parentesco médio, e em informações oriundas de marcadores moleculares, busca avaliar a influência dessas abordagens na acurácia e na eficiência da seleção da população em estudo. O estudo foi realizado em colaboração entre a UENF e a EMBRAPA Recursos Genéticos, sendo parte da continuidade do programa de melhoramento genético da goiabeira iniciado na UENF.

2 REVISÃO DE LITERATURA

2.1 A cultura da goiaba

A goiabeira (*Psidium guajava* L.), da família Myrtaceae, é uma frutífera tropical amplamente cultivada nas regiões tropicais e subtropicais do mundo. Sua origem é atribuída ao norte da América do Sul (Landau et al., 2010). A dispersão para outras regiões tropicais foi intensificada com o movimento dos navegantes europeus, que levaram mudas para as colônias africanas e asiáticas. A goiaba é uma das frutas que coloca o Brasil em uma posição dentre os maiores produtores de frutas no mundo, juntamente com a Índia, o Paquistão e o México (Pereira et al., 2017).

O gênero *Psidium* que compreende cerca de 150 espécies de arbustos e pequenas árvores. Dentre elas, a *P. guajava* L. é a mais cultivada e amplamente distribuída. A família Myrtaceae possui mais de 3.000 espécies, predominantemente nos trópicos da América, Ásia e Oceania (Pereira et al., 2017).

A cultura da goiaba se adapta a uma ampla faixa de altitudes, desde o nível do mar até 1.700 m. As condições ideais de cultivo envolvem temperaturas de 25 a 28 °C, precipitações entre 800 mm e 2.000 mm anuais bem distribuídas, solos areno-argilosos profundos, bem drenados e ricos em matéria orgânica, com pH entre 5,0 e 6,5 (Barbosa et al., 2010; Pereira et al., 2017).

É uma planta de hábito perene, com copa geralmente baixa e ramificada, podendo atingir entre dois e sete metros de altura, dependendo da variedade e do sistema de manejo adotado (Barbosa et al., 2010). As folhas são simples, opostas, com margens inteiras, de coloração verde intensa e superfície pilosa na fase jovem. O sistema radicular da planta, embora superficial em sua maior parte, apresenta capacidade de explorar camadas mais profundas do solo, o que contribui para sua rusticidade e tolerância às condições de déficit hídrico moderado.

Apresenta alta eficiência fotossintética e boa resposta ao manejo da irrigação e adubação. A cultura possui comportamento adaptativo às diferentes condições ambientais, sendo capaz de manter atividade vegetativa e produtiva mesmo sob regimes climáticos variados (Barbosa et al., 2010).

As flores da goiabeira são hermafroditas, com coloração branca e organizadas isoladamente ou em pequenos grupos. A estrutura floral é composta por cinco pétalas e um número elevado de estames. A polinização é predominantemente entomófila, sendo a abelha *Apis mellifera* o principal agente polinizador; no entanto, a autofecundação também é observada com menor frequência.

A frutificação pode iniciar-se de sete a oito meses após o plantio, especialmente quando se utilizam mudas enxertadas. Os frutos são do tipo baga, com grande variabilidade em tamanho, formato, coloração da casca e da polpa, número de sementes e teor de sólidos solúveis. A casca pode variar do verde-claro ao amarelo-brilhante, enquanto a polpa pode ser branca, rosada, vermelha ou amarela, com sabor adocicado e aroma característico. O desenvolvimento do fruto segue um padrão sigmoide de crescimento, compreendendo fases de divisão celular, alongamento e maturação. A frutificação pode ocorrer de forma contínua ou ser induzida por práticas de poda e manejo hídrico (Barbosa et al., 2010).

A propagação da goiabeira pode ser sexuada, por sementes, ou vegetativa, por estaquia, enxertia ou cultura de tecidos. A enxertia por garfagem ou borbúlia sobre porta-enxertos provenientes de sementes é o método mais utilizado comercialmente, conferindo uniformidade aos pomares (Barbosa et al., 2010; Pereira et al., 2017).

A estaquia herbácea tem sido empregada com sucesso, utilizando ramos jovens com duas gemas, em câmaras de nebulização. O uso de substratos inertes, como vermiculita, e a substituição do solo como substrato, visam evitar a disseminação de patógenos como o nematoide *Meloidogyne mayaguensis* (Barbosa et al., 2010).

As práticas de manejo incluem adubação, irrigação, podas, controle fitossanitário e consorciação. A adubação deve ser baseada em análises de solo e foliares. A calagem é indicada para elevação da saturação por bases acima de 50% (Pereira et al., 2017).

A irrigação é indispensável em regiões semiáridas, com valores de evapotranspiração entre 4,5 mm e 6,1 mm por dia. Sistemas como microaspersão e gotejamento são recomendados (Rai et al., 2010). A poda é realizada para formação, frutificação, renovação e controle da copa. Existem podas de formação, frutificação (drástica ou contínua), raleio e limpeza.

O controle fitossanitário deve ser integrado. Plantas daninhas devem ser controladas por métodos manuais, mecânicos ou químicos, respeitando o estágio das plantas e a seletividade dos herbicidas. Um dos principais desafios sanitários enfrentados pelos produtores atualmente é o chamado "declínio da goiabeira". Trata-se de uma síndrome de causa complexa e multifatorial, caracterizada por sintomas como amarelecimento foliar, seca de ramos e posterior morte da planta. O declínio da goiabeira tem causado perdas expressivas em regiões produtoras, principalmente onde há histórico de uso contínuo da área e baixa renovação de material genético (Barbosa et al., 2010).

Entre as cultivares mais utilizadas no Brasil estão: Paluma, Pedro Sato, Kumagai, Rica, Século XXI e Sassaoka. Cada uma das cultivares apresentam características específicas de produtividade, cor da polpa, resistência e destino de produção (*in natura* ou indústria).

2.2 Estudo da diversidade genética da goiabeira

O estudo da diversidade genética é essencial para programas de conservação e melhoramento genético da goiabeira, especialmente diante de seu histórico de domesticação e da consequente redução da variabilidade genética em cultivares comerciais. A caracterização da diversidade permite identificar fontes de alelos favoráveis em genótipos silvestres e em variedades locais, possibilitando sua incorporação em programas de cruzamento controlado para obtenção de genótipos superiores quanto à produtividade, resistência a estresses bióticos e abióticos e qualidade de fruto. Estudos demonstram que a limitada base genética de goiabeira pode ser ampliada por meio da introdução de espécies do mesmo gênero, que apresentam elevada diversidade e características agronômicas promissoras (Kumar et al., 2020; Koura et al., 2024).

Além disso, a diversidade genética está diretamente relacionada à eficiência de estratégias de seleção assistida por marcadores, pois influencia a acurácia da predição de genótipos desejáveis e a sustentabilidade de longo prazo dos programas de melhoramento (Thakur et al., 2021). Dessa forma, compreender e explorar a variabilidade genética presente no germoplasma de *Psidium* é imprescindível para o avanço da fruticultura tropical e para o enfrentamento de desafios relacionados às mudanças climáticas e à segurança alimentar.

Estudos com marcadores moleculares demonstraram a eficácia da utilização de RAPD e SSR no mapeamento de populações segregantes e na construção de mapas genéticos. O trabalho de Padmakar e colaboradores (2015) teve como objetivo explorar o genoma da goiabeira (*P. guajava* L.) por meio da utilização de marcadores moleculares SSR (microsatélites) e RAPD, visando o avanço na construção de mapas de ligação genética. Para isso, foram desenvolvidas duas populações de mapeamento F1: uma proveniente do cruzamento entre 'Kamsari' e 'Purple Local' e outra do cruzamento entre 'Purple Local' e 'Allahabad Safeda'. Com base nos dados obtidos a partir da utilização de marcadores RAPD e SSR, foram construídos mapas de ligação parciais para os genitores 'Purple Local' e 'Allahabad Safeda' (Kumar et al., 2020).

O uso da tecnologia de sequenciamento em massa e o desenvolvimento de SNPs (*Single Nucleotide Polymorphism*) permitiu avanços significativos na avaliação

de diversidade em *Psidium*. Em uma análise abrangente envolvendo nove espécies, foi possível identificar elevada variabilidade genética entre espécies nativas e limitada diversidade em *P. guajava*. Essa menor diversidade intraespecífica da goiabeira cultivada foi atribuída ao seu histórico de domesticação e cultivo intensivo. Observou-se também proximidade genética entre *P. guajava* e a espécie tetraploide *P. guineense* (Koura et al., 2024).

Foram analisados 48 acessos de diferentes regiões do México e de outros continentes com o objetivo de desenvolver e aplicar marcadores moleculares SNP para investigar a diversidade genética da goiabeira. Foram utilizados mais de 6.000 SNPs de alta qualidade, o que representou um avanço significativo em relação à densidade de marcadores usados em estudos anteriores. Utilizando análises de componentes principais e estrutura populacional, verificou-se ausência de associação clara entre agrupamentos genéticos e origem geográfica, possivelmente devido à ampla troca de germoplasma. O trabalho destacou ainda regiões genômicas associadas às características da síndrome de domesticação, reforçando a utilidade dos SNPs para conservação e melhoramento genético da espécie (Diaz-Garcia et al., 2023).

2.3 O melhoramento genético da goiabeira

Programas de melhoramento genético da goiabeira têm sido implementados em diversos países como a Índia, a Tailândia, o México e diversos países tropicais. Tais programas têm por objetivo sanar limitações de cultivares tradicionais, como baixa produtividade, suscetibilidade às doenças e qualidade inferior dos frutos. Tais programas empregam métodos clássicos como seleção massal e hibridação, além de técnicas modernas como mutagenese induzida, poliploidia e seleção assistida por marcadores moleculares (Thomas et al., 2001; Pommer, 2012; Anushka et al., 2024).

Na Índia, onde a goiaba é cultivada desde o século XVII, há uma grande diversidade de cultivares adaptadas regionalmente. O foco dos programas indianos tem sido o desenvolvimento de cultivares anãs, com menor número de sementes,

amadurecimento uniforme, teor ajustado de pectina e maior vida de prateleira. As principais cultivares comerciais incluem 'Pant Prabhat', 'HAPSI-35' e 'Allahabad Safeda' (Anushka et al., 2024).

No Brasil, programas de melhoramento foram responsáveis pelo desenvolvimento de cultivares como 'Paluma', 'Pedro Sato', 'Kumagai', 'Século XXI' e 'Rica', sendo 'Paluma' a principal variedade comercial devido ao seu alto rendimento, tamanho de fruto e adaptabilidade às condições tropicais. Além disso, cultivares como 'Ogawa', 'Sassaoka' e a série 'Cortibel' foram obtidas por seleção de campo ou em parceria com produtores, refletindo a integração entre pesquisa e prática agrícola (Pommer, 2012).

Apesar dessas conquistas, a base genética estreita ainda representa um dos principais entraves ao progresso do melhoramento genético no país. A utilização de poucos cultivares elite limita a variabilidade disponível e aumenta a vulnerabilidade da cultura a estresses bióticos e abióticos. A ampliação dessa base por meio da introdução de germoplasma exótico e do uso de espécies silvestres como *P. cattleianum*, *P. molle* e *P. cujavillus* é recomendada para garantir maior diversidade (Thomas et al., 2001; Pommer, 2012; Anushka et al., 2024).

A utilização de ferramentas moleculares é uma estratégia eficaz para apoiar o melhoramento. Walter e colaboradores (2024) empregaram 48 pares de primers microsatélites (SSR) para caracterizar 97 genótipos, incluindo 94 de famílias de irmãos-completos e três cultivares comerciais. O estudo revelou variabilidade genética significativa, com valores de heterozigosidade esperada variando de 0,030 a 0,599 e observada entre 0,010 e 0,577. A análise UPGMA formou quatro grupos distintos, enquanto a análise bayesiana indicou dois agrupamentos principais, sugerindo o cruzamento entre grupos divergentes como estratégia para maximizar o ganho genético.

Além disso, estudos como o de Mondragón-Jacobo e colaboradores (2010) destacam o uso de marcadores genéticos na identificação de QTLs para características de interesse, como espessura do mesocarpo, resistência a doenças e características sensoriais dos frutos. O mesmo trabalho relatou o desenvolvimento

de cultivares de polpa rosada no México, enfatizando o potencial da genética para diversificar a oferta comercial da fruta.

O uso de mutações induzidas também tem sido explorado com sucesso para ampliar a variabilidade genética. A taxa de mutação em condições naturais é baixa, mas pode ser significativamente aumentada por agentes físicos e químicos, permitindo a obtenção de genótipos com características raras ou desejáveis (Anushka et al., 2024).

O futuro do melhoramento genético da goiabeira depende da integração entre metodologias tradicionais e novas tecnologias, como seleção genômica e edição gênica. A caracterização genômica ampla e a colaboração internacional entre centros de pesquisa são apontados como caminhos promissores para o desenvolvimento de cultivares mais produtivas, resilientes e adaptadas às mudanças climáticas (Mondragón-Jacobo et al., 2010; Pommer, 2012; Anushka et al., 2024).

2.4 Marcadores moleculares do tipo SNP

Os marcadores moleculares do tipo SNP (*Single Nucleotide Polymorphism*) destacam-se pela ampla distribuição no genoma, natureza codominante e elevado poder de resolução, o que os torna mais eficientes que marcadores tradicionais como RAPD, AFLP e SSR. Essa característica possibilita análises precisas de diversidade genética, mapeamento de QTLs e estudos de associação genômica, constituindo uma ferramenta robusta no estudo da variabilidade em espécies agrícolas (Das et al., 2022).

O desenvolvimento de tecnologias de genotipagem em larga escala, como DArTseq e SNP-arrays, permitiu a obtenção de milhares de marcadores de forma rápida e economicamente viável. Essas plataformas ampliaram a capacidade de análises robustas em genética de populações e melhoramento, favorecendo estimativas consistentes de diversidade genética e estrutura populacional. Na goiabeira, o uso de SNPs tem se mostrado promissor para revelar padrões de

divergência entre famílias e grupos genéticos, contribuindo para o entendimento da base genética da espécie (Koura et al. 2024).

No contexto do melhoramento genético, os SNPs são amplamente aplicados em seleção assistida por marcadores, na implementação da seleção genômica ampla (GS) e na associação genômica (GWAS). Esses marcadores, por apresentarem alta densidade e cobertura do genoma, permitem estimativas de valores genéticos preditos com maior acurácia. Trabalhos recentes em goiabeira evidenciam o potencial dos SNPs na predição genômica de caracteres produtivos e de qualidade dos frutos, confirmando seu papel central em programas de melhoramento (Thakur et al., 2021).

Além disso, os SNPs desempenham papel relevante na conservação e no manejo de recursos genéticos, já que possibilitam caracterizações detalhadas da variabilidade intra e interespecífica. Esse tipo de abordagem fornece informações fundamentais para estratégias de conservação e preservação da biodiversidade, ampliando a base genética disponível para futuros programas de melhoramento. Estudos conduzidos com populações de goiabeira demonstram que o uso desses marcadores pode orientar práticas de conservação ao revelar o grau de diferenciação genética entre acessos (Grossi et al., 2021).

2.5 Predição de valores genéticos genômicos

A predição de valores genéticos genômicos (GEBVs – *genomic estimated breeding values*) tem implicações diretas na aceleração do ganho genético, especialmente em espécies perenes ou de ciclo longo, como a goiabeira. A acurácia dessas predições depende de fatores como a arquitetura genética das características, a qualidade dos dados fenotípicos, da densidade e distribuição dos marcadores genéticos, e do modelo estatístico empregado. Nesse contexto, modelos mistos, como o BLUP (*Best Linear Unbiased Prediction*), e suas extensões genômicas, RRBLUP (*Ridge Regression BLUP*) e GBLUP (*Genomic BLUP*), têm

sido amplamente utilizados para estimar efeitos aditivos e prever o desempenho genético de indivíduos antes mesmo de sua avaliação fenotípica completa.

O modelo BLUP tradicional, fundamentado em matrizes de parentesco médio, tem sido amplamente utilizado em programas de melhoramento devido à sua capacidade de prever valores genéticos individuais considerando simultaneamente os efeitos genéticos e ambientais. Sua robustez frente a desequilíbrios amostrais o torna especialmente útil em experimentos de campo com perdas ao longo do tempo.

Bezerra e colaboradores (2020) aplicaram a metodologia REML/BLUP para estimar parâmetros genéticos em 11 famílias de irmãos completos de goiabeira e selecionar indivíduos superiores. Foram avaliadas as seguintes características: peso do fruto, comprimento do fruto, diâmetro transversal, razão comprimento/diâmetro, peso da placenta, rendimento de polpa, espessura do endocarpo, espessura do mesocarpo, peso da polpa, sólidos solúveis e pH. A divergência genética entre selecionados foi calculada pela distância euclidiana média e os resultados obtidos foram utilizados para instruir cruzamentos futuros.

O advento das tecnologias de marcadores moleculares tornou possível substituir a matriz de parentesco tradicional por informações genômicas. O modelo RRBLUP foi proposto para lidar com essa nova configuração, assumindo que todos os marcadores têm efeito pequeno, mas diferente de zero, e penalizando excessos de variância por meio de uma regressão de crista. Já o modelo GBLUP substitui a matriz de parentesco médio por uma matriz de parentesco genômico baseada nas informações dos marcadores moleculares. Um estudo desenvolvido por Javid e colaboradores (2022) teve como objetivo avaliar a eficiência de metodologias de predição genômica, com ênfase nos modelos RRBLUP e GBLUP, cujos resultados evidenciaram que o GBLUP apresentou desempenho superior, indicando ser uma ferramenta promissora para a seleção genômica em trigo.

Com o objetivo de explorar múltiplas características simultaneamente e aumentar os ganhos genéticos totais, abordagens multivariadas de seleção genômica também tem sido desenvolvidas. Em estudo de simulação baseado em populações F₂, os modelos multivariados de predição – como MPLS (*Multivariate Partial Least Squares*) – foram comparados com métodos univariados como

RRBLUP, *Bayesian* Lasso e PLS. O RRBLUP destacou-se pelo equilíbrio entre eficiência preditiva e tempo computacional, especialmente em cenários com características de baixa herdabilidade, reforçando sua aplicabilidade em programas de melhoramento voltados às características quantitativas de difícil mensuração ou influenciados por grande variação ambiental (Silva et al., 2021b).

