Genetic divergence between Brazilian Northern and Northeastern populations of Spondias mombin revealed with SNP markers

Allison Vieira Silva
Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz: Universidade de Sao Paulo Escola Superior de Agricultura Luiz de Queiroz
Caroline Bertocco Garcia
Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz: Universidade de Sao Paulo Escola Superior de Agricultura Luiz de Queiroz
Igor Araújo Santos Carvalho
Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz: Universidade de Sao Paulo Escola Superior de Agricultura Luiz de Queiroz
Wellington Ferreira Nascimento
Universidade Federal do Maranhão: Universidade Federal do Maranhao
Santiago Linorio Ferreyra Ramos
UFAM: Universidade Federal do Amazonas
Doriane Picanço Rodrigues
Universidade Federal do Amazonas
Maria Imaculada Zucchi
Agência Paulista de Tecnologia dos Agronegócios: Agencia Paulista de Tecnologia dos Agronegocios
Flaviane Malaquias Costa
Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz: Universidade de Sao Paulo Escola Superior de Agricultura Luiz de Queiroz
Alessandro Alves-Pereira
Universidade Federal do Amazonas
Carlos Eduardo de Araújo Batista
Instituto Tecnológico Vale: Instituto Tecnologico Vale Desenvolvimento Sustentavel
Edson Ferreira da Silva
Universidade Federal Rural de Pernambuco Departamento de Tecnologia Rural
Elizabeth Ann Veasey (eaveasey@usp.br)
Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz: Universidade de Sao Paulo Escola Superior de Agricultura Luiz de Queiroz https://orcid.org/0000-0002-7574-2020

Research Article

Keywords: Amazon. Conservation. Genetic structure. Population genomics. Yellow mombin

Posted Date: August 9th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3220961/v1

Abstract

Spondias mombin L. (family Anacardiaceae) is an arboreal and allogamous fruit tree native from southern Mexico to southeastern Brazil, with great potential for economic exploitation. This study aimed to evaluate the structure and genomic diversity of yellow mombin in Brazil's North and Northeast regions using SNP (Single Nucleotide Polymorphisms) markers. Significant genetic divergence was observed in the clustering analysis and dendrogram construction between the North and Northeast regions. The Mantel test identified a high positive and significant correlation ($r = 0.78$; $p < 0.001$), indicating isolation by distance. In the genetic diversity analysis, populations from the North Region presented a greater number of alleles (2.722, on average) and genetic diversity ($H_E = 0.1860$) regarding the populations from the Northeast Region (2.509 alleles and $H_E = 0.1059$). Although presenting greater genetic diversity, the North Region had a positive inbreeding coefficient (f) in three of the four studied populations, ranging from 0.0855 to 0.2421. The results obtained contribute to the understanding of the distribution of genetic variation and the conservation status of the yellow mombin populations in the two regions; they could also be used as a subsidy for developing conservation strategies and the genetic improvement of this species.

Introduction

Several fruit species with economic potential are found in the Amazon. Among these, the yellow mombin (*Spondias mombin* L., Anacardiaceae) stands out for the agro-industrial value of its fruits, which can be consumed in natura or the form of processed products (Silva et al. 2004). Yellow mombin is a semi-domesticated species, native to areas ranging from southern Mexico to southeastern Brazil (Mitchell and Daly 2015), and its centers of diversity are the Atlantic Forest and the Western Amazon, more specifically the border regions between the State of Acre, in Brazil, and Peru and Bolivia (Fonseca et al. 2017). The cultivation of this fruit tree was carried out by indigenous people in the Amazon even before the arrival of Europeans in the Americas in 1492 (Clement 1999), and since then, it has been observed quite frequently in the North, Northeast, and Midwest regions of Brazil. In the Northeast region, it was described for the first time in 1587 in the Treaty Descriptive of Brazil by the Portuguese author Gabriel Soares de Souza (Sacramento and Souza 2009).

It is an arboreal and erect species, reaching more than 20 m in height, with a trunk covered by very thick, grayish, rough, protruding bark and with cracks (Lorenzi 1992). The leaves are compound, alternate, and imparipinnate, with 5 to 11 pairs of leaflets (Sacramento and Souza 2009). *S. mombin* (2n = 2x = 30) reproduces by cross-pollination performed mainly by bees, and it also presents vegetative propagation. Furthermore, it presents protandrous dichogamy and a high value of the pollen/ovule ratio (Ramos 2009; Zortéa et al. 2019). The species is classified as andromonoecious (Ramos 2009), having staminate and hermaphrodite flowers arranged in inflorescences of the terminal pyramidal panicles type (Coradin et al. 2022).

Among the plant parts, the fruits of yellow mombin have greater economic importance and can be consumed raw or transformed into various products. The fruits are bright yellow drupes when ripe, with juicy mesocarp (Lorenzi 1992). The pulp has high levels of potassium, magnesium, phosphorus, and copper compared to other fruits consumed in Brazil; it has phenolic compounds, antioxidants, and carotenoids in greater concentration, which gives the fruit high nutritional value, and helps to prevent several diseases, including cardiovascular (Tiburski et al. 2011). Nearly every part of the tree, ranging from its fruit, leaves, seeds, stem bark, bark, and even flowers, has medicinal properties that have been used in ethnomedicine in cases of abortion, constipation, fever, gonorrhea, postpartum hemorrhage, digestive pain, diarrhea, dysentery, and wounds (Onguro et al. 2023). However, yellow mombin is still in the process of domestication and exploitation of this fruit is still carried out, in large part, through extractivism. Thus, studies

regarding the characterization of this genetic resource are extremely important for advancing research involving the species (Silva et al. 2017a).

The *ex situ* and *in situ* preservation of *Spondias* species germplasm that has commercial value is still very restricted and considered vulnerable. Also, preservation does not follow the accelerated process of genetic erosion caused by anthropic action, which may lead to the loss of agronomically valuable individuals (Silva et al. 2004). The Brazilian Amazon, for example, has been subjected to a considerable rate of deforestation (INPE 2020), with livestock, large-scale agriculture, and slash-and-burn agriculture as the main direct causes (Rivero and Almeida 2009). The model of predatory exploitation of forest resources, based on the replacement of forest areas by extensions of pasture and cultivated areas, does not consider the particularities of the various ecological spaces in the region (Vieira et al. 2008). As a result of the practice of this unsustainable model in tropical forests, there is a loss of genetic diversity and unfavorable changes in the frequency of alleles with adaptive or commercial value (O'Neill et al. 2001).

Characterizing the diversity and genetic structure of the natural populations of yellow mombin is fundamental to outlining conservation strategies and exploring the genetic resources of this species. Previous studies carried out to study the structure and genetic diversity of the species have used morphological and physical/chemical data (Silva et al. 2014; Silva et al. 2017a), as well as isoenzymatic (Silva et al. 2009) and molecular markers, such as RAPD (Random Amplified Polymorphic DNA) (Lima et al. 2011), AFLP (Amplified Fragment Length Polymorphism) (Santos and Oliveira 2008; Magalhães et al. 2013), ISSR (Inter Simple Sequence Repeat) (Silva et al. 2017b), SSR (Simple Sequence Repeat) (Cristóbal Pérez et al. 2019; Zortéa et al. 2021), nrDNA (nuclear ribosomal DNA) and SNPs (Single Nuclear Polymorphism) (Nobre et al. 2018). However, none of these studies used such a geographically extensive sample, covering three biomes, as carried out in the present study.

SNP markers, obtained through the genotyping-by-sequencing technique (GBS) (Elshire et al. 2011), are the most abundant in the genome. Also, in addition to the evaluation of neutral variation, they allow the study and identification of regions of the genome that are undergoing natural selection in the population, that is, regions directly associated with adaptation (Cortinovis et al. 2020), presenting great potential in population genomics studies. SNP markers have been used in studies with fruit tree species from the Amazon, such as the study of population genomics of cacao (*Theobroma cacao* L.) to understand the species' domestication process (Cornejo et al., 2018), the study of population structure of rubber tree [*Hevea brasiliensis* (Willd. ex Adr. de Juss.) Muell-Arg.] (Souza et al. 2018), as well as the analysis of the genomic diversity of the juçara palm (*Euterpe edulis* Mart.) by Alves-Pereira et al. (2019). The population structure and genetic diversity of species of the genus *Acrocomia* were evaluated by Díaz et al. (2021) with this marker, while the effects of artificial selection on the genetic differentiation of *Elaeis oleifera* (Kunth) Cortés accessions were studied by Leão et al. (2022). In the genus *Spondias*, SNP markers have been used so far to test the hypothesis of a hybrid origin of the species *S. bahiensis* P. Carvalho, Van den Berg and M. Machado, as well as the species *S. mombin* and *S. tuberosa* Arruda (Nobre et al. 2018).

The present study aims to characterize the diversity and genetic structure of *S. mombin* populations with a wide sampling, including two regions (North and Northeast) and three biomes (Amazon, Cerrado, and Atlantic Forest) of Brazil, from SNP markers obtained with the genotyping by sequencing (GBS) technique; this is the first study of such magnitude developed for the species. This information makes it possible to identify populations with genetic vulnerability and provide information that could subsidize the management and in situ and ex situ conservation of genetic variability and contribute to developing future genetic improvement programs.

