Diversity, genetic structure and core collection of mangaba (Hancornia speciosa) genebank

Ana Veruska Cruz Silva (ana.veruska@embrapa.br)
Embrapa Tabuleiros Costeiros https://orcid.org/0000-0002-4177-9335
Itamara Bomfim Gois
Universidade Federal de Sergipe
Adrielle Naiana Ribeiro Soares
Universidade Federal de Sergipe
Ana Silva Ledo
Embrapa Tabuleiros Costeiros

Research Article

Keywords: ex situ conservation, microsatellites, Brazilian fruits, Apocynaceae

Posted Date: November 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2163981/v1

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on July 5th, 2023. See the published version at https://doi.org/10.1007/s10722-023-01643-3.

Abstract

This work was developed with the objective of characterizing the variability and genetic structure of matrices and progenies of the Mangaba genebank of Embrapa Tabuleiros Costeiros, to define a core collection. From the matrix of seven accessions, 289 individuals were generated, evaluated by nine SSR markers. The characterization of genetic variability was performed from the estimates: Average number of alleles observed (Na = 5.63) and effective (Ne = 2.71); Shannon Information Index (I = 1.04); Heterozygosity observed (Ho = 0.45) and expected (He = 0.51); Fixation index (f = 0.13); Percent polymorphism (%P = 90.47); and, Number of private alleles (Nap = 20). The Analysis of Molecular Variance (AMOVA) identified that the largest proportion of genetic variation is present within accessions (92%). The estimates of genetic differentiation (G_{ST} and R_{ST}) were considered low (< 0.05) to moderate (0.05 to 0.15) magnitude. Multivariate analyses (PCoA, Rogers and Bayesian genetic diversity) did not discriminate the accessions according to origin. The formation of the core collection allowed retaining 94.90% of the identified alleles. The accessions present genetic variability to be explored in mangaba conservation and genetic improvement programs.

Introduction

The genus *Hancornia* is monospecific and consists of the species *Hancornia speciosa* Gomes (Jimenez et al. 2015). Popularly known as mangaba, the species belongs to the Apocynaceae family with a wide distribution in the Neotropics, being reported in Brazil, Paraguay, Bolivia and Peru (Collevatti et al. 2018; Fajardo et al. 2018). In Brazil, it can be seen from Amapá, in the North, to Paraná, in the South, in different phytophysiognomies associated with the Cerrado and the Atlantic Forest (Silva et al. 2021).

The mangaba is hermaphrodite and presents a self-incompatibility system, a mechanism that favors allogamy, being pollinated mainly by moths (Darrault and Schlindwein 2005). Its fruits have small oval seeds, which are mainly dispersed by medium and large mammals (Costa et al. 2017).

The species is used in folk medicine and different parts of the plant have been used to treat various diseases, such as gastric ulcers and hypertension (Dórea et al. 2021). Additionally, it presents edible pulp with relevant nutritional properties (Jimenez et al. 2015), factors that have contributed to its use by the local population. The fruits are consumed *in natura* and used in the preparation of jellies, juices, cookies, ice cream, playing an important role for the local economy (Collevatti et al. 2016; Chaves et al. 2020). It is one of the most important raw materials for the juice and ice cream agroindustry among the native fruits in the Northeast and Midwest of Brazil, but the supply is unsatisfactory. According to Soares et al. (2019), despite its importance for local populations and its agro-industrial potential, the exploitation of the species has been carried out to a greater extent in natural populations, that is, in an extractivist manner (Soares et al. 2019).

The form of exploitation associated with the process of fragmentation of the areas of natural occurrence makes mangaba one of the genetically endangered fruit species in Brazil. Considering its ecological and

economic importance, the collection, conservation, and characterization of available mangaba genetic resources are fundamental (Jimenez et al. 2015).

The adoption of *in situ* and *ex situ* conservation strategies is necessary for the maintenance of the remaining genetic variability of the species. The seeds are classified as recalcitrant, thus *ex situ* conservation should be accomplished through *in vitro* and/or *in vivo* field collections (Almeida et al. 2019).

Due to the reduction of natural areas and the social, economic and cultural importance of mangaba, Embrapa Tabuleiros Costeiros, a research center of Brazilian Agricultural Research Coporation (Embrapa) has been developing conservation strategies for the species for more than 15 years. The Mangaba genebank was established in 2006 and is accredited by the Ministry of the Environment (MMA) / Genetic Heritage Management Council (CGEN) as the depositary of the species in Brazil. It currently consists of 21 accessions represented by 299 plants, collected over a wide geographic range (Silva et al. 2021).

There is no improvement program, nor mangaba cultivars. In the Mangaba genebank some accessions were preliminarily identified and selected as early (Silva et al. 2019), with low seasonality (Machado et al. 2020), and as excellent sources of vitamin C and antioxidants (Silva et al. 2012; Silva et al. 2017; Santos et al. 2020; Santana et al. 2021; Silva et al. 2021).

Knowledge of the genetic variability present in a BAG is essential for the efficient management and use of this resource (Bernard et al. 2018). This characterization can be accomplished through morphological, agronomic, chemical, and molecular variables. DNA markers have been used as key tools in the molecular characterization of different accessions, as they are not influenced by environment and plant development stage, and are more accurate, efficient and reliable for the discrimination of related genotypes. The genetic diversity of the accessions of the Mangaba genabank was carried out following its expansion (Costa et al. 2011; Silva et al. 2011; Silva et al. 2019). With the need to form an area of progenies from this genebank, aiming at a breeding program, the proposal of this work was built.

Microsatellites (SSR) are important tools for characterizing genetic diversity and structure due to technical simplicity, speed, high resolving power, reproducibility, high level of polymorphism, and codominant inheritance (Kölliker et al. 2010; Silva et al. 2019). These markers have been successfully used for different purposes in mangaba, such as in studies to estimate gene flow among varieties (Collevatti et al. 2016), characterize spatial genetic structure (Costa et al., 2017), and quantify the genetic variability present in natural populations (Amorim et al. 2015; Rodrigues et al. 2015; Chaves et al. 2020), in germplasm (Silva et al. 2019) and progenies (Soares et al. 2018).

Genebanks are considered valuable reservoirs of allelic variability for different traits of economic and ecological importance that can be exploited in conservation and genetic improvement programs (Campoy et al. 2016). However, the preservation and maintenance of a large number of accesses is labor intensive and expensive. The establishment of a core collection is a strategy that allows maximizing genetic

diversity in a smaller number of accessions, reducing the number of redundant genotypes, which is an alternative for reducing maintenance costs (Pereira et al. 2020).

In view of the above, this work was carried out with the objective of characterizing the variability and genetic structure of mangaba progenies from matrices of the genebank, using SSR markers, and to define a core collection for the germplasm, with the purpose of generating information for the conservation, management and efficient use of this genetic resource.

Material And Methods

The Mangaba genebank (Fig. 1) is located in Itaporanga d'Ajuda, Sergipe state, Brazil (11º 06' 40" S and 37º 11'15" W).