A germinação pré-colheita (PHS) representa um desafio relevante para a triticultura, pois ocasiona redução significativa da qualidade e do rendimento das sementes, sendo, portanto, essencial a identificação de regiões genômicas associadas à tolerância a essa característica. Nesse contexto, Rabieyan e colaboradores realizaram um estudo de associação genômica ampla (GWAS) em 298 acessos e variedades de trigo iraniano, utilizando diferentes modelos de predição genômica (RRBLUP, GBLUP e BRR). Os autores identificaram milhares de pares de marcadores significativos distribuídos nos genomas B, A e D, além de associações relevantes entre marcadores e características como cor da semente, atividade da α -amilase e PHS. O modelo RRBLUP apresentou maior eficiência na detecção dos efeitos genéticos em comparação a GBLUP e BRR, demonstrando potencial para ser aplicado em programas de seleção genômica do trigo (Rabieyan et al., 2022).

O trabalho de Susmitha e colaboradores teve como objetivo destacar o papel dos estudos de associação genômica ampla (GWAS) como ferramenta poderosa para dissecar caracteres competitivos em leguminosas. A revisão evidencia a relevância dos GWAS na identificação de associações entre SNPs e caracteres fenotípicos, aplicáveis à seleção genômica (GS) por modelos como rrBLUP, gBLUP, Bayesianos e Random Forest. Além disso, discutiu o uso de diferentes populações de mapeamento, como painéis de germoplasma e populações multiparentais (MAGIC, NAM, MCILs), ressaltando que o avanço da genotipagem de alto rendimento, da fenômica e de técnicas computacionais tem ampliado as perspectivas para o melhoramento competitivo das leguminosas (Susmitha et al., 2023).

2.6 Melhoramento genético da goiabeira na UENF

O Programa de Melhoramento Genético da Goiabeira da Universidade Estadual do Norte Fluminense Darcy Ribeiro (PMGG-UENF) foi iniciado em 2007, com a coleta de 20 acessos de *P. guajava* L. provenientes de pomares comerciais e plantas de polinização aberta nos municípios de São João da Barra e Bom Jesus do Itabapoana, no estado do Rio de Janeiro (Pessanha et al., 2011). Essa etapa de pré-melhoramento marcou o início de um processo contínuo que visava gerar variabilidade genética e estruturar populações para avaliação e seleção com foco em produtividade e qualidade de frutos.

Após os estudos de diversidade genética e análises citogenéticas com esses acessos, foram escolhidos genitores com alto grau de heterozigosidade para composição dos cruzamentos. Como resultado, formaram-se 17 famílias de irmãos-completos, avaliadas por três safras sucessivas quanto a características agrônomicas e de qualidade de fruto (Quintal et al., 2017).

Paralelamente, Campos et al. (2016) avaliaram 138 acessos do banco de germoplasma da UENF utilizando descritores morfoagronômicos e físico-químicos, aplicando a metodologia Ward-MLM para estimar a divergência genética. Essa abordagem auxiliou na escolha de genitores para cruzamentos controlados.

Com as 17 famílias segregantes estabelecidas, foram conduzidas análises baseadas em modelos mistos (REML/BLUP) para estimar parâmetros genéticos e selecionar as famílias superiores. A metodologia permitiu considerar simultaneamente os efeitos fixos e aleatórios do experimento e foi essencial para a obtenção de estimativas acuradas de herdabilidade, valores genotípicos e repetibilidade (Quintal et al., 2017).

Com base nos resultados dessas avaliações, o programa deu origem a novas estratégias, incluindo o desenvolvimento de populações endogâmicas e cruzamentos dirigidos entre irmãos-completos. Ambrósio et al. (2021) avaliaram 18 famílias endogâmicas obtidas por autofecundação de genótipos superiores e

constatarem que o índice de seleção multiplicativo foi o mais eficiente na identificação de famílias com desempenho superior em produtividade e qualidade dos frutos.

Santos e colaboradores caracterizaram, com 33 marcadores microssatélites, populações segregantes de *Psidium* visando resistência a *Meloidogyne enterolobii*. As análises revelaram diversidade genética moderada, especialmente nos cruzamentos *P. guineense* × *P. cattleyanum* e *P. guajava* × *P. cattleyanum*, e estruturaram as populações em três grupos distintos. Os marcadores SSR foram eficientes na discriminação dos híbridos, destacando genótipos promissores para avanço no melhoramento da goiabeira (Santos et al., 2020).

O fluxograma, apresentado na Figura 1, ilustra a trajetória do Programa de Melhoramento Genético da Goiabeira da UENF desde sua fundação em 2008, com a coleta inicial de acessos e formação das primeiras progênies, até as etapas mais recentes de avaliação fenotípica e molecular. Atualmente, o programa encontra-se na etapa representada pelo quadrante inferior central do fluxograma, que trata do estudo de diversidade e seleção genômica (GW) em população de irmãos-completos. Nesta fase ocorre a incorporação de dados genotípicos de marcadores do tipo SNP na seleção de indivíduos, sendo estratégica para o desenvolvimento da seleção genômica (GW) e para o aumento da acurácia na predição de valores genéticos.

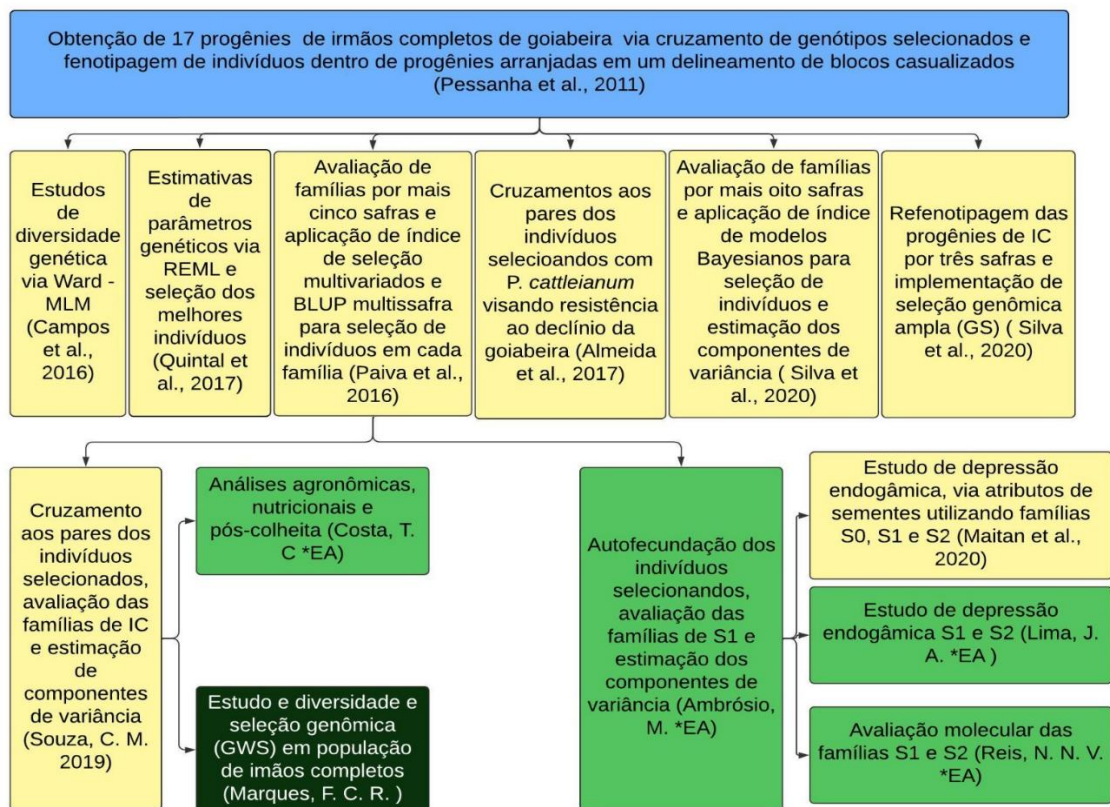


Figura 1: Fluxograma das etapas do programa de melhoramento genético de goiabeiras da UENF.

Fonte: Liberation Sans

3 OBJETIVOS

3.1 Geral

O objetivo foi avaliar a diversidade genética de 11 famílias de irmãos completos de goiabeira com marcadores SNP e aplicar metodologias de seleção genômica, comparando modelos com e sem informações genômicas para verificar o impacto dos dados genotípicos na acurácia da predição de valores genéticos.

3.2 Específicos

- Caracterizar a variabilidade fenotípica entre 247 genótipos de goiabeira pertencentes às 11 famílias de irmãos completos com base em descritores morfológicos e agronômicos.
- Avaliar a diversidade genética molecular desses genótipos por meio de marcadores SNP obtidos por genotipagem de alta densidade.
- Estimar os valores genéticos dos indivíduos pertencentes às famílias de irmãos completos de goiabeira, utilizando os modelos mistos BLUP tradicional, baseado unicamente em informações fenotípicas e matriz de parentesco médio.
- Aplicar os modelos RRBLUP e GBLUP para a predição de valores genéticos genômicos.

4 TRABALHOS

4.1 Prospecting variability in full-sib families of *Psidium guajava* L. through phenotypic information and SNP markers

Reproduzido com a permissão:

Marques, F. C. R., Viana, A. P., Costa, T. C., Silva, F. A., Silva Júnior, O. B. and Grattapaglia, D. (2025). Prospecting variability in full-sib families of *Psidium guajava* L. through phenotypic information and SNP markers. *Bragantia*, 84, e20240087. <https://doi.org/10.1590/1678-4499.20240087>

ABSTRACT: The species *P. guajava* L., commonly known as guava, is distinguished in the genus *Psidium* for its economic importance and medicinal properties. This study aimed to assess the population structures in 247 guava genotypes, which were organized into 11 fullsib families (FSF), and to explore the differentiation of these FSF based on phenotypic traits and single nucleotide polymorphism molecular markers. The study utilized a randomized block design to evaluate 10 phenotypic variables and 7,784 polymorphic markers, derived from leaf samples through genetic sequencing using DArTseq technology. The diversity among individuals was quantified using the Mahalanobis' generalized distance. The UPGMA clustering method identified three homogenous clusters in the families. Excoffier's analysis of molecular variance (AMOVA) demonstrated significant variability both between and in these clusters, with inter-group variability accounting for about 20% of the total variation. The genetic distance between individuals, calculated using molecular markers, was determined by the arithmetic complement of the unweighted similarity index. Ward's method delineated five subpopulations. AMOVA further verified significant variability both between and in these subpopulations, attributing 89% of the total variation to differences between the groups formed. Eleven FSF could not be differentiated based on the analyzed data. Nevertheless, the analyses confirmed substantial genetic diversity, endorsing the continuity of the breeding program through selection and future crosses.

Key words: full-sib families, genetic distance, genetic diversity.

INTRODUCTION

The guava (*P. guajava* L.) belongs to the family Myrtaceae and the subclass Rosidae. It is a perennial species. This means pollination and fruiting can be done in the same individual plant, because it has separate male and female flowers. The leaves are simple, opposite, and aromatic. The flowers are white with numerous stamens; they are solitary or in small clusters in the leaf axils.

Originating from South America, the guava was spread across African and Asian territories by European explorers. Presently, this crop thrives in various tropical and subtropical zones worldwide, owing to its remarkable adaptability to the diverse soil and climatic conditions prevalent in these areas (Angulo-López et al. 2021).

India and Pakistan collectively account for nearly half of the global guava yield. Alongside these nations, Brazil, Mexico, and Venezuela significantly contribute to the production of the fruit on the American continent (Yousaf et al. 2021). In 2022, guava cultivation in Brazil encompassed an area of 22,684 hectares, yielding 564,764 t, with the state of Rio de Janeiro producing 16,657 t, ranking it as the seventh-largest guava producer in the country (IBGE 2023).

The favorable tropical climate and fertile soil in Rio de Janeiro provide optimal conditions for guava farming. The strategic location of the state near major urban centers, along with its developed port and export facilities, ensures the swift distribution of produce. Additionally, the economic diversity of Rio de Janeiro, spanning tourism to manufacturing, creates a supportive environment for investments in research and technological advancements to enhance guava production.

Guava is appreciated both fresh and in-processed products, with consumers often preferring seedless varieties or those with few, soft seeds. Additionally, traits such as sweetness, color, pulp consistency, appearance, skin thickness, and nutritional value are highly valued. Improving these qualities is the goal of various breeding programs. In Brazil and around the world, these programs also aim to

enhance fruit uniformity, extend shelf life, and increase resistance to pests, such as guava decline, as well as to water stress (Thakur et al. 2021).

In this framework, plant breeding plays a fundamental role in agricultural production, enabling the enhancement of commercially valuable traits and the development of cultivars tailored to the specific conditions of the region.

Given the perennial nature of guava, breeding strategies must account for the variations observed in the lengthy production cycles, early yield among plants, and environmental influences. The integration of molecular markers with advancements in computational technology bridges traditional breeding methods, which rely solely on phenotypic observations, with contemporary molecular techniques, thereby refining the accuracy of genetic parameter estimation and prediction (Silva et al. 2022).

There is little scientific research evaluating genetic diversity in association with phenotypic variability in guava, using genetic markers for both geographic association studies and germplasm evaluation, in Brazil and abroad. This is partly due to the use of geographically close populations and the limitation of molecular markers, which are scarce and have low density across the genome, as in the case of RAPD markers (Diaz-Garcia and Padilla-Ramírez 2023).

There is also a significant limitation in studies on the evolution of guava and its genetic improvement potential due to the absence of a reference genome. However, efforts for chromosome-level analyses are underway. Feng et al. (2021), for example, presented a chromosome-level genome assembly of guava while investigating metabolic pathways related to fruit softening.

Among molecular markers, single nucleotide polymorphisms (SNPs) are distinguished for their precision, high resolution, and compatibility with automated processing on next-generation sequencing (NGS) platforms (Grossi et al. 2021). According to the scope of NGS platforms, DArTseq technology is particularly noted for its broad genomic coverage and efficiency, making it a preferred choice for diversity studies and offering considerable economic advantages (Koura et al. 2024).

This study aimed to assess the population structure in a group of 247 guava genotypes, organized into 11 full-sib families (FSF), utilizing phenotypic data and information from SNP molecular markers.

The studied population consists of individuals with phenotypic variability observed in the field. This article presents a multifaceted population analysis methodology to determine whether this variability is also reflected in the genetic markers, considering that the individuals analyzed have a high degree of relatedness and that the parents are relatively close geographically. Additionally, it is estimated that previous interventions, such as selection and controlled crossings carried out by the Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF) guava breeding program, may have significantly influenced the genotypic variability of the population.

MATERIAL AND METHODS

Plant material

The genotypes comprising the study population were obtained in 2008, with the beginning of the UENF genetic improvement program. During a pre-breeding stage, 20 accessions of *P. guajava* L. were selected from commercial orchards and open-pollinated plants in the municipalities of São João da Barra and Bom Jesus do Itabapoana, in the state of Rio de Janeiro, southeastern Brazil.

After cytogenetic and diversity studies using RAPD markers, conducted by Pessanha et al. (2011), individuals with a considerable degree of heterozygosity were selected for crossing. These crosses resulted in 17 full-sibling families. Characteristics related to fruit quality and productivity of this new segregating population were evaluated over three production periods by Quintal et al. (2017). In this study, the mixed model methodology was used to select the parents of 11 full-sibling families, which are the focus of the study presented in this article, as shown in Table 1.

Table 1. Kinship information of the 11 full-sib families generated from selection by Quintal et al. (2017). The numbers separated by dots represent families and individuals, respectively. Roman numerals represent the blocks of experimental design. (Campos dos Goytacazes, RJ, Brazil).

Family	Kinship	Nº. of individuals	Family	Kinship	Nº. of individuals
1	8.4 II x 5.4 II	24	7	12.11 II x 12.4 II	23
2	10.5 I x 12.4 II	23	8	12.11 II x 3.11 II	19
3	12.11 II x 8.4 I	22	9	12.11 I x 7.9 I	24
4	4.5 I x 2.6 I	22	10	8.1 II x 12.4 II	22
5	12.11 I x 8.4 I	23	11	12.4 II x 12.11 II	21
6	12.4 x 12.11	24			

The research was conducted at the experimental site of UENF, situated in the Antônio Sarlo Agricultural School in Campos dos Goytacazes, RJ, Brazil (21°08'02"S; 41°40'47"W). The area was characterized by a sub-humid and dry tropical climate, with an average annual temperature fluctuating between 22 and 25°C, and annual precipitation ranging from 1,200 to 1,300 mm.

The guava plants were planted with a spacing of 1.2 m between each plant and 3.5 m between rows. All cultivation practices, including pruning, were performed following standard guava cultivation guidelines (Silva et al. 2016; Semiárido Ministério da Agricultura & Abastecimento, 2010).

A randomized block experimental design was adopted, comprising three replicates for each family. Each experimental unit included eight plants from the same family. The assessment was made on five randomly selected fruits per plant, specifically those with a yellowish-green skin indicative of the third stage of maturation, defined by a color angle ranging from 112 to 108°h (Azzolini et al. 2004).

Phenotypic variables

The evaluation of the plants involved analyzing five fruits for the following traits:

- Individual fruit weight (FW): determined using a semi-analytical scale, with the weight expressed in g;

- Fruit length (FL): measured along the longitudinal axis of the fruits with a digital caliper, with the length expressed in mm;
- Fruit diameter (FD): assessed at the equatorial region of the fruits using a digital caliper, with the diameter expressed in mm.;
- Pulp weight (PPW): calculated as the weight of the fruit excluding the placenta, measured with a semi-analytical scale and expressed in g;
- Placenta weight (PLW): focused on the seed-concentrated region, determined using a semi-analytical scale;
- Endocarp thickness (ET): measured in the seed-concentrated region with a digital caliper, with the thickness expressed in mm;
- Mesocarp thickness (MT): pertains to the area from the fruit peel to the beginning of the placenta, measured using a digital caliper and expressed in mm;
- Peel thickness (PT): relates to the fruit peel region, measured with a digital caliper and expressed in mm;
- Soluble solids content (BRIX): assessed using an Atago No. 1 digital refractometer, with results expressed in degrees Brix (°Brix);
- Total production (PROD): calculated by summing the weights of all fruits produced by each evaluated plant, measured using an electronic scale and expressed in g.

Molecular variables

Genetic material from 247 guava specimens was collected for genotyping. The high cost of producing specific probes for guava makes it unfeasible to obtain detailed genotypic information for this species. However, considering that guava and eucalyptus belong to the same family, Myrtaceae, and share a relatively close taxonomic level, there is a moderately close relationship between these plants. This genetic proximity suggests that they share common characteristics due to their ancestry within the family. Given the availability of Brazilian Agricultural Research Corporation (Embrapa), SNP probes and the high probability of random and uniform hybridization across the guava genome, the Axiom 72K ESAI Eucalyptus SNP Array Initiative (Embrapa) was used to obtain the necessary genotypic information. Genomic DNA extraction was performed on young, fully matured leaves, adhering to

the procedure described by Inglis et al. (2018). For molecular marker identification, samples underwent analysis utilizing the diversity arrays technology (DArTseq) sequencing genotyping approach, refined by DArT P/L in Canberra, Australia. The analysis resulted in the identification of 72,202 SNP markers within the population. These markers were subjected to four quality filters, that were excluded:

- Markers with a minor allele frequency of 5% or less;
- Markers with over 10% missing data;
- Markers displaying zero variation (uniform across all evaluated individuals);
- Duplicate markers.