Materials and methods

Sampling, DNA extraction and quantification

Young leaves of yellow mombin adult individuals were randomly collected from nine populations in the North and Northeast Regions of Brazil: four in the North Region, all in the States of Amazonas-AM (Amazon Biome); five in the Northeast Region, including two populations in the State of Pernambuco-PE, one in Paraíba-PB, one in Bahia-BA, all from the Atlantic Forest Biome; and one population from the State of Maranhão-MA, located in the Cerrado Biome (Table 1; Fig. 1). The collections were carried out along roadsides far from the urban centers of cities. It was possible to visualize that, although uninhabited, the sampling areas showed clear signs of anthropic action, with little presence of large trees.

DNA was extracted from leaves using the protocol described by Inglis et al. (2018) with some modifications, including three to four prewashes using sorbitol wash buffer (100 mM Tris-HCl pH 8.0, 0.35 M Sorbitol, 5 mM EDTA pH 8.0, 1% (w/v) Polyvinylpyrrolidone (average molecular weight 40,000; PVP-40). Sample quantification was estimated using GelRed dye under UV in 1% agarose gels, comparing specific and known concentrations of lambda phage DNA (5, 20, 40, 80, and 160 ng), and validated with a high sensitivity fluorometer (Invitrogen), the Qubit 4, which presents a high level of accuracy. After quantification, DNA samples were normalized to a concentration of 20 ng μL^{-1} for GBS library preparation.

Table 1
Yellow mombin (*Spondias mombin*) populations used in the study including their origin (municipality/State), number of individuals sampled, biome, and geographic coordinates.

Population	Municipality-State[a]	Nº individuals	Biome	Latitude	Longitude
NV	Novo Airão-AM	5	Amazon	2°37'20,8"S	60°57'01,7"W
IR	Iranduba-AM	2	Amazon	3°17'03,4"S	60°11'16,2"W
PF	Presidente Figueiredo-AM	5	Amazon	2°03'15,5"S	60°01'41,8"W
SV	Silves-AM	3	Amazon	04⁰14'41,0"S	42⁰17'40,0"W
CP	Chapadinha-MA	3	Cerrado	3°44'43,0"S	43°22'20,3"W
PD	Paudalho-PE	6	Atlantic Forest	7º53'48,0"S	35º10’47,0”w
SM	São Lourenço da Mata-PE	6	Atlantic Forest	8º00'10,0"S	35º01'04,0"W
AR	Areia-PB	6	Atlantic Forest	6°59'19,6"S	35°44'11,6"W
MS	Mata de São João-BA	5	Atlantic Forest	12°29'19,6"S	37°59'42,8"W
Total		41			

[a] States = AM – Amazonas, PE - Pernambuco, PB - Paraíba, and BA - Bahia

Assembly of the genomic library and identification of SNPs

After the first normalization using agarose gel electrophoresis, the samples were quantified once again and normalized as needed. After digestion tests, the two enzymes to be used in yellow mombin were selected, P*st*I (specific cutting enzyme) and M*se*I (common cutting enzyme). The samples, now digested, had the barcode-type adapters attached to the fragments generated in the digestion. Biding was performed with the aid of the enzyme NEB T4 DNA ligase #M0202 in NEB Buffer4 buffer in the presence of ATP. To ensure success in the ligation reaction, T4 ligase was used at a higher concentration. After the binding reaction, each sample is bound to a specific barcode. Barcodes are sequences of up to six nucleotides that differ from each other in at least one base and that serve as identifiers, allowing the multiplexing of several samples, since each sample has a specific barcode and can be identified later. At

the end of multiplexing, the multiplex (set of identified samples) was amplified by PCR (polymerase chain reaction). The enrichment of the fragments was carried out using primers complementary to the adapters that were used in the preparation of the GBS library of the 41 yellow mombin samples. The sequencing of PCR products was performed on the HiSeq2500 Illumina platform.

SNP markers were obtained using the Stacks v. 1.42 software (Catchen et al. 2011). The samples were initially submitted to the process_radtags stage, where the sequences undergo quality control and sequences that could compromise the veracity of the analyses are eliminated. In this first step, the demultiplex is also performed, where the sequence of each individual is properly separated according to the different *barcodes*. After the first step (filtering and demultiplexing), the retained sequences were analyzed by the ustacks component with the parameters -m 3 (which defines the number of reads needed to form a stack), -M 2 (specifies the number of mismatches allowed among stacks to be grouped into a locus), and -N 2 (the maximum distance allowed between primary stacks and secondary reads to be inserted into stacks). The next step was to create a catalog with all the possible identified loci and the presence of these loci in each of the individuals. This step was performed by cstacks with the parameter -n 2 (the parameter determines the number of mismatches between the loci of the samples for the formation of the catalog). After the formation of the catalog, two components were used to cross-reference information between the loci obtained for each individual and the loci of the catalog (sstacks) and remove the loci with less probability (rxstacks, --lnl_lim − 10). The population component was used for the final filtering of SNP markers. Only one SNP per tag was retained, with depth > = 3X, MAF (minimum allele frequency) > = 0.01, SNP present in at least 60% of samples within populations, SNP present in at least five of the populations.

Statistical analysis

SNP markers were analyzed with R software (R Core Team 2020) from specialized packages. Genetic structuring analyses were performed using DAPC (Discriminant Analysis of Principal Components) (Jombart et al. 2010). DAPC was performed using the *adegenet* package (Jombart 2008) in the R software. The analysis aims to identify groups based on genetic structure, minimizing variation within groups and maximizing variation between groups. The number of regional groups (population collection sites) is not provided in the analysis, and the optimal number of groups, which explains the data, is calculated by the K-means method, where several values of k (groups) and several groupings are tested by using the Bayesian Information Criterion (BIC). The generated results are provided in a curve of BIC values as a function of k, where the ideal value of groups is taken as the lowest value of the curve, from which the curve returns to its crescent. Clustering analyses were carried out from phylogenetic trees constructed by the *Neighbor-Joining* method without rooting, generated based on the genetic distance of Nei (1972) with the aid of the ape package (Paradis and Schliep 2019).

Estimates of the basic parameters of genetic diversity, such as total number of alleles (*A*), observed heterozygosity (H_O), expected heterozygosity (H_E), and inbreeding coefficient (*f*), were generated with the aid of the hierfstat package (Goudet and Jombart 2020). The Molecular Variance Analysis (AMOVA) was conducted using the poppr package (Kamvar et al. 2014, 2015). Both the basic parameters and the AMOVA were carried out in two situations, considering the regional populations collected, and the groups formed from the DAPC analysis. With the help of the hierfstat package, a pairwise *F*st matrix was generated, and the graphical visualization of the matrix was generated with the help of the corrplot package (Wei and Simko 2017). The Mantel test was performed to assess the correlation between genetic divergence between populations based on the pairwise *F*st value and the pairwise geographic distance of the populations in the ade4 package (Chessel et al. 2004; Dray and Dufour 2007; Dray et al. 2007; Bougeard and Dray 2018; Thioulouse et al. 2018). The geographic distance matrix, generated from geographic coordinates, was built with the aid of the geodist package (Padgham and Sumner 2020).

Results

Genetic diversity and structure of *Spondias mombin* populations

Structure and genetic diversity studies were carried out with 2,003 loci corresponding to SNP markers for 41 individuals of yellow mombin. The DAPC analysis was performed using 15 principal components, which explained 77% of the total variation, and two discriminant functions. The optimal number of groups obtained from the *K-means* analysis corresponds to three groups (Fig. 2a and 2b). Group G1, in red, is composed of 25 individuals from the populations of Paudalho-PE (PD), São Lourenço da Mata-PE (SM), Areia-PB (AR), Mata de São João-BA (MS), one accession from the population of Chapadinha-MA (CP) and one accession from the Presidente Figueiredo-AM (PF) population. Group G2, in green, is composed of 14 individuals collected in the State of Amazonas, at Iranduba (IR), Silves (SV), Presidente Figueiredo (PF), and Novo Airão (NV) populations. Group G3, in blue, comprises two of the three individuals from the Chapadinha-MA population. The DAPC analysis with K = 2 grouped the Chapadinha-MA population with the other northeastern populations in the G1 group (Suppl. Fig. S1).

Groups G1 and G2 are genetically closer than the G3 group, which, although its individuals are from a population of the Northeast Region, was isolated from the other two groups. In the DAPC analysis, with only one discriminant function, it was possible to visualize the grouping result in only one dimension (Fig. 2b). Groups G1 and G2 show an overlap area, with a greater distribution amplitude in group G2, demonstrating greater genetic variation within this group than others. In the membership probability analysis, based on the data generated in the DAPC analysis and with optimal k = 3, there were three groups observed, with group 1 (G1 - red) represented mostly by populations from the Northeast, group 2 (G2 - green) by the populations of the North, and group 3 (G3 - blue) by the population of Maranhão (Fig. 3). In this analysis, the separation of individuals by macroregions was explicit. However, two exceptions were observed: an accession from the State of Amazonas (PF) and one of the three individuals from the Chapadinha population of Maranhão within group G1. In the G2 group, only individuals from the North Region, represented by populations collected in the State of Amazonas, were found. The G3 group was isolated from the others, formed by two of the three individuals from the Chapadinha population of Maranhão.

