REVISTA DE
Biología Tropical

https://doi.org/10.15517/rbt.v69i3.45810

Spondias mombin (Anarcadiaceae): molecular characterization and conservation

Kelli Évelin Müller Zortéa[1*,2,3]; https://orcid.org/0000-0003-0545-6130
Ana Aparecida Bandini Rossi[1,2,3]; https://orcid.org/0000-0002-8318-5375
Auana Vicente Tiago[1,3]; https://orcid.org/0000-0001-9556-9491
Elisa dos Santos Cardoso[1,3]; https://orcid.org/0000-0001-9477-1743
Joyce Mendes Andrade Pinto[4]; https://orcid.org/0000-0002-9484-1868
Eulalia Soler Sobreira Hoogerheide[4]; https://orcid.org/0000-0003-0944-3898

1. Programa de Pós-Graduação em Biodiversidade e Biotecnologia (Rede Bionorte), Universidade do Estado de Mato Grosso - Carlos Alberto Reyes Maldonado, Alta Floresta, Mato Grosso, Brasil; kellimullerz@gmail.com (Correspondence*), anabanrossi@unemat.br, auanavt@gmail.com, elisabyo@gmail.com
2. Programa de Pós-Graduação em Genética e Melhoramento de Plantas, Universidade do Estado de Mato Grosso - Carlos Alberto Reyes Maldonado, Alta Floresta, Mato Grosso, Brasil.
3. Laboratório de Genética Vegetal e Biologia Molecular, Universidade do Estado de Mato Grosso - Carlos Alberto Reyes Maldonado, Alta Floresta, Mato Grosso, Brasil.
4. Embrapa Agrossilvipastoril, Sinop, Mato Grosso, Brasil; joyce.andrade@embrapa.br, eulalia.hoogerheide@embrapa.br

ABSTRACT

Introduction: The fruit of the yellow mombin (*Spondias mombin* L.) is notable due to its sensory and functional qualities. However, there is little knowledge regarding the genetic diversity of this species, and this would aid the implantation of the cultivation of the fruit as a crop, since current production is based on extractivism.
Objective: Evaluate the diversity and genetic structure of natural populations of *S. mombin* in the state of Mato Grosso, Brazil, through microsatellite molecular markers in order to assist in the implementation of conservation strategies and the collection of genetic resources.
Methods: A total of 139 *S. mombin* individuals were sampled in ten natural populations. PCR amplifications were performed with seven fluorescence-marked microsatellite primers. Genetic diversity was evaluated by the number of alleles, expected (*He*) and observed heterozygosity (*Ho*), polymorphic information content (PIC), fixation index (*f*), rare and exclusive alleles. The genetic structure was evaluated using AMOVA, UPGMA dendrogram and Bayesian statistical analysis.
Results: 46 alleles were amplified, which had an average of 6.6 alleles per locus. *He* was higher than *Ho* and *f* was positive, indicating the presence of inbreeding. The PIC ranged from 0.048 to 0.700, and only two loci were poorly informative. We found 27 rare alleles and 16 unique alleles. The largest component of variation was intrapopulational (90.64 %). The estimated gene flow was 1.99, which indicates that there is no genetic isolation between populations, and justifies the F_{ST} value (0.0963). The ten populations were grouped into two groups, and two populations constituted an isolated group. The Mantel test demonstrated that the genetic structure is not related to the geographic distance between populations.
Conclusion: There is genetic diversity in the populations of *S. mombin*, which must be conserved *in situ* or *ex situ*, due to the diversity they present and because they are promising sources for collection of germplasm.

Key words: conservation; genetic resources; genetic variability; microsatellite; yellow mombin.

Brazil has a high potential for the production of numerous varieties of fruits due to its territorial extension, varied types of soils, and climatic conditions (Carvalho et al., 2017). It also has a great diversity of native fruits, many of them still unknown or little exploited commercially, but that have increasing market value, among which is the yellow mombin (*Spondias mombin* L.) (Freitas & Mattietto, 2013). This tree species belongs to the family Anacardiaceae and its fruit stands out for its sensory and functional qualities (Pinto et al., 2003). In addition, the fruit has high levels of carotenoids, tannins and phenolic compounds, and is considered a natural source of antioxidants (Freitas & Mattietto, 2013; Mattietto et al., 2010). *S. mombin* is among the Brazilian species that has the greatest economic potential (Lederman et al., 2008); however, it is in a domestication phase, and information that would assist in the deployment of viable commercial orchards is scarce (Souza et al., 2006). The type of exploitation, the lack of commercial crops and the high loss of fruits during harvest and transport cause annual variations in the harvests, which directly influence its industrialization and functioning of the fruit processing industries (Santana, 2010).

The domestication and incorporation of native species into productive systems require information on genetic variation and knowledge of the size and distribution of genetic variability in natural populations (Costa et al., 2011). In addition, knowledge of the genetic structure of populations and their levels of genetic diversity are important in order to understand how selection acts as a function of adaptability and how the effects of environmental fragmentation may influence population dynamics, thus helping to guide more efficient breeding and conservation programs (Estopa et al., 2006; Sujii et al., 2015).

Genetic diversity can be verified through morphological, agronomic, molecular characteristics, among others (Dardengo et al., 2021; Giles et al., 2018; Giustina et al., 2017; Santos et al., 2020). However, molecular techniques make it possible to accelerate the process of analysis of variability and selection especially when working with perennial species, since they do not require the plant to complete its reproductive cycle, do not suffer interference from the environment and, in addition, have high efficiency for material discrimination (Ferreira & Grattapaglia, 1998).

Recently, some studies have contributed to the characterization of the *Spondias* genus through molecular phylogeny (Silva et al., 2015; Machado et al., 2015), and much information is now available in GenBank. For *S. mombin*, the assembly of the complete chloroplast genome (Santos & Almeida, 2019) and its mitochondria (Martins et al., 2019) has already been obtained. The genetic diversity of *S. mombin* has been evaluated by means of isoenzymes (Gois et al., 2009), morphological traits of the fruits (Silva et al., 2017a), molecular markers, RAPD (Lima et al., 2011) and ISSR (Silva et al., 2017b). However, no studies are available regarding the diversity and genetic structure based on microsatellite molecular markers.

Microsatellites are one of the most polymorphic classes of molecular markers available today and have advantages over other markers based on PCR (polymerase chain reaction) because they are codominant, highly reproducible, require small amounts of DNA, and have high resolution power and high levels of polymorphism (Caixeta et al., 2016). The greatest limitation to the use of this type of marker is the lack of initiators available for all species, since the marker is specific (Arnold et al., 2002) as well as the great amount of work required for the development and isolation of loci containing these markers (Zane et al., 2002). However, the flanking regions of microsatellites are generally conserved between nearby species or genera (Arnold et al., 2002; Caixeta et al., 2016), thus enabling the transferability of these markers between some species. Aguilar-Barajas et al. (2014) were able to select 14 polymorphic microsatellites for *Spondias radlkoferi* Donn. Sm., of which 12 were successfully transferred to *S. mombin*, which enabled its use in diversity studies of this species.