Imputations of 2.75% of missing data were conducted by substituting with the allele most prevalent in the individual. After quantity filtering, a total of 7,874 markers remained for each individual (Cruz 2013; Cruz 2016).

Statistical approach

The genetic variability between and within the designated subpopulations (families or groups) was assessed using Excoffier's molecular analysis of variance (AMOVA) (Excoffier et al. 1992; Cruz et al. 2020; Huang et al. 2021; Abbaszadeh et al. 2023). AMOVA employs a model that allocates variability contributions to the effects under investigation (Eq. 1).

$$Y_{ij} = \mu + P_i + D_{ij} \quad (1)$$

where:

- Y_{ij} : variability observed between individual pair J within population i;
- μ : constant term;
- P_i : effect of subpopulation i; and
- D_{ij} : dissimilarity effect between individual pair j within subpopulation i.

The measure of proportionality, Φ_{ST} , was calculated from the ratio $\frac{\sigma_p^2}{\sigma_T^2}$, in which σ_p^2 represents the variance between groups, σ_i^2 is the associated variance within groups; and $\sigma_T^2 = \sigma_p^2 + \sigma_i^2$. Therefore, Φ_{ST} indicates the contribution of molecular variance between the subpopulations studied relative to the total molecular variance.

A permutation test with 999 iterations was employed to establish confidence intervals for the analyzed statistics, aiming to test the null hypothesis that assumes no variance within and between the populations under study.

To quantify genetic distances between the FSF, Mahalanobis' generalized distance was employed (Eq. 2) as the chosen measure of dissimilarity, based on data from the phenotypic variables examined:

$$D_{ii'}^2 = \delta' \varphi^{-1} \delta = \sum_{j=1}^n \sum_{j'=1}^n \omega_{jj'} d_j d_{j'} \quad (2)$$

where:

- $D_{ii'}^2$: Mahalanobis' generalized distance between individual i and i' ;
- $\delta' = [d_1, d_2, \dots, d_v]$, in which v is the number of phenotypic variables analyzed. $\delta' = [d_1, d_2, \dots, d_v]$, is the difference between the mean values of variable v for families i and i' ;
- φ : g matrix of residual variances and covariances; and
- $\omega_{jj'}$: element of the inverse matrix of residual variances and covariances at row j and column j .

Additionally, the S_j statistic proposed by Singh (1981) was computed to ascertain the relative contribution of each variable toward explaining the observed divergence (Eq. 3):

$$S_j = \frac{\sum_{j'=1}^p \omega_{jj'} d_j d_{j'}}{\sum_{j=1}^p S_j} \times 100 \quad (3)$$

The analysis of molecular data included assessments of heterozygosity, inbreeding coefficients, and Polymorphism Information Content (PIC). Observed heterozygosity is calculated as the ratio between the number of heterozygous individuals and the total number of individuals in the population. Expected heterozygosity, on the other hand, can be calculated from allele frequencies and is equal to twice the product of the frequencies of the reference and alternative alleles. Inbreeding is determined as the arithmetic complement of the ratio between observed and expected heterozygosity. The PIC calculation takes into account the allele frequencies at a given locus p_i to determine how informative that locus is (Eq. 4).

$$PIC = 1 - \sum p_i^2 - \sum \sum 2p_i^2 \cdot p_j^2 \quad (4)$$

Dissimilarity between two individuals was quantified using the arithmetic complement of the unweighted similarity index (Eq. 5):

$$S_{ii'} = \frac{1}{2L} \sum_{j=1}^L c_j \quad (5)$$

where: $S_{ii'}$: similarity index between individuals i and i' ; L : number of loci being studied; c_j : number of alleles common between individuals i and i' at loci j (Table 2).

Table 2. Count of alleles in common for each locus

c_j	Alleles
2	AA and AA; AB and AB; BB and BB
1	AA and AB; AB and BB
0	AA and BB

The unweighted similarity index ranges from 0 to 1. A dissimilarity matrix was constructed, with each d_{ij} element representing the distance between individuals i and j (Eq. 6):

$$D_{n \times n} = \begin{cases} d_{ij} = 0; \text{ if } i = j \\ d_{ij} = (1 - S_{ij}); \text{ if } i \neq j \end{cases} \quad (6)$$

For clustering and analyzing subpopulations, the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Ward's method (D^2) were employed. The UPGMA method was used due to the characteristics of the phenotypic data presented and its low sensitivity to the presence of outliers. Conversely, the Ward method was chosen because it provides greater internal homogeneity and has a tendency to combine groups with few elements, a desired characteristic given the low variability presented by the SNP information.

Data management, including the organization of phenotypic data and molecular markers, was facilitated by the 'Tidyverse' suite of packages for R, utilizing the 'dplyr' package. Plots were generated using the ggplot2 package (Wickham et al. 2019). Mahalanobis' generalized distance calculations and clustering techniques were performed using the 'Multivariate Analysis' package, also for R (Azevedo 2021). Dendrogram creation was accomplished with the 'Dendextend' package (Galili 2015).

Finally, AMOVA analyses were conducted using the ‘genetic diversity’ module of Genes software (Cruz 2013; Cruz 2016).

RESULTS AND DISCUSSION

Available variability based on phenotypic observations.

A AMOVA analysis was implemented considering the 11 FSF and individuals in the FSF as hierarchical levels (Table 3).

Table 3. Result of Excoffier's AMOVA analysis.

Source of variation	DF	SS	MS	Estimate	%
Between FSF (σ_p^2)	10	13.6628	1.3663	0.0531 ^{ns}	18.6796
Within FSF (σ_t^2)	224	51.8239	0.2314	0.2314 ^{ns}	81.3204
Total	234	65.4867	1.5977	0.2845 ^{ns}	100

DF – degrees of freedom; SS – sum of squares; MS – mean square; ns – estimate not significant according to the permutation test with 999 iterations.

The Φ_{ST} statistic calculated was 0.1868^{ns}. The permutation test indicates that the variability between and in the 11 FSF was not statistically significant, leading to the conclusion that differentiation between the 11 FSF under investigation is not feasible.

In contrast, a study by Bernardes et al. (2022) reported a moderate yet significant level of differentiation between three populations of *P. guajava* L., sourced from five restinga regions (coastal plains covered by marine deposits) in Espírito Santo state, Brazil. It is important to note that the differences observed in the study by Bernardes et al. (2022) can be attributed, among other factors, to the greater geographic scope of their sampling locations, as opposed to the geographically very close origins of the 11 FSF examined in this study. The accessions that gave rise to the population studied here were collected in two sampling areas approximately 100 km apart. This proximity contributed to homogeneity and the resulting impossibility of distinguishing between families. The observed low genetic diversity in and between the 11 FSF can be linked to two cycles of selection. The first cycle phase was

informed by data from producers in the orchards, while the subsequent phase involved multi-crop evaluations, a multivariate assessment index, and the application of mixed models (Paiva et al. 2016).

Moreover, the expected uniformity in commercial orchards simplifies management practices. Furthermore, the existence of kinship between families contributes significantly to the AMOVA analysis confirming the impossibility of differentiating the FSF based on the variables analyzed.

Still addressing genetic diversity analysis, the relative contribution of phenotypic variables to diversity was quantified using the statistical approach proposed by Singh (1981). Variables such as fruit length, weight, and diameter; endocarp thickness; pulp weight; and production accounted for over 85% of the observed variability (Fig. 1).

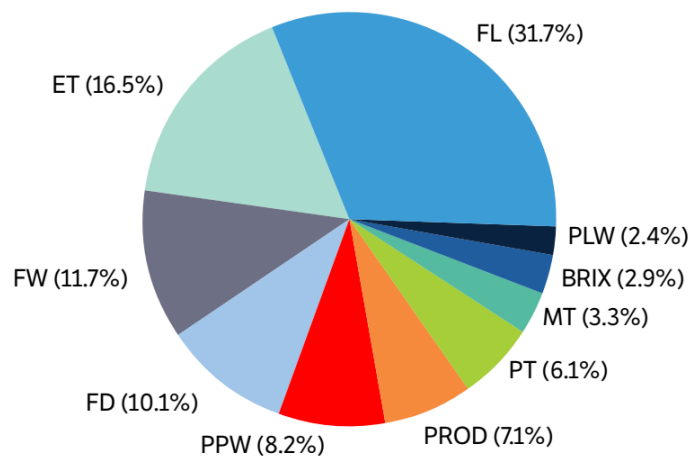


Figure 1. Contribution of the 10 variables studied to explain the diversity among the 11 guava full-sib families.

FL: fruit length; PLW: placenta weight; BRIX: soluble solids content; MT: mesocarp thickness; PT: peel thickness; PROD: total production; PPW: pulp weight; FD: fruit diameter; FW: individual fruit weight; ET: endocarp thickness.

This finding aligned with results presented by Singh et al. (2015), who employed Mahalanobis' generalized distance to assess genetic distances among 35 guava genotypes, highlighting the significant role of morphological traits of the fruit in the diversity observed. These findings highlight the significance of morphological traits in the selection and breeding of guava genotypes, aiming at the development of better adapted and productive cultivars.

It is equally important to acknowledge that selection cycles and artificial crossings have varied impacts on the variables under study. Consequently, the contribution of guava fruit morphological traits to the observed diversity will differ across different phases of breeding programs.

The dissimilarity matrix was derived from Mahalanobis' generalized distance, and the dendrogram illustrates the three groups formed using the UPGMA hierarchical clustering method and Mojena's cut-off technique (Mojena 1977), setting k at 1.25 (Fig. 2). Group I included families 1, 3, 5, 7, 8, and 9, while Group II comprised families 2 and 10, and Group III consisted of families 4, 6, and 11.

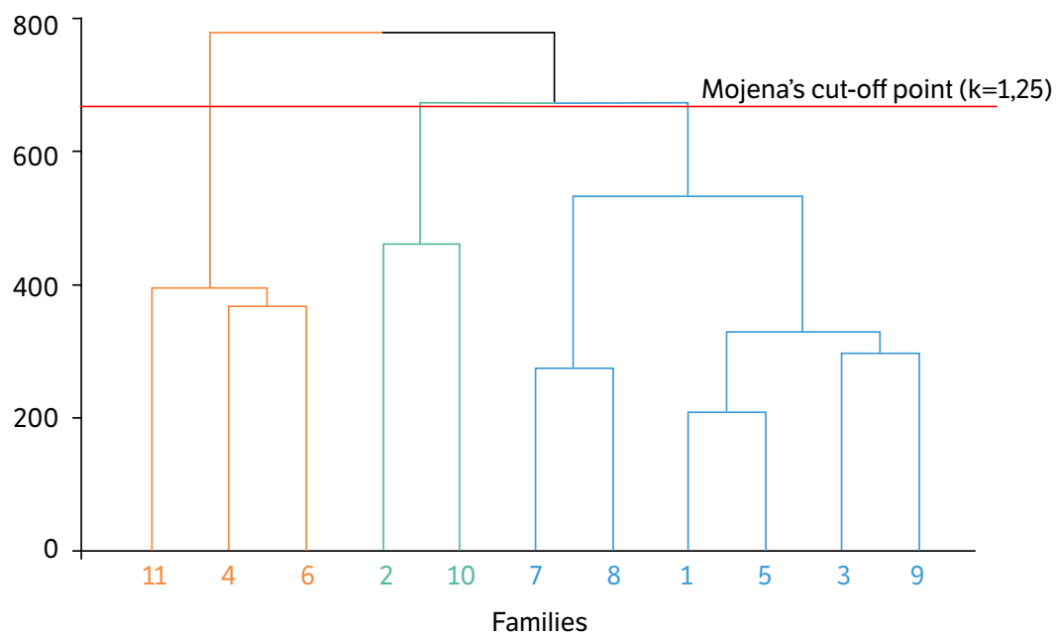


Figure 2. Dendrogram obtained by Mahalanobis' generalized distance using the unweighted pair group method with arithmetic mean (UPGMA) methodology, clustering the 11 guava full-sib families into three groups. Distinct colors in the dendrogram indicate different groups. Blue represents Group I, green represents Group II, and magenta represents Group III.

Ambrósio et al. (2022) also applied the UPGMA technique to group 61 *P. guajava* L. genotypes into three more homogenous clusters. The clustering observed in both their study and this research indicated the presence of, albeit limited, genetic diversity, presenting opportunities for exploration in subsequent selection and crossing cycles.

Crossing between individuals from genetically distinct groups maximizes vigor in their progeny, making the information on group formation essential in parent selection for future crosses to optimize the use of heterosis. An AMOVA analysis was

conducted considering the groups formed via UPGMA methodologies and the individuals in these clusters as hierarchical levels (Table 4).

Table 4. Result of Excoffier's analysis of molecular variance for the three groups formed using the unweighted pair group method with arithmetic mean (UPGMA) approach.

Source of variation	DF	SS	MS	Estimate	%
Between groups (σ_p^2)	2	9.0362	4.5181	0.0613*	20.128
Within groups (σ_t^2)	232	56.4505	0.2433	0.2433*	79.872
Total	234	65.4867	4.7614	0.3046	100

DF: degrees of freedom; SS: sum of squares; MS: mean square; *5% level of significance in the permutation test.

The derived Φ_{ST} statistic stood at 0.20128*. AMOVA indicated that over 20% of the total variation originates from differences between the groups, as delineated by the findings. The permutational test further validated the significance of these variability estimates at the 5% significance level.

The detection of significant variability between the clusters formed through the UPGMA clustering technique substantiates the presence of genetic diversity in the population under study. The maintenance of variability and differentiability among the formed groups indicates the viability of continuing the genetic improvement program with the studied population, suggesting that new cycles of controlled selection and crossing could be effectively applied. However, to optimize the process and ensure result accuracy, further studies will be needed to estimate genetic parameters and variance components. These data are crucial for refining selection strategies and achieving sustainable progress in the genetic improvement of the population.

The AMOVA analysis was applied considering the 11 FSF and the individuals in these FSF as hierarchical levels. The analysis utilized the arithmetic complement of the unweighted similarity index between loci as the measure of similarity (Table 5).

Table 5. Result of Excoffier's molecular analysis of variance for the 11 guava full-sib families considering data from single nucleotide polymorphisms molecular markers.

Source of variation	DF	SS	MS	Estimate	%
---------------------	----	----	----	----------	---

Between groups (σ_p^2)	10	0.7915	0.0792	0.0009ns	1.4157
Within groups (σ_i^2)	236	14.1267	0.0599	0.0599ns	98.5843
Total	246	14.9182	0.1391	0.0607ns	100

DF: degrees of freedom; SS: sum of squares; MS: mean square; ns: estimate not significant according to the permutation test with 999 iterations.

The calculated Φ_{ST} statistic was 0.014157^{ns} , suggesting that roughly 1% of the total variance is attributable to differences among the 11 FSF. The permutational test confirmed the lack of significance of these estimates. Therefore, based on the generated distance matrix, the 11 FSF were not differentiable.

These AMOVA findings aligned with the observed low genetic variability, which mirrors the earlier noted unfeasibility of distinguishing FSF based on phenotypic variables. Furthermore, the impact of kinship in the FSF became particularly evident when examining molecular data (Table 1).

The investigation into heterozygosity, inbreeding, and PIC provided an overview of the genetic makeup of the studied families (Fig. 3). The outcomes pointed to a generally low genetic variability among families, with family 8 being an exception due to its significantly lower values of both expected and observed heterozygosity and PIC in comparison to the other families.

Notably, families 7, 9, and 11 exhibited greater disparity between expected heterozygosity under Hardy-Weinberg equilibrium and the observed heterozygosity, though these variances did not surpass 0.007. The presence of inbreeding, as measured by the F index, was markedly evident in families 7, 8, 9, and 11, with family 7 showing a pronounced value of 0.4. These findings suggested a degree of genetic uniformity in the examined families.

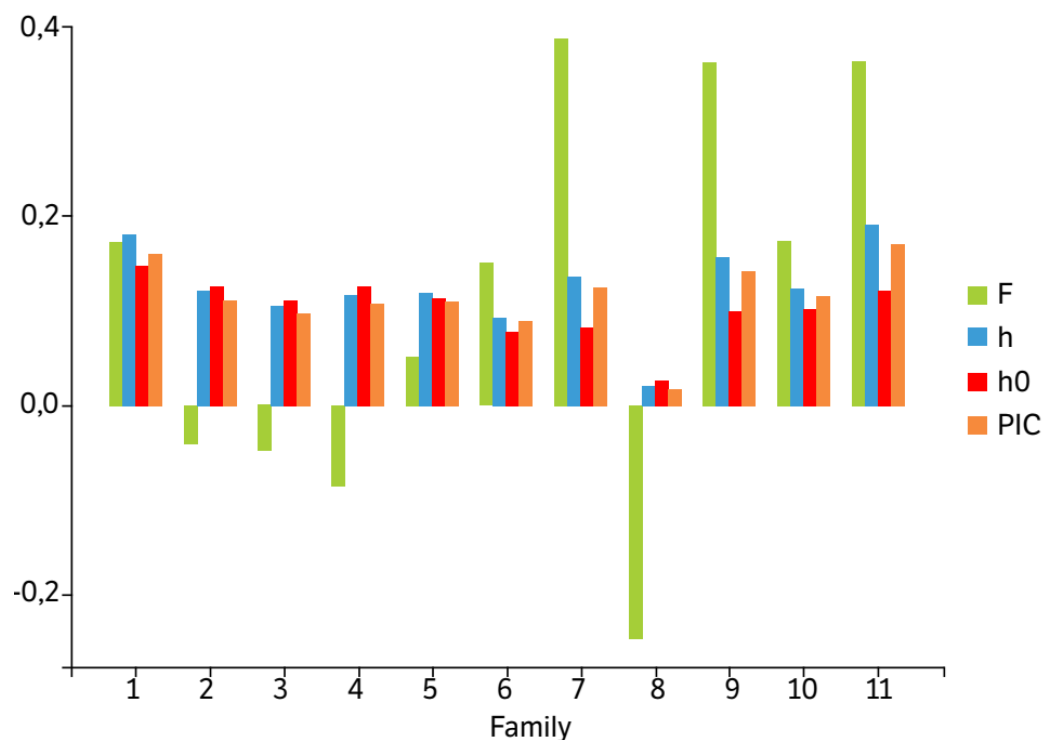


Figure 3. Results of inbreeding, heterozygosity, and polymorphism information content analyses for the 11 guava full-sib families, obtained using GENES software.