Cluster analysis was performed using a genetic distance tree built using the *Neighbor-Joining* method (Fig. 4). The tree was built without defined rooting, not for evolutionary inferences but for understanding how the groups defined based on the DAPC analysis relate to each other and where a pattern of grouping by collection sites was observed. Five distinct groups were formed. Groups I (populations of Bahia, BA), II (populations of Pernambuco, PE), and III (populations of Paraíba, PB), which are equivalent to group G1 in the DAPC analysis, formed three well-defined groups, without individuals from other states. Group IV, which is equivalent to group G3 in the DAPC analysis, was formed by two of the three individuals from the Maranhão (MA) population and was placed between the group of populations from the Northeast (G1) and group V, formed by populations from Amazonas (AM), equivalent to G2. In the dendrogram, also constructed using the *Neighbor-Joining* method (Suppl. Fig. S2) and Nei's distance (Nei 1972), it is possible to observe the behavior of each individual regarding the groups identified by the DAPC analysis. In this analysis, the third individual of Chapadinha, MA, classified in the G1 group, is genetically very close to the other two Chapadinha individuals of group G3 in the DAPC.

In the pairwise analysis of genetic divergence (*F*st), high genetic differentiation was found between all groups, with values greater than 0.25 in all scenarios according to the scale for *F*st presented by Hartl and Clark (1998). The highest *F*st values were identified between groups G1 and G3 (0.37), and G2 and G3 (0.36), and the lowest *F*st value (0.33) was observed between groups G1 and G2 (Table 2). AMOVA provided similar results for the DAPC groups and collected populations, both significant at 5% probability (Table 3). In both scenarios, the highest percentage of variation occurred within populations and groups, 61.4% and 59.3%, respectively.

Table 2
Result of pairwise *F*st analysis between groups identified by DAPC analysis for *Spondias mombin* species. Group G1 is formed by the populations of Paudalho-Pernambuco, São Lourenço da Mata-Pernambuco, Mata de São João-Bahia, and Areia-Paraíba. Group G2 is composed by the populations of Iranduba, Novo Airão, Presidente Figueiredo and Silves, all from the State of Amazonas. Group G3 is formed by two individuals from the population of Chapadinha-Maranhão.

F_{st}	G1	G2	G3
G1	0.00		
G2	0.33	0.00	
G3	0.37	0.36	0.00

Table 3
Results generated by AMOVA to identify the sources of genetic variation among and within the groups identified in the DAPC and the collected populations of *Spondias mombin.*

Source of variation	Degree of freedom	Sums of square	Coefficient of variation	Percent variation	F statistics
Among populations	8	12,636.86	259.25	38.56	0.3856*
Within populations	32	13,215.64	412.99	61.44	
Among DAPC groups	2	2,481.53	104.30	40.69	0.4070*
Within DAPC groups	38	5,775.89	151.99	59.31	

- *Significant at *p*-value < 0.05.

The pairwise *F*st matrix, generated with all the collected populations, enabled the identification of the greatest genetic distance when comparing the populations from the Northeast region with those from the North region (Table 4; Fig. 5). When the Northeastern populations were compared to each other, the populations of São Lourenço da Mata (SM) and Paudalho (PD), both from the State of Pernambuco, presented the lowest *F*st value (0.06). The highest *F*st value within the Northeast region was obtained when the populations of Areia (AR), Paraíba, and Mata de São João (MS), Bahia, were compared, generating an *F*st of 0.21, with a genetic divergence considered high (Hartl and Clark 1998). Within the North region populations, the lowest *F*st value (0.05) was obtained when comparing the Silves (SV) and Presidente Figueiredo (PF) populations, indicating low genetic divergence. The greatest genetic divergence between the Amazonian populations was obtained when Iranduba (IR) and Novo Airão (NV) were compared to each other, with a *F*st value of 0.22, which indicates a high genetic divergence between the populations.

The geographic distance matrix constructed from the geographic coordinates presents distance values between the populations varying from 21.4 km to 2,934.4 km (Table 4). The Mantel test identified a positive and significant correlation (r = 0.78; p-value < 0.001) between the geographic distance among populations and the values of genetic divergence (*F*st), showing the occurrence of isolation by distance for the yellow mombin populations.

Table 4

On the upper diagonal, the pairwise matrix of geographical distance was calculated between the populations, and on the lower diagonal, the *F*st values were calculated between the populations. The distance values shown are in kilometers. The populations are represented by the abbreviations AR (Areia-Paraíba), CP (Chapadinha-Maranhão), IR (Iranduba-Amazonas), MS (Mata de São João-Bahia), NV (Novo Airão-Amazonas), PD (Paudalho-Pernambuco), PF (Presidente Figueiredo-Amazonas), SM (São Lourenço da Mata-Pernambuco), and SV (Silves-Amazonas).

Populations	AR	CP	IR	MS	NV	PD	PF	SM	SV
AR		919.04	2,741.17	656.84	2,837.75	117.73	2,749.53	137.37	2,553.47
CP	0.46		1,869.07	1,133.84	1,957.92	1,016.66	1,861.16	1,037.87	1,672.22
IR	0.51	0.62		2,647.72	112.19	2,816.34	137.16	2,835.82	205.18
MS	0.21	0.44	0.50		2,755.27	594.25	2,689.56	593.78	2,487.38
NV	0.47	0.49	0.23	0.44		2,914.74	120.44	2,934.39	285.75
PD	0.16	0.43	0.49	0.15	0.45		2,829.20	21.36	2,631.60
PF	0.41	0.44	0.19	0.39	0.16	0.38		2,849.09	202.19
SM	0.17	0.43	0.49	0.15	0.45	0.06	0.39		2,651.36
SV	0.39	0.43	0.21	0.37	0.17	0.36	0.08	0.37	

Two scenarios were considered to obtain the basic diversity parameters. The first considered the groups identified in the DAPC analysis, and the second considered the collection sites in the definition of populations (Table 5). The G2 group had the highest number of alleles (*A* = 3,547) and the greatest genetic diversity (*He* = 0.2031), and it was 37% more genetically diverse than the G1 group and 70% more diverse than the G3 group, presenting 649 more alleles than the G1 group and 1307 more than the G3 group. The result presented by the G2 group agrees with the inference obtained in the DAPC density analysis, showing a wider distribution, followed by the G1 group. Group G3 had the lowest number of alleles (*A* = 2,240) and low genetic diversity (*He* = 0.0594). However, the G2 group had the highest value for the inbreeding coefficient (*f* = 0.2280), almost double in relation to the G1 group. The G3 group had a negative value (*f* = -0.8529).

Table 5
Results of the genetic diversity analysis carried out between the groups identified in the DAPC and the populations of yellow mombin (*Spondias mombin*). The number of individuals (N), the total number of alleles (*A*), the observed heterozygosity (*Ho*), the expected heterozygosity (*He*), and the inbreeding coefficient (*f*) were analyzed for the populations: PD (Paudalho-Pernambuco), SM (São Lourenço da Mata-Pernambuco), MS (Mata de São João-Bahia), AR (Areia-Paraíba), CP (Chapadinha-Maranhão), from the Northeast region; and NV (Novo Airão-Amazonas), IR (Iranduba-Amazonas), PF (Presidente Figueiredo-Amazonas), and SV (Silves-Amazonas), from the North region.

Groups	***N***	***A***	***Ho***	***He***	***F***
G1	25	2,898	0.1128	0.1279	0.1187
G2	14	3,547	0.1568	0.2031	0.2280
G3	2	2,240	0.1101	0.0594	-0.8529
Average	-	2,895	0.1266	0.1301	-0.5062
Populations	*N*	*A*	*Ho*	*He*	*F*
PD	6	2,624	0.1205	0.1184	-0.0177
SM	6	2,624	0.1254	0.1178	-0.0649
AR	6	2,492	0.0833	0.0990	0.1581
MS	5	2,562	0.1213	0.1170	-0.0371
CP	3	2,244	0.1100	0.0606	-0.8146
Average (NE)	-	2,509	0.1121	0.1059	-0.0587
IR	2	2,447	0.1790	0.1094	-0.6362
SV	3	2,778	0.1873	0.2092	0.1045
NV	5	2,806	0.1509	0.1650	0.0855
PF	5	2,856	0.1505	0.1986	0.2421
Average (N)	-	2,722	0.1669	0.1860	0.1025

In the genetic diversity analysis considering the populations, the result was close to that obtained with the DAPC groups. Populations from the North region presented, on average, higher values of genetic diversity (*A* = 2,722; *He* = 0.1860) when compared to the populations from the Northeast region (*A* = 2,509; *He* = 0.1056) (Table 5). The Iranduba population presented divergent values from the other populations in the Amazonas State for the number of alleles, genetic diversity, and inbreeding coefficient, presenting observed heterozygosity superior to the expected heterozygosity. The Northeast Region populations presented similar values for the diversity parameters, except for the population of Chapadinha which had the lowest number of alleles and *He* among all the evaluated populations. The inbreeding coefficient had a positive value (0.1581) in the Areia population and negative value in the other populations.