Due to an increase in the demand for the fruits of *S. mombin* and the lack of knowledge that would assist in the implementation of commercial plantations in the future, especially with regard to understanding of the genetic diversity that the species possesses, this study aimed to evaluate, by means of microsatellite molecular markers, the diversity and genetic structure of natural populations of *S. mombin*, in the state of Mato Grosso, Brazil.

MATERIALS AND METHODS

Study area: The study was conducted in 10 municipalities in the state of Mato Grosso. In each municipality, an area in which the species naturally occurred was selected. Each selected area was evaluated as a natural population of *S. mombin*. We sampled 139 individuals distributed as follows: Alta Floresta (AF) 19, Apiacás (AP) 7, Cáceres (CC) 15, Colíder (CL) 14, Marcelândia (MR) 18, Nobres (NB) 8, Nova Bandeirantes (NA) 18, Porto Estrela (PE) 15, Tangará da Serra (TS) 17 and Vila Bela da Santíssima Trindade (VB) 8. All the sampled areas have strong anthropogenic influence, with isolated trees in the middle of the pasture matrix, in the urban matrix or in disconnected forest fragments. The sampled populations of *S. mombin* come from the three biomes that exist in the state of Mato Grosso, namely the Amazon, Pantanal and Cerrado biomes, according to the delimitation of biomes proposed by the Ministry of the Environment (2018).

The Amazon is one of the biomes with the greatest richness of species in the world, but it is also marked by environmental degradation. The Amazon region where sample collections were carried out has mostly an Am-type climate, with defined dry and rainy seasons, a mean annual temperature above 26 °C and annual rainfall capacity of between 2 500 to 3 100 mm (Alvares et al., 2013).

The Pantanal biome is a large wetland mainly located in Brazil. It has natural resources of great importance, but is one which has been put at risk due to unsustainable land use and occupation practices, such as the conversion of natural vegetation into cultivated areas and pastures (Bergier et al., 2018). The region where the sample collections were carried out has an Aw-type climate, with annual precipitation ranging from 1 300 to 1 600 mm and a mean annual temperature of 24 to 26 °C (Alvares et al., 2013).

The Cerrado is the biome that occupies 20 % of the Brazilian territory; however, it is extremely anthropized and much of the original cover has been replaced by pasture areas or temporary crops such as soybeans, maize and rice (Duarte & Leite, 2020). The sample collection region has an Aw-type climate, with annual precipitation ranging from 1 900 to 2 200 mm and a mean annual temperature of 24 °C (Alvares et al., 2013).

Collection of plant material and DNA extraction: Sampling of populations was carried out by randomly prioritizing individuals of reproductive age and with greater geographical distance from each other. Leaf material was collected from each selected individual with a preference for young leaves, without damage and/or signs of disease. Control material was herborized and deposited in the Southern Amazon Herbarium - HERBAM, Mato Grosso State University, Alta Floresta, MT, according to the methodology of Fidalgo and Bonomi (1989) and under the registry numbers: 15 289, 15 290, 15 291, 15 292 and 15 293.

The total genomic DNA was extracted from approximately 100 mg of the foliar material following the CTAB method described by Doyle and Doyle (1987), with modifications for the species: an increase in the concentration of Polyvinylpyrrolidone (PVP) from 1 to 2 % and of β-mercaptoethanol from 0.2 to 1.8 %, an addition of 0.4 % proteinase K in the extraction buffer, reduction of the incubation time in a water bath from 60 minutes to 30 minutes, while maintaining the temperature at 65 °C.

Genotyping: Amplifications were performed via PCR using seven microsatellite markers developed by Aguilar-Barajas et al. (2014) for *Spondias radlkoferi* and with proven

transfer to *S. mombin*. The reactions were performed in a final volume of 10 μl, containing 1.5 μL buffer 10x (100 mM Tris-HCl pH 8.3; 500 mM KCl), 1.5 mM MgCl2, 0.2 mM of each dNTP, 0.5 μM of each primer (F and R); 0.12 μL of Taq DNA polymerase (5 U/μL); 20 ng of DNA and ultrapure water. Amplifications were conducted using an Aeris™ thermocycler (Esco®) following the method described by Aguilar-Barajas et al. (2014).

The primers were marked using different fluorochromes to allow the analysis of the PCR products through duplex and triplex systems in an automatic sequencer (model ABI 3 730, Applied Biosystems). The amplified fragments were detected using the software GeneMarker V. 2.6.3. The fragment size information was used to assemble the matrix with each primer pair for each individual, which was subjected to analysis of the diversity and genetic structure.

Data analysis: The genetic diversity of *S. mombin* populations was characterized as the total number of alleles, number of alleles per locus (Na), observed heterozygosity (*Ho*) and expected (*He*) in Hardy-Weinberg equilibrium and polymorphic information content (PIC). The inbreeding coefficient (*f*) was also calculated according to the method by Weir and Cockerham (1984) for each population and for each locus. These analyses were performed with the aid of Power Marker software v. 3.25 (Liu & Muse, 2005). Alleles were classified according to their frequency and occurrence. As for the frequency, alleles with a frequency equal to or less than 0.05 were considered rare, and those that were common were those with a frequency greater than 0.05 (Sebbenn, 2003). Regarding occurrence, alleles were considered to be widely distributed if they were found in more than 25 % of the populations, and of local distribution if they were found in less than 25 % of the analyzed populations (Sebbenn, 2003). Alleles that occurred in only one population, regardless of frequency, were considered exclusive. For each population, the rare and exclusive alleles were classified. The classification proposed by Sebbenn (2003) was used for local rare, widely distributed rare, local common and widely distributed common alleles.

The analysis of molecular variance (AMOVA) was used to infer the genetic structure of populations and was performed according to Excoffier et al. (1992) with the aid of Arlequin 3.01 software (Excoffier et al., 2006). The gene flow (Nm) was obtained in the PopGene program 1.32 (Yeh et al., 1999), via the equation: $Nm = 0.25 (1 - FST)/FST$. The matrix of the genetic distance values of Nei et al. (1983) among populations generated by the Power Marker Program v. 3.25 was imported into the GENES program (Cruz, 2016) for the construction of the dendrogram using the UPGMA method (Unweighted Pair-Group Method with Arithmetic Averages). This grouping method was chosen because it was the one that best represented the genetic variation under study, based on the value of the cophenetic correlation coefficient (CCC), stress and distortion. To verify whether there was a correlation between genetic similarity and geographical distance in the different populations analyzed, a Mantel test was performed, with 10 000 permutations, using the Genes software (Cruz, 2016).

