REVISTA BRASILEIRA DE FRUTICULTURA
Published by the Brazilian Society of Fruit Crops
ISSN 0100-2945

Genetics And Plant Breeding - Original Article - Edited by: Willian Krause

Genetic variability and population structure in guava full-sib families via microsatellite markers

Fernando Henrique de Barros Walter[1], Natan Ramos Cavalcante[1], Alexandre Pio Viana[1], Eileen Azevedo Santos[2], Débora Souza Mendes[1], Julie Anne Vieira Salgado de Oliveira[1], Helaine Christine Cancela Ramos[1], Marcela Santana Bastos Boechat[1]

[1]Center for Agricultural Sciences and Technologies, State University of North Fluminense Darcy Ribeiro, Campos dos Goytacazes, RJ, Brazil;
[2] Center for Agro-Environmental Research, Studies and Development, State University of Mato Grosso, Tangará da Serra, MT, Brazil;
Corresponding author: *natancavalcante2@hotmail.com*

Abstract: Guava production is a promising activity with great prominence in several regions of Brazil; however, a major obstacle faced by producers is the low number of available cultivars. The present study proposes to estimate and analyze genetic structure and variability, through molecular traits, aiming at the future development of new cultivars. Ninety-four genotypes from 11 full-sib families and the cultivars Paluma, Pedro Sato, and Cortibel 1 were selected for DNA extraction, totaling 97 genotypes. For molecular characterization, 48 pairs of microsatellite primers were used. This information was used to estimate the parameters of genetic diversity, genetic distance, genotype clustering, and the genetic structure of the population. The use of molecular information revealed the existence of genetic variability between the genotypes of the full-sib families and the cultivars. The average number of alleles per locus was 2,542. Expected heterozygosity values ranged from 0.030 to 0.599, averaging 0.401. Observed heterozygosity ranged from 0.010 to 0.577, averaging 0.293. Based on the UPGMA hierarchical clustering method, four groups were formed and crossing is recommended between individuals from groups 1 and 2. Bayesian analysis allowed the distinction of genotypes into only two groups, due to the individuals sharing most of the genomic regions analyzed.

Index terms: molecular marker, plant breeding, *Psidium guajava*.

Variabilidade genética e estrutura da população em famílias de irmãos-completos de goiabeira via marcadores microssatélites

Resumo: A produção da goiabeira é promissora e tem grande destaque em diversas regiões do Brasil; entretanto, um grande entrave que os produtores enfrentam é

Rev. Bras. Frutic., v.46, e-638 DOI: *https://dx.doi.org/10.1590/0100-29452024638*
Received 15 Feb, 2024 ● Accepted 07 Aug, 2024 ● Published Sep/Oct, 2024. Jaboticabal - SP - Brazil.

All the contents of this journal, except where otherwise noted, is licensed under a *Creative Commons Attribution 4.0 International (CC BY 4.0).*

o baixo número de cultivares disponíveis. O presente estudo teve por objetivos estimar e analisar a estrutura e a variabilidade genética por meio de características moleculares, a fim de, futuramente, desenvolver novoas cultivares. Foram selecionados para extração de DNA 94 genótipos de 11 famílias de irmãos-completos, e os cultivares Paluma, Pedro Sato e Cortibel 1, totalizando 97 genótipos. Para a caracterização molecular, foram utilizados 48 pares de iniciadores microssatélites. Com essas informações, foram estimados os parâmetros de diversidade genética, a distância genética, o agrupamento dos genótipos e a estrutura genética da população. O uso de informações moleculares mostrou haver variabilidade genética entre os genótipos das famílias de irmãos-completos e os cultivares. O número médio de alelos por loco foi 2,542. Os valores de heterozigosidade esperada variaram de 0,030 a 0,599, apresentando média de 0,401. A heterozigosidade observada variou de 0,010 a 0,577, com média de 0,293. Com base no agrupamento pelo método hierárquico UPGMA, foi constatada a formação de quatro grupos, sendo indicados cruzamentos de indivíduos dos grupos 1 e 2. A análise bayesiana permitiu a distinção dos genótipos em apenas dois grupos, devido aos indivíduos compartilharem a maioria das regiões genômicas analisadas.

Termos para indexação: marcador molecular, melhoramento de plantas, *Psidium guajava*.

Introduction

Myrtaceae is considered one of the largest botanical families worldwide (FRANZON et al., 2009), comprising 130 genus and over 3600 species (GOVAERTS et al., 2012). Dispersal centers are found in Australia, Southeast Asia, Tropical and Temperate Americas, plus a small representation in Africa (THORNHILL et al., 2015).

Within the family Myrtaceae, the genus *Psidium* is one of the most important, representative, and exploited (FRANZON et al., 2009). Representatives of this genus play an important ecological role thanks to their food (fruit), timber, and ornamental purposes. Additionally, they hold great potential for commercial exploitation and use in the pharmaceutical industry due to their high levels of antioxidant substances and essential oils (BEZERRA et al., 2006; FRANZON et al., 2009).

Guava (*Psidium guajava* L.) is the main representative of the genus owing to its economic value for industry and commerce (FRANZON et al., 2009). Brazil stands out in guava production, most of which is consumed in the national territory (INCAPER, 2023). The fruit is appreciated mainly for its flavor and aroma, besides being considered one of the most complete for human health for being balanced in nutrients, with high levels of vitamin C, sugars, mineral salts, and fiber (FACHI et al., 2018).

However, there is a low number of cultivars available and adapted to producing regions, with a total of 18 cultivars registered in the RNC (National Registry of Cultivars) (MAPA, 2023). In this context, when well conducted, supported, and directed towards the search for practical results, plant breeding is one of the best alternatives for the development of the agribusiness sector, as it enables the release of cultivars adapted to the different producing regions of Brazil.

In view of this scenario, the State University of Northern Rio de Janeiro (UENF) has been developing a breeding program whose purpose is to select promising genotypes to release high-yielding cultivars (PESSANHA et al., 2011; OLIVEIRA et al., 2014; QUINTAL et al., 2017; GOMES et al., 2017; SANTOS et al., 2020; SOUSA et al., 2020; AMBRÓSIO et al., 2021).

Based on the knowledge of genetic variability within the population, one can more assertively select divergent parents for crossing, thereby generating segregating populations with increased genetic variability (CRUZ et al., 2012). The study of genetic variability is

defined as the process by which variations among individuals, groups of individuals, or populations are assessed through specific methods or a combination of methods, across various datasets. This variability can be assessed using morphological, physiological, biochemical, cytogenetic, and molecular traits for estimation (MOHAMMADI et al., 2003).

Nonetheless, for new cultivars to be developed, it is extremely important to know the genetic variability of the material to be explored. One of the techniques that aid breeders is the use of molecular markers, which save time and accelerate the response in breeding programs (SANTOS et al., 2020).