Discussion

The results clearly showed the separation between the North and Northeast regions of Brazil regarding population genetic structure and a trend of grouping yellow mombin populations by State. The pairwise *F*st matrix constructed with the groups identified in the DAPC analysis showed Fst > 0.25 in all combinations between the three groups, indicating that they had segregation in the DAPC and cluster analyses. *F*st values of this magnitude indicate very high genetic differentiation between populations (Hartl and Clark 1998). Also, the pairwise *F*st matrix among the nine populations studied allowed the identification of the lower divergence between populations from the same region, both among populations from the Northeast and among those from the North. The *F*st values between populations of the same macro-region indicate genetic divergence of a moderate (between 0.05 and 0.15) to a high degree (between 0.15 and 0.25) (Hartl and Clark 1998). The opposite occurs, however, when comparing a population from the Northeast region with a population from the North region of the country. In all scenarios in which a population from the Northeast region is compared with a population from the North region, the values obtained indicate very high genetic divergence, with *F*st ranging from 0.36 to 0.50. Garcia (2023) and Nascimento et al. (2021) also observed this clear separation between the North and Northeast regions when studying populations of bacuri tree (*Platonia insignis* Mart.) with SNP and SSR markers, respectively.

The population from Chapadinha, although located in the Northeast Region of Brazil, presents similar *F*st values when compared with other populations in the Northeast region and with populations in the North Region, showing the highest value of genetic divergence when compared with the population from Iranduba (0.62). A possible explanation is that this population was the only one sampled in the Cerrado biome, which may also have led to adaptations and differentiation through evolutionary processes of natural selection or genetic drift. The discrimination between Cerrado and Amazon populations of the fruit tree *P. insignis* based on eight nuclear (ncSSR) and three chloroplast (cpSSR) microsatellite markers was also observed by Nascimento et al. (2021), including the municipality of Chapadinha among the Cerrado populations.

AMOVA showed that the greatest source of genetic variation is found among individuals within (61.4%) than among populations (38.6%), and within (59.3%) than among DAPC groups (40.7%). Higher intrapopulation than interpopulation genetic variation is expected in allogamous perennial species since most variation occurs between individuals within populations (Hamrick et al. 1992; Paschoa et al. 2018; Francisconi et al. 2021). Yellow mombin presents sexual and asexual (vegetative) reproduction. It is an allogamous species and depends on cross-pollination for greater reproductive success, showing protandrous dichogamy and a high value for the pollen/ovule ratio (Ramos 2009).

Yellow mombin fruits are appreciated as juice and drinks and represent an important source of income in many regions. A study by Silva et al. (2009) with populations of this species in Northeast Brazil suggested that commercialization and the exchange of fruits among communities are important mechanisms acting in the gene flow between distant populations of this species. The dispersion of fruits over long distances are factors that favor a greater concentration of genetic variation within populations of tree species (Hamrick et al. 1992). Although intra-population variation was the major source of variation detected in the AMOVA, the percentage of inter-population variation was also quite expressive. F statistics indicate significant genetic divergence between the DAPC groups and the populations, results similar to those identified by Magalhães et al. (2013), who analyzed yellow mombin populations in the Amazon using AFLP markers and identified 47.2% of the variation between populations and 52.8% within the studied populations. Two studies evaluating populations of *S. mombin* originating from the State of Mato Grosso, in the Midwest region of Brazil, also observed that most of the genetic variation was found within populations: 77.4% by Silva et al. (2017) with ISSR markers and 90.6% by Zortéa et al. (2021) with SSR markers. In studies carried out with other tropical fruit species, Surapaneni et al. (2013) also observed that 62.2% of the genetic variation in *Mangifera indica* L. originated from variation between individuals within populations, while Santos et al. (2019)

observed that 78% of the variation in *Anacardium occidentale* L. originated between individuals within populations. Therefore, the execution of conservation actions or the use of the genetic diversity of yellow mombin in the north and northeast of Brazil must consider greater diversity within populations or regions than between them.

Considering the genetic diversity parameters, the data show the North region had the highest values for genetic diversity identified in the study, which can be expected as this region encompasses the center of diversity for the species (Fonseca et al. 2017). In a study carried out with six species of tropical tree angiosperms, seven of the nine populations presented *He* values higher than those found by Hamrick et al. (1992), identifying a value of 0.106 for the average genetic diversity presented by the species, with the caveat that the markers used in that study (isoenzymes) are different from those used in this study (SNPs). It is noteworthy that all populations in the North region had higher values than those found by Hamrick et al. (1992) for genetic diversity (*He*), ranging from 0.109 to 0.209.

Two yellow mombin populations (Chapadinha and Areia) showed low genetic diversity values that may be related to the reduction in the number of individuals in these populations, which appear to suffer more intense anthropic action. The Chapadinha population, however, from the State of Maranhão, presented the observed heterozygosity (*Ho*) higher than the expected heterozygosity (*He*), which indicates the recent introduction of alleles in the population. The Areia population is a highland swamp located in the State of Paraíba that also showed a genetic diversity value below the expected average and, in addition, had the highest inbreeding coefficient (f = 0.1581), the only one positive among the Northeast populations. The high-altitude microclimate of the region (highland swamps) could interfere with the local flora. However, this population has great potential for supplying genes related to water stress. Areia is a municipality located in an altitude swamp, and its predominant biome is the Atlantic Forest, despite having smaller areas of Caatinga (Moreira 1989; Barbosa et al. 2004). Thus, the northeastern semi-arid region is expected to significantly influence its flora. Therefore, we consider it extremely important to maintain diversity in this population, increasing the chances of obtaining the greatest possible number of genes adapted to such adverse edaphoclimatic conditions.

In a study carried out by Silva et al. (2009) with four yellow mombin populations from Pernambuco assessed with nine isoenzyme loci, the values for genetic diversity ranged from 0.530 to 0.574. Using SSR markers, Zortéa et al. (2019) found values ranging from 0.199 to 0.486, with a mean of *He* = 0.335, for populations in Mato Grosso, Midwest region. A study with three populations from the north of the State of Mato Grosso based on 14 ISSR primers identified genetic diversity values ranging from 0.20 to 0.24 (Silva et al. 2017b). Lins Neto et al. (2013), based on 11 ISSR markers, observed values of genetic diversity ranging from 0.27 to 0.34 for *Spondias tuberosa* L. The values for genetic diversity observed in the present study are lower than those previously identified for the species. This fact may be related to the larger number of loci obtained using SNP markers (a total of 2,003 SNP loci) compared to other studies, making the results more accurate. In a comparative study between SNPs and microsatellite markers, Fischer et al. (2017) found, in the same sample, mean values of genetic diversity (*He*) of 0.378 when using SSR markers and 0.148 when using SNPs. Lemopoulos et al. (2019), also in a comparative study, identified a mean value of genetic diversity of 0.597 with SSR and 0.11 when using SNPs. According to Fischer et al. (2017), SNP markers, derived from NGS technology, proved to be superior to SSR markers in carrying out genomic studies for analyses of genetic diversity between and within populations, concluding that a few thousand SNPs are sufficient to carry out robust analyzes and not biases of genetic diversity. Lemopoulos et al. (2019) concluded that SNP markers (in large amounts) are more informative and provide greater resolution even in studies performed with small and isolated populations. Garke et al. (2012) performed a comparative study between SNPs and SSR in identifying genetic structure among populations. They concluded that 70 SNP markers had the same performance as 29 SSR markers in discriminant analyses based on PCA. Furthermore, in the same study, the authors concluded that a set with more than 250 SNP markers performed better than a set of 29 SSR markers in cluster analysis and in identifying genetic mixing between populations.

Considering the inbreeding coefficient, except for the Iranduba population, all other Amazonian populations showed positive values for the inbreeding coefficient. The preservation policies of the Amazon Forest have been weakened each year with a 9.5% increase in the deforestation rate between 2019 and 2020 (INPE 2020), increasing the state of alert for the yellow mombin populations in the region, which, despite having greater genetic diversity, have a higher inbreeding coefficient and risk of genetic erosion.

Brazil has the highest number of *S. mombin* accessions maintained in germplasm banks, and all accessions are maintained *ex situ* (Santana 2010; Portal Alelo Recursos Genéticos 2022). However, only one accession from the State of Amazonas was verified in the existing germplasm banks. There is an urgent need to include accessions from the northern region in the existing germplasm bank or create new banks for the conservation of the genetic diversity of that region, maximizing the chances of preserving the greatest possible genetic diversity of the species. Highlighting that in the present study, the populations of the State of Amazonas presented the highest indexes of genetic diversity, mainly within populations. However, the creation of an *in situ* germplasm bank is more indicated. *In situ* preservation favors the maintenance of fauna and flora in the area to be preserved, allows the species to co-evolve in the habitat under the same selection pressures to which its pollinators and pests are subjected (Ramos et al. 2015). Considering the wide area of occurrences of the yellow mombins, the creation of protected areas and the expansion of existing ones where this species occurs may be the most economical alternative. According to the results obtained in this study, the species presents a genetic structure regarding geographic distance, with a high correlation between geographic distance and genetic differentiation.