The Structure program (Pritchard et al., 2000), based on Bayesian statistics, was used to infer the number of groups (k). Twenty runs were performed for each k value (1 to 13), 200 000 burn-ins and 500 000 Markov chain Monte Carlo (MCMC) simulations. The criteria proposed by Pritchard & Wen (2004) and also the criterion proposed by Evano et al. (2005) were used to define the most probable K in relation to those proposed, and the results were uploaded to the Structure Harvester site (Earl & von Holdt, 2012).

RESULTS

Genetic diversity of *S. mombin*: The seven microsatellites used from *S. mombin* amplified a total of 46 alleles, ranging from 5 to 10 and averaging 6.6 alleles per locus. The PIC ranged from 0.048 (SPO21) to 0.700 (SPO40)

with an average of 0.317. *He* was higher than *Ho* for six loci and presented means of 0.335 and 0.300 respectively. The fixation index was positive and different from 0, indicating a higher number of homozygotes in the loci and presence of inbreeding. Only locus SPO18 presented a negative value for *f* (Table 1).

The mean number of alleles found per population was 18.3, among which the Cáceres population (CC) presented the highest number of alleles. The highest number of heterozygotes, estimated by the fixation index, was found in Nobres (NB) and the highest number of homozygotes in the Tangará da Serra population (TS). This result for TS is also observed by the lower *Ho* value. The PE and TS populations

presented a higher rate of inbreeding compared to the other populations (Table 2).

In the evaluation of the general allelic frequency for the 10 populations, 10 local rare alleles, 13 local common alleles and 23 widely distributed common alleles were found. By evaluating the allelic frequency by populations, we found 27 rare alleles (frequency < 0.05), distributed in 8 of the 10 populations of *S. mombin* analyzed. The population Caceres (CC) and Alta Floresta (AF) had the highest numbers of rare alleles (Table 3). In all, 16 unique alleles were found, that is, they appeared in only one of the 10 populations. The highest number of exclusive alleles was found in the AF population (Table 3).

TABLE 1

Estimation of the genetic diversity of *S. mombin* obtained in seven microsatellite loci

Locus	Repeat motif	N	Na	PIC	*He*	*Ho*	*f*
SPO3	$(TTAA)_6$	134	5	0.185	0.193	0.164	0.153
SPO12	$(CTT)_{10}$	135	10	0.250	0.256	0.178	0.310
SPO18	$(CT)_9$	126	6	0.294	0.316	0.350	-0.103
SPO21	$(CT)_6$	124	5	0.048	0.048	0.024	0.495
SPO22	$(AT)_9$	137	7	0.470	0.500	0.445	0.111
SPO31	$(AAT)_9$	136	6	0.272	0.294	0.287	0.028
SPO40	$(ATGT)_7$	99	7	0.700	0.742	0.646	0.134
Mean		127.28	6.6	0.317	0.335	0.300	0.112

N: number of individuals used in the analysis, except for missing data; Na: number of alleles; PIC: polymorphism information content; *He*: expected heterozygosity; *Ho*: observed heterozygosity; *f*: fixation index.

TABLE 2

Estimation of the genetic diversity of *S. mombin* populations obtained using seven microsatellite loci

Population	N	Na	PIC	*He*	*Ho*	*f*
AF	19	23	0.340	0.380	0.354	0.095
AP	7	17	0.368	0.486	0.833	-0.245
CC	15	24	0.326	0.359	0.349	0.073
CL	14	21	0.305	0.375	0.500	-0.076
MR	18	20	0.279	0.314	0.344	-0.064
NB	8	15	0.465	0.211	0.298	-0.351
NA	18	16	0.482	0.251	0.284	-0.091
PE	15	17	0.228	0.263	0.227	0.187
TS	17	16	0.224	0.250	0.169	0.357
VB	8	14	0.174	0.199	0.235	-0.115
Total/Species	139	46	0.317	0.335	0.300	0.112

N: number of individuals used in the analysis, except for missing data; Na: number of alleles; PIC: polymorphism information content; *He*: expected heterozygosity; *Ho*: observed heterozygosity; *f*: fixation index.

TABLE 3

Rare and unique alleles found in the populations of *S. mombin* for the 7 loci microsatellites

Population	Rare alleles	Exclusive alleles
AF	6	4
AP	0	2
CC	7	2
CL	3	3
MR	2	1
NB	2	0
NA	2	0
PE	3	3
TS	2	1
VB	0	0
Total	27	16

Genetic structure of natural populations of *S. mombin*: The analysis of molecular variance showed that there was significant genetic differentiation (P < 0.000) among populations and that the largest component of variation was within populations (90.64 %) (Table 4). The estimated gene flow was 1.99, which indicates that there is no genetic isolation among

populations. The F_{ST} value (0.0963) demonstrates that there is little genetic differentiation among populations. Thus, it appears that the gene flow maintained low differentiation among populations.

The UPGMA dendrogram provided the formation of two groups, the first group (GI) formed by 8 populations and the second group (GII) by 2 populations. The populations for Alta Floresta (AF) and Marcelândia (MR) were isolated in GII (Fig. 1).

The Bayesian analysis obtained by the Structure program, grouped the 139 individuals of *S. mombin* into two groups (k = 2). The contribution of each population to the groups generated by the Structure program can be verified in Fig. 2A and Fig. 2B. The populations of Alta Floresta (AF) and Marcelândia (MR) presented greater contribution to the red group and the rest of the populations presented greater contribution to the green group. This grouping is in accordance with the UPGMA dendrogram, where the AF and MR populations are more genetically similar, and form a
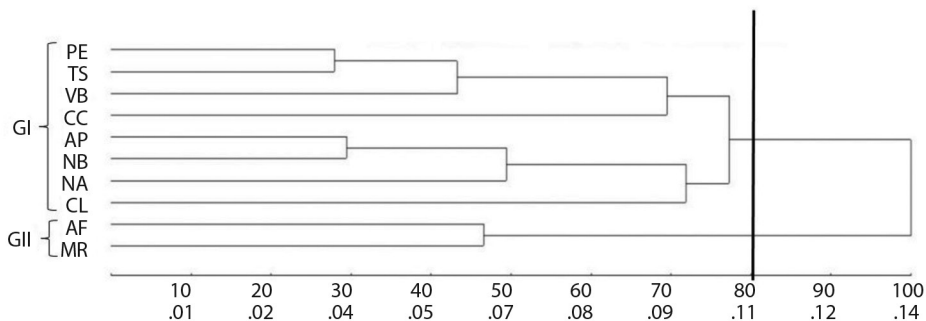

Fig. 1. UPGMA dendrogram generated from the matrix of genetic distance by Nei et al. (1983), and based on data from seven microsatellite loci for 10 natural populations of *S. mombin*. CCC = 0.78; distortion = 3.79 %; stress = 19.47 %.

