



Investigating genetic diversity in sapucaia using inter simple sequence repeat markers

R.C. Borges¹, F.M.G. Santos², M.C.C. Maia³, P.S.C. Lima³ and S.E.S. Valente⁴

¹Departamento de Agronomia, Universidade Federal do Piauí, Campus Ininga, Teresina, PI, Brasil

²UNINOVAFAPI, Teresina, PI, Brasil

³Empresa Brasileira de Pesquisa Agropecuária, EMBRAPA Meio-Norte, Teresina, PI, Brasil

⁴Departamento de Biologia, Universidade Federal do Piauí, Campus Ininga, Teresina, PI, Brasil

Corresponding author: S.E.S. Valente

E-mail: svalente@ufpi.edu.br

Genet. Mol. Res. 15 (3): gmr.15038565

Received February 17, 2016

Accepted March 28, 2016

Published August 18, 2016

DOI <http://dx.doi.org/10.4238/gmr.15038565>

Copyright © 2016 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. Sapucaia is a tree species originating from the Brazilian Amazon and is widely distributed in Brazil, especially in the mid-north region (Piauí and Maranhão states). Its seeds are rich in calories and proteins, and possess great potential for commercialization. Little is known about the genetic variability in the germplasm of most *Lecythis* species. Here, 11 inter-simple sequence repeat primers were used to estimate the genetic variability among 17 accessions, and to determine the levels of genetic variation and the standards of population structure in sapucaia. The accessions were obtained from the active germplasm bank (AGB) of Embrapa Meio-Norte, Teresina, PI, Brazil, and corresponded to four occurrence areas. Ninety-six loci were analyzed

among the studied individuals. High variation was found at the species level, where the percentage of polymorphic bands was 94.79%, Nei's genetic diversity (h) was 0.3110, and Shannon's index (I) was 0.4732. In the analyzed populations, the percentage polymorphism ranged from 20.83 to 94.79%, Nei's genetic diversity ranged from 0.0863 to 0.2969, and Shannon's index ranged from 0.1260 to 0.4457. Significant genetic differentiation was detected among the populations ($\Phi_{ST} = 10.66\%$); however, the greatest genetic differentiation was found within the populations (89.34%), between which there was an intermediate level of gene flow ($N_m = 1.10$). Accessions BGS 2 and BGS 4 were the most divergent, whereas accessions BGS 14 and BGS 15 were the most similar. Therefore, sapucaia analyzed from the AGB present an elevated level of genetic diversity and may have potential use in genetic breeding programs.

Key words: Chestnut tree; Genetic variability; Germplasm bank; *Lecythis pisonis*

INTRODUCTION

Sapucaia (*Lecythis pisonis* Cambess) is a popular species in Brazil, which belongs to the family Lecythidaceae. This chestnut tree originates from the Amazon and is distributed widely throughout Brazil. Its nuts are an important source of calories (average: 736.87 kcal/100 g), are rich in calcium, phosphorus, and manganese, and can be consumed by humans and animals as a source of lipids and proteins (Souza et al., 2008). Several parts of the plant have been used in popular medicine, and recent studies have revealed the presence of compounds that have moderate cytotoxic activity (Silva et al., 2012). This plant is used in the wood industry, in the production of cosmetics, for ornamental cultivation, and in the production of handmade objects (Revilla, 2002).

Brazilian chestnut trees have potential as an extractive product that can be exported. The exploitation of chestnuts is now the main economic activity in the Amazon region and sustains many families (Sujii et al., 2013). Sapucaia seeds can be lucrative because, in their natural environment, this species produces up to 10 tons/ha per year (Revilla, 2002). Therefore, the intensity of exploitation has irreparably harmed the Brazilian flora, and has endangered many species of native fruit trees (Ribeiro and Rodrigues, 2006), which has highlighted the importance of preserving sapucaia.