Conclusions

The *S. mombin* species presents a pattern of genetic structure of isolation by distance, with limited gene flow between the North and Northeast regions but present within populations of the same region. Greater genetic diversity was found in populations from the North compared to the Northeast region. On the other hand, populations from the North region showed higher values for the inbreeding coefficient when compared to populations from the Northeast region. From a conservation point of view, the species has the potential for *in situ* conservation with populations that show reasonable levels of genetic diversity. Curators of *ex situ* germplasm banks for the species should consider inserting accessions from different regions of the country in the collection, mainly from the state of Amazonas, which presented higher levels of genetic diversity, increasing efficiency in the conservation of the genetic diversity of the species. The study was carried out using a broad sampling of the geographic distribution of the species in the country, and the wide coverage of genomic diversity contributed significantly to the understanding of the current state of conservation of natural populations of the species and will enable breeders to devise strategies for its conservation and genetic improvement.

Declarations

Statements & Declarations

Acknowledgements

The authors would like to thank Dr. Charles Roland Clement, from the Instituto Nacional de Pesquisas da Amazônia, for helping us with the plant collections.

Funding

This study was financially supported by *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP, 2019/04100-6). Scholarships were provided by FAPESP (2019/15544-2 to IASC and 2021/04698-9 to FMC), by *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), (309445/2020-5 to EAV; to AVS) and by *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES) (to CBG).

Author contributions

Planning and design of research: AVS, CBG, WFN, MIZ, and EAV; Material preparation, field collection expedition, and laboratory analysis: AVS, CBG, WFN, SLFR, IASC, EFS, DPR, and CEB; Statistical analysis: AVS, CBG, FMC, and AAP; The first draft of the manuscript was written by AVS and EAV; All authors commented on previous versions of the manuscript and contributed to the final manuscript.

Data and materials availability

The datasets generated and analyzed during the current study are available in the Mendeley repository [https://data.mendeley.com/datasets/kzmzrdw5fr/1]

Competing interests

The authors declare no competing interests.

Ethical approval

This research was registered in the National System for the Management of Genetic Heritage and Associated Traditional Knowledge (SisGen) (registration nº A3AF200). It does not contain any studies with human participants or animals performed by any authors.

References

1. Alves-Pereira A, Novello M, Dequigiovanni G et al (2019) Genomic diversity of three Brazilian native food crops based on Double-Digest Restriction Site-Associated DNA sequencing. Trop Plant Biol 12:268–281. https://doi.org/10.1007/s12042-019-09229-z
2. Barbosa MRV, Agra MF, Sampaio EVSB, Cunha JP, Andrade LA (2004) Diversidade florística na Mata do Pau-Ferro, Areia, Paraíba. In: Pôrto KC, Cabral JJP, Tabarelli M (eds) (orgs) Brejos de altitude em Pernambuco e Paraíba: história natural, ecologia e conservação. Ministério do Meio Ambiente, Brasília, pp 111–121
3. Bougeard S, Dray S (2018) Supervised multiblock analysis in R with the ade4 package. J Stat Softw 86:1–17. https://doi.org/10.18637/jss.v086.i01
4. Catchen JM, Amores A, Hohenlohe P et al (2011) Stacks: Building and genotyping loci de novo from short-read sequences. G3. Genes Genom Genet 1:171–182. https://doi.org/10.1534/g3.111.000240
5. Chessel D, Dufour AB, Thioulouse J (2004) The ade4 package - I: One-table methods. R News 4:5–10. https://journal.r-project.org/articles/RN-2004-002/
6. Clement CR (1999) 1942 and the loss of Amazonian crop genetic resources: the relation between domestication and human population decline. Econ Bot 53:188–202. https://www.jstor.org/stable/4256179
7. Coradin L, Camillo J, Vieira ICG, Brasília (2022) https://ala-bie.sibbr.gov.br/ala-bie/species/369904
8. Cornejo OE, Yee M-C, Dominguez V et al (2018) Population genomic analyses of the chocolate tree, *Theobroma cacao* L., provide insights into its domestication process. Commun Biol 1:1–12. https://doi.org/10.1038/s42003-018-0168-6

9. Cortinovis G, Frascarelli G, Di Vittori V, Papa R (2020) Current state and perspectives in population genomics of the common bean. Plants 9:330. https://doi.org/10.3390/plants9030330
10. Cristóbal-Pérez EJ, Fuchs EJ, Harvey N, Quesada M (2019) Isolation and characterization of microsatellites loci in *Spondias purpurea* (Anacardiaceae) and cross amplification in congeneric species. Mol Biol Rep 46:5581–5585. https://doi.org/10.1007/s11033-019-04968-4
11. Díaz BG, Zucchi MI, Alves-Pereira A et al (2021) Genome-wide SNP analysis to assess the genetic population structure and diversity of *Acrocomia* species. PLoS ONE 16:1–24. https://doi.org/10.1371/journal.pone.0241025
12. Dray S, Dufour A (2007) The ade4 Package: Implementing the duality diagram for ecologists. J Stat Softw 22:1–20
13. Dray S, Dufour A, Chessel D (2007) The ade4 Package – II: Two-Table and K-Table Methods. R News 7(2):47–52. https://cran.r-project.org/doc/Rnews/
14. Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:1–10. https://doi.org/10.1371/journal.pone.0019379
15. Fischer MC, Rellstab C, Leuzinger M et al (2017) Estimating genomic diversity and population differentiation - an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. BMC Genomics 18:1–15. https://doi.org/10.1186/s12864-016-3459-7
16. Fonseca N, Machado C, de F, Júnior JF S, et al (2017) Umbú, cajá e espécies afins. IICA/PROCISUR, Argentina. http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1096489
17. Francisconi AF, Alves RP, Clement CR, Dequigiovanni G, Carvalho IAS, Veasey EA (2021) Genetic structure and diversity identify incipient domestication of Piquiá [*Caryocar villosum* (Aubl.) Pers.] along the lower Tapajós River, Brazilian Amazonia. Genet Resour Crop Evol 68:1487–1501. https://doi.org/10.1007/s10722-020-01078-0
18. Garcia CB (2021) Genômica populacional do bacurizeiro (*Platonia insignis* Mart.). Dissertation, Luiz de Queiroz College of Agriculture, University of São Paulo
19. Gärke C, Ytournel F, Bed'Hom B et al (2012) Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. Anim Genet 43:419–428. https://doi.org/10.1111/j.1365-2052.2011.02284.x
20. Goudet J, Jombart T (2020) Hierfstat: Estimation and Tests of Hierarchical F-Statistics. R package version 0.5-7. https://CRAN.R-project.org/package=hierfstat
21. Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. New For 6:95–124. https://link.springer.com/content/pdf/10.1007/BF00120641.pdf
22. Hartl D, Clark A (1998) Principles of Population Genetics. Oxford University Press, Oxford
23. Inglis PW, Marilia de Castro RP, Resende LV, Grattapaglia D (2018) Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. PLoS ONE 13:1–14. https://doi.org/10.1371/journal.pone.0206085
24. INPE - Instituto Nacional de Pesquisas Espaciais, Projeto PRODES (2019) : Monitoramento da Floresta Amazônica Brasileira por Satélite, http://www.obt.inpe.br/prodes/. Accessed 20 December 2020
25. Jombart T (2008) Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. https://doi.org/10.1093/bioinformatics/btn129
26. Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genet 11:94. https://doi.org/10.1186/1471-2156-11-94
27. Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet 6:1–10. https://doi.org/10.3389/fgene.2015.00208