Seven accessions with matrices in fruiting were selected for the production of the progenies, which were evaluated together with the 289 progenies generated, totaling 296 accessions (seven matrices and 289 progenies) (Table 1).

Table 1
Accessions, procedence and number of progenies present in the Mangaba genebank of Embrapa Tabuleiros Costeiros. Itaporanga d'Ajuda, Sergipe, Brazil.

Acession/Code	Procedence	Nº of progeny
Preguiça - PR	Indiaroba, SE	17
Lagoa Grande - LG	Mata de São João, Bahia	59
Pontal - PT	Indiaroba, Sergipe	56
Barra de Itariri - BI	Conde, Bahia	32
Terra Caída - TC	Indiaroba, Sergipe	52
Água Boa - AB	Salvaterra, Pará	55
Costa Azul - CA	Jandaíra, Bahia	18
		289
Total	7 matrices + 289 progenies	296

To obtain the progenies, fruits were collected from each access used as matrix. The fruits were pulped, and the seeds treated with 2.5% sodium hypochlorite solution for 3 minutes. After, they were washed and dried in the shade. For seedling production, the seeds were sown in polyethylene bags containing washed sand and identified according to the matrix accession (Fig. 2). At six months after sowing, leaves were taken from each progeny for genomic DNA extraction. The collected leaves were stored at -80 ºC until DNA extraction.

DNA extraction was performed using the method described by Doyle and Doyle (1990) modified by Alzate-Marin et al. (2009). The extracted DNA was diluted in 50 µL of TE and its quantification was performed by spectrophotometry using Nanodrop 2000c (Thermo Scientific, USA). Nine pairs of microsatellite primers (Table 2) developed by Rodrigues et al. (2015) were used to characterize genetic diversity and structure.

Table 2
Annealing temperature (AT), repeat motifs, allele amplitude (pb) and fluorescence of microsatellites used to characterize the diversity and genetic structure of mangaba genebank of Embrapa Tabuleiros Costeiros. bp – base pairs.

Primer	AT (ºC)	Repeat motifs	Allele amplitude (bp)	Fluorescence
HS01	56	(GCA)6(TC)20(GCA)8	250 a 310	HEX
HS03	56	(CT)5(CT)5	120 a 180	6-FAM
HS05	56	(GA)15(TGC)6	200 a 300	HEX
HS06	54	(GA)14	100 a 150	HEX
HS08	52	(CA)6(CT)17	200 a 250	6-FAM
HS10	56	(CT)14(CT)8	100 a 200	HEX
HS16	54	(GA)12	100 a 150	6-FAM
HS27	54	(GA)14	100 a 150	6-FAM
HS33	56	(AG)24	80 a 120	6-FAM

In each amplification reaction a total of 13 µL of solution containing: 9 ng DNA, 0.25 mg mL^{-1} BSA (bovine serum albumin), 0.2 µM of each primer (25 pmol), 1x PCR buffer (10X), 0.25 mM dNTPs mix (2.5 mM), 1 U of Taq DNA polymerase and sterile ultrapure water.

The amplifications were performed in a Veriti 96 Fast thermal cycler (Applied Biosystems/USA). The PCR reaction consisted of an initial step at 94°C for 1 minute, followed by 35 cycles (95°C for 1 minute for denaturation of the DNA strands, primer annealing temperature (Table 2) for 1 minute, and 72°C for DNA strand extension for 1 minute), and a final extension step performed at 72°C for 20 minutes. To test for contamination by exogenous DNA, a negative control containing all PCR components except DNA (replaced by water) was included in each experiment.

The amplification results were visualized by electrophoresis on a 2% (w/v) agarose gel stained with ethidium bromide (0.5 µg mL^{-1}) (Sambrook et al., 1989) and photodocumented under ultraviolet light (Kodac Gel Logic 200 Imaging System). For size comparison of the amplified fragments the DNA marker 1 Kb Ladder (Invitrogen) was used.

For the analyses of the fragments obtained by each SSR primer pair used, 1 µL of each reaction was mixed with 10 µL of HiDi formamide (Applied Biosystems, Foster City, CA) and 1 µL of the internal marker

carboxy-X-rhodamine (ROX), developed by Brondani and Grattapaglia (2001). The resulting solution was denatured for 5 minutes at 95°C. Then, fragment separation was performed in an ABI 3730 automated DNA analyzer (Applied Biosystems, Foster City, CA). Fluorescence peak detection and genotyping were performed with the program Genemapper version 4.1 (Applied Biosystems). Allele size was adjusted for allele classes defined by the AlleloBin program (Prasanth and Chandra, 2006).

The statistics allele frequency, number of alleles per *locus*, expected and observed heterozygosity, and PIC (Polymorphic Information Content) were estimated for each SSR primer pair using Power Marker 3.25 software (Liu and Muse 2005).

The genetic variability of the accessions was characterized by estimates of the number of alleles observed (Na) and effective (Ne); Shannon Information Index (I); Heterozygosity observed (Ho) and expected (He); calculated according to the proportions expected by Hardy-Weinberg Equilibrium (Nei, 1978); Fixation index (f); Percentage polymorphism (%P); and; Number of private alleles (Nap).

Analysis of Molecular Variance (AMOVA) was performed to estimate the genetic variance existing between and within accessions, and the level of significance was determined at 9,999 permutations. Genetic differentiation among accessions (G_{ST}) (Nei 1973), which corresponds to the proportion of genetic variation within accessions compared to the total genetic variation, was estimated and its significance tested using 10,000 bootstraps. Additionally, Nei genetic distance was estimated among the accessions. Analyses were performed using Genalex 6.5 software (Peakall and Smouse 2012).

Genetic distance among the accessions was assessed using Rogers (1972) genetic distance and visualized by constructing a dendrogram using the UPGMA (Unweighted Pair Group Method with Arithmetic Means) algorithm. The analysis was performed with the help of the poppr package (Kamvar et al. 2014) for R (R Core Team 2016). Ten thousand bootstraps were performed to infer on the reliability of the clusters. The software FigTree 1.4.1 was used to format the obtained dendrogram. Principal coordinate analysis (PCoA), at the individual level, was performed using Genalex 6.5 software (Peakall and Smouse 2012).

Bayesian analysis was performed to estimate the genetic structure of the accessions using Structure v.2.3.4 software (Pritchard et al. 2000). Genetic clustering values (k) ranging from 1 to 7 (number of accessions) were tested, and for each k, 10 independent repetitions were performed. Each repetition consisted of a burn in period of 50,000 iterations, followed by 100,000 MCMC (Markov Chain Monte Carlo) iterations, assuming the admixture ancestry model and uncorrelated allele frequency. The number of gene groups (k) was identified by the ΔK method described by Evanno et al. (2005), implemented in Structure Harvester software (Earl and Vonholdt 2012). Accessions with membership values lower than 0.8 were considered to be of mixed ancestry.

The core collection is a subsample of the germplasm and is established to represent the genetic diversity present in the collection in a smaller number of accessions. The maximum length sub tree function implemented in DARwin 6.0.14 (Perrier and Jacquemoud-Collet 2006) was used to identify the

accessions to compose the mangaba core collection. This function eliminates accessions that are considered redundant and allows the selection of accessions that retain the most genetic variation. Ten thousand bootstraps were performed to infer on the reliability of the result.