The plants were recoded to facilitate data entry into the GENES program. The list of codes is presented in Table 2. Ward's method was applied to square the distances. By adopting the cut off technique proposed by Mojena (1977), five unique groups were identified (Fig. 4). Group I comprised genotypes G241, G242, G244, and G245, while Group II included G234, G235, G236, G237, G238, G240, and G243. Genotypes G22, G32, G33, G35, G44, G58, G61, and G199 formed Group III, and Group IV consisted of G5, G23, G26, G31, G34, G36, G37, G38, G40, G41, G48, G49, G52, G60, and G239. The remaining 212 genotypes were categorized into Group V.

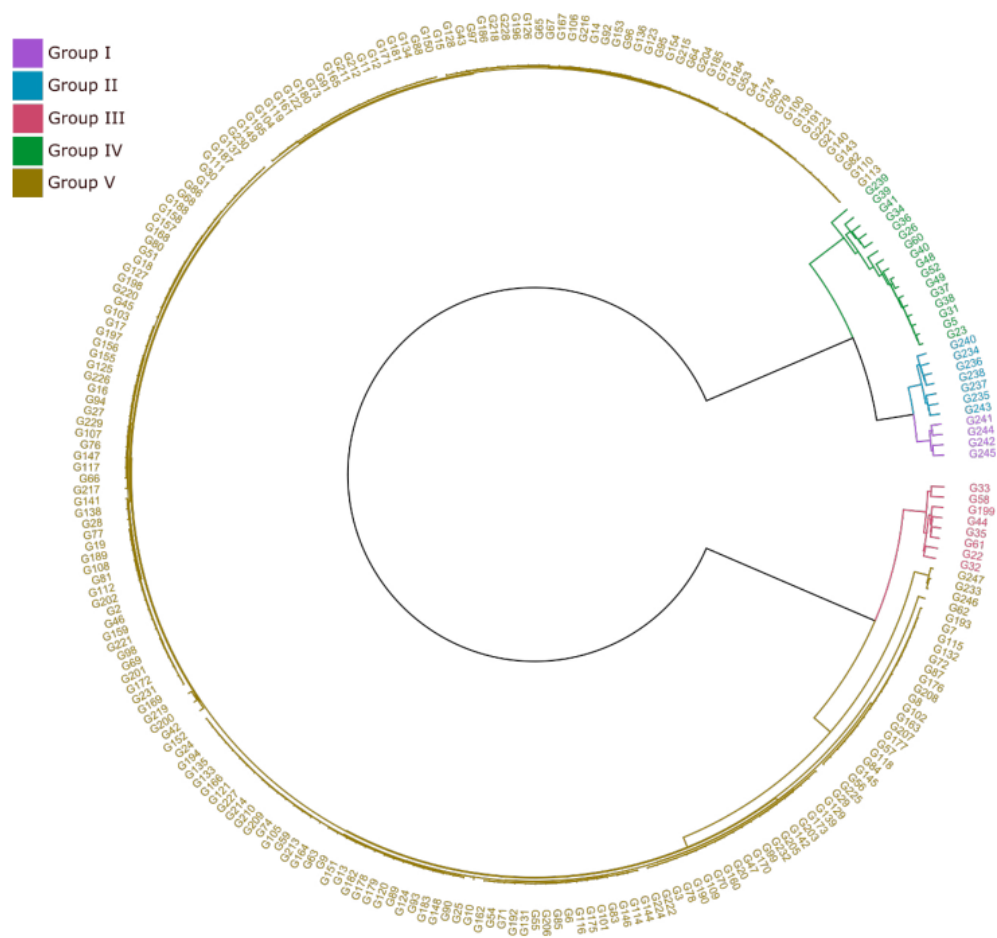


Figure 4. Dendrogram obtained with the dissimilarity matrix of the 247 genotypes studied. The clustering method used was Ward's D2. The groups were formed based on the cut-off point proposed by Mojena (1977) with $K = 1.25$.

The consolidation of over 85% of the individuals into a single group reinforces the genetic homogeneity of the population, as evidenced by genetic marker data. This aligned with Laviola et al. (2018), who also reported limited genetic diversity in the Brazilian *Jatropha Curcas* (Barbados nut) germplasm bank using SNP markers. However, the aim of the present study diverged by focusing on understanding population structure to enhance selection and hybridization processes in the guava breeding program.

As observed by Pierre et al. (2022), who applied SNP markers and Ward's clustering to separate African cassava cultivars into three main clusters, this study confirmed the presence of genetic diversity and the potential for organizing genotypes into distinct genetic groups.

This insight informs the selection of individuals to maximize heterosis, as the genetic distance between groups is likely to enhance vigor and progeny performance in crosses. Thus, the molecular data analysis complements phenotypic variable assessments, suggesting a link between genetic structure and the observed traits. Excoffier's AMOVA was carried out considering the groups formed by Ward's methodology as hierarchical levels (Table 6).

Table 6. Results of Excoffier's analysis of molecular variance analysis for the four groups formed by Ward's methodology using data from single nucleotide polymorphisms-type molecular markers (Excoffier et al. 1992).

Source of variation	DF	SS	MS	Estimate	%
Between groups (σ_p^2)	4	10.401	2.6003	0.1669**	89.9389
Within groups (σ_i^2)	242	4.5171	0.0187	0.0187**	10.0611
Total	246	14.9182	2.619	0.1855	100

** 1% level of significance in the permutation test. DF: degrees of freedom; SS: sum of squares; MS: mean square; *1% level of significance in the permutation test.

The Φ_{ST} statistic of 0.8994 attested that over 89% of the variability of the studied population is attributable to differences among the groups formed by Ward's methodology. This significant reduction in subpopulations, from 11 families to five groups, confirmed the unfeasibility of distinguishing families based on variability.

The observed discrepancy between the 11 families and the identified clusters, indicating a low correlation between familial classification and exhibited variability, may be attributed to the commercial orchard origins of these families. In contrast, studies employing larger and geographically diverse samples often revealed clusters that correlate significantly with the regional origins of the accessions (Serba et al. 2019). This observation aligned with the findings of Diaz-Garcia and Padilla-Ramírez (2023), who reported considerable genetic diversity among 48 *P. guajava* L. accessions sourced from 11 different Mexican states, with eight accessions from various countries including Bolivia, Brazil, Colombia, Cuba, Honduras, India, and South Africa.

It is also noteworthy that these families emerged from two selection and controlled crossing cycles, significantly narrowing the geographic distance among parental accessions, with many sharing one or two parents (Table 1).

CONCLUSION

There was no significant differentiation between the 11 families based on phenotypic or molecular data, implying that division into families does not accurately reflect the observed diversity. Nevertheless, a clear population structure was identified through UPGMA and Ward's analyses. These results offer valuable information for the forthcoming phases of the UENF guava breeding program, highlighting the importance of considering population structure in the development of breeding strategies.

REFERENCES

- Abbaszadeh, M., Sheidai, M., Koozdar, F. and Shafieizargar, A. (2023). Population and landscape genetic studies in Citrus tangerine Tanaka. *Genetic Resources and Crop Evolution*, 70, 2695-2711. <https://doi.org/10.1007/s10722-023-01596-7>
- Ambrósio, M., Viana, A. P., Cruz, D. P., Preisigke, S. C., Cavalcante, N. R., Gonçalves Júnior, D. H., Amaral, B. D., Amaral Júnior, A. T., Santana, J. G., Crevelari, J. A., Daher, R. F. and Souza, Y. P. (2022). Categories of variables in analysis of genetic diversity in S1 progenies of *Psidium guajava*. *Scientific Reports*, 12, 22367. <https://doi.org/10.1038/s41598-022-26950-0>
- Angulo-López, J. E., Flores-Gallegos, A. C., Torres-León, C., Ramírez-Guzmán, K. N., Martínez, G. A. and Aguilar, C. N. (2021). Guava (*Psidium guajava* L.) fruit and valorization of industrialization by-products. *Processes*, 9(6), 1075. <https://doi.org/10.3390/pr9061075>
- Azevedo, A. M. (2021). *MultivariateAnalysis: pacote para análise multivariada*. Available at: <https://cran.r-project.org/web/packages/MultivariateAnalysis/MultivariateAnalysis.pdf>. Accessed on: February 3, 2025.

Azzolini, M., Jacomino, A. P. and Spoto, M. H. F. (2004). Estádios de maturação e qualidade pós-colheita de goiabas “Pedro Sato” 1. *Revista Brasileira de Fruticultura*, 26, 29-31. <https://doi.org/10.1590/S0100-29452004000100009>

Bernardes, C. O., Tuler, A. C., Canal, D., Carvalho, M. S., Ferreira, A. and Ferreira, M. S. F. (2022). Genetic diversity and population structure of psidium species from restinga: a coastal and disturbed ecosystem of the brazilian atlantic forest. *Biochemical Genetics*, 60, 2503-2514. <https://doi.org/10.1007/s10528-022-10222-7>

Cruz, C. D. (2013). GENES - A software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum*, 35, 271-276. <https://doi.org/10.4025/actasciagron.v35i3.21251>

Cruz, C. D. (2016). Genes Software – extended and integrated with the R, Matlab and Selegen. *Acta Scientiarum*, 38, 547-552. <https://doi.org/10.4025/actasciagron.v38i4.32629>

Cruz, C. D., Ferreira, F. M. and Pessoni, L. A. (2020). *Biometria aplicada ao estudo da diversidade genética* (2. Ed.). Viçosa: Viçosa.

Diaz-Garcia, L. and Padilla-Ramírez, J. S. (2023). Development of single nucleotide polymorphism markers and genetic diversity in guava (*Psidium guajava* L.). *Plants People Planet*, 5, 58-69. <https://doi.org/10.1002/ppp3.10295>

Excoffier, L., Smouse, P. E. and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131, 479-491. <https://doi.org/10.1093/genetics/131.2.479>

Feng, C., Feng, C., Lin, X., Liu, S., Li, Y., and Kang, M. (2021). A chromosome-level genome assembly provides insights into ascorbic acid accumulation and fruit softening in guava (*Psidium guajava*). *Plant Biotechnology Journal*, 19, 717-730. <https://doi.org/10.1111/pbi.13498>

Galili, T. (2015). Dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, 31, 3718-3720. <https://doi.org/10.1093/bioinformatics/btv428>

Grossi, L. L., Fernandes, M., Silva, M. A., Bernardes, C. O., Tuler, A. C., Santos, P. H. D., Ferreira, A. and Ferreira, M. F. S. (2021). DArTseq-derived SNPs for the genus *Psidium* reveal the high diversity of native species. *Tree Genetics & Genomes*, 17, 23. <https://doi.org/10.1007/s11295-021-01505-y>

Huang, K., Wang, T., Dunn, D. W., Zhang, P., Sun, H. and Baoguo, L. I. (2021). A generalized framework for AMOVA with multiple hierarchies and ploidies. *Integrative Zoology*, 16, 33-52. <https://doi.org/10.1111/1749-4877.12460>

[IBGE] Instituto Brasileiro de Geografia e Estatística (2023). Produção de goiaba. Brazil: IBGE. Available at: <https://www.ibge.gov.br/explica/producao-agropecuaria/goiaba/br>. Accessed on: Mar. 16, 2023.

Inglis, P. W., Castro, R. P. M., Resende, L. V. and Grattapaglia, D. (2018). Fast and inexpensive protocols for consistent extraction of highquality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One*, 13, 0206085. <https://doi.org/10.1371/journal.pone.0206085>

Koura, A. A., Kena, A. W., Annor, B., Adejumobi, I. I., Maina, F., Maazou, A. R., Razakou, I. B. Y. A., Attamah, P., Kusi, P., Boukar, O. and Akromah, R. (2024). DArTseq-based genome-wide SNP markers reveal limited genetic diversity and highly structured population in assembled West African cowpea germplasm. *Scientific African*, 23, e02065. <https://doi.org/10.1016/j.sciaf.2024.e02065>

Laviola, B. G., Alves, A. A., Rosado, T. B., Bhering, L. L., Formighieri, E. F. and Peixoto, L. A. (2018). Establishment of new strategies to quantify and increase the variability in the Brazilian *Jatropha* genotypes. *Industrial Crops and Products*, 117, 216-223. <https://doi.org/10.1016/j.indcrop.2018.03.003>

Mojena, R. (1977). Hierarchical grouping methods and stopping rules: an evaluation. *The Computer Journal*, 20, 359-363. <https://doi.org/10.1093/comjnl/20.4.359>

- Paiva, C. L., Viana, A. P., Santos, E. A., Freitas, J. C. O. and Amaral Junior, A. T. (2016). Genetic gain estimated by different selection criteria in guava progenies. *Bragantia*, 75, 418-427. <https://doi.org/10.1590/1678-4499.477>
- Pessanha, P. G. D. O., Viana, A. P., Amaral Júnior, A. T., Souza, R. M., Teixeira M. C. and Pereira, M. G. (2011). Avaliação da diversidade genética em acessos de *Psidium* spp. via marcadores RAPD. *Revista Brasileira de Fruticultura*, 33, 129-136. <https://doi.org/10.1590/S0100-29452011000100018>
- Pierre, N., Wamalwa, L. N., Muir, W. M., Simon, B., Kanju, E., Ferguson, M. E., Ndavi, M. M. and Tumwegamire, S. (2022). Genetic diversity of local and introduced cassava germplasm in Burundi using DArTseq molecular analyses. *PLoS One*, 17, e0256002. <https://doi.org/10.1371/journal.pone.0256002>
- Quintal, S. S. R., Viana, A. P., Campos, B. M., Vivas, M. and Amaral Júnior, A. T. (2017). Selection via mixed models in segregating guava families based on yield and quality traits. *Revista Brasileira de Fruticultura*, 39, e-866. <https://doi.org/10.1590/0100-29452017866>
- Serba, D. D., Muleta, K. T., Amand, P., Bernardo, A., Bai, G., Perumal, R. and Bashir, E. (2019). Genetic diversity, population structure, and linkage disequilibrium of Pearl Millet. *The Plant Genome*, 12, 180091. <https://doi.org/10.3835/plantgenome2018.11.0091>
- Silva, F. A., Viana, A. P., Cavalcante, N. R., Amaral Júnior, A. T. and Pereira, M. G. (2022). Strategies and practical aspects in guava (*Psidium guajava* L.) breeding: 12 years of experiments. *Functional Plant Breeding Journal*, 4, 1-12. <https://doi.org/10.35418/2526-4117/v4n2a1>
- Silva, M. J. R., Tecchio, M. A., Domiciano, S., Leonel, S. and Balestrero, R. I. (2016). Phenology, yield and fruit quality of 'Paluma' guava tree at different pruning times. *Ciência e Agrotecnologia*, 40, 317-325. <https://doi.org/10.1590/1413-70542016403033315>
- Singh, D. (1981). The relative importance of characters affecting genetic divergence. *Indian Journal of Genetics e Plant Breeding*, 41, 237-245.

Singh, D., Gill, M. I. S., Boora, R. S. and Arora, N. K. (2015). Genetic diversity analysis in guava (*Psidium guajava*) on the basis of morphological and physico-chemical traits. *Indian Journal of Agricultural Sciences*, 85, 678-683. <https://doi.org/10.56093/ijas.v85i5.48504>

Thakur, S., Yadav, I. S., Jindal, M., Sharma, P. K., Dhillon, G. S., Boora, R. S., Arora, N. K., Gill, M. I. S., Chhuneja, P. and Mittal, A. (2021). Development of genome-wide functional markers using draft genome assembly of guava (*Psidium guajava* L.) cv. Allahabad Safeda to expedite molecular breeding. *Frontiers in Plant Science*, 12, 708332. <https://doi.org/10.3389/fpls.2021.708332>

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. and Yutani, H. (2019). Welcome to Tidyverse. *Journal of Open-Source Software*, 4, 1686. <https://doi.org/10.21105/joss.01686>

Yousaf, A. A., Abbasi, K. S., Ahmad, A., Hassan, I., Sohail, A., Qayyum, A. and Akram, M. A. (2021). Physico-chemical and nutraceutical characterization of selected indigenous guava (*Psidium guajava* L.) cultivars. *Food Science and Technology*, 41, 47-58. <https://doi.org/10.1590/fst.35319>

4.2 Modelagem mista via marcadores SNPs na implementação da seleção genômica em uma população de goiabeiras

RESUMO

A seleção genômica tem se mostrado uma ferramenta promissora para acelerar o melhoramento genético de espécies perenes, como a goiabeira, permitindo a predição de valores genéticos genômicos mais precisos e econômicos. Neste estudo, foram comparadas metodologias de modelagem mista de BLUP tradicional (utilizando matriz de parentesco médio), RRBLUP (usando os marcadores moleculares como variáveis de regressão na composição do valor fenotípico) e GBLUP (que utiliza uma matriz de parentesco genômico, baseada na informação dos marcadores genéticos), visando avaliar o efeito da inserção de informações moleculares na capacidade preditiva e eficiência de seleção genômica para as características massa, comprimento e diâmetro do fruto, massa da polpa e da placenta, espessura do endocarpo, do mesocarpo, e da casca, teor de sólidos solúveis (BRIX) e produção total no período de uma safra. As 11 famílias de irmãos-completos utilizadas no estudo foram obtidas no âmbito do Programa de Melhoramento Genético da Goiabeira da UENF (Universidade Estadual do Norte Fluminense), iniciado com a coleta de acessos em pomares comerciais e plantas de polinização aberta dos municípios de São João da Barra e São Francisco de Itabapoana, no estado do Rio de Janeiro. A partir da seleção de genitores divergentes e heterozigotos, realizaram-se cruzamentos controlados e sucessivos ciclos de avaliação e seleção, resultando nas famílias atualmente estudadas. Os resultados indicaram desempenho superior do modelo BLUP tradicional para todas as variáveis, além de destacarem a efetividade da métrica Rank-Biased Overlap (RBO), na comparação entre os rankings de indivíduos selecionados pelos diferentes modelos de predição.

Palavras-chave: Ganho de seleção; Melhoramento de frutíferas; Predição de valores genéticos; *Rank-Biased Overlap*.

INTRODUÇÃO

A goiabeira (*Psidium guajava* L.) é uma frutífera originária da região tropical das américas, mais especificamente na região que abrange desde o México até o noroeste da Argentina (Diaz-Garcia et al., 2023). Estudos genéticos realizados por Arévalo-Marín e colaboradores (2024) indicam que o processo de domesticação da espécie iniciou-se a partir da Amazônia brasileira, espalhando-se para os Andes peruanos, norte da América do Sul, América Central e México. Após sua domesticação, a goiabeira foi rapidamente disseminada por exploradores portugueses e espanhóis para várias regiões tropicais no mundo (Kumari et al., 2018).