Genetic variability within a species is crucial to ensure its adaptive potential in face of environmental adversities and anthropic pressure. It is necessary to quantify the genetic variability within and among populations in order to evaluate how the species face those adversities, persist, and reproduce over time (Ribeiro and Rodrigues, 2006). The existence of genetic variability is a prerequisite for genetic breeding programs, because it provides the basis for the identification of divergent individuals, helping the breeder to select for more promising combinations that are favorable to crossing (Faleiro, 2007; Ramalho et al., 2012).

Currently, few studies have investigated the genetic diversity in natural populations of sapucaia. Analyses based on phenotypic characteristics have revealed great genetic variability in the germplasm of this species (Souza et al., 2008). However, studies based on morphological

characteristics may present erroneous results, since most of the phenotypic characteristics are influenced by the environment (Borém and Caixeta, 2009).

Molecular markers have been widely used in studies of genetic diversity and in breeding programs. These markers have several advantages over morphologic indicators. For example, they are phenotypically neutral, they generate an unlimited number of genetic polymorphisms, they are applicable to all types of vegetative material across all development stages, they permit the correct identification of varieties without the need for many characteristics, and they are free of epistatic effects (Faleiro, 2007; Freitas et al., 2007).

Dominant markers, such as inter-simple sequence repeats (ISSRs), have been employed successfully in studies of genetic diversity and population structuring. The results are compatible with those obtained from studies utilizing codominant markers (Madesis et al., 2014). ISSRs are based on the polymerase chain reaction (PCR) technique, which involves the amplification of DNA segments between two microsatellite regions oriented in the opposite direction (Gupta et al., 1994; Zietkiewicz et al., 1994). These markers have the potential to be integrated in breeding programs. They have been used in the characterization and maintenance of germplasm to estimate the extension of genetic diversity in several species, to resolve problems related to phylogeny, for genomic mapping, and in studies of evolutionary biology (Vijayan, 2005; Souza et al., 2012; Yu et al., 2014). However, ISSR markers have not been used in genetic studies of *L. pisonis*.

The objective of the present study was to use ISSR markers to study the genetic diversity, genetic distance, and genetic structure among 17 accessions of *L. pisonis* originating from four areas in the states of Piauí and Maranhão, Brazil.

MATERIAL AND METHODS

Genetic material

The study was carried out with 17 accessions of sapucaia forming part of the active germplasm bank (AGB) of Embrapa Meio-Norte (Table 1), established in Teresina, PI (05°05'S, 42°49'W). These accessions were collected from four areas of occurrence in the states of Piauí and Maranhão. Accession BGS 13 was collected in the Campus of Universidade Federal do Piauí (05°05'13"S, 42°48'41"W) and was not considered to represent a population of Teresina. Therefore, it was only used to analyze groupings among the accessions.

Extraction of the genomic DNA

Genomic DNA was extracted from the leaves of collected plants using the Qiagen DNeasy Plant Mini Kit (Qiagen, Germany), following the manufacturer protocol. DNA quality was determined using the Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific, USA). DNA quantification was performed by electrophoresis on 0.8% agarose gel with 0.5X TBE. Standard bands were visualized by the addition of 10,000X GelRed™ (Uniscience, USA) using bromophenol blue as a carrier, observed under ultraviolet light (UV), and photographed with MiniBIS Pro® equipment. For DNA quantification, 5-μL samples and 100 ng/μL DNA marker (phage λ) were used to estimate the concentration of DNA per μL and, consequently, the total amount of DNA extracted.

Table 1. Accessions of *Lecythis pisonis* used in this study and their places of collection.

Area/population	Accession No.	City of collection	Brazilian state
1	BGS1	José de Freitas	Piauí
	BGS2	José de Freitas	Piauí
	BGS3	José de Freitas	Piauí
	BGS4	José de Freitas	Piauí
	BGS6	José de Freitas	Piauí
	BGS7	José de Freitas	Piauí
	BGS8	José de Freitas	Piauí
	BGS10	José de Freitas	Piauí
	BGS11	José de Freitas	Piauí
	BGS12	José de Freitas	Piauí
2	BGS14	Altos	Piauí
	BGS15	Altos	Piauí
	BGS16	Altos	Piauí
	BGS17	Altos	Piauí
3	BGS18	Timon	Maranhão
	BGS19	Timon	Maranhão
4	BGS13	Teresina	Piauí