Results And Discussion

A total of 98 alleles were detected using nine SSR markers (Table 3). The number of alleles per locus ranged from six (HS03, HS10 and HS27) to 20 (HS01), with an average of 10.89, and fragment size ranged from 103 to 343 bp. The expected heterozygosity (He) ranged from 0.04 (HS03) to 0.87 (HS01), and the observed heterozygosity (Ho) ranged from 0.01 (HS03) to 0.72 (HS06). The HS03 marker showed the lowest heterozygosity value (0.04), indicating that the analyzed accessions did not show diversity for this locus. The PIC values ranged from 0.04 (HS03) to 0.86 (HS01).

Table 3
Maximum frequence (*f*max), number of alleles observed (Na), expected heterozygosity (He), observed heterozygosity (Ho) and polymorphic information content (PIC) for nine SSR markers.

SSR markers	*f*max	Na	He	Ho	PIC
HS01	0,23	20	0.87	0.67	0.86
HS03	0,98	6	0.04	0.01	0.04
HS05	0,34	14	0.83	0.40	0.82
HS06	0,30	13	0.78	0.72	0.75
HS08	0,40	7	0.68	0.68	0.62
HS10	0,68	6	0.49	0.31	0.44
HS16	0,48	17	0.66	0.69	0.61
HS27	0,58	6	0.56	0.15	0.48
HS33	0,40	9	0.76	0.43	0.72
Averange	0.49	10.89	0.63	0.45	0.59

The PIC value represents the probability of detecting polymorphism between two random samples (Ismail et al. 2019). With the exception of primer HS03, the primers showed high discrimination power, with PIC values higher than 0.44 (HS10). Considering the PIC formula, the observed values are dependent on the number of detected alleles and their relative frequency (Guzmán et al. 2020). Therefore, the observation of one or two alleles with high frequency will contribute to low PIC values, as could be observed for primer HS03 that presented one allele with frequency equal to 0.98 (Table 3).

The use of nine SSR primers allowed the characterization of the variability and genetic structure among the progenies and seven matrix accessions of the Mangaba genebank. The number of alleles per locus detected (six to 20) is indicative of the allelic richness of the population, and was considered sufficient to meet the objective of the present study since for SSR markers the detection of two to seven alleles per locus is considered satisfactory (Aljumaili et al. 2018). The high heterozygosity values detected for the SSR markers used may be related to the reproductive system (interbreeding) of the species (Table 3).

The average number of alleles per locus is an indication of genetic diversity, and ranged from 3.67 (CA) to 7.22 (PT) (Table 4). The average was lower than that observed by Collevatti el al. (2018) in a study with 28 natural mangaba subpopulations (9.6) using SSR markers. The effective number of alleles was lower than the observed number of alleles, suggesting that many alleles are rare ($p < 0.05$) or have low frequency ($0.05 > p < 0.25$) (Viegas et al. 2011). The accessions that presented the lowest number of alleles were PR and CA, which may be related to the smaller number of individuals analyzed (20). The effective number of alleles was lower than the observed number of alleles, ranging from 1.98 (CA) to 3.46 (BI). And, unique alleles were observed for five accessions (Table 4).

Table 4
Estimates of genetic variability parameters for mangaba genebank accessions of Embrapa Tabuleiros Costeiros.

Accession	Na	Ne	I	Ho	He	*f*	%P	Nap
PR	3.89	2.33	0.85	0.53	0.44	-0.19	77.78	0
LG	5.67	3.02	1.04	0.41	0.51	0.25	100	1
PT	7.22	3.20	1.22	0.45	0.59	0.28	100	6
BI	6.78	3.46	1.31	0.48	0.62	0.27	88.89	3
TC	5.78	2.22	0.91	0.43	0.45	0.02	88.89	4
AB	6.44	2.76	1.19	0.48	0.58	0.14	100	6
CA	3.67	1.98	0.76	0.37	0.39	0.04	77.78	0
Average	5.63	2.71	1.04	0.45	0.51	0.13	90.47	20

Na: Average number of alleles per locus; Ne: Effective number of alleles per locus; I: Shannon index; Ho: Observed heterozygosity; He: Expected heterozygosity; f: fixation index; %P: Polymorphism percentage; Nap: Number of private alleles.

Considering the Shannon Index, there was lower genetic diversity in access CA (0.76), and higher in access BI (1.31) (Table 4). The high values observed for the Shannon Index indicate the existence of high genetic variability in the materials. The estimated values were higher than those observed by Santos et al. (2017), studying 36 mangaba accessions using ISSR markers (0.28 to 0.42).

Ho ranged from 0.37 (CA) to 0.53 (PR), with a mean of 0.45, and He ranged from 0.39 (CA) to 0.62 (BI), with a average of 0.51 (Table 4). This index refers to the proportion of accessions that are heterozygous for a given locus. With the exception of the PR access, the others presented He > Ho, according to the Hardy-Weinberg Equilibrium. This result is an indication of heterozygote deficiency in the population, and suggests the occurrence of crossing between related individuals (Bernard et al. 2018), and can be proven by the estimated values of the Fixation Index (f) for these accessions (greater than zero). The estimates for Ho and He were lower than expected for allogamous species (0.63 and 0.65, respectively) and for long-lived species (0.63 and 0.68, respectively) (Nybom 2004). The estimates for Ho were lower than those reported by Costa et al. (2017), studying mangaba genotypes using SSR markers (0.679 to 0.714), and similar to those observed by Chaves et al. (2020), also for mangaba genotypes evaluated using SSR markers (0.428 to 0.581).

The access PR presented higher Ho than He. This result is an indication of excess heterozygosity in this progeny (Yun et al. 2020), confirming the result obtained for the Fixation Index (f = -0.19) (Table 4). The Fixation Index is one of the most important parameters in population genetics, as it presents the balance between homozygotes and heterozygotes present in the population (Pereira et al. 2020). The average value for the Fixation Index was 0.13, indicating a low level of inbreeding for the accessions evaluated.

Mangaba has a self-incompatibility mechanism (Darrault and Schlindwein 2005), which favors cross-fertilization and reduces the occurrence of inbreeding. Thus, the excess of heterozygotes observed for the PR access can be explained by the reproductive system of the species and, probably, the deficit of heterozygotes observed for the other accesses occurred due to crossing between related individuals. Biparental inbreeding was reported to be the cause of the high values observed for the endogamy coefficient in natural populations of mangaba sampled in the Midwest region of Brazil (Costa et al. 2017).

The percentage of polymorphic loci was higher than 75% (Table 4), confirming the presence of genetic variability for the accessions evaluated. The results were higher than those observed in remaining mangaba populations (73.77%) (Silva et al. 2017), which may be related to the marker (SSR), which is considered more informative than ISSR, and the number of individuals evaluated (296). The high genetic variability detected for the accessions evaluated is often related to species that have wide geographic distribution (Al Salameen et al. 2018), as is the case with mangaba.