TABLE 4

Analysis of molecular variance (AMOVA) of the 10 populations of *S. mombin* studied from 7 microsatellite markers

Source of variation	D.f.	SS	CV	TV (%)	*p* value
Among populations	9	31.686	0.112	9.36	< 0.000
Within populations	268	257.646	1.084	90.64	
Total	277	289.332	1.196		

D.f.: Degrees of freedom; SS: sum of squares; CV: coefficient of variation; TV: total variation; *p*: are the probabilities of having a component of variance greater than the values observed at random. The probabilities were calculated by 1 023 random permutations. $F_{ST} = 0.0963$.
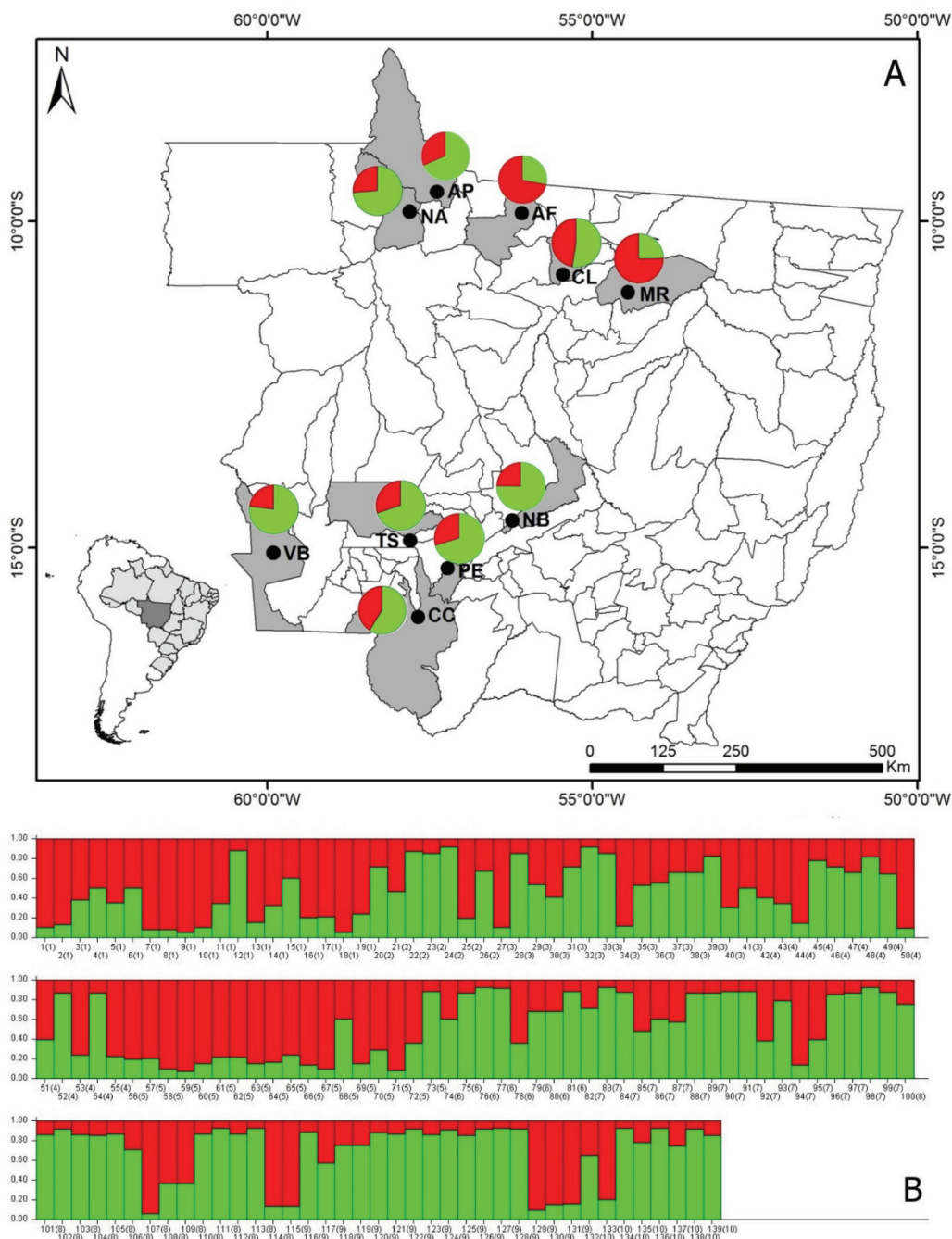
Fig. 2. Grouping analysis and geographical distribution of the groups generated by the Structure program. **A.** Each population is represented by a graph, the colors indicate the proportion of contribution of each *Spondias mombin* population to the groups formed in the Structure program (K = 2). **B.** Distribution of the two genetic groups generated by the Structure program. The vertical lines along the X axis represent the individuals and the colored segments along the Y axis demonstrate the coefficient of association of each individual assigned to each of the inferred K. Numbers within parentheses correspond to populations. 1: AF (Alta Floresta); 2: AP (Apiacás); 3: CC (Cáceres); 4: CL (Colider); 5: MR (Marcelândia); 6: NB (Nobres); 7: NA (Nova Bandeirantes); 8: PE (Porto Estrela); 9: TS (Tangará da Serra); 10: VB (Vila Bela da Santíssima Trindade).

TABLE 5

Genetic distance matrix* by Nei et al. (1983) and geographical distances** (Km)
among the 10 natural populations of *S. mombin*

	AF	AP	CC	CL	MR	NB	NA	PE	TS	VB
AF	----	0.146	0.147	0.188	0.066	0.128	0.095	0.150	0.131	0.167
AP	154	----	0.134	0.116	0.148	0.042	0.083	0.140	0.126	0.098
CC	705	718	----	0.107	0.146	0.096	0.101	0.093	0.098	0.105
CL	126	253	632	----	0.168	0.079	0.112	0.164	0.156	0.093
MR	220	359	655	102	----	0.107	0.087	0.170	0.136	0.163
NB	528	570	225	432	439	----	0.058	0.105	0.100	0.069
NA	192	62	677	270	376	540	----	0.096	0.074	0.101
PE	650	674	68	570	588	157	630	----	0.040	0.068
TS	588	584	130	524	560	174	544	104	----	0.055
VB	714	680	261	685	733	403	620	298	231	----

*Genetic distance above the diagonal. **Geographical distances (Km) below the diagonal.

separate group from the other populations. This structure demonstrates that genetic variation between populations is lower than within populations and agrees with the result of AMOVA.