Among the various classes of molecular markers available today, microsatellites (SSR) stand out as one of the most useful for assessing gene flow, genetic diversity, and population structure in plant species (XIAO et al., 2014). Additional characteristics of these markers include high reproducibility, multiallelic nature, codominant inheritance, high informative power, and wide dissemination throughout the genome (OLIVEIRA et al., 2006; OLIVEIRA; SILVA, 2008). This implies that they can precisely capture the genetic contributions from both parents, offering valuable information for breeding programs (MIAH et al., 2013). Multiple studies have employed SSR markers to characterize the genetic diversity in guava (KUMAR et al., 2020; MA et al., 2020; PARDO et al., 2023).

The present study proposes the application of microsatellite markers as molecular tools to estimate and analyze the genetic variability of genotypes from full-sib families, aiming at decision-making regarding the advancement of the guava breeding program.

Material and Methods

Genetic material and experimental conditions

The material used for DNA extraction was collected in the experimental area of the Antônio Sarlo State Agricultural Technical School, located in the municipality of Campos dos Goytacazes, state of Rio de Janeiro, Brazil (21º45’ S, 41º20’ W, 11 m above sea level). The local climate is characterized as subhumid and dry tropical, with an average annual temperature ranging from 22 to 25 °C and average annual precipitation between 1200 and 1300 mm.

Ninety-four genotypes from 11 full-sib families and the cultivars Paluma, Pedro Sato, and Cortibel 1 were selected for DNA extraction, totaling 97 genotypes (Table 1). The choice of these 94 genotypes was based on the production of the previous season and the cultivars used to compare the genetic structure.

Table 1. Origin and identification of the 11 full-sib families, the three cultivars (Paluma, Pedro Sato, and Cortibel belonging to the Active Germplasm Bank - BAG), and the total genotypes of each progeny.

Description	Crossings	Total number of genotypes	Number of selected gentypes
Family 1	8.4 II x 5.4 II	24	4
Family 2	10.5 I x 12.4 II	23	16
Family 3	12.11 II x 8.4 I	22	5
Family 4	4.5 x 2.6	22	10
Family 5	12.11 I x 8.4 I	23	12
Family 6	12.4 x 12.11	24	4
Family 7	12.11 II x 12.4 II	23	11
Family 8	12.11 II x 3.11 II	19	3
Family 9	12.11 x 7.9	24	3
Family 10	8.1 II x 12.4 II	22	17
Family 11	12.4 II x 12.11 II	21	9
Paluma	BAG da UENF	2	1
Pedro Sato	BAG da UENF	2	1
Cortibel 1	BAG da UENF	2	1
Total			**97**

Description of crossings: 8.4 II x 5.4 II – Family: 8, plant: 4 and repetition: 2 X Family: 5, plant: 4 and repetition 2.

The information obtained from the genotyping of the 97 genotypes was used to obtain the genetic structure of the population and estimate the parameters of genetic diversity between the genotypes.

Extraction and quantification of genomic DNA

Genomic DNA was extracted and quantified in the DNA Markers Section of the Plant Breeding Laboratory of the Center for Agricultural Sciences and Technologies at the State University of Northern Rio de Janeiro (LMGV/CCTA/UENF), located in Campos dos Goytacazes - RJ, Brazil.

In the extraction process, young leaves were collected individually from each genotype using the standard CTAB method with modifications (DOYLE; DOYLE, 1990).

Then, the DNA was quantified by analysis on 1% agarose gel with 1X TAE buffer (Tris, Sodium Acetate, EDTA, pH 8.0), using the 100-bp lambda (λ) marker (100 ng/μL^{-1}) (Invitrogen, USA), and stained with a mixture of Gel RedTM and Blue Juice (1:1). The images were captured by the Mini Bis Pro gel-documenting system (Bio-Imaging Systems). Based on the obtained images, the DNA concentration was estimated in comparison with the 100-bp marker and the DNA samples were diluted to the working concentration of 10 ng/μL^{-1}.

Primer screening

To test the polymerase chain reaction (PCR) conditions, a total of 192 pairs of primers (GuavaMap, 2008) designed to amplify SSR loci in *P. guajava* were tested, with a temperature gradient ranging from 48 to 60 ºC. Based on the indicated temperature, 2 ºC more and 2 ºC less were tested and the amplification with the best visualization was selected. After screening, a set of 48 primers was selected for the amplification reactions (Table 2).

Table 2. Sequence of the 48 pairs of microsatellite primers used in the analysis of the 97 genotypes of *Psidium guajava*.

Locus	Sequence (5' 3')		Annealing temperature	Reference
mPgCIR030	F: CTCAAAGCACTATCATGTCG	R: CCTTGTGGGTTCTCTTTTG	55 °C	GuavaMap 2008
mPgCIR032	F: CGCCTTTCGTAAAAGAAGT	R: TCATATACTCGGACAAAACG	54 °C	GuavaMap 2008
mPgCIR034	F: CTTTTAAGCACGTTTGACCT	R: GCGTATTAACAGCTTGGAGA	55 °C	GuavaMap 2008
mPgCIR035	F: TTGACCTGGCATTAACAGA	R: CATTGGGAAAGGGAAGAA	52 °C	GuavaMap 2008
mPgCIR036	F: ATGACTTTGGAAGAGGATGA	R: AGCAGTTACAGTTGCTCGTT	52 °C	GuavaMap 2008
mPgCIR039	F: GCTCACCTTACTCATTCAGC	R: CTGTTGCTAAGAGCTTTCGT	52 °C	GuavaMap 2008
mPgCIR040	F: TGAATCTCCAGTGTCTTATCG	R: TGATTTCAACTGCGTATGTC	56 °C	GuavaMap 2008
mPgCIR044	F: TTCCAGGTCTATTGGATGTC	R: GGGGACACAAAACTTCATTA	56 °C	GuavaMap 2008
mPgCIR046	F: ATAGAACGCCATGTTACCAA	R: CAGGCTTATCTGTTACACCA	55 °C	GuavaMap 2008
mPgCIR048	F: TGCAGCTTCTTCAATGTTC	R: AAAACTTGGCAACGTCAGT	55 °C	GuavaMap 2008
mPgCIR049	F: TCCTTTCCCACTTTTCATC	R: ACAAGCTCCACACGTTACTC	52 °C	GuavaMap 2008
mPgCIR090	F: ACGATGCAAGCTCTCCTAC	R: CTTTAGGGCTCTTCTTGAAA	56 °C	GuavaMap 2008
mPgCIR091	F: GCGGTGGATTTGAATTTAG	R: CCAAGTAACCCACAACAATA	52 °C	GuavaMap 2008
mPgCIR092	F: GTAAGTTTTCCGCATCCTG	R: ATATTGCACAGGGTGGTATC	52 °C	GuavaMap 2008
mPgCIR096	F: ACGCTGCAAACGATACTAAT	R: AACTCACACGAGCACAGAG	56 °C	GuavaMap 2008
mPgCIR097	F: GACCTCAGTAGTTCAGCATGT	R: TAGAGTGGACGGGAGGAG	52 °C	GuavaMap 2008
mPgCIR106	F: GACTCGACAGGAAGGTCTC	R: CAGCTTTGTATATCGCAAGA	56 °C	GuavaMap 2008
mPgCIR108	F: AGGACCTCACAGAAGTTCAC	R: CGCTGTTTACACTGTCGTT	52 °C	GuavaMap 2008
mPgCIR109	F: AATTTCCACAGATCACAAGG	R: GGCATCTCCATCAAATACAT	52 °C	GuavaMap 2008
mPgCIR133	F: CGATCTTGGAATGTAAGAGG	R: TGGATTTGCAGGTTCTATCT	54 °C	GuavaMap 2008
mPgCIR134	F: AAACAAGTGGGAAAACCAG	R: CAAAAGTCCAGCAAGAAAAC	50 °C	GuavaMap 2008
mPgCIR135	F: CAGATAGCAAAACTGCCTCT	R: ATATCCCTCTCGCCTTCTT	56 °C	GuavaMap 2008
mPgCIR138	F: GGTGAGACCACTGAGTTCC	R: GAAAGACCAATGAAATCGAC	56 °C	GuavaMap 2008