The estimate for Nei's genetic distance (Table 5) among the accessions ranged from 0.098 to 0.607. The lowest genetic distance of Nei was observed between the PT and TC accessions (0.098), and the highest, between the AB and LG accessions (0.607). The genetic differentiation among accessions (GST) ranged from 0.040 (PT and TC) to 0.201 (LG and CA) (Table 5).

Table 5
Nei's genetic distance (above the diagonal) and GST (below the diagonal) between mangaba accessions.

	PR	LG	PT	BI	TC	AB	CA
PR	-	0.478	0.198	0.340	0.188	0.324	0.229
LG	0.162	-	0.261	0.239	0.574	0.607	0.589
PT	0.071	0.08	-	0.182	0.098	0.172	0.155
BI	0.105	0.069	0.043	-	0.250	0.309	0.312
TC	0.087	0.186	0.040	0.085	-	0.144	0.153
AB	0.110	0.154	0.047	0.074	0.058	-	0.341
CA	0.111	0.201	0.064	0.108	0.079	0.125	-

The values observed for Nei and GST genetic distance corroborate with the estimates obtained for AMOVA (Table 6), in which the smallest proportion of genetic variation (8%) was detected among accessions, while the largest proportion (92%) was observed within accessions. In general, it is observed for species with cross-fertilization, such as mangaba, that 10–20% of the genetic variation is found between populations and that for autogamous species this value is higher than 50% (Al Salameen et al. 2018). This pattern was also observed in a study conducted on natural populations of mangaba using RAPD markers (Fajardo et al. 2018).

Table 6
Molecular analysis of variance (AMOVA) among the seven mangaba accessions.

Source of Variation	GL	SQ	QM	Variância	%	p-Valor	R_{ST}
between accessions	6	1880524.440	313420.740	3230.208	8%	0.001**	0.076
within accessions	605	23650842.429	39092.302	39092.302	92%		
Total	611	25531366.869		42322.510	100%		
** Significant at 1% probability.							

The estimated value for the R_{ST} statistic was 0.076 (Table 6), proving the existence of moderate genetic differentiation among the accessions evaluated. The lowest genetic distance of Nei was observed between the accessions PT and TC (0.098, Table 5), which may be associated with the origin of these

accessions (Indiaroba, Sergipe, Brazil). The G_{ST} values between PT and BI (0.043); PT and TC (0.040); and PT and AB (0.047) are considered low according to the classification proposed by Wright (1978). The other values observed are considered moderate. The presence of private alleles is an indication of differentiation among accessions and demands strategies for the conservation of accessions possessing these alleles. The PR and CA accessions showed no private alleles (Table 2).

The principal coordinates analysis (Fig. 3), performed to evaluate the distribution of genetic variability among the accessions, did not allow us to distinguish them according to origin. The first two principal coordinates explained 22.07% of the total genetic variance of the 296 accessions, with 13.66% explained by coordinate 1 and 8.41% by coordinate 2.

The genetic distance between the accessions was estimated using Rogers' coefficient (1972) and ranged from 0.0 (between accessions BIP2.3 and BIP2.5) to 1.0 (between accessions PTP2.14 and LGP1.10; PTP2.14 and PTP2.11; PTP2.14 and PTP4; PTP2.14 and ABP2.1) (Fig. 4).

The population genetic structure analysis based on Bayesian statistics allowed the identification of two clusters (k = 2) (Fig. 5). The first cluster was composed of 197 progenies, generated from arrays of the accessions PR, PT, BI, TC, AB, and CA. The second cluster was composed of 90 accessions, generated from the LG, PT, and BI accessions. A total of nine accessions (membership values less than 80% for the two clusters detected) showed mixed ancestry.

Knowledge of the genetic structure of the germplasm is essential for the design of efficient strategies for the conservation and genetic improvement of the species. Cluster identification allows for the selection of genitors for breeding programs, which can contribute to increased genetic diversity and potential gain from selection (Campoy et al. 2016).

Population genetic structure analysis based on Bayesian statistics was used to infer on the ancestry of the accessions from the molecular information (Bernard et al. 2018). This analysis did not discriminate the accessions according to their origin, confirming the results obtained for the principal coordinates analysis (Fig. 2) and Rogers genetic distance analysis (Fig. 3). These results indicate that there is no correlation between the molecular data and the geographical origin of the analyzed accessions (Ismail et al., 2019). It was observed that 57 LG accessions, 17 PT accessions and 15 BI accessions were grouped in the same cluster with the analysis in Structure (green color, Fig. 4) and considering the first principal coordinate (Fig. 2), these accessions are grouped on the negative side. Analyzing the dendrogram (Fig. 3), these accessions are also grouped together (blue, yellow, and green).

The creation/maintenance of a germplasm bank presents logistical and economical limitations. Thus, the creation of a core collection, which represents most of the genetic diversity present in the BAG in a smaller number of accessions, is an efficient way to reduce costs (Campoy et al. 2016; Bernard et al. 2018) and increase the efficiency of the design of conservation and genetic improvement strategies for the species. The maximum length sub tree function of the DARwin 6.0.14 software was used iteratively to eliminate redundant accessions, based on the molecular data, and allowed the selection of 225

accessions to compose the core collection of the Mangaba BAG. The seven accessions had representatives in the composition of the core collection, being 6.67% PR, 18.22% LG, 20.89% AB, 11.55% BI, 6.67% CA, 20% PT, and 16% TC (Fig. 6). The selected accessions retained 94.90% of the detected alleles.

The three approaches used to study the genetic structure of mangaba accessions (Structure, PCoA, and Dendrogram) indicated that the accessions used as a matrix have the same genetic background and share common alleles among them (Ahmed et al. 2021). Moreover, as the species mangaba presents a self-incompatibility mechanism (Darrault and Schlindwein 2005), which favors allogamy, the maintenance of the parent accessions in the same experimental field (BAG) contributed to the occurrence of gene flow and, consequently, to the genetic similarity observed among the progenies.

The use of SSR molecular markers allowed the identification of genetic variability within and between progenies and matrices of the accessions of the Mangaba genebank, and contributed to the selection of materials to compose the core collection of this BAG, implemented in the field 15 months after sowing (Fig. 7).

Additionally, data related to agronomic and morphological characterization should be used to support the formation of this core collection, since the combination of this information contributes to the design of more efficient strategies for the use of this genetic resource.

Conclusion

Mangaba progenies from the matrices of seven accessions of the Mangaba genebank show genetic diversity within and among accessions. This diversity can be exploited for conservation and direct use of this genetic resource, and the formation of the core collection is considered essential for the rational and economic management of the collection. The identification of clusters by analyzing the genetic structure of germplasm allows the efficient design of crossings (selection of parents) for breeding programs, corroborating the increase in genetic diversity and gain with selection. Additionally, the data obtained can subsidize studies related to the identification of genes associated with traits of economic interest, through the correlation between the molecular marker data with the agronomic/morphological.

Declarations

Acknowledgements

Vania Cristina Rennó Azevedo for laboratory support in Embrapa Cenargen and Marcela C. Muniz.

Funding

This work was supported by CNPq Grant numbers 313273/2021-9.