28. Kamvar ZN, Tabima JF, Gr NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and / or sexual reproduction. PeerJ 2:e281. https://doi.org/10.7717/peerj.281
29. Leão AP, Filho JAF, Pereira VM, Alves AA, Souza Júnior MT (2022) Genomic characterization of SNPs for genetic differentiation and selection in populations from the American oil palm [*Elaeis oleifera* (Kunth) Cortés] germplasm bank from Brazil. Divers 14:270. https://doi.org/10.3390/d14040270
30. Lemopoulos A, Prokkola JM, Uusi-Heikkilä S et al (2019) Comparing RADseq and microsatellites for estimating genetic diversity and relatedness — Implications for brown trout conservation. Ecol Evol 9:2106–2120. https://doi.org/10.1002/ece3.4905
31. Lima ATB, Desouza VAB, Gomes RLF, Lima PSC (2011) Molecular characterization of cajá, *Spondias mombin* (Anacardiaceae), by RAPD markers. Genet Mol Res 10:2893–2904. https://doi.org/10.4238/2011.November.25.1
32. Lins Neto EMF, de Oliveira IF, Britto FB, de Albuquerque UP (2013) Traditional knowledge, genetic and morphological diversity in populations of *Spondias tuberosa* Arruda (Anacardiaceae). Genet Resour Crop Evol 60:1389–1406. https://doi.org/10.1007/s10722-012-9928-1
33. Lorenzi H (1992) Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil. Plantarum, Nova Odessa
34. Magalhães MA, Lopes MTG, Coelho LM et al (2013) Genetic diversity within and among populations of Cajazeira. Rev Ciências Agrar - Amaz J Agric Environ Sci 56:61–67. https://doi.org/10.4322/rca.2013.082
35. Mitchell JD, Daly DC (2015) A revision of *Spondias* L. (Anacardiaceae) in the neotropics. PhytoKeys 55:1–92. https://doi.org/10.3897/phytokeys.55.8489
36. Moreira ERF (1989) Mesorregiões e microrregiões da Paraíba: delimitação e caracterização. GAPLAN, João Pessoa
37. Nascimento WF, Dequigiovanni G, Ramos SLF et al (2021) Nuclear and chloroplast microsatellites reveal high genetic diversity and structure in *Platonia insignis* Mart., an endangered species native to the Brazilian Amazon. Acta Bot Bras 35:432–444. https://doi.org/10.1590/0102-33062020abb0324
38. Nei M (1972) Genetic distances between populations. Am Nat 106:283–292. https://www.jstor.org/stable/2459777
39. Nobre LLM, Santos JDO, Leite R, Almeida C (2018) Phylogenomic and single nucleotide polymorphism analyses revealed the hybrid origin of *Spondias bahiensis* (family Anacardiaceae): de novo genome sequencing and comparative genomics. Genet Mol Biol 41:878–883. http://dx.doi.org/10.1590/1678-4685-GMB-2017-0256
40. O'Neill GA, Dawson I, Sotelo-Montes C et al (2001) Strategies for genetic conservation of trees in the Peruvian Amazon. Biodivers Conserv 10:837–850. https://doi.org/10.1023/A:1016644706237
41. Ogunro OB, Oyeyinka BO, Gyebi GA, Batiha GE (2023) Nutritional benefits, ethnomedicinal uses, phytochemistry, pharmacological properties and toxicity of *Spondias mombin* Linn: a comprehensive review. J Pharm Pharmacol 75:162–226. https://doi.org/10.1093/jpp/rgac086
42. Padgham M, Sumner MD (2020) Geodist: Fast, Dependency-Free Geodesic Distance Calculations. R package version 0.0.6. https://CRAN.R-project.org/package=geodist
43. Paradis E, Schliep K (2019) Ape 5.0: An environment for modern phylogenetics and evolutionary analyses. R Bioinf 35:526–528. https://doi.org/10.1093/bioinformatics/bty633
44. Paschoa RP, Christ JA, Valente CS et al (2018) Genetic diversity of populations of the dioecious *Myrsine coriacea* (Primulaceae) in the Atlantic Forest. Acta Bot Bras 32:376–385. https://doi.org/10.1590/0102-33062017abb0355
45. Portal Alelo Recursos Genéticos (2022) https://alelo.cenargen.embrapa.br/. Accessed 20 December 2022

46. Ramos MC (2009) Ecologia da polinização de taperebá (*Spondias mombin* L., Anacardiaceae) em área de floresta secundária no município de Santo Antônio do Tauá, Pará, Brasil. Dissertation, Universidade Federal do Par&#225
47. Ramos SRR, Silva Júnior JF, Bustamante PG, Mota DMM, Queiróz MA (2015) Preservação in situ, conservação ex situ e uso dos recursos genéticos por agricultores e comunidades tradicionais no Brasil. In: Veiga RFA, Queiróz MA (eds) Recursos Fitogenéticos. Editora UFV, Viçosa, pp 269–274
48. R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/
49. Rivero S, Almeida O (2009) Pecuária e desmatamento: uma análise das principais causas diretas do desmatamento da Amazônia. Nov Econ 19:41–66. https://doi.org/10.1590/S0103-63512009000100003
50. Sacramento CK, Souza FX (2009) Cajá. In: Santos-Serejo JA, Dantas JLL, Sampaio CV, Coelho YS (eds) Fruticultura tropical: espécie regionais e exóticas. Embrapa Informação Tecnológica, Brasília, pp 83–106
51. Santana FF (2010) Caracterização de genótipos de cajazeiras. Dissertation, Universidade Estadual Paulista
52. Santos JO, Mayo SJ, Bittencourt CB, de Andrade IM (2019) Genetic diversity in wild populations of the restinga ecotype of the cashew (*Anacardium occidentale*) in coastal Piauí, Brazil. Plant Syst Evol 305:913–924. https://doi.org/10.1007/s00606-019-01611-4
53. Santos CAF, Oliveira VR (2008) Inter-relações genéticas entre espécies do gênero *Spondias* com base em marcadores AFLP. Rev Bras Frutic 30:731–735
54. Silva BM, Rossi AAB, Dardengo JDFE et al (2017a) Genetic divergences between *Spondias mombin* (Anacardiaceae) genotypes found through morphological traits. Rev Biol Trop 65:1337. https://doi.org/10.15517/rbt.v65i4.25765
55. Silva BM, Rossi AAB, Tiago AV et al (2017b) Genetic diversity of Cajazeira (*Spondias mombin* L.) in three geographic regions. Genet Mol Res 16. https://doi.org/10.4238/gmr16018946
56. Silva CA, Costa PR, Detoni JL et al (2014) Divergência genética entre acessos de cajazinho (*Spondias mombin* L.) no norte do Espírito Santo. Rev Ceres 61:362–369. https://doi.org/10.1590/S0034-737X2014000300010
57. da Silva EF, Martins LSS, de Oliveira VR (2009) Diversity and genetic structure in cajá tree (*Spondias mombin* l.) populations in northeastern Brazil. Rev Bras Frutic 31:171–181. https://doi.org/10.1590/s0100-29452009000100024
58. Silva JF, Fernandes Bezerra JE, Eliezer Lederman I et al (2004) Collecting, ex situ conservation and characterization of “cajá-umbu” (*Spondias mombin* x *Spondias tuberosa*) germplasm in Pernambuco State, Brazil. Genet Resour Crop Evol 51:343–349. https://doi.org/10.1023/B:GRES.0000023462.99118.00
59. Souza LM, Dos Santos LHB, Rosa JRBF et al (2018) Linkage disequilibrium and population structure in wild and cultivated populations of rubber tree (*Hevea brasiliensis*). Front Plant Sci 9:1–10. https://doi.org/10.3389/fpls.2018.00815
60. Surapaneni M, Vemireddy LR, Begum H et al (2013) Population structure and genetic analysis of different utility types of mango (*Mangifera indica* L.) germplasm of Andhra Pradesh state of India using microsatellite markers. Plant Syst Evol 299:1215–1229. https://doi.org/10.1007/s00606-013-0790-1
61. Thioulouse J, Dray S, Dufour A, Siberchicot A, Jombart T, Pavoine S (2018) Multivariate Analysis of Ecological Data with ade4. Springer, New York
62. Tiburski JH, Rosenthal A, Deliza R et al (2011) Nutritional properties of yellow mombin (*Spondias mombin* L.) pulp. Food Res Int 44:2326–2331. https://doi.org/10.1016/j.foodres.2011.03.037
63. Vieira ICG, Toledo PM, Silva JMC, Higuchi H (2008) Deforestation and threats to the biodiversity of Amazonia. Braz J Biol 68:949–956. https://doi.org/10.1590/S1519-69842008000500004

64. Wei T, Simko V (2017) R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). https://github.com/taiyun/corrplot
65. Zortéa KÉM, Rossi AAB, Bispo RB et al (2019) Meiotic behavior and pollen viability of *Spondias mombin* L.: Native fruit species of the Amazon. Floresta e Ambient 26:1–8. https://doi.org/10.1590/2179-8087.037518
66. Zortéa KÉM, Rossi AAB, Tiago AV et al (2021) *Spondias mombin* (Anarcadiaceae): Molecular characterization and conservation. Rev Biol Trop 69:1023–1036. https://doi.org/10.15517/rbt.v69i3.45810