Table 2. Continuation.

Locus	Sequence (5' 3')		Annealing temperature	Reference
mPgCIR139	F: ATAATCCCCTCCCATAACTA	R: CCAACTCAACATGAGAAGC	54 °C	GuavaMap 2008
mPgCIR140	F: GTGGTGAAGGAGTAAAGCTG	R: GCAGTATAAAGCAACAGATGG	56 °C	GuavaMap 2008
mPgCIR142	F: GTCGTGGTTTCCCAAAATA	R: TACTGACTTCCCACACTCTTG	52 °C	GuavaMap 2008
mPgCIR148	F: CATACAGAGTCGGATGGTTT	R: GCTGCTGGTCTTAAAGCTAA	56 °C	GuavaMap 2008
mPgCIR150	F: CCTAGTGACTCGAAGCAATC	R: TTGAGCCCTAGCATAGACAG	55 °C	GuavaMap 2008
mPgCIR151	F: GGCTTTCCGTATGCTTAAT	R: AGACTGTGGTGGTTATCTGG	54 °C	GuavaMap 2008
mPgCIR153	F: GCCTCTGGTAAATCTGTTGA	R: ACATACGGATCAAGTCCAAA	56 °C	GuavaMap 2008
mPgCIR154	F: CTTCAGCTACAGCCTTTCC	R: GGAGAAAGCAGAAATTCCA	54 °C	GuavaMap 2008
mPgCIR164	F: TCCCGTAAGTTGTTCTGTTT	R: CGTAAGAGATCGTGAAGGAG	56 °C	GuavaMap 2008
mPgCIR166	F: CTTTCCCATCAAAACGTAAG	R: CCAATTCATGCACTTAGACA	56 °C	GuavaMap 2008
mPgCIR167	F: CAAGTTTGAAGGTGATGTAGAC	R: GTTATTTGCGGTTTGCAG	55 °C	GuavaMap 2008
mPgCIR174	F: GCCACTGTGTAAGAGGATTG	R: ATTGTGGGAGATTGGAGAC	56 °C	GuavaMap 2008
mPgCIR176	F: TTCTCAATGATAGGTTACGG	R: ATGACTATTCTCCACCAGAT	52 °C	GuavaMap 2008
mPgCIR179	F: GGGTCTCGACTAAAGAAGGA	R: CCTCCATTTGCATCAACTTT	52 °C	GuavaMap 2008
mPgCIR185	F: AACGCATCTGGCATTGAT	R: CCTTGGTCTCCCTCTTACTC	52 °C	GuavaMap 2008
mPgCIR190	F: GAAGATTGGGAACCTAAGAA	R: AGAAATGGGTTGTGGAGAG	52 °C	GuavaMap 2008
mPgCIR200	F: CCTTGCTTTGGTGAGGTC	R: GCTAATTCAGTCCTTCCAACT	53 °C	GuavaMap 2008
mPgCIR203	F: ATGAAGGCATTACCTAAGAC	R: ACCCTATTAACCCTTAGCAA	54 °C	GuavaMap 2008
mPgCIR205	F: ACCTCTCCAGCTCTACACG	R: GAGGTTGTCGAAGGTTGAT	50 °C	GuavaMap 2008
mPgCIR209	F: CTAAAGCCACATCCAGCA	R: CTAACATTTGCCTTCTACAGC	56 °C	GuavaMap 2008
mPgCIR220	F: AGAGCAGTGGTTGCTATTTT	R: CCCATCTCTTACTTTTCTTGTG	56 °C	GuavaMap 2008
mPgCIR221	F: CTAAGCCTGAAGTCCCAAAT	R: CCTCTTCTAAAGGCAACGAC	56 °C	GuavaMap 2008
mPgCIR230	F: CACATTTGCTCCTGATTTTC	R: GCTCTTCAACGACCATCTT	56 °C	GuavaMap 2008
mPgCIR235	F: TCTTCTGGACAACAAATTCC	R: GATTTAAGGCAAAGGATCG	52 °C	GuavaMap 2008
mPgCIR240	F: CGAATGTCCAAGATTCAGTT	R: CCTTCTTCATCTCAGCCTTT	55 °C	GuavaMap 2008

Polymerase chain reaction (PCR) and genotyping

Polymerase chain reactions were performed in Applied Biosystems/Veriti 96 well thermocyclers, in a 35-cycle program, observing the following temperatures and durations: 94 ºC for 4 min (initial denaturation); 94 ºC for 2 min (cyclic denaturation); specific temperature of each primer, in ºC, for 1 min (annealing); 72 ºC for 2 min (cyclical extension); 72 ºC for 10 min (final extension); and 4 ºC forever. The final volume was 13 µL of each sample, consisting of 2 µL of DNA (5 ng/µL), 1.50 µL of 10X Buffer (NH_4SO_4), 1.5 µL of $MgCl_2$ (25 mM), 1.5 µL of dNTPs (2 mM), 1 µL of primer (R+F) (5 µM), and 0.12 µL of Taq DNA polymerase (5 U/µL) (Invitrogen, Carlsbad, California, USA).

The PCR products were diluted at a ratio of 6 µL of sample to 18 µL of Buffer E from the DNF 900 kit and subjected to capillary electrophoresis (Fragment Analyzer - AATI), in which amplified fragments of 35 to 500 pb are separated with a resolution of approximately 2 bp. Each run lasted 2 h and 20 min, under a voltage of 8 kW.