Competing Interests

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

Data Avalilability

All relevant data generad durinf this study are included in this manuscript.

References

1. Ahmed W, Feyissa T, Tesfaye K, Farrakh S (2021) Genetic diversity and population structure of date palms (*Phoenix dactylifera* L.) in Ethiopia using microsatellite markers. J Genet Eng Biotechnol 19. https://doi.org/10.1186/s43141-021-00168-5
2. Al Salameen F, Habibi N, Kumar V, Amad SA, Dashti J, Doaij BA (2018) Genetic diversity and population structure of *Haloxylon salicornicum* moq. in Kuwait by ISSR markers. PLoS ONE 13:e0207369. https://doi.org/10.1371/journal.pone.0207369
3. Aljumaili SJ, Rafii MY, Latif MA, Zaharah SZ, Ibrahim W, Miah G (2018) Genetic diversity of aromatic rice germplasm revealed by SSR markers. Biomed Res Int 7658032. https://doi.org/10.1155/2018/7658032
4. Almeida GQ, Chaves LJ, Vieira MC, Ganga RMD (2019) Agronomic evaluation of a *Hancornia speciosa* Gomes germplasm collection from the Brazilian Cerrado. Crop Breed Appl Biotechnol 19:8–14. https://doi.org/10.1590/1984-70332019v19n1a02
5. Alzate-Marin AL, Guidugli MC, Soriani HH, Martinez CA, Mestriner MA (2009) An efficient and rapid DNA minipreparation procedure suitable for PCR/SSR and RAPD analyses in tropical forest tree species. Brazilian Arch Biol Technol 52:1217–1224. https://doi.org/10.1590/S1516-89132009000500020
6. Amorim JAE, Mata LR, Lédo AS, Azevedo VCR, Silva AVC (2015) Diversity and genetic structure of mangaba remnants in states of northeastern Brazil. Genet Mol Res 14:823–833. https://doi.org/10.4238/2015.February.2.7
7. Bernard A, Barreneche T, Lheureux F, Dirlewanger E (2018) Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. PLoS ONE 13:e0208021. https://doi.org/10.1371/journal.. pone.0208021
8. Brondani RPV, Grattapaglia D (2001) Cost-effective method to synthesize a fluorescent internal DNA standard for automated fragment sizing. Biotechniques 31:793–800. https://doi.org/10.2144/01314ST06
9. Campoy JA, Lerigoleur-balsemin E, Christmann H, Beauvieux R, Girollet N, Garcia JQ, Dielewanger E, Barreneche T (2016) Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. BMC Plant Biol 16:49. https://doi.org/10.1186/s12870-016-0712-9
10. Chaves LJ, Ganga RMD, Guimarães RA, Caldeira AJR (2020) Quantitative and molecular genetic variation among botanical varieties and subpopulations of *Hancornia speciosa* Gomes

(Apocynaceae). Tree Genet Genomes 16. https://doi.org/10.1007/s11295-020-01444-0

11. Collevatti RG, Olivatti AM, Telles MPC, Chaves LJ (2016) Gene flow among *Hancornia speciosa* (Apocynaceae) varieties and hybrid fitness. Tree Genet Genomes 12. https://doi.org/10.1007/s11295-016-1031-x
12. Collevatti RG, Rodrigues EE, Vitorino LC, Ribeiro MSL, Chaves LJ, Telles MPC (2018) Unravelling the genetic differentiation among varieties of the Neotropical savanna tree *Hancornia speciosa* Gomes. Ann Bot 122:973–984. https://doi.org/10.1093/aob/mcy060
13. Costa CF, Collevatti RG, Chaves LJ, Lima JS, Soares TN, Telles MPC (2017) Genetic diversity and fine-scale genetic structure in *Hancornia speciosa* Gomes (Apocynaceae). Biochem Syst Ecol 72:63–67. https://doi.org/10.1016/j.bse.2017.03.001
14. Costa TS, Silva AVC, Ledo AS, Santos ARF, Silva Júnior JF (2011) Diversidade genética de acessos do banco de germoplasma em Sergipe. Pesq Agrop Bra 46:499–508. https://doi.org/10.1590/S0100-204X2011000500007
15. Darrault RO, Schlindwein C (2005) Limited Fruit Production in *Hancornia speciosa* (Apocynaceae) and Pollination by Nocturnal and Diurnal Insects 1. Biotropica 37:381–388. https://doi.org/10.1111/j.1744-7429.2005.00050.x
16. Dórea ACS, Almeida DD, Santos AA, Gaspar LMAC, Silva GA, Alves LL, Padilha FF (2021) *Hancornia speciosa* extract presents antimicrobial action against *Pseudomonas aeruginosa* and *Staphylococcus aureus* Extract shows antimicrobial activity / O extracto de Hancornia speciosa apresenta acção antimicrobiana contra Pseudomonas aeruginosa e Staphy. Brazilian J Dev 7:49817–49835. https://doi.org/10.34117/bjdv.v7i5.29980
17. Doyle JJ, Doyle JL (1990) Isolation of Plant DNA from Fresh Tissue. Focus (Madison) 12:13–15
18. Earl DA, vonHoldt BM (2012) Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7
19. Evanno G, Regnaut S, Goudet J, Costa DF, Chagas KPT, Vieira FA (2005) (2018) Genetic diversity in natural populations of *Hancornia speciosa* Gomes: Implications for conservation of genetic resources. Ciência e Agrotecnologia 42:623–630. https://doi.org/10.1590/1413-70542018426019018
20. Guzmán FA, Moore S, Vicente MC, Jahn MM (2020) Microsatellites to enhance characterization, conservation and breeding value of *Capsicum* germplasm. Genet Resour Crop Evol 67:569–585. https://doi.org/10.1007/s10722-019-00801-w
21. Ismail NA, Rafii MY, Mahmud TMM, Hanafi MM, Miah G (2019) Genetic diversity of torch ginger (*Etlingera elatior*) germplasm revealed by ISSR and SSR Markers. Biomed Res Int 2019: 5904804. https://doi.org/10.1155/2019/5904804
22. Jimenez H, Martins L, Montarroyos A, Silva Júnior JF, Alzate-Marin AL, Moraes Filho RM (2015) Genetic diversity of the Neotropical tree *Hancornia speciosa* Gomes in natural populations in Northeastern Brazil. Genet Mol Res 14:17749–17757. https://doi.org/10.4238/2015.December.21.48