The genetic structure of *S. mombin* is not determined by the geographical structure, since the most genetically similar populations are not the closest geographically speaking. This result is supported by the Mantel test, which revealed a negative correlation (-0.158) and was not significant between the matrices of genetic similarity and geographical distance. To better illustrate this result, one can cite the example of the AF population that in the analyses is grouped with the MR population that is geographically further than the AP, CL and NA populations (Fig. 2A and Table 5).

DISCUSSION

The polymorphism found in all the analyzed loci indicate the presence of genetic diversity in the analyzed populations. The mean number of alleles per locus found in this study was higher than that found by Aguilar-Barajas et al. (2014) for the same loci in 20 individuals from *S. mombin* in Mexico. These authors tested the transfer of SSR primers developed for *S. radlkoferi* in *S. mombin* and found an average of 6.4 alleles per locus, with a range of 5 to 7 alleles per locus. The increase

in the mean of alleles per locus in this study may be a result of the sizable genetic difference between the individuals evaluated and the greater geographical distance between the areas sampled.

The expected and observed heterozygosity were lower than that found by Aguilar-Barajas et al. (2014) for the same loci. They were also inferior to those found by Silva et al. (2009) when analysis was performed by means of isoenzymes for the genetic diversity of four populations of *S. mombin* in the Rainforest Atlantic Zone of Pernambuco, Brazil. The lower rate of heterozygosity found for *S. mombin* in relation to other tropical tree species may be related to forest fragmentation and population size reduction. These factors can affect genetic processes such as genetic drift, gene flow, selection, reproduction systems, in addition to causing reproductive isolation, increased spatial structure within the population and consequently loss of genetic variability (Carvalho et al., 2010; Young et al., 1996). All the populations of *S. mombin* analyzed in this study were in fragmented areas with significant anthropogenic action, and some individuals were isolated in a pasture or urban matrix.

The frequency of heterozygotes in the population is also a measure of diversity (Nei, 1973), and a heterozygote presents variability due to carrying two different alleles in the

same locus. Observed heterozygosity (*Ho*) is the mean frequency of heterozygous individuals per locus among the sample components (Shimizu et al., 2000) and is highly influenced by the reproductive system, with allogamous species, such as *S. mombin,* generally presenting higher values (Faraldo et al., 2000). The expected heterozygosity (*He*) is the average frequency of heterozygotes, per locus, expected by the Hardy-Weinberg equilibrium (Shimizu et al., 2000) and in this case, there is no influence of the reproductive system. In the analyzed populations of *S. mombin*, the *Ho* was higher than the *He* in 6 of the 10 populations and, when *Ho* is higher than *He*, it is understood that there are more heterozygotes in the population than expected by the Hardy-Weinberg equilibrium (Gois et al., 2009). This result demonstrates there is an important genetic diversity in the populations studied.

In all, one locus and six populations had negative fixation rates, which indicates an excess of heterozygotes. However, the positive fixation index in most loci and in four populations, as well as in the average for the species, indicates the presence of inbreeding, even if in a low proportion, due to the excess of homozygotes. The fixation index (*f*) ranges from -1 to 1, in which positive values indicate excess homozygotes and negative values indicate excess heterozygotes (Azevedo, 2007). Inbreeding is a consequence of the reduction in the number of reproductive individuals in the population, which induces self-fertilization and crossbreeding between related individuals (Young et al., 1996; Reis et al., 2011). Since *S. mombin* is a preferentially allogamous species, with a low rate of self-fertilization and with a tendency to self-incompatibility (Carneiro & Martins, 2012; Oliveira et al., 2012), it is unlikely that only self-fertilization is responsible for inbreeding, and may be caused by crossbreeding between related individuals or by preferential mating.

The way seeds are dispersed may favor the formation of a spatial genetic structure, with more closely related individuals (Sobierajski et al., 2006). It was found during the collections

that there was a large amount of endocarps under the individuals, as well as young individuals of *S. mombin* in a nearby radius. The agoutis, tapirs, bats and monkeys are considered dispersants of seeds of *S. mombin* (Heithaus et al., 1975; Henry et al., 2000; Smythe, 1970), however, no studies were conducted on the distance of seed dispersion performed by these animals. In this case, if they consume the pulp of the fruit and leave the endocarp near the mother tree and the seeds germinate giving rise to a new individual, a spatial genetic structure is formed that allows the crossing between related individuals increasing inbreeding. In addition, there is the possibility of seeds germinating from fruits that fall close to the mother plant, without the need for a disperser to interfere with the process.

Another factor that can contribute to this is pollination or pollinating agents of the species. Turner et al. (1982) found that when pollinators perform pollination preferably in neighboring or nearby plants, high levels of inbreeding are found, even in plants that have self-incompatibility. Small bees, mainly *Apis mellifera* and *Scaptotrigona tubiba* (Carneiro & Martins, 2012; Oliveira et al., 2012), carry out pollination of *S. mombin*. This type of bee calculates energy expenditures for resource collection and opts for shorter displacements in search of energy savings (Bizotto & Santos, 2015). Thus, pollinators may prefer to forage nearby plants so as not to increase energy expenditure, thus assisting in fertilization among related individuals.

The mean of the PIC showed that the loci analyzed were moderately informative, two being poorly informative (PIC < 0.25), four moderately informative (PIC ranging between 0.25 and 0.50) and one highly informative (PIC > 0.50), according to the classification of Botstein et al. (1980). Although they are not all highly informative, it was possible, through these loci, to verify the genetic diversity of *S. mombin* and find rare and unique alleles.

The Caceres (CC) and Alta Floresta (AF) populations presented the highest number of rare alleles and deserve more attention

regarding conservation since, according to Raposo et al. (2007), populations that have a higher proportion of rare alleles are more sensitive to loss of genetic diversity due to genetic drift. The AF population also presented a greater number of exclusive alleles, thus deserving greater attention for conservation. In addition to rare and exclusive alleles, 13 common alleles of local distribution were identified. These alleles are also important for the conservation of the species because, supposedly, they are the ones that ensure adaptation to specific environments and, thus, can have great value in cases of sudden environmental changes or even as a source of genetic variability for resistance to attack by pests and diseases (Sebbenn, 2003).

The UPGMA dendrogram demonstrated that the populations formed two groups, one with the MR and AF populations and the other containing the other populations. The cophenetic correlation showed an association of 78 % (CCC = 0.78) between the distances obtained in the dissimilarity matrix and the cophenetic matrix. According to Rohlf (1970), CCC values greater than 0.70 reflect good agreement between matrices. The dendrogram data were confirmed by the structure grouping, where AF and MR were again presented in the same group.