Statistical analysis of molecular data

The observations obtained by amplification of the 48 SSR markers were converted into a numerical code for each allele per locus. This numerical matrix was developed by assigning values from 1 to the maximum number of alleles per locus, as described next: for a locus that has three alleles, homozygous forms (A1A1, A2A2, and A3A3) were represented by the numbers 11, 22, and 33; and heterozygotes (A1A2, A1A3, and A2A3) by

12, 13, and 23. From this numerical matrix, three indices were tested, namely, the unweighted index, the weighted index, and the Smouse Peakall index (PEAKALL; SMOUSE, 2012). Based on the highest cophenetic correlation, the weighted index was applied and analyses were carried out using GENES software (CRUZ, 2013). After the distance matrix was obtained, clustering was performed via dendrogram using the UPGMA (Unweighted Pair-Group Method with Arithmetic Mean) method in Mega software version 6 (KUMAR et al., 2008). The number of markers was estimated using GENES software (CRUZ, 2013) and the graph was plotted using SigmaPlot software (SYSTAT SOFTWARE, 2013).

The amount of genetic variability of the 97 genotypes was estimated using Genalex 6.5 software (PEAKALL;SMOUSE, 2012), based on the following parameters: number of alleles per polymorphic locus (NA), observed heterozygosity (Ho), expected heterozygosity (He), information index (I), and fixation index (f).

The information index, known as the Shannon Index (I) is used to indicate diversity and can be calculated as shown below:

$$\mathrm{I} = -1.\sum[\mathrm{Pi.Ln(Pi)}]$$

Where Pi: allele frequency for each of the alleles in question.

Ho is the proportion of heterozygous individuals observed in a studied population:

$$\mathrm{Ho} = \frac{\mathrm{Nx}}{\mathrm{N}}$$

Where Ho: observed heterozygosity; Nx: number of heterozygotes; and N: total number of individuals in the sample.

He can be defined as an estimated fraction of all individuals that could be heterozygous for a locus. It is estimated by the following formula:

$$\mathrm{He}=1-\sum \mathrm{P}_i^2$$

Where He = expected heterozygosity; = frequency of allele i.

The Fixation Index (F), estimates the average inbreeding, can be estimated as follows:

$$\mathrm{F}= 1-\frac{\mathrm{He}}{\mathrm{He}}$$

Where Ho: observed heterozygosity, i.e. the proportion of N samples that are heterozygous in a given locus; and He: proportion of expected heterozygosity under random mating.

Analysis of the genetic structure of the population

To access the structure of the 97 genotypes, a method based on Bayesian clustering algorithms was applied, using STRUCTURE software version 2.3.4 (PRITCHARD et al., 2000). For this purpose, the "admixture model" and independent allele frequencies were adopted, using a burn-in period of 250,000, followed by an extension (Markov Chain Monte Carlo) of 750,000 repetitions. Ten simulations were performed with k ranging from 1 to 10 for better stability of the number of clusters.

The Δk statistical test was performed using "Structure Harvester" software, based on the criteria established by Evanno et al. (2005). This criterion is based on the mean and standard deviation of the LnP(D) estimated in each of the 10 interactions per k. The Δki values were estimated using the following formula:

$$\Delta k_i = ABS\frac{(k_i + 1 - (2^{*}k_i) + k_i - 1)}{\text{Standard deviation of } k_i}$$

where i: number of simulated groups (i= 1 to 10); and ABS: the module.

This Δk value is estimated for each k, and the one with the highest value is selected. After choosing the optimal Δk, the simulation with the lowest value of LnP(D) is chosen among the 10 simulations made to obtain it. Each color of the generated graph represents a possible group of structured individuals.

Results and discussion

In the analysis of the genetic variability between the evaluated genotypes, the number of alleles per locus ranged from 2 to 4, averaging 2.542, in a total of 112 alleles for the 48 evaluated loci (Table 3). Coser et al. (2012) studied the genetic diversity of 28 genotypes of *P. guajava* and reported that the 24 SSR loci analyzed generated two to five alleles per locus, with an average of 2.7. Kherwar et al. (2018) evaluated 36 guava varieties, including wild species, and found a greater number of alleles per locus, with an allele variation of 2 to 7 alleles and an average of 3.682.

Among the 48 analyzed microsatellite loci, 30 would be sufficient to allow the evaluated genotypes to be clustered (Figure 1). The high number of markers used in this study is because the genetic material analyzed originates from previous selections (PESSANHA et al., 2011; QUINTAL et al., 2017) and is structured as biparental crosses with related parents, thus reducing the genetic base of the population.

The use of the Shannon Index (Table 3) as a measure of population diversity is based on values from 0 to 1, and the closer it is to zero, the lower the genetic diversity. Considering all the analyzed genotypes, the values of this index ranged from 0.080 (mPgCIR092) to 1.000 (mPgCIR235). The average value of 0.606 revealed moderate variability in this population, which is sufficient to allow the continuity of the breeding program.

Expected heterozygosity (He) values ranged from 0.030 (mPgCIR092) to 0.599 (mPgCIR235), averaging 0.401. Observed heterozygosity (Ho) ranged from 0.010 (mPgCIR048 and mPgCIR205) to 0.577 (mPgCIR154), averaging 0.293. Of the 48 loci analyzed, only 14 had Ho values greater than He, indicating a moderate number of heterozygous individuals for the evaluated loci. This was already expected since these are related genotypes with common parents.

Table 3. Diversity measures for the 48 microsatellite markers in the 97 analyzed genotypes.