23. Kamvar ZN, Tabima JF, Gr̈unwald NJ (2014) Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2014:1–14. https://doi.org/10.7717/peerj.281
24. Kölliker R, Rosellini D, Wang Z (2010) In: Boller B, Posselt UK, Veronesi F (eds) Development and Application of Biotechnological and Molecular Genetic Tools BT - Fodder Crops and Amenity Grasses. Springer New York, New York, NY, pp 89–113
25. Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. https://doi.org/10.1093/bioinformatics/bti282
26. Machado CA, Silva AVC, Silva Júnior JF, Ledo AS (2020) Fenologia de acessos do Banco Ativo de Germoplasma de mangaba da Embrapa Tabuleiros Costeiros. Aracaju: Embrapa Tabuleiros Costeiros, 18p. Available at: https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1127956/fenologia-de-acessos-do-banco-de-germoplasma-de-mangaba-da-embrapa-tabuleiros-costeiros Accessed 10 Jun 2022
27. Nei M (1978) Estimation of Average Heterozygosity and Genetic Distance from A Small Number of Individuals. Genetics 89:583–590
28. Nei M (1973) Analysis of Gene Diversity in Subdivided Populations. Proc Nat Acad Sci 70:3321–3323
29. Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol 13:1143–1155. https://doi.org/10.1111/j.1365-294X.2004.02141.x
30. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinforma Appl 28:2537–2539. https://doi.org/10.1093/bioinformatics/bts460
31. Pereira VM, Ferreira Filho JA, Leão AP, Vargas LHG, Farias MP, Rios SA, Cunha RNV, Formighieri AF, Alves AA, Souza Júnior MT (2020) American oil palm from Brazil: Genetic diversity, population structure, and core collection. Crop Sci 60:3212–3227. https://doi.org/10.1002/csc2.20276
32. Perrier X, Jacquemoud-Collet JP (2006) DARwin software. Available at: http://darwin.cirad.fr/darwin Accessed 02 Feb 2022
33. Prasanth VP, Chandra S (2006) ALS-Binary and BinGeno: A C Program for converting allele size in microsatellite markers to 0–1 (Binary Data) Available at: <https://www.researchgate.net/publication/270648182_ALS-Binary_and_BinGeno_A_C_Program_for_converting_allele_size_in_microsatellite_markers_to_0-1_Binary_Data Accessed 02 Feb 2022
34. Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945–959. 10.1093/genetics/155.2.945
35. Rodrigues A, Yamaguishi A, Chaves L, Coelho ASG, Lima JS, Telles MPC (2015) Development of microsatellite markers for *Hancornia speciosa* Gomes (Apocynaceae). Genet Mol Res 14:7274–7278. https://doi.org/10.4238/2015.July.3.2

36. Rogers JS (1972) Measures of genetic similarity and genetic distance. Stud Genet Univ Texas Publ 7213:145–153
37. Sambrook J, Fritschi EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York, NY
38. Santana JGS, Santos OS, Freitas LS, Soares FA, Muniz EM, Ledo AS, Silva AVC (2021) Phenotypic characterization of the germplasm bank of mangaba (*Hancornia speciosa* Gomes), a unique Brazilian native fruit, with emphasis on its high Vitamin C content. New Z J Crop Hortic Sci 49:361–373. https://doi.org/10.1080/01140671.2021.1898990
39. Santos PS, Freitas LS, Santana JGS, Muniz EN, Rabbani ARC, Silva AVC (2017) Genetic diversity and the quality of Mangabeira tree fruits (*Hancornia speciosa* Gomes-Apocynaceae), a native species from Brazil. Sci Hort 226:372–378. https://doi.org/10.1016/j.scienta.2017.09.008
40. Santos PS, Freitas LS, Muniz EN, Santana JGS, Silva AVC (2021) Phytochemical and antioxidant composition in accessions of the mangaba active germplasm bank. Revista Caatinga, 34: 228–235, 2020. https://doi.org/10.1590/1983-21252021v34n123rc
41. Silva AVC, Amorim JAE, Melo M, Ledo AS, Rabbani ARC (2017) Genetic Diversity of Remaining Populations of Mangaba (*Hancornia speciosa* Gomes) in Restingas of Brazil. J Agric Sci 9:46–52. https://doi.org/10.5539/jas.v9n2p46
42. Silva AVC, Vitória MF, Azevedo VCR, Cardoso MN, Soares ANR, Silva Júnior JF, Ledo, AS (2019) Genetic divertsity of the mangaba genebank using microsatellites. Gen Mol Res 18:1–10. https://doi.org/10.4238/gmr18108
43. Silva AVC, Oliveira JMSP, Cardoso MN, Nascimento ALS, Soares TSN, Silva Júnior JF, Ledo AS, Muniz EN (2021) Collection, ex situ conservation and characterization of mangaba (*Hancornia speciosa* Gomes) germplasm in coastal lowland of Northeastern Brazil. Genet Resour Crop Evol 68:2441–2453. https://doi.org/10.1007/s10722-021-01141-4
44. Silva AVC, Nascimento ALS, Soares ANR, Rabbani ARC, Silva Junior JF, Ledo AS (2019) Identification and preliminar characterization of mangaba premature fruiting accession (*Hancornia speciosa* - Apocynaceae). agro@mbiente on-line, v. 13, p.115–128
45. Silva AVC, Amorim JAE, Vitória MF, Ledo AS, Rabbani ARC (2017) Characterization of trees, fruits and genetic diversity in natural populations of mangaba. Ciência e Agrotecnologia 41:255–262. https://doi.org/10.1590/1413-70542017413048416
46. Silva AVC, Lédo AS, Muniz EM, Santos JS, Feitosa RB, Almeida CS (2012) Postharvest characterization of mangaba (*Hancornia speciosa* Gomes) from natural populations in Sergipe, Brazil. Acta Horticulturae, v. 945, p. 263-265.10.17660/ActaHortic.2012.945.35
47. Silva BSR, Sant'Ana GC, Chaves CL, Androcioli LG, Ferreira RV, Sera GH, Charmetant P, Leroy T, Pot D, Domingues DS, Pereira LFP (2019) Population structure and genetic relationships between Ethiopian and Brazilian *Coffea arabica* genotypes revealed by SSR markers. Genetica 147:205–216. https://doi.org/10.1007/S10709-019-00064-4

48. Soares ANR, Silva AVC, Muniz EN, Melo MFV, Santos PS, Ledo AS (2019) Biometry, Emergence and Initial Growth of Accessions and Mangaba Progenies. J Agric Sci 11:436–448. https://doi.org/10.5539/jas.v11n4p436
49. Soares ANR, Clivati D, Melo MFV, Gitzendanner MA, Soltis D, Soltis P, Silva Júnior JF, Ledo AS, Silva AVC (2018) Genetic diversity of accessions and first generation progeny of the mangaba genebank. Amer J Plant Sci 9:1618–1629. https://doi.org/10.4236/ajps.2018.98117
50. Viegas MP, Silva CLSP, Moreira JP, Cardin LT, Azevedo VCR, Ciampi AY, Freitas MLM, Moraes MLT, Sebbenn AM (2011) Diversidade genética e tamanho efetivo de duas populações de *Myracrodruon urundeuva* Fr. All., sob conservação ex situ. Rev Árvore 35:769–779. https://doi.org/10.1590/s0100-67622011000500002
51. Wright S (1978) Evolution and the Genetics of Populations. Vol. 4. Variability within and among Natural Populations. University of Chicago Press, Chicago
52. Yun SA, Son H-D, Im H-T, Kim S-C (2020) Genetic diversity and population structure of the endangered orchid *Pelatantheria scolopendrifolia* (Orchidaceae) in Korea. https://doi.org/10.1371/journal.pone.0237546

Figures

Figure 1

Mangaba Genebank of Embrapa Tabuleiros Costeiros. Itaporanga d'Ajuda, Sergipe, Brazil.