The geographic distance of the populations is not related to the genetic similarity between them, which was verified by the Mantel test. Despite the absence of a relationship between the genetic structure and the geographical distances for the species, the presence of exclusive alleles in the populations may indicate that this differentiation process is underway and that time and evolutionary mechanisms have not yet been sufficient to differentiate populations at geographical levels. In addition, the species presents alleles with high frequency, which decrease the differentiation between populations.

The value of gene flow revealed that populations are not genetically isolated, since it was greater than 1. According to Wright (1931), Nm values greater than 1 are sufficient to avoid random loss of alleles within the population.

However, this information should be evaluated with care, since due to the high distance between most populations, current gene flow via pollen or seed dispersal is unlikely. Thus, this value may refer to the historical gene flow between populations over time, according to the distribution dynamics of the species in the studied area, which helped to make the allelic frequencies similar amongst themselves.

The populations of *S. mombin* in the state of Mato Grosso present genetic diversity and are promising sources of genetic resources that can be selected, collected and propagated. An example of this is the presence of local common alleles that indicate adaptation to specific environments. Furthermore, a study of genetic divergence between *S. mombin* genotypes through morphological characteristics carried out by Silva et al. (2017a) in the region where the AF, MR and NA populations are located has already demonstrated the potential of some sampled genotypes for use in breeding, due to the presence of variation in traits of interest. These data reinforce the importance of maintaining the genetic diversity of the populations studied.

Genetic diversity is higher at the intrapopulational level, thus, *in situ* or *ex situ* conservation of individuals from each of the studied populations is proposed, in order to enable the maintenance of genetic diversity and effective conservation of *S. mombin*. It is hoped that this knowledge regarding the genetic diversity of *S. mombin* will assist in the implementation of conservation strategies, since it reveals how this diversity is distributed in space and it signals how the loss of genetic variability has occurred in the species.

RESUMEN

Spondias mombin (Anarcadiaceae): caracterización molecular y conservación

Introducción: El fruto amarillo del jobo o yuplón (_Spondias mombin_ L.) destaca por sus cualidades sensoriales y funcionales. Sin embargo, existe poco conocimiento sobre la diversidad genética de esta especie, lo que ayudaría a la implantación del cultivo del fruto como cultivo, ya que la producción actual se basa en el extractivismo.
Objetivo: Evaluar la diversidad y estructura genética de poblaciones naturales de _S. mombin_ en el estado de Mato Grosso, Brasil, a través de marcadores moleculares microsatélites para ayudar en la implementación de estrategias de conservación y recolección de recursos genéticos.
Métodos: Se muestrearon un total de 139 individuos de _S. mombin_ en diez poblaciones naturales. Las amplificaciones por PCR se realizaron con siete cebadores de microsatélites marcados con fluorescencia. La diversidad genética se evaluó por el número de alelos, heterocigosidad esperada (_He_) y observada (_Ho_), contenido de información polimórfica (PIC), índice de fijación (f), alelos raros y exclusivos. La estructura genética se evaluó mediante AMOVA, dendrograma UPGMA y análisis estadístico bayesiano.
Resultados: Se amplificaron 46 alelos, los cuales tenían un promedio de 6.6 alelos por locus. Fue más alto que _Ho_ y f positivo, lo que indica la presencia de endogamia. El PIC osciló entre 0.048 y 0.700, y solo dos loci fueron poco informativos. Encontramos 27 alelos raros y 16 alelos únicos. El mayor componente de variación fue intrapoblacional (90.64 %). El flujo de genes estimado fue de 1.99, lo que indica que no hay aislamiento genético entre poblaciones y justifica el valor de FST (0.0963). Las diez poblaciones se agruparon en dos grupos y dos poblaciones constituyeron un grupo aislado. La prueba de Mantel demostró que la estructura genética no está relacionada con la distancia geográfica entre poblaciones.
Conclusión: Existe diversidad genética en las poblaciones de _S. mombin_, la cual debe ser conservada _in situ_ o _ex situ_, por la diversidad que presentan y porque son fuentes promisorias para la recolección de germoplasma.

Palabras clave: conservación; recursos genéticos; variabilidad genética; microsatélite; jobo, yuplón.

REFERENCES

Aguilar-Barajas, E., Sork, V. L., González-Zamora, A., Rocha-Ramírez, V., Arroyo-Rodríguez, V., & Oyama, K. (2014). Isolation and characterization of polymorphic microsatellite loci in _Spondias Radlkoferi_ (Anacardiaceae). _Applications in Plant Sciences, 2_(11), 1–3. https://doi.org/10.3732/apps.1400079

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. _Meteorologische Zeitschrift, 22_(6), 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Arnold, C., Rossetto, M., McNally, J., & Henry, R. J. (2002). The application of SSRs characterized for grape (_Vitis vinifera_) to conservation studies in Vitaceae. _American Journal of Botany, 89_(1), 22-28. https://doi.org/10.3732/ajb.89.1.22

Azevedo, V. C. (2007). _Desenvolvimento e aplicações de microssatélites, análise de cpDNA e modelagem computacional para estudos da estrutura e dinâmica genética de maçaranduba_ – Manilkara huberi _(Ducke) Chev. Sapotaceae_ (Doctoral thesis). Universidade de Brasília, Brasil.

Bergier, I., Assine, M. L., McGlue, M. M., Alho, C. J. R., Silva, A., Guerreiro, R. L., & Carvalho, J. C. (2018). Amazon rainforest modulation of water security in the Pantanal wetland. _Science of the Total Environment, 619-620_, 1116–1125. https://doi.org/10.1016/j.scitotenv.2017.11.163

Bizotto, L. A., & Santos, R. S. S. (2015). Dinâmica de voo e coleta de recursos por _Apis mellifera_ em pomar de macieira. _Enciclopédia Biosfera, 11_(21), 3499–3506.

Botstein, D., White, R. L., Skolmick, H., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism. _American Journal of Human Genetics, 32_(3), 314–331. https://doi.org/10.1007/BF00266542

Caixeta, E. T., Oliveira, A. C. B., Brito, G. G., & Sakiyama, N. S. (2016). Tipos de marcadores moleculares. En A. Borém & E. T. Caixeta (Eds.), _Marcadores moleculares_ (1st ed., pp. 9–93). Editora UFV.