Locus	Na	I	He	Ho	F
mPgCIR030	2.000	0.648	0.456	0.255	0.440
mPgCIR032	2.000	0.584	0.395	0.021	0.947
mPgCIR034	3.000	0.762	0.515	0.458	0.109
mPgCIR035	4.000	0.229	0.090	0.021	0.770
mPgCIR036	3.000	0.703	0.487	0.340	0.301
mPgCIR039	3.000	0.399	0.211	0.196	0.073
mPgCIR040	2.000	0.594	0.404	0.313	0.227
mPgCIR044	2.000	0.693	0.500	0.547	-0.095
mPgCIR046	3.000	0.701	0.485	0.072	0.851
mPgCIR048	3.000	0.715	0.498	0.010	0.979
mPgCIR049	3.000	0.563	0.356	0.344	0.033
mPgCIR090	2.000	0.261	0.135	0.042	0.692
mPgCIR091	2.000	0.542	0.356	0.381	-0.071
mPgCIR092	2.000	0.080	0.030	0.031	-0.016
mPgCIR096	2.000	0.691	0.497	0.557	-0.119
mPgCIR097	2.000	0.690	0.497	0.505	-0.016
mPgCIR106	3.000	0.119	0.042	0.022	0.492
mPgCIR108	2.000	0.617	0.426	0.385	0.095
mPgCIR109	3.000	0.661	0.445	0.474	-0.065
mPgCIR133	3.000	0.496	0.286	0.021	0.926
mPgCIR134	2.000	0.675	0.482	0.532	-0.104
mPgCIR135	2.000	0.542	0.356	0.031	0.913
mPgCIR138	3.000	0.741	0.507	0.436	0.139
mPgCIR139	4.000	0.753	0.510	0.533	-0.046
mPgCIR140	2.000	0.687	0.494	0.467	0.055
mPgCIR142	3.000	0.547	0.341	0.299	0.124
mPgCIR148	2.000	0.587	0.398	0.463	-0.165
mPgCIR150	2.000	0.673	0.480	0.490	-0.019
mPgCIR151	2.000	0.693	0.500	0.402	0.196
mPgCIR153	3.000	0.640	0.390	0.323	0.172
mPgCIR154	3.000	0.631	0.418	0.577	-0.381
mPgCIR164	2.000	0.263	0.137	0.147	-0.080
mPgCIR166	4.000	0.670	0.416	0.441	-0.059
mPgCIR167	2.000	0.608	0.417	0.308	0.263
mPgCIR174	2.000	0.693	0.500	0.253	0.494
mPgCIR176	2.000	0.692	0.498	0.522	-0.048
mPgCIR179	2.000	0.640	0.448	0.344	0.232
mPgCIR185	3.000	0.522	0.319	0.021	0.935
mPgCIR190	4.000	0.891	0.523	0.074	0.858
mPgCIR200	2.000	0.439	0.268	0.213	0.207
mPgCIR203	2.000	0.687	0.493	0.406	0.177
mPgCIR205	3.000	0.766	0.506	0.010	0.981
mPgCIR209	2.000	0.529	0.345	0.340	0.014
mPgCIR220	3.000	0.498	0.261	0.255	0.022
mPgCIR221	2.000	0.667	0.474	0.438	0.077
mPgCIR230	3.000	0.921	0.544	0.457	0.159
mPgCIR235	3.000	1.000	0.599	0.261	0.564
mPgCIR240	2.000	0.685	0.492	0.032	0.936
Average	**2.542**	**0.606**	**0.401**	**0.293**	**0.275**

NA: number of alleles; I: Shannon Index; He: expected heterozygosity; Ho: observed heterozygosity; and F: fixation index.

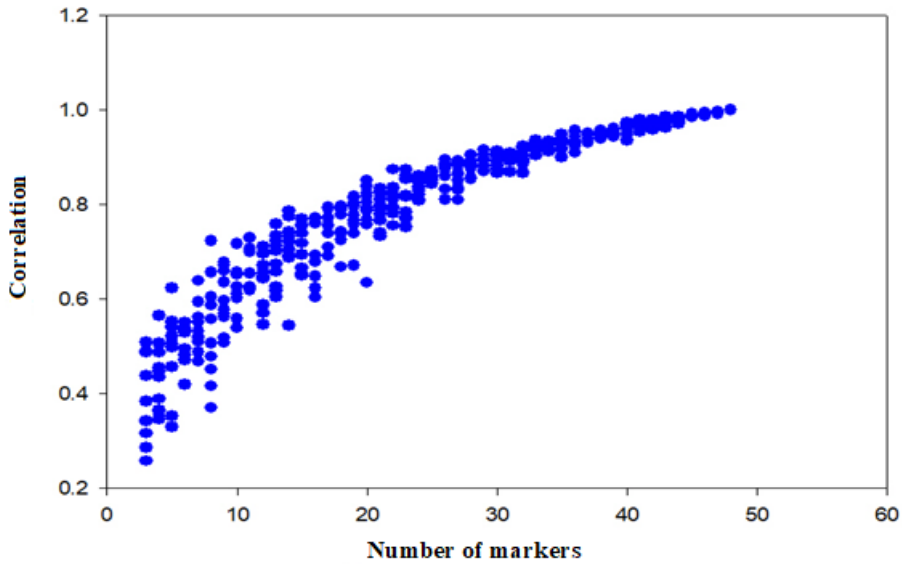

Figure 1. Optimal number of microsatellite markers analyzed in 97 genotypes of *Psidium guajava*.

As for the Fixation Index (F) (Table 3), which estimates the average inbreeding, the variation between loci ranged from -0.381 (mPgCIR154) to 0.981 (mPgCIR205), averaging 0.275. Fourteen of the 48 loci showed a negative F, which occurs when Ho values are higher than He. This is expected in random crosses and indicates excessive heterozygosity, meaning the alleles for these loci are not being fixed by inbreeding.

Two approaches were used to identify divergent groups. The first considered the distance estimated by the weighted index and the grouping of individuals by the UPGMA hierarchical method (Figure 2). Based on this strategy, the number of groups was defined as four, using the criterion of Mojena (1977), which considers cutoffs at 79% to 87% dissimilarity and k=1.25 (Cruz, Ferreira, Pessoni, 2011).

Group I contained the largest set of genotypes, with 73 individuals (75.26%), in which cultivar Cortibel 1 was present. Group II consisted of 22 individuals (22.68%). Groups III and IV contained cultivars Pedro Sato and Paluma, respectively (1.03% each group). Therefore, to maintain genetic diversity, crosses between genotypes from the full-sib families of groups I and II are indicated, as they contribute to delaying the increase in population inbreeding. Kareem et al. (2018) evaluated 37 guava accessions and obtained four groups formed. The researchers suggested that accessions collected in the same geographic region or breeding program tend to cluster together, indicating that despite the wide distribution of guava in the tropical world, the exchange of germplasm between regions has been limited.

The second clustering strategy followed the Bayesian method based on the criteria of Evanno et al. (2005), assuming that the analyzed individuals have a fraction of their genome inherited from their ancestors. The alleles detected for the 48 loci studied were used to draw inferences about the genetic structure of full-sib families of guava and the three most cultivated varieties in Brazil: Paluma, Pedro Sato, and Cortibel 1. The highest value of ΔK (Figure 3) was obtained when two clusters were formed, suggesting that maximum structuring was observed when the sample was divided into two groups, which were well structured (Figure 4). Seventy genotypes were allocated to group 1, one of which was cultivar Cortibel 1. Group 2 was composed of 27 genotypes, two of which were cultivars Paluma and Pedro Sato.