DNA amplification and electrophoresis

Eleven ISSR primers developed by the University of British Columbia (UBC), Canada, were selected to characterize the 17 studied accessions (Table 2). Amplifications were carried out in reactions containing 10 μ L 1X buffer [20 mM Tris-HCl, pH 8.0; 0.1 mM EDTA; 1 mM DTT; 50% glycerol (v/v)], 2.0 mM MgCl₂, 0.8 mM dNTP (dATP, dCTP, dGTP, and dTTP), 0.8 μ M each primer, 1 U Taq DNA polymerase (Invitrogen, Thermo Fisher Scientific, USA), and 1 μ L genomic DNA (7 ng/ μ L). PCR amplifications were performed in a Veriti 96-Well Thermal Cycler (Applied Biosystems®, USA) with an initial denaturation step of 4 min at 94°C; 40 cycles of 60 s at 94°C (denaturation), annealing at a primer-dependent temperature for 1 min, extension at 72°C for 1 min and a final extension of 5 min at 72°C.

Table 2. Inter-simple sequence repeat primers used in the molecular characterization of sapucaia accessions.

Primer	Tm (°C)	Ta (°C)	Sequence (5'-3')
UBC 807	47.0	52.0	(AG) ₈ T
UBC 810	45.4	52.0	(GA) ₈ T
UBC 825	51.4	55.0	(AC) ₇ A
UBC 826	52.8	56.0	(AC) ₈ C
UBC 836	48.9	54.0	(GA) ₈ Y ^a A
UBC 840	47.4	52.0	(GA) ₈ Y ^a T
UBC 841	48.5	50.0	(GA) ₈ Y ^a C
UBC 842	48.8	50.0	(GA) ₈ Y ^a G
UBC 845	48.1	46.0	(CT) ₈ R ^b G
UBC 856	52.8	59.0	(AC) ₈ Y ^a A
UBC 864	43.6	43.0	(ATG) ₅

^aY: C/T; ^bR: A/G. Tm: melting temperature; Ta: annealing temperature.

The products of each amplification were separated on a 1.5% agarose gel in 0.5X TBE buffer and the gels were stained with GelRed™ 1X (Uniscience) by electrophoresis for 4 h (110 V). The bands were visualized by coloring the gels with GelRed™ and through the use of UV light to determine the presence of standard and experimental bands. These standards were analyzed to determine the number of bands based on the resolution and degree of amplification

through visual evaluation, and were compared to molecular DNA indicator Ladders, 100 and 50 bp in size (Invitrogen, Thermo Fisher Scientific).

Statistical analysis

The individuals were genotyped according to the absence (0) or presence (1) of a band. From these data, two binary matrices were generated: one with 16 individuals, which was used to analyze diversity and genetic structuring, and one with 17 accessions, which was used to analyze genetic similarity among the accessions. In all analyses, it was assumed that the loci were dominant and were in Hardy-Weinberg equilibrium. From that assumption, we observed the percentage of polymorphic loci (PPB), rate of Nei's genetic diversity (h , also known as expected heterozygosity), Shannon's index (H), and genetic differentiation coefficient (G_{ST}). For this analysis, the POPGENE software, version 1.31 (Yeh et al., 1999) was used. Nei's genetic distance (Nei, 1987) was used to create a dendrogram using the unweighted pair group method with arithmetic mean (UPGMA) in the TFPGA software (Miller, 1997) with the Taylor expansion (Lynch and Milligan, 1994). The consistency of each node was calculated with 1000 bootstraps.

The Jaccard coefficient (Rohlf, 1992) was used to obtain a genetic similarity matrix that grouped the 17 sapucaia accessions by the UPGMA method with the help of the PAST 1.34 program (Hammer et al., 2001). The stability of the groupings was verified with 1000 bootstraps and adjustments between the similarity matrix and the dendrogram were evaluated by the cophenetic correlation coefficient (r) (Cruz and Carneiro, 2012).