Carneiro, L. T., & Martins, C. F. (2012). Africanized honey bees pollinate and preempt the pollen of _Spondias mombin_ (Anacardiaceae) flowers. _Apidologie, 43_(4), 474–486. https://doi.org/10.1007/s13592-011-0116-7

Carvalho, A. C. M., Freitas, M. L. M., Moraes, S. M. B., Moraes, M. L. T., Stranghetti, V., Alzate-Marin, A. L., & Sebbenn, A. M. (2010). Diversidade genética, endogamia e fluxo gênico em pequena população fragmentada de *Copaifera langsdorffii*. *Revista Brasileira de Botânica, 33*(4), 599–606. https://doi.org/10.1590/S0100-84042010000400008

Carvalho, C., Kist, B. B., Santos, C. E., Treichel, M., & Filter, C. F. (2017). *Anuário brasileiro da fruticultura 2017*. Editora Gazeta Santa Cruz.

Costa, T. S., Silva, A. V. C., Lédo, A. S., Santos, A. R. F., & Silva Júnior, J. F. (2011). Diversidade genética de acessos do banco de germoplasma de mangaba em Sergipe. *Pesquisa Agropecuária Brasileira, 46*(5), 499–508. https://doi.org/10.1590/S0100-204X2011000500007

Cruz, C. D. (2016). Genes Software – extended and integrated with the R, Matlab and Selegen. *Acta Scientiarum Agronomy, 38*(4), 547–552. https://doi.org/10.4025/actasciagron.v38i4.32629

Dardengo, J. F. E., Rossi, A. A. B., Oliveira, L. O., Pena, G. F., Rivas, L. H., Silva, C. J., & Rufatto, F. P. (2021). Structure and genetic diversity of *Theobroma speciosum* (Malvaceae) and implications for Brazilian Amazon conservation. *Rodriguésia, 72*(1), e02022018. https://doi.org/10.1590/2175-7860202172023

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochem Bulletin, 19*(1), 11–15.

Duarte, T. E. P. N, & Leite, L. B. (2020). Cidades médias no Cerrado Brasileiro: desafios para a conservação da biodiversidade. *Terra Plural, 14*(1), e2013420. https://doi.org/10.5212/terraplural.v.14.2013420.006

Earl, D. A., & von Holdt, B. M. (2012). Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conservation Genetics Resources, 4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Estopa, R. A., Souza, A. M., Moura, M. C. O., Botrel, M. C. G., Mendonça, E. G., & Carvalho, D. (2006). Diversidade genética em populações naturais de candeia (*Eremanthus erythropappus* (DC.) MacLeish). *Scientia Forestalis, 70*, 97–106.

Evano, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology, 14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., Laval, G., & Schneider, S. (2006). *Arlequin 3.01. An integrated software package for population genetics data analysis*. University of Berne.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics, 131*(2), 479–491.

Faraldo, M. I. F., Silva, R. M., Ando, A., & Martins, P. S. (2000). Variabilidade genética de etnovariedades de mandioca em regiões geográficas do Brasil. *Scientia Agricola, 57*(3), 499–500. https://doi.org/10.1590/S0103-90162000000300020

Ferreira, M. E., & Grattapaglia, D. (1998). *Introdução ao uso de marcadores moleculares em análise genética*. Embrapa Recursos Genéticos e Biotecnologia.

Fidalgo, O., & Bononi, V. L. R. (1989). *Técnicas de coleta, preservação e herborização de material botânico*. Instituto de Botânica de São Paulo.

Freitas, D. G. C., & Mattietto, R. A. (2013). Ideal sweetness of mixed juices from Amazon fruits. *Ciência e Tecnologia de Alimentos, 33*(S1), 148–154. https://doi.org/10.1590/S0101-20612013000500022

Giles, J. A. D., Partelli, F. L., Ferreira, A., Rodrigues, J. P., Oliosi, G., & Lima e Silva, F. H. (2018). Genetic diversity of promising 'conilon' coffee clones based on morpho-agronomic variables. *Annals of the Brazilian Academy of Sciences, 90*(2), 2437–2446. http://dx.doi.org/10.1590/0001-3765201820170523

Giustina, L. D., Rossi, A. A. B., Vieira, F. S., Tardin, F. D., Neves, L. G., & Pereira, T. N. S. (2017). Variabilidade genética em genótipos de teca (*Tectona grandis* Linn. F.) baseada em marcadores moleculares ISSR e caracteres morfológicos. *Ciência Florestal, 27*(4), 1311–1324. https://doi.org/10.5902/1980509829894

Gois, I. B., Silva-Mann, R., & Ferreira, R. A. (2009). Variabilidade genética de *Spondias lutea* L. em uma população do baixo São Francisco sergipano, por meio de isoenzimas. *Scientia Forestalis, 37*(81), 55–60.

Heithaus, E. R., Fleming, T. H., & Opler, P. A. (1975). Foraging patterns and resource utilization in seven species of bats in a seasonal tropical forest. *Ecology, 56*(4), 841–854. https://doi.org/10.2307/1936295

Henry, O., Feer, F., & Sabatier, D. (2000). Diet of the Lowland Tapir (*Tapirus terrestris* L.) in French Guiana. *Biotropica, 32*(2), 364–368.

Lederman, I. E., Lira Júnior, J. S., & Silva Júnior, F. S. (2008). *Spondias no Brasil: umbu, cajá e espécies afins*. Empresa Pernambucana de Pesquisa Agropecuária.

Lima, A. T. B., Souza, V. A. B., Gomes, R. L. F., & Lima, P. S. C. (2011). Molecular characterization of cajá, *Spondias mombin* (Anacardiaceae), by RAPD markers. *Genetics and Molecular Research, 10*(4), 2893–2904. http://dx.doi.org/10.4238/2011.November.25.1

Liu, K., & Muse, S. (2005). PowerMarker: Integrated analysis environment for genetic marker data.

1035

Bioinformatics, 21(9), 2128–2129. https://doi.org/10.1093/bioinformatics/bti282

Machado, M., Carvalho, P., & van den Berg, C. (2015). Domestication, hybridization, speciation, and the origins of an economically important tree crop of *Spondias* (Anacardiaceae) from the Brazilian Caatinga dry forest. *Neodiversity, 8*, 8–49. https://doi.org/10.13102/neod.81.2

Martins, G., Balbino, E., Marques, A., & Almeida, C. (2019). Complete mitochondrial genomes of the *Spondias tuberosa* Arr. Cam and *Spondias mombin* L. reveal highly repetitive DNA sequences. *Gene, 720*, 144026. https://doi.org/10.1016/j.gene.2019.144026

Mattietto, R. A., Lopes, A. S., & Menezes, H. C. (2010). Caracterização física e físico-química dos frutos da cajazeira (*Spondias mombin* L.) e de suas polpas obtidas por dois tipos de extrator. *Brazilian Journal of Food Technology, 13*(3), 156–164. https://doi.org/10.4260/BJFT2010130300021

Ministry of the Environment. (2018). *Download de dados geográficos*. http://mapas.mma.gov.br/i3geo/datadownload.htm

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, 70*(12), 3321–3323.