Figures

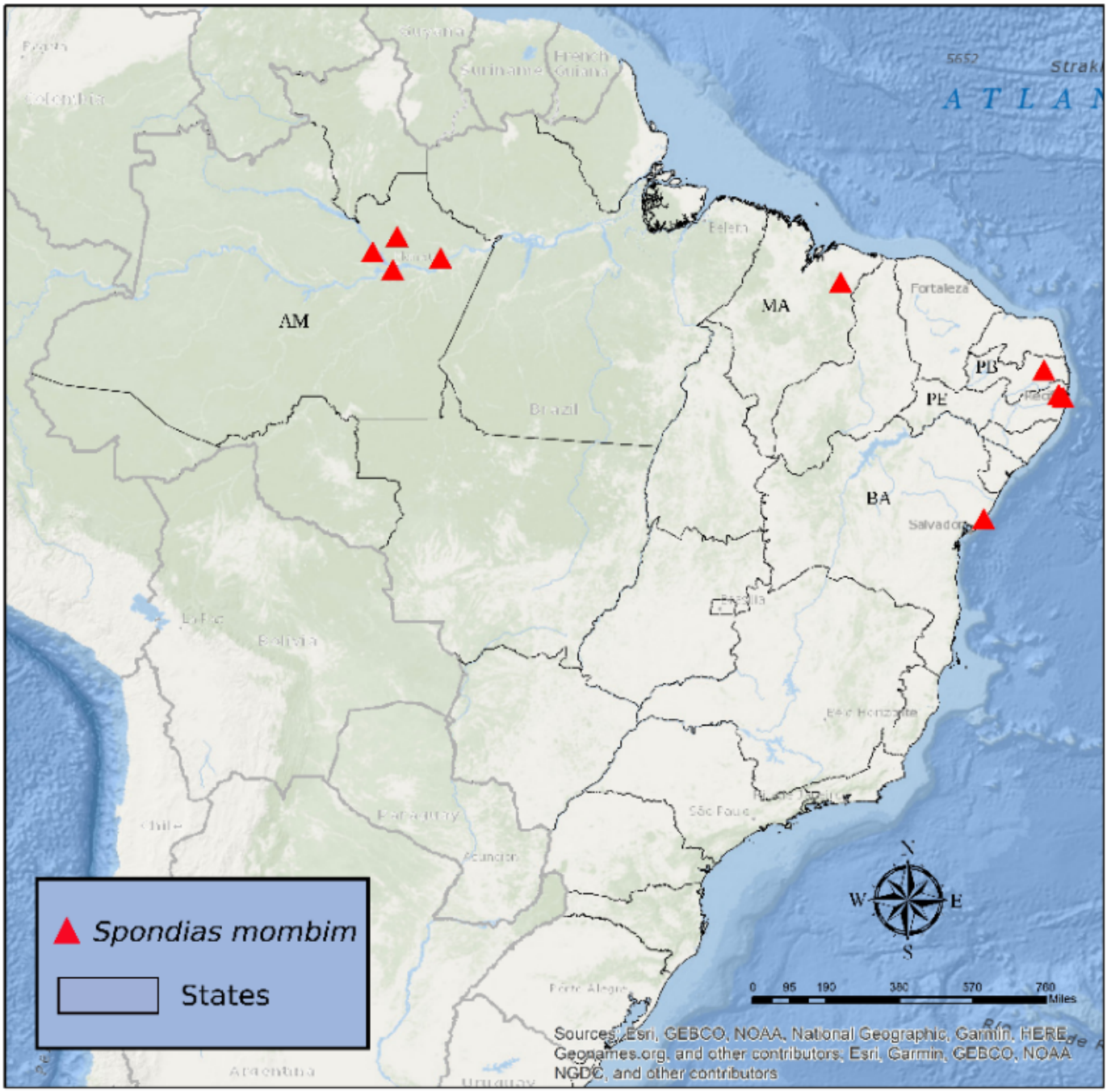

Figure 1

Map with the sampling locations of *Spondias mombin* populations in the States of Amazonas - AM, Maranhão - MA, Pernambuco - PE, Paraíba - PB, and Bahia - BA.

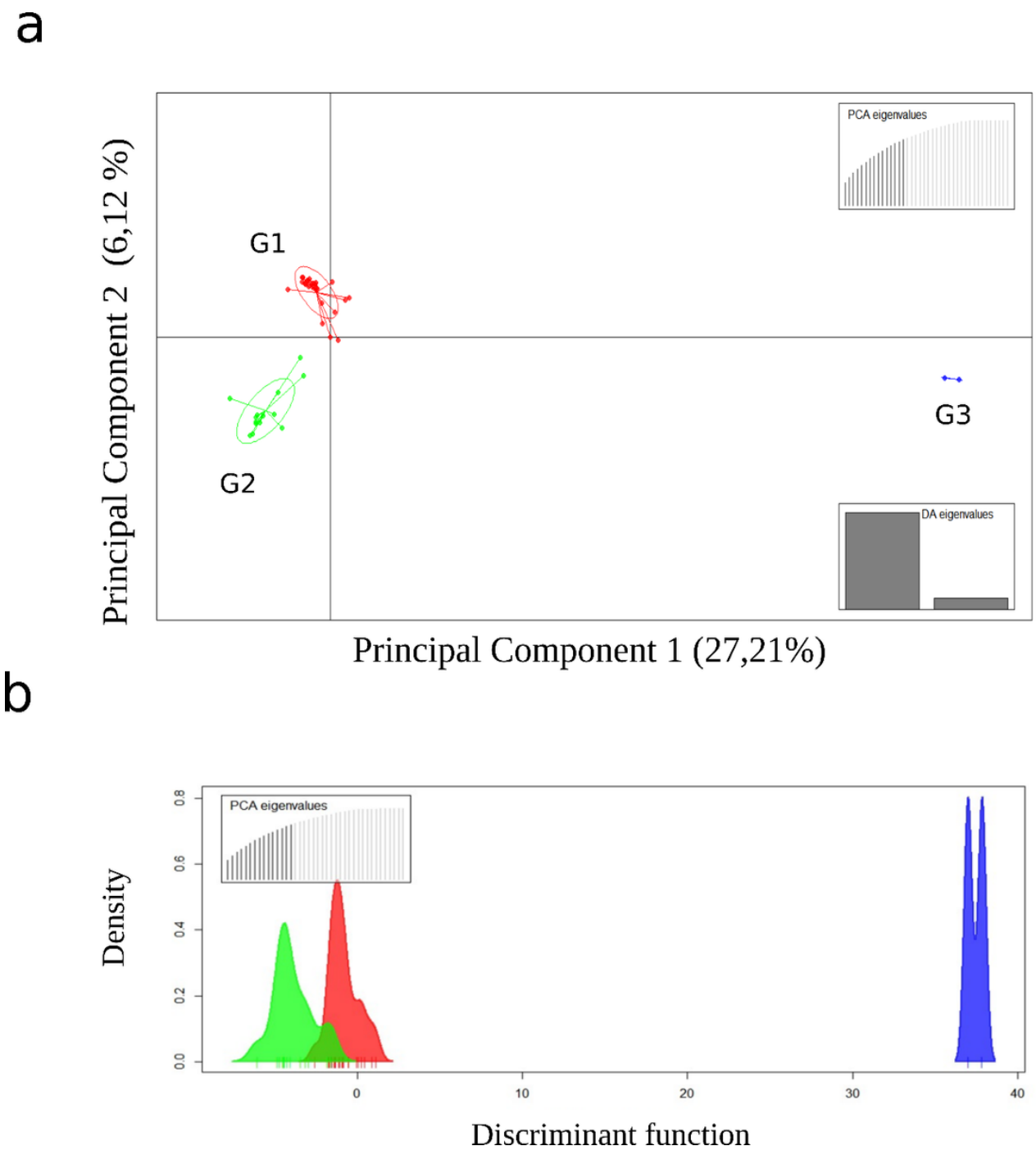

Figure 2

Discriminant Analysis of Principal Component (DAPC) performed with 41 yellow mombin (*Spondias mombin*) individuals and 2,003 SNP markers. a) A scatter plot based on two discriminant functions, with the three groups identified by the K-means method. b) DAPC from the first discriminant function to better understand the density and overlap between groups. The G1 group is represented in red, the G2 group is represented in green, and the G3 group is represented in blue.

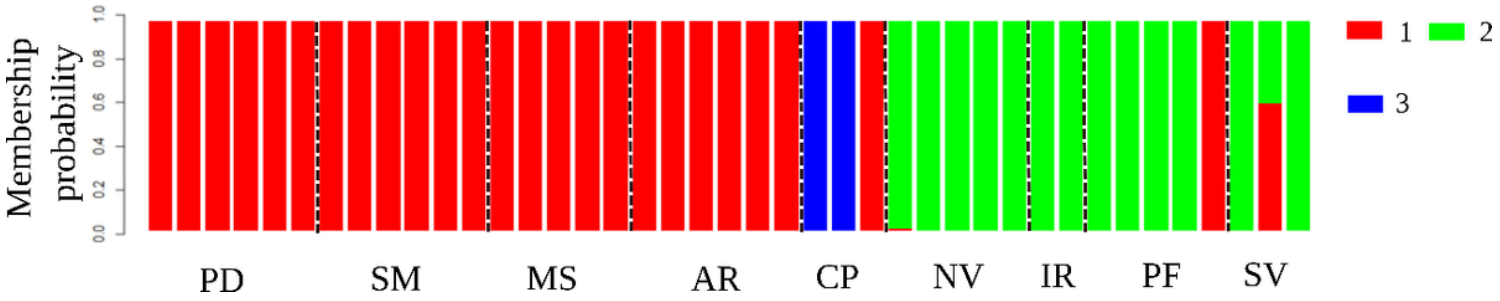

Figure 3

Clustering probability analysis according to the results generated in the DAPC. Each delimited bar represents an individual; the yellow mombin (*Spondias mombin*) populations are represented by the interval formed between the dotted lines, such as PD (Paudalho - Pernambuco), SM (São Lourenço da Mata - Pernambuco), MS (Mata de São João - Bahia), AR (Areia - Paraíba), CP (Chapadinha - Maranhão), NV (Novo Airão -Amazonas), IR (Iranduba - Amazonas), PF (Presidente Figueiredo - Amazonas), and SV (Silves - Amazonas).