Figure 2

Fruits, processing and mangaba progenie formation.

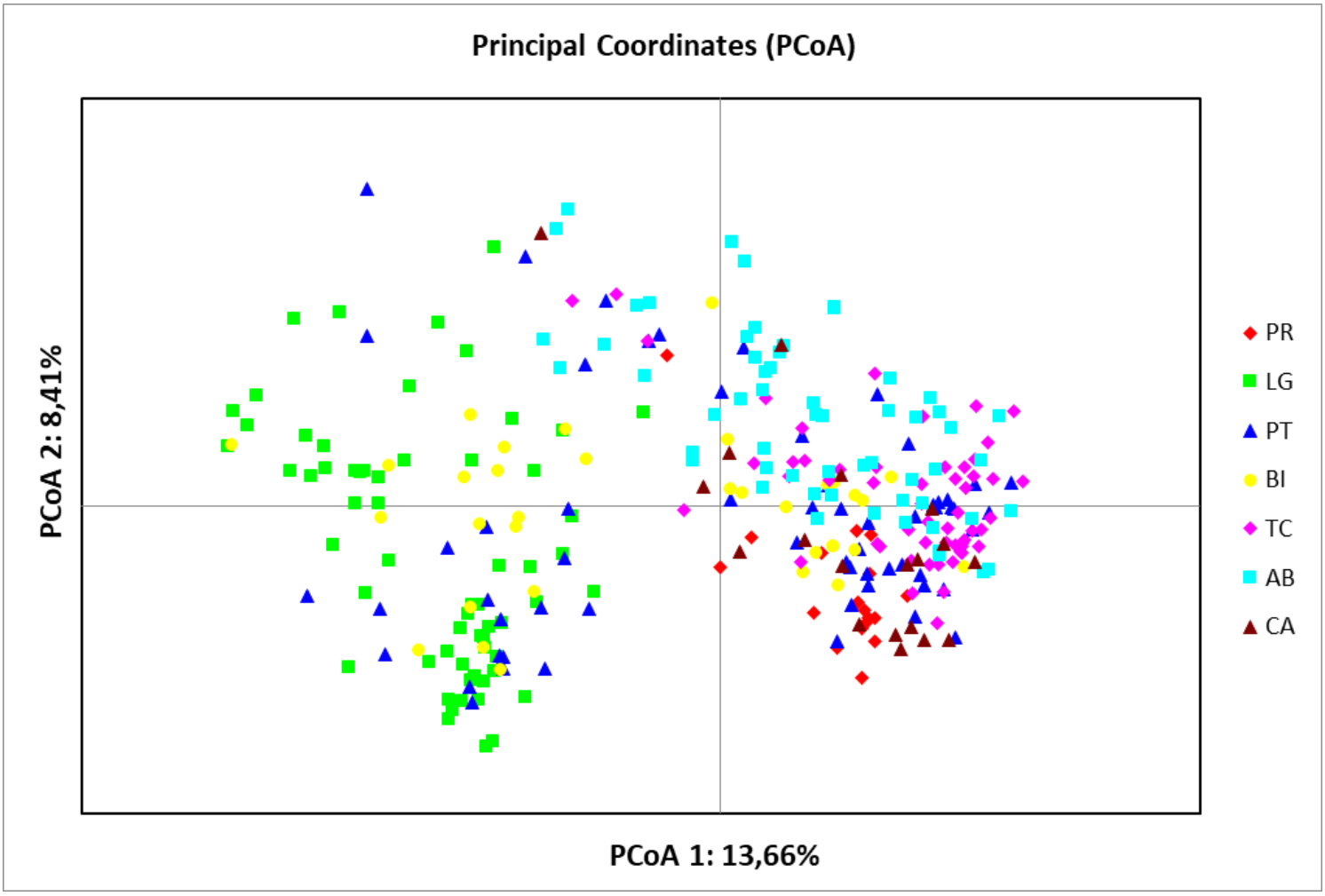

Figure 3

Principal Coordinate Analysis among mangaba accessions.

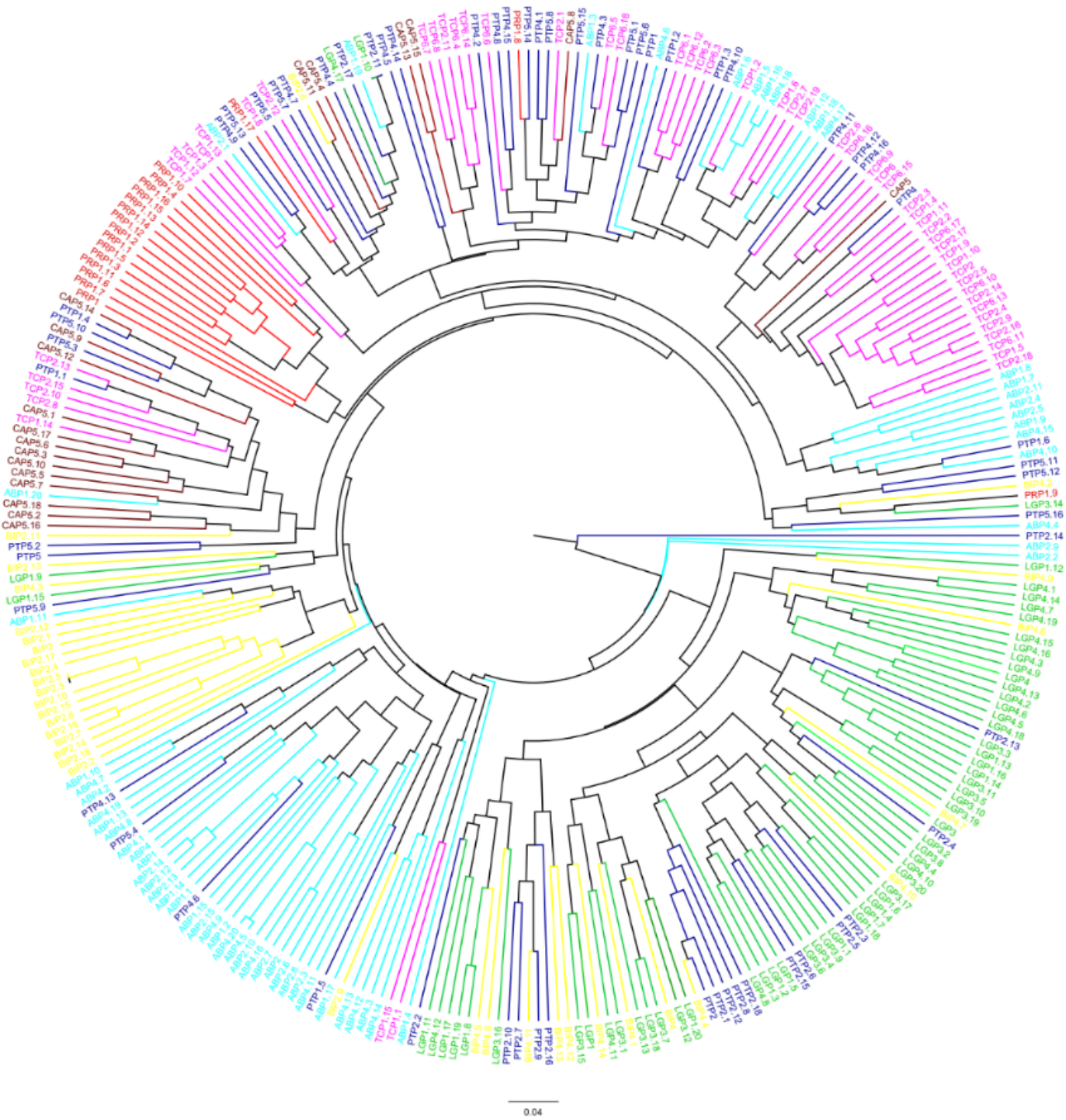

Figure 4

Dendrogram obtained using the UPGMA clustering method based on the Rogers's genetic distance (Rogers, 1972) between mangaba accessions.