Nei, M., Tajima, F., & Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution, 19*(2), 153–170.

Oliveira, M. O., Souza, F. X., & Freitas, B. M. (2012). Abelhas visitantes florais, eficiência polinizadora e requerimentos de polinização na cajazeira (*Spondias mombin*). *Revista Acadêmica: Ciências Agrárias e Ambientais, 10*(3), 277–284. https://doi.org/10.7213/academica.7711

Pinto, W. S., Dantas, A. C. V. L., Fonseca, A. A. O., Ledo, C. A. S., Jesus, S. C., Calafange, P. L. P., & Andrade, E. M. (2003). Caracterização física, físico-química e química de frutos de genótipos de cajazeiras. *Pesquisa Agropecuária Brasileira, 38*(9), 1059–1066. https://doi.org/10.1590/S0100-204X2003000900006

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945–959.

Pritchard, J. K., & Wen, W. (2004). *Documentation for structure software: Version 2*. University of Chicago.

Raposo, A., Martins, K., Ciampi, A. Y., Wadt, L. H. O., & Veasey, E. A. (2007). Diversidade genética de populações de andiroba no Baixo Acre. *Pesquisa Agropecuária Brasileira, 42*(9), 1291–1298. http://dx.doi.org/10.1590/S0100-204X2007000900011

Reis, R. V., Oliveira, E. J., Viana, A. P., Pereira, T. N. S., Pereira, M. G., & Silva, M. G. M. (2011). Diversidade genética em seleção recorrente de maracujazeiro-amarelo detectada por marcadores microssatélites. *Pesquisa Agropecuária Brasileira, 46*(1), 51–57. http://dx.doi.org/10.1590/S0100-204X2011000100007

Rohlf, J. (1970). Adaptive Hierarchical Clustering Schemes. *Systematic Zoology, 19*(1), 58–82.

Santana, F. F. (2010). *Caracterização de genótipos de cajazeira* (Doctoral thesis). Universidade Estadual Paulista, Brazil. https://repositorio.unesp.br/handle/11449/105213

Santos, V., & Almeida, C. (2019). The complete chloroplast genome sequences of three *Spondias* species reveal close relationship among the species. *Genetics and Molecular Biology, 42*(1), 132–138. http://dx.doi.org/10.1590/1678-4685-GMB-2017-0265

Santos, J. F. L., Rossi, A. A. B., Pena, G. F., Tiago, A. V., Zortéa, K. E. M., Cardoso, E. S., Pedri, E. C. M., Santos, I. C. B., Santos, P. H. A. D., Santos, D. B., & Santos, I. R. B. (2020). Variability of fruits and seeds of *Byrsonima crassifolia* (Malpighiaceae) genotypes cultivated in northern Mato Grosso State, Brazil. *Genetic and Molecular Research, 19*(2), gmr18620. http://dx.doi.org/10.4238/gmr18620

Sebbenn, A. M. (2003). Número de populações para conservação genética *in situ* de espécies arbóreas. *Revista do Instituto Florestal, 15*(1), 45–51.

Shimizu, J. Y, Jaeger, P., & Sopchaki, S. A. (2000). Variabilidade genética em uma população remanescente de araucária no Parque Nacional do Iguaçu, Brasil. *Boletim de Pesquisa Florestal, 41*, 18–36.

Silva, J. N., Costa, A. B., Silva, J. V., & Almeida, C. (2015). DNA *barcoding* and phylogeny in neotropical species of the genus *Spondias*. *Biochemical Systematics and Ecology, 61*, 240–243. http://dx.doi.org/10.1016/j.bse.2015.06.005

Silva, E. F., Martins, L. S. S., & Oliveira, V. R. (2009). Diversity and genetic struture in cajá tree (*Spondias mombin* L.) populations in northeastern Brazil. *Revista Brasileira de Fruticultura, 31*(1), 171–181. http://dx.doi.org/10.1590/S0100-29452009000100024

Silva, B. M., Rossi, A. A. B., Dardengo, J. F. E., Tiago, P. V., Silveira, G. F., & Souza, S. A. M. (2017a). Genetic divergences between *Spondias mombin* (Anacardiaceae) genotypes found through morphological traits. *Revista de Biología Tropical, 65*(4), 1337–1346. http://dx.doi.org/10.15517/rbt.v65i4.25765

Silva, B. M., Rossi, A. A. B., Tiago, A. V., Schmitt, K. F. M., Dardengo, J. F. E., & Souza, S. A. M. (2017b). Genetic diversity of Cajazeira (*Spondias mombin* L.) in three geographic regions. *Genetics and Molecular Research, 16*(1), gmr16018946. http://dx.doi.org/10.4238/gmr16018946

Smythe, N. (1970). Relationships between fruiting seasons and seed dispersal methods in a neotropical forest. *The American Naturalist, 104*(935), 25–35.

Sobierajski, G. R., Kageyama, P. Y., & Sebbenn, A. M. (2006). Sistema de reprodução em nove populações de *Mimosa scabrella* Bentham (Leguminosaceae). *Scientia Forestalis, 71*, 37–49.

Souza, F. X., Costa, J. T. A., & Lima, R. N. (2006). Características morfológicas e fenológicas de clones de cajazeira cultivados na Chapara do Apódi, Ceará. *Revista Ciência Agronômica, 37*(2), 208–215.

Sujii, P. S., Martins, K., Wadt, L. H. O., Azevedo, V. C. R., & Solferini, V. N. (2015). Genetic structure of *Bertholletia excelsa* populations from the Amazon at different spatial scales. *Conservation Genetics, 16*(4), 955–964. https://doi.org/10.1007/s10592-015-0714-4

Turner, M. E, Stephens, J. C., & Anderson, W. W. (1982). Homozygosity and patch structure in plant populations as a result of nearest-neighbor pollination. *Proceedings of the National Academy of Sciences, 79*(1), 203–207. https://doi.org/10.1073/pnas.79.1.203

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution, 38*(6), 1358–1370. https://doi.org/10.2307/2408641

Wright, S. (1931). Evolution in mendelian populations. *Genetics, 16*(2), 97–159.

Yeh, F. C., Yang, R. C., & Boyle, T. (1999). *POPGENE: Microsoft Windows-based freeware for population genetic analysis* (Version 1.32, software). https://sites.ualberta.ca/~fyeh/popgene_download.html

Young, A., Boyle, T., & Brown, T. (1996). The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution, 11*(10), 413–418. https://doi.org/10.1016/0169-5347(96)10045-8

Zane, L., Bargelloni, L., & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology, 11*(1), 1–16. https://doi.org/10.1046/j.0962-1083.2001.01418.x