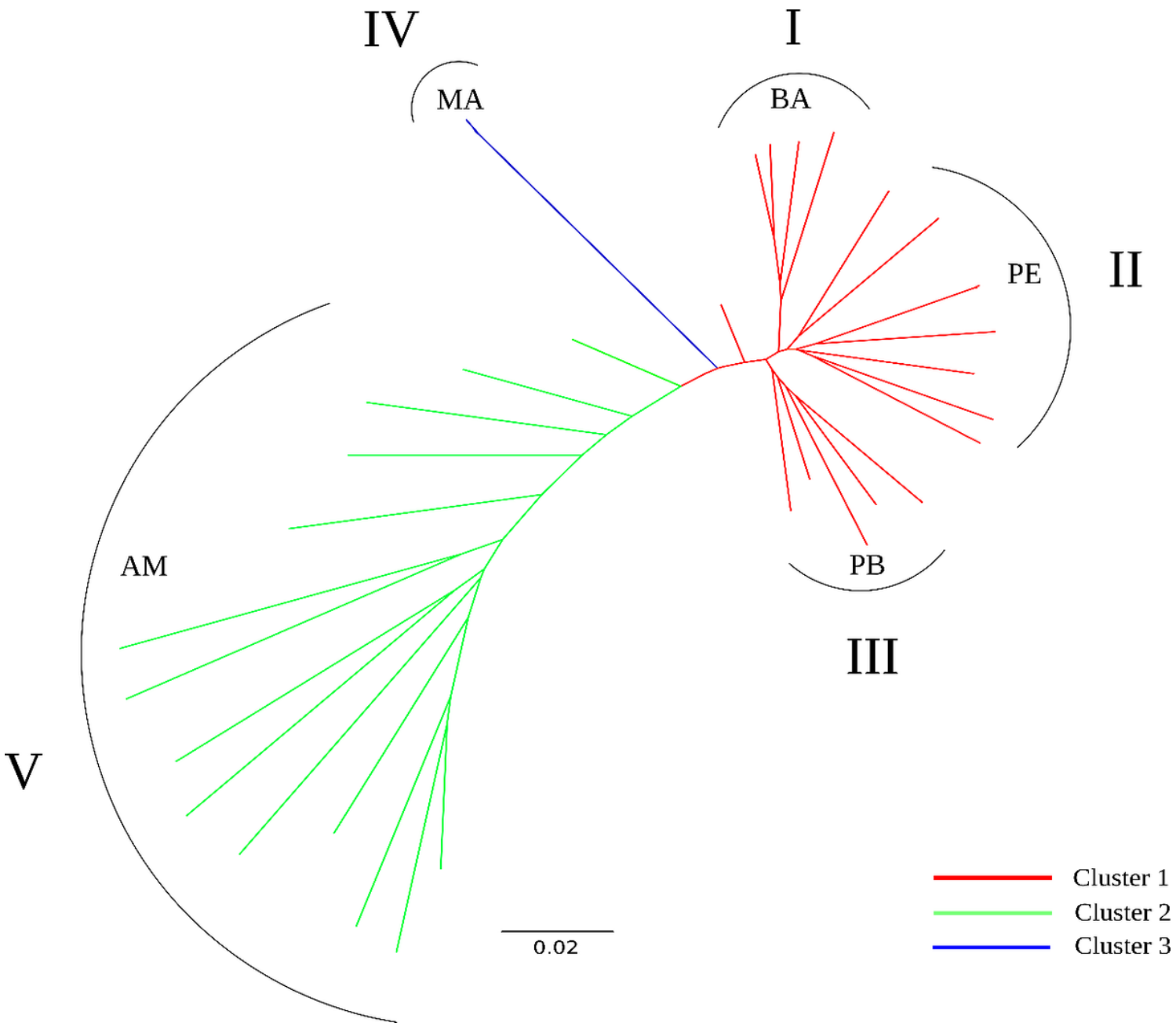

Figure 4

Cluster analysis from the phylogenetic tree built by the *Neighbor-Joining* method based on the groups identified by the DAPC analysis for the yellow mombin (*Spondias mombin*) populations collected in the States of Amazonas (AM), Maranhão (MA), Bahia (BA), Pernambuco (PE) and Paraíba (PB). The Roman numerals (I, II, III, IV, and V) represent the groups obtained in this analysis, while the colors represent the groups identified by the DAPC analysis.

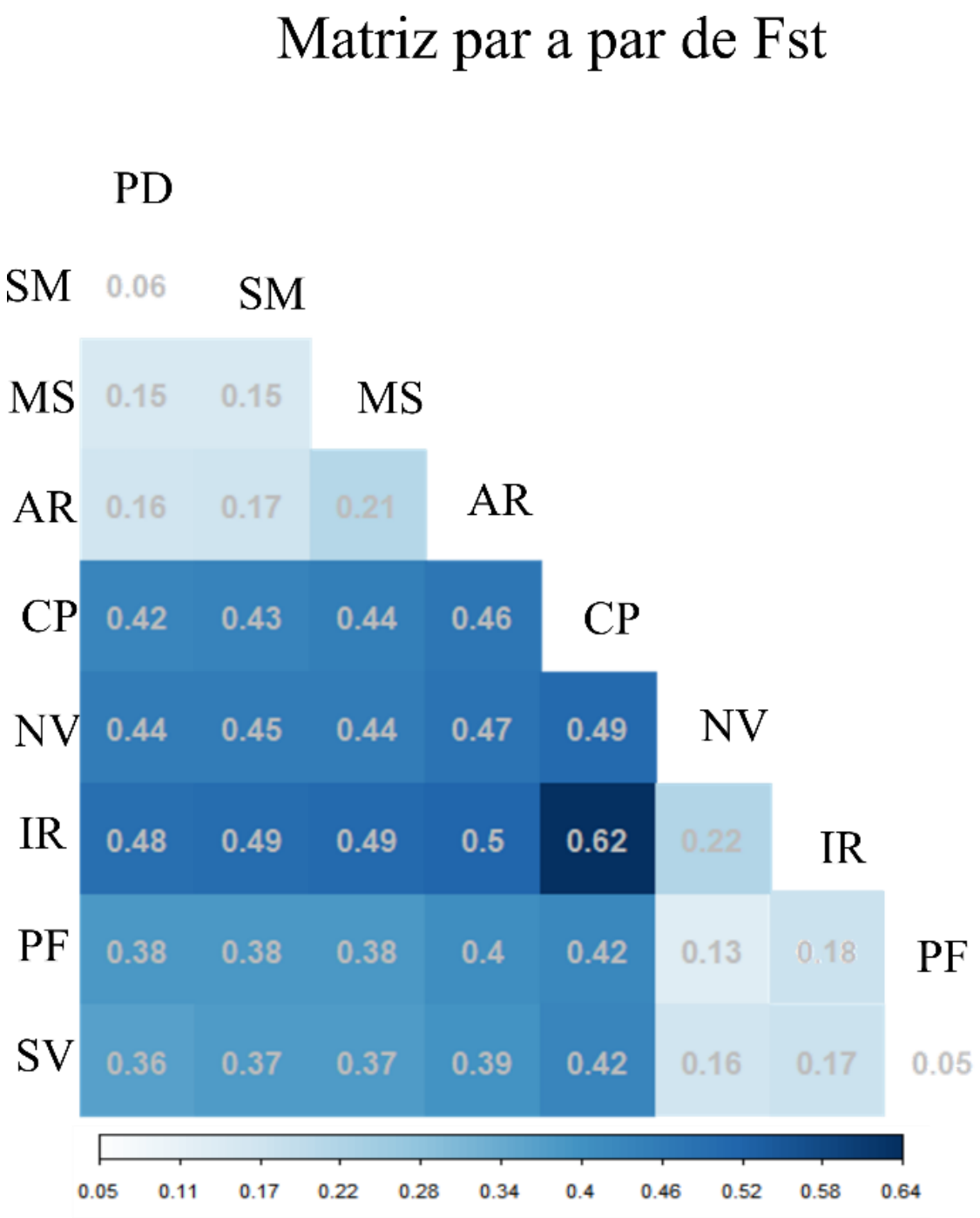

Figure 5

Results obtained from the pairwise analysis of *F*st between yellow mombin (*Spondias mombin*) populations. All values are significant at *p*-value < 0.05. The populations are represented by the abbreviations PD (Paudalho-Pernambuco), SM (São Lourenço da Mata-Pernambuco), MS (Mata de São João-Bahia), AR (Areia-Paraíba), CP (Chapadinha-Maranhão), NV (Novo Airão-Amazonas), IR (Iranduba-Amazonas), PF (Presidente Figueiredo-Amazonas), and SV (Silves-Amazonas).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SilvaSupplementaryMaterial.docx