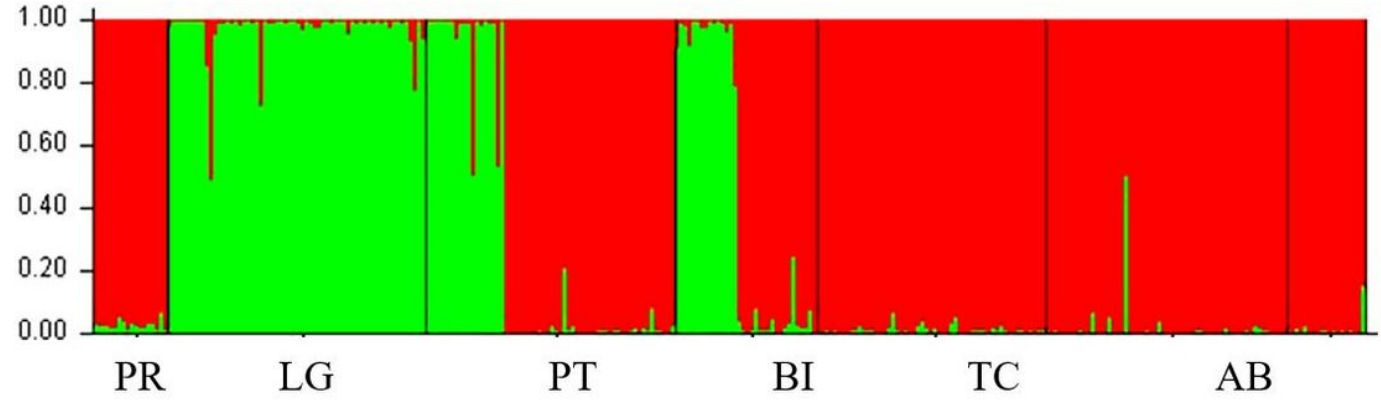

Figure 5

Estimated population structure for 296 mangaba accessions (k=2).

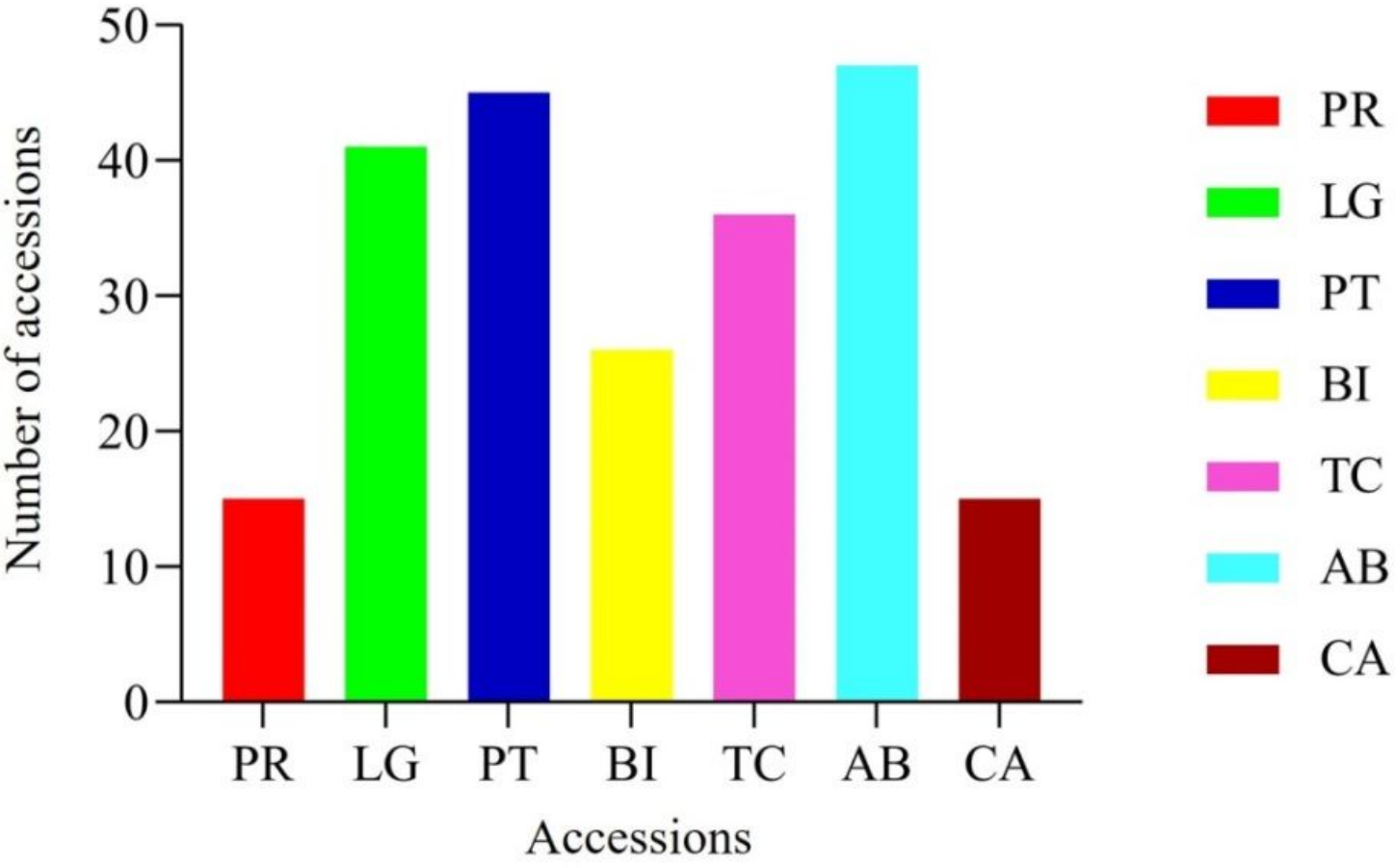

Figure 6

Number of accessions selected to compose the Mangaba Core Collection.

Figure 7

Production of mangaba progenies and implementation of the core collection area. Itaporanga d'Ajuda, Sergipe, Brazil.