To analyze structure and genetic differentiation, 16 individuals were placed in order of hierarchy, forming two groups (Piauí and Maranhão), which were subdivided into three populations, two from Piauí (José de Freitas and Altos) and one from Maranhão (Timon). The G_{ST} coefficient was used as an indicator of genetic differentiation. To evaluate the genetic differentiation among groups, the index, θ_p , was calculated (Weir and Cockerham, 1984) (using the TFPGA software with the Taylor expansion). Genetic variation within and among the populations was evaluated by analysis of molecular variance (AMOVA) using the Arlequin software version 3.01 (Excoffier et al., 2005). Gene flow (N_m) among populations was estimated indirectly using the G_{ST} index, according to the equation $N_m = 0.25 (1 - G_{ST}) / G_{ST}$ (McDermott and McDonald, 1993).

Bayesian inference with the Structure software version 2.3.3 (Pritchard et al., 2000) was used to evaluate the existence of population genetic structure. The number of simulated populations was $K = 4$, based on the best value of ΔK (Evanno et al., 2005). There were six simulations and 10 repetitions for each K value. A mixed model with correlated allele frequencies was used (Falush et al., 2003). The program was run with 30,000 burn-ins and 1,000,000 Markov Chain Monte Carlo interactions, and was analyzed in Structure, using the option USEPOPINFO (Falush et al., 2007). This model assumes that each individual originated, with high probability, in the geographic region from which they were sampled, but with a small probability remaining that they were an immigrant or were descended from a recent immigrant. The GENSBACK = 3 parameter was used, to ensure that each individual could not belong to the original population or have a father, grandfather, or great grandfather in another population. The MIGRPRIOR = 0.05 function was used to ensure, with a probability of 95%, that the individuals belonged to the population before mixing. The program was run with 1,000,000 interactions after 30,000 burn-ins.

RESULTS

Genetic diversity

For the 17 accessions of *L. pisonis*, the 11 selected ISSR primers generated 96 clear and reproducible bands, of which 91 were polymorphic (94.79%). Figure 1 shows the profile of the UBC 841 primer bands. Each primer generated 5-11 bands, which ranged from 200 to 2500 bp in size (Table 3).

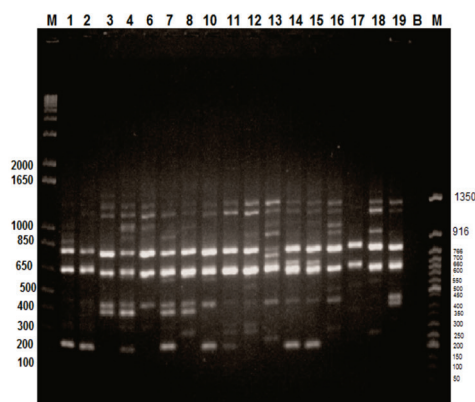


Figure 1. Products of amplification, with the use of UBC 841 ISSR primer, obtained for 19 accessions of sapucaia that are part of the collection of Germplasm of Embrapa Meio-Norte, established in Teresina, PI, Brazil. Lane M (left): DNA ladder of 100kb; Lane M (right): DNA ladder of 50 kb; lanes 1-12: accessions from the population of José de Freitas (Piauí); lanes 13-17: accessions from the population of Teresina (Piauí); lanes 18-19: accessions from the population of Timon (Maranhão); lane B: negative control DNA.

Table 3. Analyzed primers, total amplified loci, total polymorphic loci, and polymorphism percentage generated by ISSR primers.

Primer	Total loci	Polymorphic loci	Polymorphism (%)
UBC 807	11	11	100.00
UBC 810	11	11	100.00
UBC 825	9	9	100.00
UBC 826	11	11	100.00
UBC 836	5	4	80.00
UBC 840	11	9	81.20
UBC 841	8	8	100.00
UBC 842	8	7	87.50
UBC 845	6	6	100.00
UBC 856	11	11	100.00
UBC 864	5	4	80.00
Total	96	91	94.79
Mean	8.7	8.3	

In the studied populations, PPB ranged from 87.50% (José de Freitas) to 20.83% (Timon) with an average of 54.51% (Table 4). At the species level, the average genetic diversity indexes were $h = 0.3110$ and $I = 0.4732$. Greater genetic diversity was found for individuals from José de Freitas. Values for genetic diversity indexes in the Piauí group were much lower than for those in the Maranhão group (Table 4).