A goiaba é uma das frutas tropicais mais cultivadas no Brasil, destacando-se pelo alto valor nutricional, com destaque para fibras, ácido ascórbico e compostos fenólicos. Devido à sua natureza climatérica e alta perecibilidade, é amplamente empregada como matéria-prima em sucos, polpas, doces e purês. A goiaba também possui ampla relevância medicinal. Suas folhas são usadas na medicina tradicional para prevenir e tratar diversas doenças (Menezes et al., 2021). Os compostos extraídos de frutos, cascas e raízes apresentam propriedades antimicrobianas, hepatoprotetoras, antidiabéticas e anti-inflamatórias (Chu et al., 2022).

O melhoramento genético vem desempenhando um papel crucial na agricultura, impactando a segurança alimentar, a sustentabilidade e a viabilidade econômica. Particularmente, no que diz respeito à goiaba, os programas de melhoramento genético promovem o aumento do rendimento e qualidade dos frutos a partir do desenvolvimento de variedades de goiabeiras com frutas de qualidade superior, aumentando a produtividade, resistência a doenças e melhorando a adaptabilidade das plantas às mudanças ambientais (Kumar et al., 2023).

A utilização de BLUPs (*Best Linear Unbiased Predictions*), obtidos a partir de dados fenotípicos e matrizes de parentesco médio, desempenha um papel central na seleção de indivíduos para cruzamentos controlados, ao permitir a estimativa acurada dos valores genéticos aditivos e, conseqüentemente, a maximização do progresso genético ao longo das gerações (Maan et al., 2023).

Contudo, quando se trata de espécies perenes, os programas de melhoramento genético enfrentam desafios como longos ciclos de produção, porte elevado das plantas, alta heterozigosidade, prolongado período juvenil, complexidade de caracteres quantitativos, dependência de avaliações fenotípicas demoradas, restrição da variabilidade genética e baixa capacidade de resposta a mudanças climáticas (Jaiswal, 2024).

Os marcadores moleculares surgem como ferramentas inovadoras, permitindo a identificação direta de variações genéticas no nível do DNA. Com isso, é possível acelerar o desenvolvimento de novas cultivares, reduzir custos operacionais e aprimorar a precisão das previsões genéticas, tornando o melhoramento mais eficiente e estratégico (Valdisser et al., 2017; Inglis et al., 2018; Grossi et al., 2021; Koura et al., 2024).

Os marcadores de polimorfismo de nucleotídeo único (SNP) são ferramentas essenciais no melhoramento genético de plantas tropicais e perenes, devido à sua alta densidade, codominância, custo-efetividade e facilidade de automação. Eles permitem identificar genes associados a características agronômicas importantes (Das et al., 2022).

A seleção genômica (GW) é uma metodologia de melhoramento genético que utiliza informações de milhares de marcadores moleculares distribuídos ao longo do genoma para prever o valor genético de indivíduos, mesmo antes da manifestação de suas características fenotípicas. Duas importantes metodologias estatísticas são empregadas nesse sentido: *Ridge Regression Best Linear Unbiased Predictor* - Regressão de Crista Melhor Preditor Linear Não Viesado (RRBLUP) e *Genomic Best Linear Unbiased Predictor* - Melhor Previsor Genômico Linear Não Viesado (GBLUP).

Diante disso, o estudo aqui apresentado foi desenvolvido a partir de goiabeiras inseridas no programa de melhoramento genético da UENF. O objetivo foi verificar os efeitos do uso de informações dos marcadores genéticos do tipo SNP, desenvolvido pela EMBRAPA para uso no eucalipto (*Axiom 72K ESAI Eucalyptus SNP Array Initiative*), na seleção genômica das goiabeiras, no segundo ciclo de seleção e com reduzida variabilidade genética.

MATERIAL E MÉTODOS

Material Vegetal

Foram avaliados 247 genótipos de goiabeira organizados em 11 famílias de irmãos completos. O experimento foi conduzido na área experimental da Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), no Colégio Agrícola Antônio Sarlo, localizado em Campos dos Goytacazes, RJ, Brasil (21°08'02" S; 41°40'47" W). A região apresenta clima tropical subúmido e seco, com temperaturas médias anuais entre 22 e 25 °C e precipitação média anual de 1200 a 1300 mm. Os acessos que deram origem às famílias avaliadas foram coletados em pomares comerciais dos municípios de São João da Barra e São Francisco de Itabapoana, no estado do Rio de Janeiro (Quintal et al., 2017).

Tabela 1: Informações de parentesco das 11 famílias de irmãos completos geradas a partir da seleção por Quintal et al. (2017), Campos dos Goytacazes, RJ, Brasil.

Família	Parentesco*	Família	Parentesco*
1	8.4 II x 5.4 II	7	12.11 II x 12.4 II
2	10.5 I x 12.4 II	8	12.11 II x 3.11 II
3	12.11 II x 8.4 I	9	12.11 x 7.9
4	4.5 x 2.6	10	8.1 II x 12.4 II
5	12.11 I x 8.4 I	11	12.4 II x 12.11 II
6	12.4 x 12.11		

* A coluna de parentesco indica os genitores que deram origem às famílias estudadas.

Condução do Experimento

O plantio seguiu-se em três fileiras de goiabeiras. As mudas foram cultivadas com um espaçamento de 1,2 metros entre cada planta e 3,5 metros entre fileiras. Foi implementado um sistema de irrigação por gotejamento. A adubação seguiu os resultados obtidos através da análise de solo, sendo realizada em três aplicações sucessivas de ureia (123 g), superfosfato simples (315 g) e cloreto de potássio (87 g) por planta, em intervalos de 30 dias após a poda para frutificação. O manejo da cultura e as podas foram executados conforme as orientações técnicas para o cultivo da goiabeira. O experimento utilizou um delineamento em blocos casualizados com três repetições. Cada bloco experimental continha oito plantas de

cada família, sendo as variáveis avaliadas em cinco frutos escolhidos aleatoriamente. Estes frutos apresentavam casca com coloração verde-amarelada e um ângulo de cor entre 112° e 108°, caracterizando o estágio 3 de maturação (Azzolini et al., 2004).

Variáveis Avaliadas

Das 247 plantas que passaram pelo processo de genotipagem, apenas 235 produziram frutos durante o período de avaliação. Foram avaliados cinco frutos por planta e consideradas as seguintes variáveis: Massa individual do fruto (MASSA), medida com uma balança semi-analítica e apresentada em gramas (g); Comprimento do fruto (COMP), avaliado na direção longitudinal com um paquímetro digital, com valores expressos em milímetros (mm); Diâmetro transversal do fruto (DIA), correspondente ao comprimento da região equatorial do fruto, medido com um paquímetro digital e expresso em milímetros (mm); Massa da polpa (MPO), extraída a polpa e pesada com uma balança semi-analítica, os resultados foram expressos em gramas (g); Massa da placenta (MPL), correspondente à área onde estão concentradas as sementes, avaliada com uma balança semi-analítica; Espessura do endocarpo (EE), medida na região das sementes com um paquímetro digital e expressa em milímetros (mm); Espessura do mesocarpo (EM), definida como a camada entre a casca e a placenta, avaliada com um paquímetro digital e registrada em milímetros (mm); Espessura da casca (EC), correspondente à camada externa do fruto, mensurada com um paquímetro digital e expressa em milímetros (mm); Teor de sólidos solúveis (SST), determinado com um refratômetro digital Atago nº 1 e expresso em graus Brix (°Brix). A variável Produção total (PROD) foi calculada somando as massas de todos os frutos de cada planta, utilizando uma balança eletrônica e expressando os valores em gramas (g).

Marcadores Moleculares e Genotipagem

Foram coletadas amostras de 247 plantas de goiabeira e submetidas à genotipagem utilizando o SNP array Axiom 72K ESAI Eucalyptus SNP Array Initiative, desenvolvido pela Embrapa. O DNA genômico foi extraído de folhas jovens completamente desenvolvidas, conforme o protocolo descrito por Inglis et al. (2018).

A identificação dos marcadores moleculares foi realizada por meio da tecnologia de genotipagem por sequenciamento DArTseq (Diversity Arrays Technology), desenvolvida e otimizada pela empresa DArT P/L em Canberra, Austrália (www.diversityarrays.com).

Foram identificados 72.202 marcadores. Posteriormente, os dados passaram por filtros de qualidade, que promoveram a eliminação de SNPs com frequência do alelo menos comum (MAF) inferior ou igual a 5 %, o descarte de marcadores com mais de 10 % de dados ausentes, a remoção de SNPs que não apresentavam variação e que apresentavam duplicatas, mantendo-se neste último filtro, apenas um marcador. Os dados ausentes foram imputados utilizando o alelo mais frequente em cada indivíduo, resultando em 7.874 marcadores informativos.

Análise de Dados

Foram utilizadas três metodologias distintas de predição de valores genéticos: BLUP tradicional, RRBLUP e GBLUP. O BLUP tradicional baseia-se exclusivamente em informações de pedigree, estimando os valores genéticos aditivos a partir da matriz de parentesco médio, sendo amplamente aplicado em programas de melhoramento por permitir a separação dos efeitos genéticos dos efeitos ambientais. O RRBLUP, por sua vez, incorpora dados genômicos por meio da regressão ridge, tratando simultaneamente milhares de marcadores SNP como variáveis regressoras, sob a premissa de que todos os loci contribuem de forma igual e independente para a variância genética. Já o GBLUP substitui a matriz de parentesco tradicional pela matriz de parentesco genômico, construída diretamente a partir das informações dos marcadores moleculares.

Em todas essas metodologias, a predição dos valores genéticos foi realizada no contexto de modelos mistos (Eq. 1), nos quais os efeitos fixos e os efeitos aleatórios são estimados de forma conjunta por meio da solução das equações de Henderson. Essa metodologia permite obter estimativas não viesadas e mais precisas, especialmente quando se trabalha com grande volume de dados genômicos, garantindo maior eficiência na seleção de indivíduos superiores.

$$y = Xb + Zu + e \quad (1)$$

Onde:

- Xb é a parcela do modelo referente aos efeitos fixos:
 - X : é a matriz de incidência dos efeitos fixos (blocos, tratamentos ou covariáveis como densidade de plantio ou ano de experimento).
 - b : é o vetor de efeitos fixos.
- Zu é a parcela referente aos efeitos aleatórios (genéticos aditivos, de dominância, epistáticos entre outros):
 - Z : matriz de incidência que relaciona os dados fenotípicos aos efeitos genéticos individuais estimados.
 - u é o vetor de efeitos aleatórios genéticos.
- e (erro residual):
 - Representa os desvios ou variações não explicadas pelo modelo.
 - Inclui influências ambientais e outros fatores aleatórios não contabilizados.

As equações matriciais de modelos mistos, apresentadas inicialmente por Henderson na década de 1940, foram utilizadas na estimativa dos efeitos fixos e na predição dos valores genéticos individuais (Eq. 2):

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \quad (2)$$

Onde:

- G a matriz de covariância dos efeitos genéticos ($G = A\sigma_u^2$ no BLUP tradicional e $G = H\sigma_u^2$ no GBLUP);
- $\lambda = \frac{\sigma_e^2}{\sigma_u^2}$ é a razão entre as variâncias residual e genética;
- A é a matriz de parentesco médio (BLUP tradicional) ou a matriz genômica G (GBLUP).

Na metodologia RRBLUP a matriz de parentesco não é usada e a matriz Z é composta pelas informações das marcas SNPs. Assim sendo, a matriz $\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1}\lambda \end{bmatrix}$ é substituída por $\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda I \end{bmatrix}$ onde I é a matriz identidade. Assume-se então que todos os SNPs contribuem igualmente para a variabilidade

genética. As estimativas dos componentes de variância se deram pelo método REML (Restricted Maximum Likelihood).

As análises foram realizadas com o auxílio dos softwares R e RSTUDIO. A resolução das equações de modelos mistos se deu através da biblioteca rrBLUP (Endelman, 2011).

Foram utilizados 6 parâmetros para contrastar as modelagens utilizado as abordagens BLUP, GBLUP e RRBLUP:

- σ_a^2 é a variância genética.
- σ_e^2 é a variância do erro.
- $h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$ herdabilidade no sentido restrito.
- $\hat{r}_g$ é a capacidade preditiva.
- $\hat{b}$ é o viés de predição.
- MSE é o erro quadrático médio.

A seleção se deu a partir do ranqueamento dos indivíduos, em função de cada uma das variáveis analisadas, selecionando os trinta mais bem colocados. Foram computados os valores fenotípicos, os BLUPs obtidos, os valores genéticos preditos e ganho de seleção.

Foram selecionados os 30 melhores indivíduos para cada característica analisada, sendo que, para as características massa, comprimento e diâmetro do fruto, massa da polpa, espessura do mesocarpo, brix e produção, foram selecionados os 30 indivíduos que apresentaram os maiores valores de BLUPs. Para as variáveis espessura da casca, do endocarpo e massa da placenta foram selecionados os 30 indivíduos que apresentaram menor valores de BLUPs.

Os ranqueamentos foram comparados usando a métrica chamada *Rank-Biased Overlap* (RBO) (Eq. 3) que mede a similaridade enquanto permite que os elementos no topo do ranking tenham mais peso no cálculo (Corsi et al., 2024).

$$RBO = (1 - p) \sum_{d=1}^D \frac{|A_{1:d} \cap B_{1:d}|}{d} p^d \quad (3)$$

Onde:

- $A_{1:d}$ e $B_{1:d}$ são os elementos nos rankings A e B até a profundidade d ;

- $|A_{1:d} \cap B_{1:d}|$ é o número de elementos em comum entre os rankings $A_{1:d}$ e $B_{1:d}$;
- d é a profundidade do ranking (de 1 até o tamanho do maior ranking D);
- p é o fator de decaimento que pondera as primeiras posições, geralmente utilizado entre 0,8 e 0,98;
 - Valores altos de p dão mais peso às primeiras posições do ranking (no trabalho aqui apresentado foi utilizado o valor de 0,9);
 - $1 - p$ é um fator de normalização.

Quanto mais próximo de 1 for o valor de RBO, maior a similaridade entre os ranqueamentos. Valores próximos de 0 (zero) indicam ranqueamentos bem diferentes.

RESULTADOS E DISCUSSÃO

A Tabela 2 apresenta a comparação entre os modelos BLUP tradicional, RRBLUP e GBLUP na predição dos valores genéticos genômicos para dez características. Modelos que não se ajustaram, resultando em herdabilidade nula, foram desconsiderados.

Tabela 2: Comparação das três abordagens utilizadas na previsão dos valores genéticos genômicos.

Modelos	Par.	Variáveis									
		M	C	D	Ec	Em	Ee	Mpo	Mpl	SST	P
BLUP tradicional	h^2	0,85	0,99	0,87	0,19	0,13	0,80	0,67	0,20	0,10	0,19
	σ_a^2	2502,99	102,56	46,68	0,07	0,47	20,89	1170,43	112,32	0,26	1372024
	σ_e^2	425,98	0,01	6,73	0,29	3,20	5,18	567,90	441,56	2,46	5912477
	$\hat{r}_{gg}$	0,81	0,46	0,44	0,51	0,46	0,49	0,80	0,68	0,25	0,65
	$\hat{b}$	-0,21	0,03	-0,22	-1,41	-2,84	-0,27	-0,51	-2,30	-2,36	-1,43
	MSE	1239,19	254,72	182,63	0,41	9,59	61,90	891,76	424,18	10,18	5294007
rrBLUP	h^2	0,91	0,91	0,91	0,04	0,00	0,92	0,00	0,65	0,07	0,00
	σ_a^2	14370,3	468,09	262,55	0,01	-	149,39	-	889,22	0,18	-
	σ_e^2	1484,80	45,12	27,45	0,35	-	12,56	-	475,41	2,65	-
	$\hat{r}_{gg}$	0,59	0,32	0,34	0,17	-	0,35	-	0,49	0,18	-
	$\hat{b}$	-0,63	-0,33	-0,84	-3,31	-	-0,43	-	-1,82	-3,95	-
	MSE	2307,10	291,85	203,40	0,49	-	70,42	-	480,23	10,38	-
GBLUP	h^2	0,80	0,81	0,80	0,02	0,00	0,84	0,00	0,43	0,03	0,00
	σ_a^2	6120,42	205,81	113,26	0,01	-	67,74	-	366,95	0,10	-
	σ_e^2	1564,58	47,32	28,73	0,35	-	13,00	-	482,03	2,65	-

$\hat{r}_{gg}$	0,57	0,31	0,33	0,17	-	0,36	-	0,46	0,18	-
$\hat{b}$	-0,69	-0,38	-0,92	-3,36	-	-0,49	-	-1,96	-3,86	-
<i>MSE</i>	2388,51	293,14	204,88	0,49	-	70,46	-	489,57	10,38	-

Variáveis analisadas: M - massa do fruto, C - comprimento do fruto, D - diâmetro do fruto, Ec - espessura da casca, Em - espessura do mesocarpo, Ee - espessura do endocarpo, Mpo - massa da polpa, Mpl - massa da placenta, SST - sólidos solúveis totais e Produção total. Modelos não ajustados, com herdabilidade zero, não foram contabilizados. Os parâmetros analisados foram: herdabilidade estimada (h^2), variância genética aditiva (σ_a^2), variância residual (σ_e^2), capacidade preditiva ($\hat{r}_{gg}$), vies de predição ($\hat{b}$) e erro quadrado médio (MSE).

Os valores de herdabilidade estimados pelos modelos para cada variável analisada foram classificados em três faixas: baixa ($h^2 \leq 0,3$), moderada ($0,3 < h^2 \leq 0,6$) e alta ($h^2 > 0,6$) (Azevedo et al., 2021). As metodologias BLUP tradicional, RRBLUP e GBLUP apresentaram baixa herdabilidade para espessura da casca, espessura do mesocarpo, teor de sólidos solúveis (Brix) e produção total. Esse resultado indica forte influência ambiental e baixa proporção da variância genética aditiva nesses caracteres. Resultados como esse podem estar associados à complexidade poligênica, ao efeito de interações genótipo $\times$ ambiente e à maior suscetibilidade a erros de mensuração. Dessa forma, esses caracteres mostram-se menos responsivos à seleção direta, exigindo estratégias indiretas ou maior precisão experimental para ganhos consistentes. A massa da polpa também apresentou baixa herdabilidade nos modelos RRBLUP e GBLUP, assim como a massa da placenta no modelo BLUP tradicional. Para esta última característica, apenas o GBLUP apresentou herdabilidade moderada.

Por outro lado, as três metodologias indicaram alta herdabilidade para massa, comprimento e diâmetro do fruto, bem como para a espessura do endocarpo. Além disso, o modelo RRBLUP também apresentou alta herdabilidade para a massa da placenta.