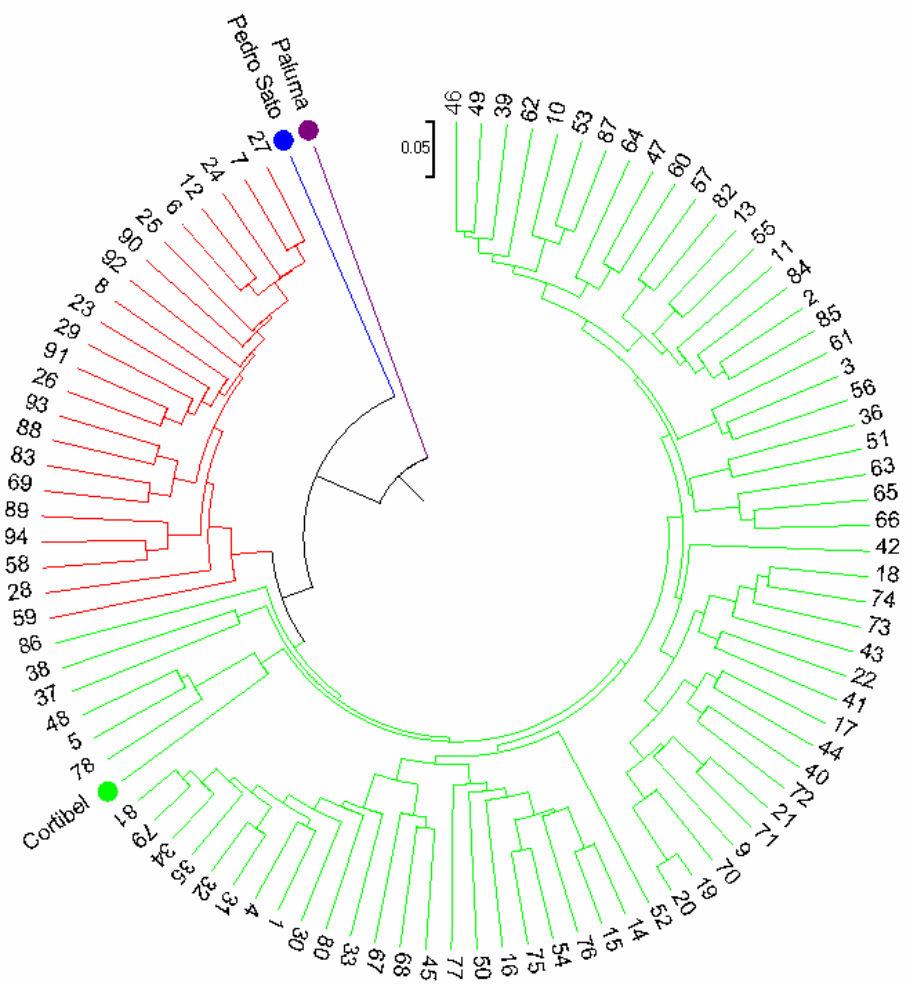

Figure 2. Dendrogram of genetic dissimilarity between 97 guava genotypes, obtained by the UPGMA method using SSR markers.

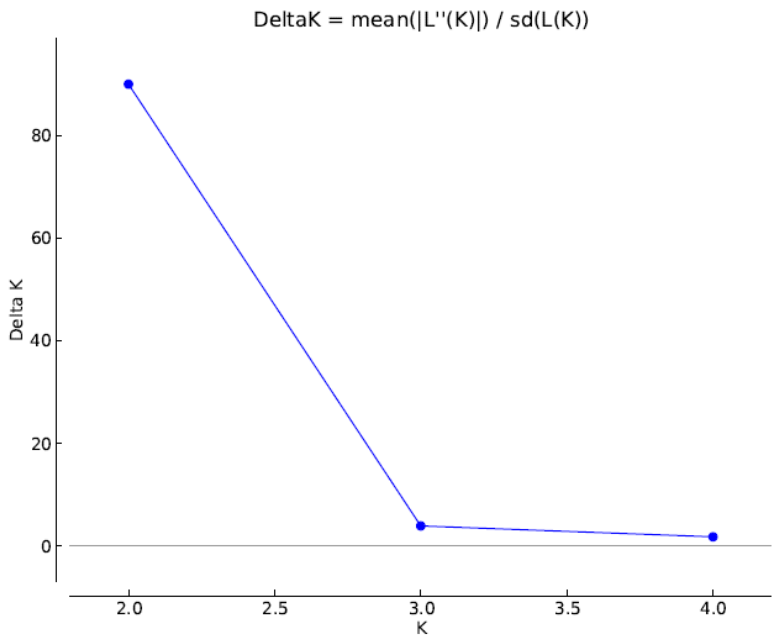

Figure 3. ΔK peak plot indicating the optimal number of genetic clusters for Bayesian analysis, obtained using Structure software v. 2.3.4.

By comparing the two strategies, we observe that the hierarchical method discriminated the Paluma and Pedro Sato cultivars, placing them in distinct groups and separating them from the rest of the population. Bayesian statistics, on the other hand, grouped these cultivars with the red cluster, which contains the smallest number of individuals. This difference is expected, since the methods make use of different criteria to define divergent groups. We also found that most individuals (green group) have alleles shared with variety Cortibel, which was observed by the two methods used. Conversely, a smaller number of individuals (red group) have ancestry and/or greater similarity with varieties Paluma and Pedro Sato.

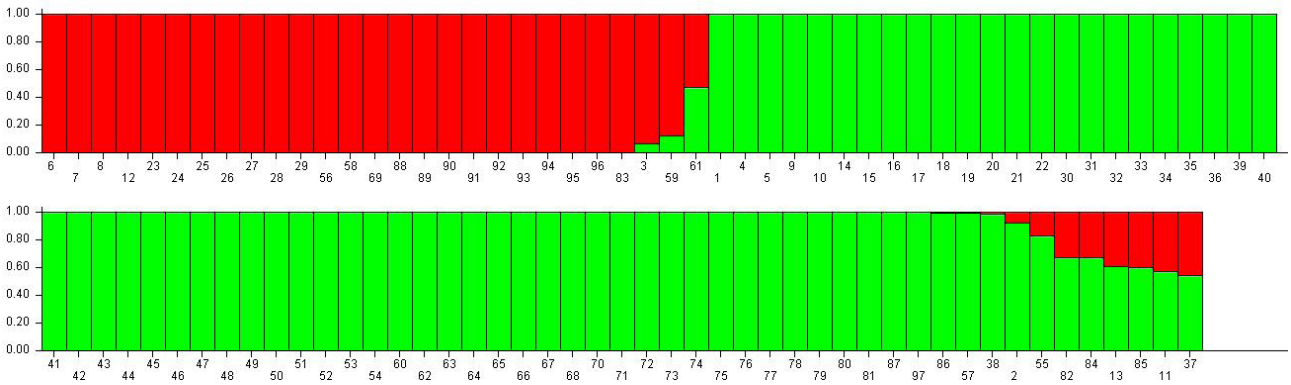

Figure 4. Clustering via Bayesian inference of 94 genotypes from 11 full-sib families and three cultivars: Paluma (95), Pedro Sato (96), and Cortibel 1 (97).
Genotypes are represented on the horizontal line, and each genetic group is represented by a color.