Observa-se que os modelos RRBLUP e GBLUP não se ajustaram adequadamente para espessura do mesocarpo, massa da polpa e produção total, resultando em herdabilidade próxima de zero ($h^2 \leq 0,01$). Esse comportamento sugere baixa variabilidade genética, indicando que os marcadores não captaram variação suficiente para estas características, possivelmente devido à forte influência ambiental e à homogeneidade genética entre os materiais avaliados.

A metodologia RRBLUP proporcionou as maiores estimativas de herdabilidade para as características de massa, comprimento e diâmetro do fruto, espessura do endocarpo e massa da placenta, em comparação com as demais abordagens avaliadas. Esses resultados indicam que o uso de marcadores genéticos como variáveis regressoras, com efeitos distribuídos de forma uniforme sobre a variância genética, pode favorecer a predição dos valores genéticos genômicos para essas características. Embora valores mais elevados de herdabilidade estejam associados, em geral, a maior acurácia na predição, essa relação não é determinística, uma vez que a acurácia também depende de fatores como o tamanho da população de treinamento, a densidade dos marcadores e a estrutura genética da população avaliada.

A estimativa da herdabilidade é um componente fundamental na seleção de modelos para predição genômica, pois fornece uma medida da proporção da variância fenotípica atribuída à variância genética aditiva em uma determinada população. Ao comparar metodologias de predição, a herdabilidade funciona como um indicador de quão bem cada modelo consegue capturar a variação genética subjacente aos dados fenotípicos, permitindo avaliar a eficácia e robustez do modelo na seleção. Além disso, herdabilidades elevadas estão associadas a maior potencial de ganho genético, sendo, portanto, um critério útil para priorizar a escolha de metodologias de predição (Srivastava et al., 2023).

A avaliação do desempenho dos modelos foi feita por meio da capacidade preditiva (r_{gg}), que é obtida através da correlação entre os valores genômicos preditos e observados como descrito em (Resende et al., 2012), viés de predição ($\hat{b}$) e MSE (*Mean Squared Error* - Erro Quadrático Médio). O BLUP tradicional apresentou a maior capacidade preditiva e menores valores de MSE para todas as características, indicando que se ajustou melhor aos dados. Além disso, esse modelo demonstrou viés de predição mais próximo de zero para comprimento e diâmetro do fruto, além da espessura do endocarpo. Os modelos RRBLUP e GBLUP também apresentaram viés próximo de zero para diâmetro do fruto e espessura do endocarpo.

No presente estudo, BLUP tradicional apresentou desempenho superior aos demais para todas as variáveis avaliadas. Esse resultado contrasta com o

observado por Peixoto e colaboradores (2024). O estudo avaliou como diferentes valores de herdabilidade e números de QTL influenciam a predição do valor genético por métodos de seleção genômica, utilizando uma população F_2 simulada com 1.000 indivíduos. Foram comparados quatro métodos (RR-BLUP, GBLUP, Bayes B e RKHS), e os resultados mostraram que Bayes B e RR-BLUP foram superiores na maioria dos cenários. Constatou-se ainda que a herdabilidade apresentou relação linear positiva com a habilidade preditiva fenotípica e quadrática positiva com a genotípica, enquanto o número de QTL não teve relação com essas habilidades. A divergência nos resultados pode ser atribuída às diferenças estruturais entre as populações avaliadas: enquanto o estudo citado utilizou uma população F_2 simulada, com ampla variabilidade genética, a população utilizada neste trabalho foi composta por famílias de irmãos completos, cuja estrutura genética é mais restrita e a menor variabilidade.

A Tabela 3 apresenta os indivíduos selecionados com base nos maiores ganhos genéticos preditos para cada característica avaliada, considerando os modelos BLUP tradicional, RRBLUP e GBLUP; ressalta-se que, para as variáveis espessura do mesocarpo, massa da polpa e produção total, os modelos genômicos (RRBLUP e GBLUP) não se ajustaram adequadamente, impossibilitando sua aplicação nessas características.

Tabela 3: Tabela com os 30 indivíduos selecionados com base nos melhores ganhos genéticos preditos para cada característica avaliada, considerando os modelos BLUP tradicional, RRBLUP e GBLUP.

MASSA DO FRUTO						COMPRIMENTO DO FRUTO					
BLUP TRADICIONAL		RRBLUP		GBLUP		BLUP TRADICIONAL		RRBLUP		GBLUP	
Planta	Ganho (%)	Planta	Ganho (%)	Planta	Planta	Planta	Ganho (%)	Planta	Ganho (%)	Planta	Ganho (%)
R1T9P1	106,97	R1T2P4	55,52	R1T2P4	52,09	R1T10P5	39,11	R1T2P4	16,83	R1T2P4	19,06
R1T10P5	90,62	R1T5P5	46,47	R1T5P5	43,44	R1T9P1	36,81	R2T10P5	14,09	R2T10P5	16,27
R3T3P6	82,42	R1T9P1	42,20	R3T9P7	37,86	R3T3P6	34,59	R2T2P7	12,92	R1T10P5	15,02
R1T2P4	78,11	R3T9P7	39,24	R1T10P5	34,81	R2T10P6	33,10	R2T5P7	12,33	R1T5P5	14,35
R2T10P6	74,75	R2T2P7	37,13	R2T10P5	32,92	R2T10P5	31,83	R1T5P5	11,94	R2T10P8	13,89
R1T2P1	72,07	R2T5P7	35,63	R2T5P7	31,46	R1T2P4	30,96	R1T10P5	11,67	R2T2P7	13,57
R2T9P2	69,54	R1T10P5	34,52	R2T2P7	30,40	R2T10P3	29,90	R2T10P6	11,46	R2T5P7	13,32
R1T2P3	67,24	R2T10P6	33,55	R2T10P2	29,57	R2T10P8	28,92	R2T10P8	11,28	R3T10P7	13,10
R2T10P3	65,45	R2T10P5	32,71	R3T10P7	28,84	R2T1P1	28,05	R2T10P3	11,01	R1T10P1	12,82

R2T3P6	63,81	R1T11P7	31,99	R1T9P1	28,21	R2T3P6	27,34	R2T10P2	10,79	R2T10P6	12,57
R1T1P3	62,10	R2T10P2	31,39	R1T10P1	27,68	R1T2P3	26,70	R3T10P7	10,57	R2T10P2	12,36
R2T10P2	60,61	R2T10P3	30,81	R2T1P5	27,20	R2T5P7	26,15	R3T9P7	10,37	R3T9P7	12,16
R2T10P5	59,33	R2T1P5	30,28	R2T10P8	26,78	R1T1P3	25,64	R3T1P5	10,18	R3T10P4	11,97
R1T3P3	58,18	R3T10P7	29,77	R1T11P7	26,43	R2T2P7	25,17	R3T10P3	10,02	R2T10P4	11,80
R1T9P2	57,03	R1T10P1	29,31	R2T10P3	25,94	R2T10P4	24,71	R3T10P1	9,88	R2T1P5	11,64
R2T5P7	55,87	R2T10P8	28,87	R2T10P6	25,51	R1T2P1	24,31	R1T3P2	9,74	R2T10P3	11,50
R3T3P3	54,74	R3T10P4	28,39	R3T3P6	25,10	R1T3P2	23,94	R1T10P1	9,62	R1T3P2	11,36
R2T2P7	53,72	R1T3P3	27,92	R3T10P4	24,73	R1T3P6	23,62	R2T1P5	9,52	R2T10P7	11,23
R2T1P1	52,74	R3T10P1	27,48	R3T10P6	24,37	R1T3P3	23,30	R1T10P4	9,42	R3T10P3	11,11
R1T5P5	51,84	R3T10P6	27,05	R2T10P4	23,98	R1T3P5	22,99	R3T10P4	9,33	R1T3P5	10,99
R3T3P4	51,02	R3T3P6	26,65	R1T3P5	23,62	R3T3P3	22,67	R2T10P4	9,25	R3T1P5	10,88
R3T3P2	50,26	R3T3P3	26,28	R2T9P2	23,25	R1T9P7	22,36	R1T10P6	9,16	R3T10P5	10,78
R1T9P7	49,56	R2T3P6	25,94	R1T3P8	22,91	R2T4P6	22,08	R1T3P5	9,08	R3T10P1	10,68
R2T10P8	48,89	R1T3P5	25,62	R1T10P6	22,57	R2T10P2	21,79	R1T1P4	9,00	R1T10P4	10,58
R1T1P5	48,25	R1T10P6	25,32	R3T10P1	22,27	R1T9P2	21,51	R3T10P5	8,91	R1T10P6	10,49
R3T1P4	47,65	R2T9P2	25,04	R3T10P5	21,98	R1T5P5	21,21	R1T9P1	8,83	R1T1P4	10,40
R3T9P7	47,08	R1T7P5	24,76	R1T7P5	21,70	R2T9P7	20,94	R3T10P8	8,76	R3T10P6	10,31
R1T7P7	46,47	R1T3P8	24,50	R2T10P1	21,43	R2T5P4	20,67	R2T10P7	8,69	R3T10P8	10,23
R2T9P5	45,89	R3T1P5	24,24	R1T3P2	21,18	R1T8P8	20,41	R1T3P3	8,59	R2T5P8	10,14
R2T5P6	45,33	R3T10P5	23,98	R3T1P5	20,95	R3T9P7	20,16	R2T5P8	8,49	R3T3P6	10,05
DIÂMETRO DO FRUTO						ESPESSURA DA CASCA					
BLUP TRADICIONAL		RRBLUP		GBLUP		BLUP TRADICIONAL		RRBLUP		GBLUP	
Planta	Ganho (%)	Planta	Ganho (%)	Planta	Planta	Planta	Ganho (%)	Planta	Ganho (%)	Planta	Ganho (%)
R2T9P2	20,01	R1T2P4	14,60	R1T2P4	11,84	R1T4P8	-14,36	R3T9P2	-2,64	R3T9P2	-6,26
R3T3P6	19,45	R1T5P5	13,70	R1T5P5	10,95	R2T3P5	-14,34	R3T7P4	-2,63	R3T7P4	-6,23
R1T9P1	19,10	R3T9P7	13,20	R3T9P7	10,36	R1T4P2	-14,32	R2T11P8	-2,60	R2T11P8	-6,20
R1T10P5	18,74	R2T10P2	12,44	R2T10P2	9,70	R1T4P7	-14,28	R1T9P6	-2,52	R1T9P6	-6,12
R2T10P2	18,52	R2T2P7	11,94	R1T10P1	9,29	R3T3P1	-14,24	R3T10P2	-2,38	R3T10P2	-5,99
R1T2P1	18,28	R1T10P1	11,57	R2T10P5	8,92	R2T3P7	-14,16	R2T11P7	-2,26	R2T11P7	-5,87
R2T3P6	17,97	R2T1P5	11,25	R1T10P5	8,65	R2T4P2	-14,01	R3T11P1	-2,15	R1T1P4	-5,77
R2T2P2	17,72	R1T7P5	10,99	R2T1P5	8,42	R1T4P3	-13,89	R1T1P4	-2,07	R3T11P1	-5,69
R1T2P4	17,50	R1T10P5	10,79	R2T2P7	8,24	R1T4P5	-13,77	R2T7P7	-2,00	R3T1P7	-5,62
R1T1P3	17,29	R1T9P1	10,60	R2T10P8	8,04	R3T4P1	-13,68	R3T1P7	-1,94	R2T7P7	-5,55
R2T9P5	17,11	R2T10P6	10,44	R1T7P5	7,88	R1T3P4	-13,60	R3T6P4	-1,87	R3T6P4	-5,49
R1T2P3	16,90	R3T7P4	10,31	R2T9P2	7,71	R2T3P3	-13,53	R3T5P3	-1,68	R3T5P3	-5,31
R1T9P2	16,71	R2T10P5	10,19	R3T7P4	7,57	R1T4P6	-13,45	R1T10P8	-1,51	R1T10P8	-5,15
R2T10P6	16,54	R2T9P2	10,09	R3T3P6	7,44	R3T3P4	-13,37	R2T1P6	-1,36	R2T1P6	-5,01
R1T3P3	16,39	R1T9P2	9,99	R3T10P4	7,32	R2T3P8	-13,27	R1T9P5	-1,23	R1T9P5	-4,88
R3T3P4	16,24	R3T9P3	9,89	R1T11P7	7,19	R3T4P2	-13,19	R2T3P7	-1,11	R2T3P7	-4,77
R3T3P2	16,05	R3T10P4	9,79	R3T10P5	7,07	R2T4P3	-13,12	R3T2P7	-1,01	R3T2P7	-4,67
R2T2P3	15,87	R2T5P7	9,70	R2T5P7	6,96	R3T3P5	-13,05	R1T3P5	-0,91	R1T3P5	-4,58
R1T1P5	15,70	R1T11P7	9,61	R3T10P6	6,87	R1T3P2	-12,98	R3T4P1	-0,82	R3T4P1	-4,50
R3T9P3	15,54	R2T3P6	9,53	R3T9P3	6,78	R1T4P4	-12,92	R3T2P6	-0,74	R3T2P6	-4,42

R1T7P5	15,38	R1T3P3	9,45	R2T10P6	6,69	R1T4P1	-12,87	R2T5P7	-0,66	R2T5P7	-4,35
R1T9P7	15,22	R2T10P8	9,38	R1T3P8	6,61	R2T4P1	-12,80	R2T4P8	-0,59	R2T4P8	-4,28
R1T2P5	15,05	R3T10P5	9,31	R1T9P2	6,53	R3T3P6	-12,73	R3T6P3	-0,50	R3T6P3	-4,20
R2T3P1	14,90	R3T10P1	9,24	R3T10P7	6,46	R1T3P1	-12,66	R2T1P7	-0,42	R2T1P7	-4,12
R2T9P1	14,76	R2T9P1	9,17	R1T5P1	6,39	R2T3P1	-12,60	R1T5P5	-0,34	R1T5P5	-4,05
R3T9P7	14,61	R1T5P1	9,11	R2T10P4	6,32	R3T3P3	-12,54	R2T4P7	-0,27	R2T4P7	-3,97
R2T2P7	14,46	R3T3P3	9,04	R3T1P5	6,26	R3T4P6	-12,49	R1T2P4	-0,19	R1T2P4	-3,90
R3T3P3	14,32	R3T3P6	8,98	R2T10P1	6,20	R3T4P5	-12,43	R1T11P7	-0,11	R1T11P7	-3,83
R2T1P1	14,18	R3T3P2	8,93	R2T10P3	6,14	R1T3P5	-12,38	R2T3P8	-0,02	R2T3P8	-3,74
R1T5P5	14,04	R2T10P3	8,87	R3T3P2	6,08	R3T3P2	-12,33	R3T1P6	0,06	R3T1P6	-3,66
ESPESSURA DO ENDOCARPO						MASSA DA PLACENTA					
BLUP TRADICIONAL		RRBLUP		GBLUP		BLUP TRADICIONAL		RRBLUP		GBLUP	
Planta	Ganho (%)	Planta	Ganho (%)	Planta	Ganho (%)	Planta	Ganho (%)	Planta	Ganho (%)	Planta	Ganho (%)
R2T11P2	-24,19	R3T11P1	-15,63	R2T11P8	-16,79	R2T6P7	-30,29	R2T11P8	-30,74	R2T11P8	-30,74
R2T6P7	-23,51	R2T11P8	-15,57	R3T6P3	-16,77	R2T6P8	-29,83	R3T11P1	-30,62	R3T11P1	-30,62
R1T4P1	-22,56	R3T6P3	-15,53	R3T11P1	-16,69	R3T6P6	-29,61	R3T6P4	-29,89	R3T6P4	-29,89
R2T6P8	-21,84	R2T11P7	-15,39	R2T11P7	-16,61	R2T6P1	-29,45	R3T9P2	-28,63	R3T9P2	-28,63
R2T7P2	-21,25	R2T6P7	-15,18	R2T6P7	-16,17	R2T6P6	-29,35	R1T9P6	-27,57	R1T9P6	-27,57
R3T4P6	-20,78	R2T7P7	-14,49	R2T7P7	-15,47	R2T6P4	-29,27	R2T7P7	-26,62	R2T7P7	-26,62
R2T6P3	-20,43	R2T6P3	-13,60	R2T6P6	-14,63	R2T6P3	-29,11	R3T6P3	-25,27	R3T6P3	-25,27
R3T6P6	-20,12	R2T6P6	-12,82	R2T6P3	-13,97	R3T6P3	-28,99	R2T7P4	-24,12	R2T7P4	-24,12
R2T8P2	-19,85	R2T6P8	-12,20	R3T11P3	-13,44	R2T6P5	-28,90	R3T1P7	-23,22	R3T1P7	-23,22
R3T6P3	-19,60	R3T6P6	-11,68	R2T6P8	-13,01	R2T6P2	-28,80	R2T11P7	-22,29	R2T11P7	-22,29
R2T6P5	-19,35	R3T11P3	-11,25	R1T4P8	-12,61	R1T6P7	-28,72	R1T1P1	-21,34	R1T1P1	-21,34
R2T6P4	-19,13	R3T4P1	-10,88	R3T9P2	-12,26	R3T6P8	-28,63	R2T6P7	-20,29	R2T6P7	-20,29
R1T4P8	-18,93	R1T4P8	-10,52	R3T6P8	-11,97	R3T6P1	-28,54	R3T7P4	-19,38	R3T7P4	-19,38
R3T11P2	-18,75	R3T9P2	-10,21	R3T6P6	-11,70	R3T6P4	-28,44	R1T5P4	-18,54	R1T5P4	-18,54
R3T4P4	-18,58	R3T4P6	-9,94	R3T6P4	-11,45	R3T6P5	-28,34	R2T4P8	-17,60	R2T4P8	-17,60
R2T2P6	-18,43	R1T4P7	-9,68	R2T7P4	-11,23	R1T6P8	-28,25	R3T6P8	-16,77	R3T6P8	-16,77
R2T11P7	-18,27	R3T6P4	-9,45	R3T1P7	-11,01	R3T6P2	-28,15	R3T6P2	-16,01	R3T6P2	-16,01
R3T4P1	-18,13	R3T1P7	-9,25	R3T4P1	-10,82	R1T6P2	-28,03	R3T6P6	-15,30	R3T6P6	-15,30
R3T11P1	-18,00	R2T6P4	-9,06	R3T6P2	-10,63	R1T6P3	-27,92	R2T3P7	-14,61	R2T3P7	-14,61
R2T11P8	-17,85	R3T6P2	-8,89	R2T6P2	-10,46	R3T6P7	-27,82	R1T6P6	-13,99	R1T6P6	-13,99
R1T4P7	-17,72	R2T7P4	-8,73	R1T4P7	-10,31	R1T6P6	-27,72	R1T6P5	-13,42	R1T6P5	-13,42
R1T8P4	-17,59	R2T6P2	-8,58	R1T9P6	-10,16	R1T6P5	-27,59	R1T7P8	-12,89	R1T7P8	-12,89
R2T6P6	-17,47	R1T6P7	-8,44	R2T11P2	-10,03	R1T6P1	-27,46	R2T6P3	-12,41	R2T6P3	-12,41
R3T7P6	-17,35	R1T4P1	-8,32	R3T4P5	-9,90	R1T6P4	-27,29	R3T6P7	-11,96	R3T6P7	-11,96
R1T11P8	-17,23	R3T6P8	-8,19	R1T6P7	-9,78	R2T11P2	-26,98	R2T6P6	-11,55	R2T6P6	-11,55
R3T5P6	-17,11	R1T6P8	-8,07	R1T6P8	-9,67	R2T7P2	-26,62	R1T7P3	-11,17	R1T7P3	-11,17
R3T6P2	-16,98	R1T4P4	-7,97	R3T4P6	-9,57	R3T7P6	-26,29	R3T6P5	-10,81	R3T6P5	-10,81
R3T11P3	-16,86	R1T9P6	-7,87	R1T7P2	-9,47	R1T11P8	-25,97	R3T2P7	-10,48	R3T2P7	-10,48
R3T6P1	-16,75	R3T4P5	-7,77	R2T6P5	-9,37	R3T11P1	-25,66	R2T6P2	-10,16	R2T6P2	-10,16
R3T4P3	-16,64	R2T4P2	-7,68	R3T6P1	-9,28	R2T11P8	-25,36	R1T6P7	-9,87	R1T6P7	-9,87
SÓLIDOS SOLÚVEIS TOTAIS (BRIX)						MODELO BLUP TRADICIONAL					