In general, although they are considered allogamous plants, in which the exchange of alleles is the main form of reproduction, a certain degree of homogeneity and a narrower genetic base are perceived, with few alleles being responsible for the differentiation of genotypes. This can be explained by the fact that these families are formed from a few selected plants with genomic regions inherited from the cultivars, which were possibly also obtained from more related genotypes. Urquía et al. (2019) reported a result similar to that of this study in an evaluation of 269 individuals of *P. guajava* distributed across three populations located on three islands of the Galápagos archipelago. The researchers found two clusters, with equivalent allelic diversity on all three islands.

To enable the continuity of the breeding program, the variability of the material must be monitored so that decision-making can be undertaken regarding the progress of the program. Phenotypic evaluations alone do not allow for an accurate detection of variability, since most of the agronomic traits are quantitative and the environment greatly influences the composition of the phenotype. Therefore, the identification of the two genetic groups is sufficient to help guide the next crosses between genotypes from different groups.

Conclusion

The SSR markers were efficient in discriminating the genotypes, which could help in new stages of the guava breeding program.

Bayesian inference and hierarchical method revealed a well-defined structure between the groups, indicating that future crosses should be performed between genotypes from different groups to maintain the genetic variability of the population.

Acknowledgments

The authors are thankful for financial sup-

port the Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (FAPERJ - E-26/010.001454/2019), Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Finance Code 001, and the National Council for Scientific and Technological Development (CNPq).

References

AMBRÓSIO, M.; VIANA, A.P.; RIBEIRO, R.M.; PREISIGKE, S.C.; CAVALCANTE, N.R.; SILVA, F.A.T.; XAVIER, G.; SOUSA, C.M.B. Genotypic superiority of Psidium Guajava S1 families using mixed modeling for truncated and simultaneous selection. **Scientia Agricola**, Piracicaba, v.78, n.2, p.1-9, 2021. *https://doi.org/10.1590/1678-992X-2019-0179*

BEZERRA, J.E.F.; LEDERMAN, I.E.; SILVA JUNIOR, J.F.; PROENÇA, C.E.B. Araçá. In: VIEIRA, R.F.; COSTA, T.S.A.; SILVA, D.B.; SANO, S.M.; FERREIRA, F.R. **Frutas nativas da região centro-oeste do Brasil**. Brasília, DF: EMBRAPA, 2006. p.42-62.

COSER, S.M.; FERREIRA, M.F.S.; FERREIRA, A.; MITRE, L.K.; CARVALHO, C.R.; CLARINDO, W.R. Assessment of genetic diversity in Psidium guajava L. using different approaches. **Scientia Horticulturae**, New York, v.148, p.223-9, 2012. *https://doi.org/10.1016/j.scienta.2012.09.030*

CRUZ, C.D. GENES – a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum**, Maringá, v.35, n.3, p.271-6, 2013. *https://doi.org/10.4025/actasciagron.v35i3.21251*

CRUZ, C.D.; FERREIRA, F.M.; PESSONI, L. **Biometria aplicada ao estudo da diversidade genética**. Viçosa: UFV; 2011.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, 2012. v.1, 514 p.

DOYLE, J.J.; DOYLE, J.L. Isolation of plant DNA from fresh tissue. **Focus**, Washington, v.12, p.13-5, 1990.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software structure: A simulation study. **Molecular Ecology**, Hoboken, v.14, n.8, p.2611-20, 2005. *https://doi.org/10.1111/j.1365-294X.2005.02553.x*

FACHI, L.R.; GARBUGIO, E.L.S.; FERREIRA, A.F.N.; MACHADO, R.F.C.; KRAUSE, W. Qualidade e correlação dos parâmetros físicos e químicos dos frutos de cultivares de goiaba. **Scientific Electronic Archives**, Snop, v.11, n.4, p.36-40, 2018.

FRANZON, R.C.; CAMPOS, L.Z.O.; PROENÇA, C.E.B.; SOUSA-SILVA, J.C. **Araças do gênero Psidium: principais espécies, ocorrência, descrição e usos**. Planaltina (DF): Embrapa Cerrados, 2009. 48 p.

GOMES, V.M.; RIBEIRO, R.M.; VIANA, A.P.; SOUZA. R.M.; SANTOS, E.A.; RODRIGUES, D.L.; ALMEIDA, O.F. Inheritance of resistance to Meloidogyne enterolobii and individual selection in segregating populations of Psidium spp. **Journal of Plant Pathology**, Pisa, v.148, p.699-708, 2017. *https://doi.org/10.1007/s10658-016-1128-y*

GOVAERTS, R.; SOBRAL, M.; ASHTON, P.; BARRIE, F.; HOLST, B.K.; LANDRUM, L.L.; MATSUMOTO, K.; MAZINE, F.F.; LUGHADHA, E.N.; PROENÇA, C.; SOARES-SILVA, L.H.; WILSON, P.G.; LUCAS, E. **World check-list of Myrtaceae**. London: Royal Botanic Gardens, 2008.

GUAVAMAP. **Improvement of Guava**: linkage mapping and QTL analysis as a basisfor marker-assisted selection. 2008. Disponível em: *https://cordis.europa.eu/project/id/15111/reporting*

INCAPER - Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural. Cadeia **produtiva da goiaba no Espírito Santo**. Vitória, ES: Incaper, 2023.

KAREEM, A.; JASKANI, M.J.; MEHMOOD, A.; KHAN, I.A.; AWAN, S.; SAJID, M.W. Morphogenetic characterization and phylogenetic relationship of Guava (Psidium guajava L.). **Revista Brasileira de Fruticultura**, Jaboticabal, v.40, n.4, p.e-069, 2018. *http://dx.doi.org/10.1590/0100-29452018069*

KHERWAR, D.; USHA, K.; MITHRA, S.V.A. Microsatellite (SSR) marker assisted assessment of population structure and genetic diversity for morpho-physiological traits in guava (Psidium guajava L.). **Journal of Plant Biochemistry and Biotechnology**, New Delhi, v.27, n.3, p.284-92, 2018. *https://doi.org/10.1007/s13562-017-0438-2*

KUMAR, C.; KUMAR, R.; SINGH, S.K.; GOSWAMI, A.K.; NAGARAJA, A.; PALIWAL, R.; SINGH, R. 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**, San Francisco, v.15, n.8, p.1-22, 2020. *https://doi.org/10.1371/journal.pone.0237538*

KUMAR, S.; NEI, M.; DUDLEY, J.; TAMURA, K. Mega: a biologist-centric software for evolutionary analysis of DNA and protein sequences. **Briefings in Bioinformatics**, Oxford, v.9, n.4, p.299-306, 2008. *https://doi.org/10.1093/bib/bbn017*

MA, Z.; LIU, S.; LIANG, Z.; XU, S.; HU, W. Analysis of genetic diversity of 45 guava germplasm evaluated using SSR markers. **International Journal of Fruit Science**, New York, v.20, n.3, p.385-93, 2020. *https://doi.org/10.1080/15538362.2019.1640168*

MAPA – Ministério da Agricultura, Pecuária e Abastecimento. **Registro nacional de cultivares - RNC**. Brasília (DF): MAPA, 2023. Disponível em: *https://sistemas.agricultura.gov.br/snpc/cultivarweb/cultivares_registradas.php*. Acesso em: 10 maio 2023.

MIAH, G.; RAFII, M.Y.; ISMAIL, M.R.; PUTEH, A.B.; RAHIM, H.A.; ISLAM, K.H.N.; LATIF, M.A. A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. **International Journal of Molecular Sciences**, Basel, v.14, n.11, p.22499-528, 2013. *https://doi.org/10.3390/ijms141122499*.

MOHAMMADI, S.A.; PRASANNA, B.M. Analysis of genetic diversity in cropplants – Salient statistical tools and considerations. **Crop Science**, Madison, v.43, n.4, p.1235-48, 2003. *https://doi.org/10.2135/cropsci2003.1235*

MOJENA, R. Hierarchical grouping methods and stopping rules: an evaluation. **The Computer Journal**, Oxford, v.20, n.4, p.359-63, 1977. *https://doi.org/10.1093/comjnl/20.4.359*

OLIVEIRA, E.J.; PÁDUA, J.G.; ZUCCHI, M.I.; VENCOVSKY, R.; VIEIRA, M.L.C. Origin, evolution and genome distribution of microsatellites. **Genetics and Molecular Biology**, Ribeirão Preto, v.29, n.2, p.294-307, 2006. *https://doi.org/10.1590/S1415-47572006000200018*

OLIVEIRA, M.S.P.; SILVA, K.J.D. Diferenciação genética entre procedências de açaizeiro por marcadores moleculares RAPD e SSR. **Revista Brasileira de Fruticultura**, Jaboticabal, v.30, n.2, p.438-43, 2008. *https://doi.org/10.1590/S0100-29452008000200029*

OLIVEIRA, N.N.S.; VIANA, A.P.; QUINTAL, S.S.R.; PAIVA, C.L.; MARINHO, C.S. Analysis of genetic distance between access of the gender Psidium by route markers ISSR. **Revista Brasileira de Fruticultura**, Jaboticabal, v.36, n.4, p.917-23, 2014. *https://doi.org/10.1590/0100-2945-413/13*

PARDO, E.; HERNÁNDEZ, K.; CAVADÍA, T. Molecular characterization of guava Psidium guajava in Cereté, Córdoba, Colombia. **Notulae Botanicae Horti Agrobotanici Cluj-Napoca**, Cluj-Napoca, v.51, n.3, p.13305, 2023. *https://doi.org/10.15835/nbha51313305*

PEAKALL, R.; SMOUSE, P.E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research an update. **Bioinformatics,** Oxford, v.28, n.19, p.2537-9, 2012. *https://doi.org/10.1093/bioinformatics/bts460*

PESSANHA, P.G.O.; VIANA, A.P.; AMARAL JÚNIOR, A.T.; SOUZA, R.M.; TEIXEIRA, M.C.; PEREIRA, M.G. Assessment of genetic diversity in access to Psidum spp. via RAPD markers. **Revista Brasileira de Fruticultura**, Jaboticabal, v.33, n.1, p.129-36, 2011. *https://doi.org/10.1590/S0100-29452011000100018*

PRITCHARD, J.K.; STEPHENS, M.; DONNELLY, P. Inference of population structureusing multilocus genotype data. **Genetics**. Austin, v.155, p. 945-59, 2000. *https://doi.org/10.1093/genetics/155.2.945*

QUINTAL, S.S.R.; VIANA, A.P.; CAMPOS, B.M.; VIVAS, M.; AMARAL JUNIOR, A.T. Analysis of structures of covariance and repeatability in guava segregant population. **Revista Caatinga**, Mossoró, v.30, n.4, p.885-91, 2017. *https://doi.org/10.1590/1983-21252017v30n408rc*

SANTOS, R.M.; VIANA, A.P.; SANTOS, E.A.; SOUZA, R.M.; ALMEIDA, O.F.; GOMES, V.M.; RODRIGUES, D.L.; SANTOS, P.R. 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. **Anais da Academia Brasileira de Ciências**, Rio de Janeiro, v.92, n.3, p.1-18, 2020. *https://doi.org/10.1590/0001-3765202020180600*

SILVA, F.A.; VIANA, A.P.; GUEDES, C.C.G.; CARVALHO, B.M.; SOUSA, C.M.B.; AMARAL, B.D.; AMBRÓSIO, M.; GLÓRIA, L.S. Impacto f Bayesian Inference on the Selection of Psidium guajava. **Scientific Reports,** London, v.10, p.1999-2008, 2020. *https://doi.org/10.1038/s41598-020-58850-6*

SOUSA, C.M.B.; RIBEIRO, R.M.; VIANA, A.P.; CAVALCANTE, N.R.; SILVA, F.A.; AMBRÓSIO, M.; AMARAL JÚNIOR, A.T. Guava breeding via full-sib family selection: conducting selection cycle and divergence btween parents and families. **Crop Breeding and Applied Biotechnology**, Londrina, v.20, n.1, p.1-9, 2020. *https://doi.org/10.1590/1984-70332020v20n1a12*

SYSTAT SOFTWARE. **SigmaPlot version 12.5**. San Jose, 2013. Disponível em: *https://sigmaplot.software.informer.com/12.5/*

THORNHILL, A.H.; HO, S.Y.; KÜLHEIM, C.; CRISP, M.D. Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. **Molecular Phylogenetics and Evolution**, Orlando, v.93, p.29-43, 2015. *https://dx.doi.org/10.1016/j.ympev.2015.07.007*

URQUÍA, D.; GUTIERREZ, B.; POZO, G.; POZO, M.J.; ESPÍN, A.; TORRES, M.L. Psidium guajava in the Galapagos Islands: Population genetics and history of an invasive species. **PloS One**, San Francisco, v.14, n.3, p.1-21, 2019. *https://doi.org/10.1371/journal.pone.0203737*

XIAO, Y.; ZHOU, L.; XIA, W.; MASON, A. S.; YANG, Y.; MA, Z.; PENG, M. Exploiting transcriptome data for the development and characterization of gene-based SSR markers related to cold tolerance in oil palm (Elaeis guineensis). **BMC Plant Biology**, London, v.14, n.1, p.384-96, 2014. *https://doi.org/10.1186/s12870-014-0384-2*