BLUP TRADICIONAL		RRBLUP		GBLUP		Esp, do mesocarpio		Massa da Polpa		Produção Total	
Planta	Ganho (%)	Planta	Ganho (%)	Planta	Ganho (%)	Planta	Ganho (%)	Planta	Ganho (%)	Planta	Ganho (%)
R1T9P3	3,94	R3T10P6	2,82	R3T10P6	0,45	R1T8P3	10,94	R1T9P1	68,12	R3T2P5	55,15
R2T9P2	3,74	R3T10P4	2,82	R2T10P5	0,44	R3T11P5	8,57	R1T10P5	62,63	R3T7P3	54,99
R3T9P4	3,67	R2T10P5	2,81	R2T10P8	0,44	R2T7P5	7,68	R1T2P4	60,67	R2T8P5	54,72
R3T9P3	3,62	R2T10P8	2,81	R3T10P4	0,44	R1T9P1	7,22	R2T9P2	59,07	R3T2P6	54,40
R3T1P2	3,57	R1T3P3	2,81	R1T10P5	0,43	R2T7P4	6,92	R1T2P1	57,45	R2T2P2	52,81
R2T9P8	3,50	R2T10P6	2,81	R3T1P5	0,43	R2T9P2	6,71	R1T2P3	56,30	R3T3P2	51,59
R1T2P7	3,44	R1T10P5	2,80	R2T1P5	0,43	R2T7P8	6,54	R2T10P6	55,47	R1T7P3	50,48
R2T9P5	3,37	R3T1P5	2,80	R2T10P3	0,43	R3T9P3	6,42	R2T3P6	54,56	R1T8P5	49,62
R2T9P3	3,32	R2T1P5	2,80	R1T3P3	0,43	R2T7P7	6,31	R2T10P5	53,32	R2T2P6	48,92
R3T9P6	3,28	R2T2P1	2,80	R3T9P3	0,42	R2T11P5	6,22	R1T9P2	52,12	R3T7P2	48,32
R3T9P8	3,24	R3T1P2	2,80	R2T2P1	0,42	R3T11P3	6,15	R1T3P3	50,90	R2T10P1	47,79
R1T9P7	3,22	R3T5P2	2,80	R3T1P2	0,42	R1T7P8	6,08	R2T10P2	49,86	R1T7P2	47,21
R2T9P7	3,18	R1T10P6	2,80	R1T10P6	0,42	R1T11P8	6,01	R1T1P3	48,85	R2T10P3	46,50
R2T9P6	3,15	R2T10P3	2,80	R2T10P2	0,42	R1T7P6	5,96	R1T9P7	47,96	R3T2P4	45,86
R2T9P1	3,12	R1T2P7	2,80	R3T5P2	0,42	R2T7P6	5,90	R2T2P7	47,17	R2T2P3	45,30
R1T6P4	3,10	R1T2P6	2,79	R2T2P3	0,42	R2T11P6	5,86	R3T3P3	46,37	R2T11P8	44,77
R3T2P7	3,08	R3T9P3	2,79	R1T5P3	0,42	R2T9P5	5,81	R1T5P5	45,63	R2T7P5	44,28
R2T2P5	3,06	R2T2P4	2,79	R1T10P1	0,42	R3T3P6	5,77	R2T9P5	44,96	R1T7P1	43,74
R1T9P2	3,03	R1T2P1	2,79	R1T2P7	0,41	R2T7P1	5,72	R2T5P7	44,34	R1T3P3	43,24
R1T2P6	3,01	R2T2P5	2,79	R2T10P6	0,41	R1T7P7	5,69	R3T3P2	43,71	R3T2P8	42,77
R1T2P3	2,99	R2T10P2	2,79	R3T5P7	0,41	R3T7P4	5,65	R3T9P7	43,13	R1T11P5	42,33
R3T2P4	2,97	R3T10P3	2,79	R1T2P6	0,41	R2T3P6	5,61	R1T3P6	42,58	R2T9P8	41,92
R1T9P5	2,95	R2T2P3	2,79	R2T2P4	0,41	R1T11P3	5,58	R1T1P5	42,06	R1T2P4	41,51
R3T9P7	2,93	R3T10P5	2,79	R3T10P3	0,41	R1T7P4	5,54	R2T2P2	41,52	R2T11P2	41,13
R2T2P1	2,91	R1T5P1	2,79	R2T2P5	0,41	R2T3P1	5,51	R2T10P8	40,99	R3T2P3	40,77
R1T9P8	2,89	R1T10P1	2,79	R1T3P8	0,41	R1T7P3	5,48	R2T1P1	40,50	R3T8P3	40,38
R2T2P3	2,87	R3T5P7	2,79	R3T10P5	0,41	R1T3P8	5,46	R1T2P6	40,03	R2T11P1	40,00
R2T9P4	2,85	R1T3P8	2,79	R1T2P3	0,41	R3T7P6	5,43	R1T2P5	39,58	R2T7P1	39,62
R1T2P1	2,83	R1T5P3	2,78	R2T2P2	0,41	R1T3P3	5,41	R1T3P5	39,16	R1T3P5	39,27
R2T2P4	2,81	R2T1P8	2,78	R1T1P7	0,41	R1T3P5	5,37	R3T1P4	38,75	R2T10P4	38,94

A Tabela 4 apresenta os valores médios das variáveis avaliadas. As médias revelam que o modelo BLUP tradicional proporcionou os maiores ganhos para todas as características avaliadas. De forma geral, os resultados demonstram que o BLUP tradicional foi mais eficaz na seleção de indivíduos para a população estudada.

Tabela 4. Média dos valores das variáveis avaliadas.

Modelagens	M	C	D	Ec	Em	Ee	Mpo	Mpl	SST	P
Média da População *	154,62	72,08	61,61	1,82	12,15	35,13	117,28	35,06	12,80	3369,48
Média BLUP Trad. **	222,45	86,97	70,04	1,62	12,63	29,46	158,13	27,04	13,13	4369,19
Média RRBLUP **	189,36	79,66	65,61	1,75	-	31,79	-	28,69	12,85	-

Média GBLUP **	186,98	79,35	65,34	1,75	-	31,88	-	29,08	12,85	-
----------------	--------	-------	-------	------	---	-------	---	-------	-------	---

M - massa do fruto, C comprimento do fruto, D diâmetro do fruto, Ec - espessura da casca, Em - espessura do mesocarpo, Ee - espessura do endocarpo, Mpo - massa da polpa, Mpl - massa da placenta, SST - sólidos solúveis totais e Produção total. Modelos não ajustados, com herdabilidade zero, não foram contabilizados. * Média dos indivíduos de toda a população. ** Média dos 30 indivíduos selecionados através das metodologias de seleção.

Na comparação entre os modelos BLUP tradicional, RRBLUP e GBLUP, observou-se uma variação considerável na identificação dos indivíduos superiores, especialmente no que se refere aos ganhos genéticos preditos para as diferentes características avaliadas. De modo geral, o BLUP tradicional apresentou maiores valores de ganho genético para a maioria das variáveis, superando os modelos baseados em marcadores moleculares. Este desempenho superior pode ser atribuído à estrutura da população utilizada no presente estudo, composta por famílias de irmãos completos com reduzida variabilidade genética. Esse resultado corrobora a evidência de que, em populações com reduzida variabilidade genética a matriz de parentesco médio pode ser mais eficiente na predição do valor genético aditivo, uma vez que reflete com maior precisão a estrutura de parentesco entre os indivíduos (Coste et al., 2021) .

A similaridade entre a seleção dos 30 melhores indivíduos nos modelos BLUP, RRBLUP e GBLUP foi averiguada por meio do *Rank-Biased Overlap* (RBO). A métrica RBO tem se mostrado uma ferramenta robusta para a comparação de listas ranqueadas, especialmente em contextos em que as listas apresentam tamanhos diferentes, elementos não coincidentes ou posições com maior relevância analítica. Um exemplo notável de sua aplicabilidade é o método Phirbo que utiliza o RBO para comparar rankings baseados em similaridade de sequência entre vírus e hospedeiros bacterianos, mesmo em cenários de dados fragmentados, como ocorre em estudos metagenômicos (Zielezinski et al., 2021).

Os resultados são apresentados na Tabela 5 e mostram que as comparações entre RRBLUP e GBLUP apresentaram os maiores valores de RBO em todas as variáveis, indicando que os modelos genômicos têm maior proximidade entre si do que em relação ao modelo BLUP tradicional.

Tabela 5: Valores de RBO obtidos para a comparação de similaridade entre os ranqueamentos dos 30 indivíduos selecionados nas três abordagens de modelagem.

Variáveis	Modelos de Seleção	Rank-Biased
-----------	--------------------	-------------

		Overlap ($\rho = 0,9$)
Massa	BLUP x RRBLUP	0,27
	BLUP x GBLUP	0,24
	RRBLUP x GBLUP	0,84
Comp.	BLUP x RRBLUP	0,26
	BLUP x GBLUP	0,18
	RRBLUP x GBLUP	0,73
Diam.	BLUP x RRBLUP	0,28
	BLUP x GBLUP	0,29
	RRBLUP x GBLUP	0,88
Ec	BLUP x RRBLUP	0,13
	BLUP x GBLUP	0,13
	RRBLUP x GBLUP	0,94
Ee	BLUP x RRBLUP	0,23
	BLUP x GBLUP	0,20
	RRBLUP x GBLUP	0,75
Mpl	BLUP x RRBLUP	0,09
	BLUP x GBLUP	0,08
	RRBLUP x GBLUP	0,92
SST	BLUP x RRBLUP	0,01
	BLUP x GBLUP	0,01
	RRBLUP x GBLUP	0,95

As variáveis espessura do mesocarpo, massa da polpa e produção foram suprimidas, pois os modelos R-BLUP e GBLUP não se ajustaram aos dados apresentados.

Essa tendência foi particularmente evidente para Espessura da casca (0,94), Massa da placenta (0,92) e SST (0,95), indicando forte concordância entre RRBLUP e GBLUP na ordenação dos indivíduos para essas características (Tabela 5).

Por outro lado, a concordância entre BLUP tradicional e os modelos genômicos foi menor, com valores de RBO frequentemente abaixo de 0,30. As menores similaridades foram observadas para Brix (0,01 para BLUP x RRBLUP e BLUP x GBLUP) e Massa da Placenta (Mpl) (0,09 para BLUP x RRBLUP e 0,08 para BLUP x GBLUP), sugerindo que a seleção baseada em matriz de parentesco médio diverge consideravelmente da seleção baseada em marcadores moleculares para essas características.

Esses valores evidenciam a captura padrões distintos de variação genética entre os modelos. A principal diferença entre os modelos que utilizam matriz de parentesco médio e aqueles baseados em marcadores moleculares reside na capacidade dos marcadores moleculares de separar com maior precisão os fatores de variação fenotípica associados à herdabilidade das características estudadas.

CONCLUSÕES

- O modelo BLUP tradicional apresentou o melhor desempenho geral, com maior capacidade preditiva e menor erro quadrático médio para todas as variáveis analisadas.
- O RRBLUP foi o modelo com maiores estimativas de herdabilidade para a maioria das características, mas não se ajustou bem a variáveis com baixa variabilidade genética.
- A comparação dos indivíduos selecionados pelos modelos revelou maior similaridade entre RRBLUP e GBLUP, com altos valores de RBO, especialmente para características como Brix e espessura da casca.
- O uso da métrica *Rank-Biased Overlap* (RBO) mostrou-se eficaz para avaliar a consistência entre os rankings de seleção, sendo uma abordagem útil em estudos com diferentes modelos genéticos.

BIBLIOGRAFIA

Arévalo-Marín, E., Casas, A., Alvarado-Sizzo, H., Ruiz-Sanchez, E., Castellanos-Morales, G., Jardón-Barbolla, L., Fermin, G., Padilla-Ramírez, J. S. and Clement, C. R. (2024). Genetic analyses and dispersal patterns unveil the Amazonian origin of guava domestication. *Scientific Reports*, 14:e15755. DOI doi.org/10.1038/s41598-024-66495-y.

Azzolini, M., Jacomino, A. P. and Spoto, M. H. F. (2004). Estádios de maturação e qualidade pós-colheita de goiabas “Pedro Sato” 1. *Revista Brasileira de Fruticultura*, 26:29-31. DOI doi.org/10.1590/S0100-29452004000100009.

Azevedo, C. V. G., Val, B. H. P., Araújo, L. C. A., Juhász, A. C. P, Mauro, A. O. D. and Unêda-Trevisoli, S. H. (2021). Genetic parameters of soybean populations obtained from crosses between grain and food genotypes. *Acta Scientiarum – Agronomy*, 43:1–8. DOI doi.org/10.4025/actasciagron.v43i1.46968.

Chu, S., Zhang, F., Wang, H., Xie, L., Chen, Z., Zeng, W., Zhou, Z. and Hu, F. (2022). Aqueous extract of guava (*Psidium guajava* L.) leaf ameliorates hyperglycemia by promoting hepatic glycogen synthesis and modulating gut

microbiota. *Frontiers in Pharmacology*, 13:907702. DOI doi.org/10.3389/fphar.2022.907702.

Corsi, M. and Urbano, J. (2024). The treatment of ties in rank-biased overlap. *Proceedings of the 47th international acm sigir conference on research and development in information retrieval*, 47, Washington DC:Association for Computing Machinery, 251–260.

Coste, C. F. D., Bienvenu, F., Ronget, V., Ramirez-Loza, J. P., Cubaynes, S., and Pavard, S. (2021). The kinship matrix: inferring the kinship structure of a population from its demography. *Ecology Letters*, 24:2750–2762. DOI doi.org/10.1111/ele.13854.

Das, D. and Mishra, R. (2022). SNP markers and its impact on crop improvement. *Asian Journal of Biological and Life Sciences*, 10:539-548. DOI doi.org/10.5530/ajbls.2021.10.72.

Diaz-Garcia, L. and Padilla-Ramírez, J. S. (2023). Development of single nucleotide polymorphism markers and genetic diversity in guava (*Psidium guajava* L.). *Plants People Planet*, 5:58–69. DOI doi.org/10.1002/ppp3.10295.

Endelman, J. B., (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, 4:250–255. DOI doi.org/10.3835/plantgenome2011.08.0024.

Grossi, L. L., Fernandes, M., Silva, M. A., Bernardes, C. O., Tuler, A. C., Santos, P. H. D., Ferreira, A. and Ferreira, M. F. S. (2021). DArTseq derived SNPs for the genus *Psidium* reveal the high diversity of native species. *Tree Genetics & Genomes*, 17:23. DOI doi.org/10.1007/s11295-021-01505-y.

Inglis, P. W., Castro, R. P. M., Resende, L. V. and Grattapaglia, D. (2018). Fast and inexpensive protocols for consistent extraction of highquality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One*, 13:0206085. DOI doi.org/10.1371/journal.pone.0206085.

Jaiswal, M. (2024). Recent advancement of genetic markers and application in tree improvement programs: a review. *Plant Archives*, 24:309-318. DOI doi.org/10.51470/PLANTARCHIVES.2024.v24.no.1.044.

- Koura, A. A., Kena, A. W., Annor, B., Adejumobi, I. I., Maina, F., Maazou, A. R., Razakou, I. B. Y. A., Attamah, P., Kusi, P., Boukar, O. and Akromah, R. (2024). DArTseq-based genome-wide SNP markers reveal limited genetic diversity and highly structured population in assembled West African cowpea germplasm. *Scientific African*, 23:e02065. DOI doi.org/10.1016/j.sciaf.2024.e02065.
- Kumar, S., Singh, A., Yadav, A., Bajpai, A., Singh, N. K., Rajan, S., TRIVEDI, M. and Muthukumar, M. (2023). Identification and validation of novel genomic SSR markers for molecular characterization of guava (*Psidium guajava* L.). *South African Journal of Botany*, 155:79–89. DOI doi.org/10.1016/j.sajb.2023.02.005.
- Kumari, S., Arumugam, N., Singh, R., Srivastav, M., Banoth, S., Mithra, A. C., Arun, M. B., Goswami, A. K. and Khan, Y. J. (2018). Diversity analysis of guava (*Psidium guajava*) germplasm collection. *Indian Journal of Agricultural Sciences*, 88: 489–497. DOI doi.org/10.56093/ijas.v88i3.78740.
- Maan S. S., Brar J. S., Mittal A., Gill M. I. S. Arora N. K., Sohi H. S., Chhuneja P., Dhillon G. S., Singh N. and Thakur S. (2023). Construction of a genetic linkage map and QTL mapping of fruit quality traits in guava (*Psidium guajava* L.). *Frontiers in Plant Science*, 14:1123274. DOI doi.org/10.3389/fpls.2023.1123274.
- Menezes, F. N. D. D., De Melo, F. H. C., Vieira, A. R. S., Almeida, T. C., Lima, M. S., Aquino, J. S., Gomez-Zavaglia, A., Magnani, M. and De Souza, E. L. (2021). Acerola (*Malpighia glabra* L.) and guava (*Psidium guayaba* L.) industrial processing by-products stimulate probiotic *Lactobacillus* and *Bifidobacterium* growth and induce beneficial changes in colonic microbiota. *Journal of Applied Microbiology*, 130:1323–1336. DOI doi.org/10.1111/jam.14824.
- Peixoto L. A., Teodoro, P. E., Teodoro, L. P. R., Cruz, C. D., Bhering, L. L. (2024). Performance of four genomic selection methods using different heritability and QTL numbers. *Pesquisa Agropecuaria Brasileira*, 59:e03535. DOI doi.org/10.1590/S1678-3921.pab2024.v59.03535.
- Quintal, S. S. R., Viana, A. P., Campos, B. M., Vivas, M. and Amaral Júnior, A. T. (2017). Selection via mixed models in segregating guava families based on yield and quality traits. *Revista Brasileira de Fruticultura*, 39:e-866. DOI doi.org/10.1590/0100-29452017866.

Srivastava, A. K., Williams, S. M. and Zhang, G. (2023) Heritability estimation approaches utilizing genome-wide Data. *Current Protocols*, 3:e734. DOI doi.org/10.1002/cpz1.734.

Valdisser, P. A. M. R., Pereira, W. J., Filho, J. E. A., Müller, B. S. F., Coelho, G. R. C., Menezes, I. P. P., Vianna, J. P. G., Zucchi, M. I., Lanna, A. C., Coelho, A. S. G., Oliveira, J. P., Moraes, A. C., Brondani C., and Vianello, R. P. (2017). In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. *BMC Genomics* 18:423. DOI doi.org/10.1186/s12864-017-3805-4.

Zielezinski, A., Barylski, J. and Karlowski, W. M. (2021). Taxonomy-aware, sequence similarity ranking reliably predicts phage–host relationships. *BMC Biology*, 19:223. DOI doi.org/10.1186/s12915-021-01146-6.

5 CONCLUSÃO

A presente tese abordou de forma integrada a diversidade genética e a predição de valores genéticos genômicos em uma população experimental de goiabeira composta por 11 famílias de irmãos completos. O primeiro estudo demonstrou que, apesar da origem controlada e do grau de parentesco elevado entre os indivíduos, foi possível identificar variabilidade fenotípica para características morfoagronômicas do fruto, com maior contribuição relativa para a divergência atribuída ao peso e à massa de polpa. As análises de agrupamento, baseadas nas distâncias de Mahalanobis e aplicadas pelos métodos UPGMA e Ward, revelaram a formação de três grupos com dados fenotípicos e até cinco grupos com dados moleculares, indicando que os marcadores SNP permitiram maior detalhamento na estruturação genética da população. Contudo, a análise de variância molecular (AMOVA), conforme o método de Excoffier, indicou que 92% da variabilidade genética estão concentrados entre indivíduos dentro das famílias e apenas 8% entre famílias, confirmando a ausência de diferenciação estatisticamente significativa entre elas.

O segundo estudo permitiu avaliar o desempenho preditivo de três modelos de estimação de valores genéticos: BLUP tradicional, RRBLUP e GBLUP. Os resultados mostraram que o modelo BLUP tradicional, baseado em matriz de parentesco médio, apresentou a maior capacidade preditiva, os menores erros quadráticos médios e viés próximo da unidade para todas as características avaliadas. Os modelos RRBLUP e GBLUP não foram ajustados para espessura do mesocarpo, massa da polpa e produção total, devido às estimativas de

herdabilidade próximas de zero. Por outro lado, o modelo RRBLUP apresentou as maiores herdabilidades para várias variáveis, como massa e comprimento do fruto, espessura do endocarpo e massa da placenta. A análise de similaridade entre os rankings revelou alta concordância entre RRBLUP e GBLUP, mas baixa sobreposição entre os modelos genômicos e o BLUP tradicional.

Assim, concluiu-se que, para populações com baixa variabilidade genética, como a analisada nesta tese, o modelo BLUP tradicional tende a apresentar desempenho superior, sobretudo quando se busca acurácia na predição de valores genéticos. No entanto, os modelos genômicos mostraram consistência interna e poderão ser mais vantajosos em populações com maior diversidade ou com estruturas de parentesco mais complexas. Os resultados obtidos fornecem subsídios técnicos relevantes para a condução de futuras estratégias de seleção na goiabeira e reforçam a importância da escolha criteriosa do modelo preditivo em função da estrutura genética da população-alvo.

REFERÊNCIAS BIBLIOGRÁFICAS

- Angulo-López, J. E., Flores-Gallegos, A. C., Torres-León, C., Ramírez-Guzmán, K. N., Martínez, G. A. and Aguilar, C. N. (2021) Guava (*Psidium guajava* L.) fruit and valorization of industrialization by-products. *Processes*, 9:1075. DOI doi.org/10.3390/pr9061075.
- Anushka, K., Singh, R. K., Padhan, A. (2024) Crop improvement of guava (*Psidium guajava*) through its genetic resources. *Futuristic Trends in Biotechnology*, 3:155-162.
- Arévalo-Marín, E., Casas, A., Alvarado-Sizzo, H., Ruiz-Sanchez, E., Castellanos-Morales, G., Jardón-Barbolla, L., Fermin, G., Padilla-Ramírez, J. S. and Clement, C. R. (2024) Genetic analyses and dispersal patterns unveil the Amazonian origin of guava domestication. *Scientific Reports*, 14:15755. DOI doi.org/10.1038/s41598-024-66495-y.
- Azevedo, C. V. G., Val, B. H. P., Araújo, L. C. A., Juhász, A. C. P, Mauro, A. O. D. and Unêda-Trevisoli, S. H. (2021) Genetic parameters of soybean populations obtained from crosses between grain and food genotypes. *Acta Scientiarum Agronomy*, 43:1–8. DOI doi.org/10.4025/actasciagron.v43i1.46968.
- Azzolini, M., Jacomino, A. P. and Spoto, M. H. F. (2004) Estádios de maturação e qualidade pós-colheita de goiabas “Pedro Sato” 1. *Revista Brasileira de Fruticultura*, 26:29-31. DOI doi.org/10.1590/S0100-29452004000100009.
- Barbosa, F. R. and Lima, M. F. (2010) A cultura da goiaba. 2. ed. Brasília: Embrapa Informação Tecnológica, 180p.

Bezerra, C. M., Ribeiro, R. M., Viana, A. P., Cavalcante, N. R., Silva, F. A., Ambrósio, M., Teixeira, A., Junior, A. (2020) Guava breeding via full-sib family selection: conducting selection cycle and divergence between parents and families. *Crop Breeding and Applied Biotechnology*, 20:e256520112. DOI doi.org/10.1590/1984-70332020v20n1a12.

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*, 38:151–157. DOI doi.org/10.1590/0100-2945-258/14.

Landau, E. C.; Silva, G. A.; Moura, L.; Hirsch, A. and Guimaraes, D. P. (2020) Dinâmica da produção agropecuária e da paisagem natural no brasil nas últimas décadas: sistemas agrícolas, paisagem natural e análise integrada do espaço rural. 1. ed. Brasília: Embrapa Informação Tecnológica, 2171p.

Chu, S., Zhang, F., Wang, H., Xie, L., Chen, Z., Zeng, W., Zhou, Z. and Hu, F. (2022) Aqueous extract of guava (*Psidium guajava* L.) leaf ameliorates hyperglycemia by promoting hepatic glycogen synthesis and modulating gut microbiota. *Frontiers in Pharmacology*, 13:907702. DOI doi.org/10.3389/fphar.2022.907702.

Corsi, M. and Urbano, J. (2024) The treatment of ties in rank-biased overlap. *Proceedings of the 47th international acm sigir conference on research and development in information retrieval*, 47, Washington DC:Association for Computing Machinery, 251–260.

Coste, C. F. D., Bienvenu, F., Ronget, V., Ramirez-Loza, J. P., Cubaynes, S., and Pavard, S. (2021) The kinship matrix: inferring the kinship structure of a population from its demography. *Ecology Letters*, 24:2750–2762. DOI doi.org/10.1111/ele.13854.

Das, D. and Mishra, R. (2022) SNP markers and its impact on crop improvement. *Asian Journal of Biological and Life Sciences*, 10:539-548. DOI doi.org/10.5530/ajbls.2021.10.72.

Diaz-Garcia, L. and Padilla-Ramírez, J. S. (2023) Development of single nucleotide polymorphism markers and genetic diversity in guava (*Psidium guajava* L.). *Plants People Planet*, 5:58–69. DOI doi.org/10.1002/ppp3.10295.

Endelman, J. B., (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, 4:250–255. DOI doi.org/10.3835/plantgenome2011.08.0024.

Grossi, L. L., Fernandes, M., Silva, M. A., Bernardes, C. O., Tuler, A. C., Santos, P. H. D., Ferreira, A. and Ferreira, M. F. S. (2021) DArTseq derived SNPs for the genus *Psidium* reveal the high diversity of native species. *Tree Genetics & Genomes*, 17:23. DOI doi.org/10.1007/s11295-021-01505-y.

Inglis, P. W., Castro, R. P. M., Resende, L. V. and Grattapaglia, D. (2018) Fast and inexpensive protocols for consistent extraction of highquality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One*, 13:0206085. DOI doi.org/10.1371/journal.pone.0206085.

Jaiswal, M. (2024) Recent advancement of genetic markers and application in tree improvement programs: a review. *Plant Archives*, 24:309-318. DOI doi.org/10.51470/PLANTARCHIVES.2024.v24.no.1.044.

Javid, S., Bihamta, M.R., Omid, M., Abbasi, A.R., Alipour, H. and Ingvarsson, P.K. (2022) Genome-wide association study (GWAS) and genome prediction of seedling salt tolerance in bread wheat (*Triticum aestivum* L.). *BMC Plant Biology*, 22:581. DOI doi-org.ez81.periodicos.capes.gov.br/10.1186/s12870-022-03936-8.

Koura, A. A., Kena, A. W., Annor, B., Adejumobi, I. I., Maina, F., Maazou, A. R., Razakou, I. B. Y. A., Attamah, P., Kusi, P., Boukar, O. and Akromah, R. (2024) DArTseq-based genome-wide SNP markers reveal limited genetic diversity and highly structured population in assembled West African cowpea germplasm. *Scientific African*, 23:e02065. DOI doi.org/10.1016/j.sciaf.2024.e02065.

Kumar, C., Kumar, R., Singh, S. K., Goswami, A. K., Nagaraja, A., Paliwal, R., Singh, R. (2020) Development of novel g-SSR markers in guava (*Psidium guajava* L.) cv. Allahabad Safeda and their application in genetic diversity, population structure and cross species transferability studies. *PLoS ONE*, 15:e0237538. DOI doi.org/10.1371/journal.pone.0237538.

Kumar, S., Singh, A., Yadav, A., Bajpai, A., Singh, N. K., Rajan, S., Trivedi, M. and Muthukumar, M. (2023) Identification and validation of novel genomic SSR markers

for molecular characterization of guava (*Psidium guajava* L.). South African Journal of Botany, 155:79–89. DOI doi.org/10.1016/j.sajb.2023.02.005.

Kumari, S., Arumugam, N., Singh, R., Srivastav, M., Banoth, S., Mithra, A. C., Arun, M. B., Goswami, A. K. and Khan, Y. J. (2018) Diversity analysis of guava (*Psidium guajava*) germplasm collection. Indian Journal of Agricultural Sciences, 88:489–497. DOI doi.org/10.56093/ijas.v88i3.78740.

Maan S. S., Brar J. S., Mittal A., Gill M. I. S. Arora N. K., Sohi H. S., Chhuneja P., Dhillon G. S., Singh N. and Thakur S. (2023) Construction of a genetic linkage map and QTL mapping of fruit quality traits in guava (*Psidium guajava* L.). Frontiers in Plant Science, 14:1123274. DOI doi.org/10.3389/fpls.2023.1123274.

Menezes, F. N. D. D., Melo, F. H. C., Vieira, A. R. S., Almeida, T. C., Lima, M. S., Aquino, J. S., Gomez-Zavaglia, A., Magnani, M., Souza, E. L. (2021) Acerola (*Malpighia glabra* L.) and guava (*Psidium guayava* L.) industrial processing by-products stimulate probiotic *Lactobacillus* and *Bifidobacterium* growth and induce beneficial changes in colonic microbiota. Journal of Applied Microbiology, 130, 4:1323–1336. DOI doi.org/10.1111/jam.14824.

Mondragón-Jacobo, C., Toriz-Ahumada, L. M., Guzman-Maldonado, H. (2010) Generation of pink-fleshed guavas to diversify commercial production in central Mexico. Acta Horticulturae, 849, 333-340. <https://doi.org/10.17660/ActaHortic.2010.849.39>

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 Sciences, 10, 2:130-135. DOI doi.org/10.24154/jhs.v10i2.118.

Peixoto L. A., Teodoro, P. E., Teodoro, L. P. R., Cruz, C. D., Bhering, L. L. (2024) Performance of four genomic selection methods using different heritability and QTL numbers. Pesquisa Agropecuária Brasileira, 59:e03535. DOI doi.org/10.1590/S1678-3921.pab2024.v59.03535.

Pereira, F. M., Usman, M. M., Newton A., Nachtigal, J. C., Maphanga, O. R. M., Willemse, S. (2017) Advances in guava propagation. Revista Brasileira de Fruticultura, 39:1. DOI doi.org/10.1590/0100-29452017358.

Pessanha, P. G. D. O., Viana, A. P., Amaral Júnior, A. T., Souza, R. M., Teixeira M. C. and Pereira, M. G. (2011) Avaliação da diversidade genética em acessos de *Psidium* spp. via marcadores RAPD. Revista Brasileira de Fruticultura, 33:129-136. DOI doi.org/10.1590/S0100-29452011000100018.

Pommer, C. V. (2012) Guava world-wide breeding: major techniques and cultivars and future challenges. Acta Horticulturae, 959:81-88. DOI doi.org/10.17660/ActaHortic.2012.959.9.

Quintal, S. S. R., Viana, A. P., Campos, B. M., Vivas, M. and Amaral Júnior, A. T. (2017) Selection via mixed models in segregating guava families based on yield and quality traits. Revista Brasileira de Fruticultura, 39:e-866. DOI doi.org/10.1590/0100-29452017866.

Rabieyan, E., Bihamta, M. R., Moghaddam, M. E., Mohammadi, V., Alipour, H. (2022) Genome-wide association mapping and genomic prediction for pre harvest sprouting resistance, low α -amylase and seed color in Iranian bread wheat. BMC Plant Biology, 22:300. DOI doi.org/10.1186/s12870-022-03628-3.

Rai, M. K., Asthana, P., Jaiswal, V. S., Jaiswal, U. (2010) Biotechnological advances in guava (*Psidium guajava* L.): Recent developments and prospects for further research. Trees, 24:1–12. DOI doi.org/10.1007/s00468-009-0384-2.

Resende, M. D. V, Silva, F. F., Lopes, P. S., Azevedo, C. F. (2012) Seleção genômica ampla (GWS) via modelos mistos (REML/BLUP), inferência bayesiana (MCMC), regressão aleatória multivariada e estatística espacial. 1. ed. Viçosa: Universidade Federal de Viçosa, 291 p..

Santos, R. M., Viana, A. P., Santos, E. A., Souza, R. M., Almeida, O. F., Gomes, V. M., Rodrigues, Daniele L., Santos, P. R. (2020) Genetic structuring of segregating populations of *Psidium* spp resistant to the southern root-knot nematode by Bayesian approach as basis for the guava breeding program. Agrarian sciences, 92:1–18. DOI doi.org/10.1590/0001-3765202020180600.

Silva, F. A., Viana, A. P., Correa, C. C. G., Santos, E. A., Oliveira, J. A. V. S., Andrade, J. D. G., Ribeiro, R. M., Glória, L. S. (2021) Bayesian ridge regression shows the best fit for SSR markers in *Psidium guajava* among Bayesian models. Scientific Reports, 11:1–11. DOI 10.1038/s41598-021-93120-z.

- Silva, L. A., Peixoto, M. A., Peixoto, L. A., Romero, J. V., Bhering, L. L. (2021) Multi-trait genomic selection indexes applied to identification of superior genotypes. *Bragantia*, 80:e3621. DOI doi.org/10.1590/1678-4499.20200381.
- Srivastava, A. K., Williams, S. M., Zhang, G. Z. (2023) Heritability estimation approaches utilizing genome-wide data. *Current Protocols*, 3:e734. DOI doi.org/10.1002/cpz1.734.
- Susmitha, P., Kumar, P., Yadav, P., Sahoo, S., Kaur, G., Pandey, M. K., Singh, V., Tseng, T. M., Gangurde, S. S. (2023) Genome-wide association study as a powerful tool for dissecting competitive traits in legumes. *Frontiers in plant*, 14:1123631. DOI doi.org/10.3389/fpls.2023.1123631.
- Thakur, S., Yadav, I. S., Jindal, M., Sharma, P. K., Dhillon, G. S., Boora, R. S., Arora, N. K., Gill, M. I. S., Chhuneja, P. and Mittal, A. (2021) Development of genome-wide functional markers using draft genome assembly of guava (*Psidium guajava* L.) cv. Allahabad Safeda to expedite molecular breeding. *Frontiers in Plant Science*, 12:708332. DOI doi.org/10.3389/fpls.2021.708332.
- Thomas, G., Jacob, S. (2001) Crop improvement in guava: an overview. *Indian Journal of Plant Genetic Resources*, 14:325–330.
- Valdisser, P. A. M. R., Pereira, W. J., Filho, J. E. A., Müller, B. S. F., Coelho, G. R. C., Menezes, I. P. P., Vianna, J. P. G., Zucchi, M. I., Lanna, A. C., Coelho, A. S. G., Oliveira, J. P., Moraes, A. C., Brondani C., and Vianello, R. P. (2017) In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. *BMC Genomics* 18:423. DOI doi.org/10.1186/s12864-017-3805-4.
- Walter, F. H. B., Cavalcante, N. R., Viana, A. P., Santos, E. A., Mendes, D. S., Oliveira, J. A. V. S., Ramos, H. C. C., Boechat, M. S. B. (2024) Genetic variability and population structure in guava full-sib families via microsatellite markers. *Revista Brasileira de Fruticultura*, 46:e638. DOI doi.org/10.1590/0100-29452024638.
- Zielezinski, A., Barylski, J. and Karlowski, W. M. (2021) Taxonomy-aware, sequence similarity ranking reliably predicts phage–host relationships. *BMC Biology*, 19:223. DOI doi.org/10.1186/s12915-021-01146-6.