Table 4. Indexes of genetic diversity.

Population	<i>h</i>	<i>I</i>	PPB (%)
Species level	0.3110	0.4732	94.79
José de Freitas	0.2969	0.4457	87.50
Altos	0.2158	0.3166	55.21
Timon	0.0863	0.1260	20.83
Piauí	0.3120	0.4734	94.79
Maranhão	0.0863	0.1260	20.83

h: Nei's genetic diversity; *I*: Shannon's index; PPB: percentage of polymorphic loci.

In the grouping obtained by the UPGMA method from Nei's distance (Nei, 1987), the formation of two nodes was observed, where the individuals from José de Freitas and Altos formed a group, showing they are genetically similar. The greatest genetic distance (0.1951) was found between the individuals of Timon and José de Freitas (Table 5).

Table 5. Nei's genetic distance among the occurrence areas.

Population	Genetic distance	Genetic distance
José de Freitas/Altos	0.1336	0.8749
José de Freitas/Timon	0.1951	0.8227
Altos/Timon	0.1733	0.8409

Structure and genetic differentiation

An elevated level of heterozygosity was observed and the G_{ST} index showed that 18.50% of the genetic differentiation is found among the populations. The θ_p coefficient showed that the genetic differentiation between groups (Piauí and Maranhão) was not significant. The value for N_m indicated that there was a migrant every generation in the areas of occurrence (Table 6).

Table 6. Hierarchy structure, genetic differentiation, and gene flow.

	H_T	H_S	H_R	G_{ST}	θ_p	N_m
Estimates	0.2833	0.2309	0.1997	0.1850	0.0892 ^a	1.10
Superior (95%CI)	-	-	-	-	0.1898	-
Inferior (95%CI)	-	-	-	-	-0.0310	-
Standard deviation	0.0212	0.0130	0.0122	-	0.0557	-

H_T : total average heterozygosity; H_S : average heterozygosity within the populations; H_R : average heterozygosity within the regions; G_{ST} : genetic differentiation index among the populations; θ_p : genetic differentiation among the regions; N_m : gene flow; CI: confidence interval. ^aNot significant.

The results obtained by AMOVA revealed that the greatest percentage of genetic variability is distributed within the populations (89.34%). However, there is moderate genetic differentiation (10.66%) among populations (Table 7).

Table 7. AMOVA results for the three populations of *Lecythis pisonis* based on ISSR markers.

Variation source	Sum of squares	Variation components	Variation percentage	Fixation indexes (Φ_{ST})	P
Among populations	37.92	1.71	10.66	0.1066	0.011 ^a
Within populations	86.30	14.38	89.34		
Total	124.22	16.09			

P: Significance test (10,100 permutations); ^asignificant.

Bayesian analysis indicated that the groupings are not strongly correlated with their areas of occurrence, because individuals from the same geographic areas presented a high probability of belonging to more than one group. Among the values tested for K, K = 4 was found to be the most probable using the Structure software.

In the attribution test, considering the best value of K (K = 4), the individuals formed four groups, regardless of their geographic location (Figure 2). Thus, the individuals from José de Freitas tended to form three, genetically distinct groups, where individuals BGS 1 and BGS 2 formed group 1, individuals BGS 7, BGS 10, and BGS 11 formed group two, and group 3 possessed half of its individuals from José de Freitas and half from Timon. Two individuals from Altos (BGS 14 and BGS 15) were different from the others and formed a fourth group. BGS 17, from Altos, provided strong evidence of genome mixing, with up to 83.3% of its genome composed of alleles in frequencies similar to those belonging to groups 1 (37.6%) and 3 (45.7%). Similarly, BGS 16 presented mixed ascendency with 67.9% probability of belonging to group 3.

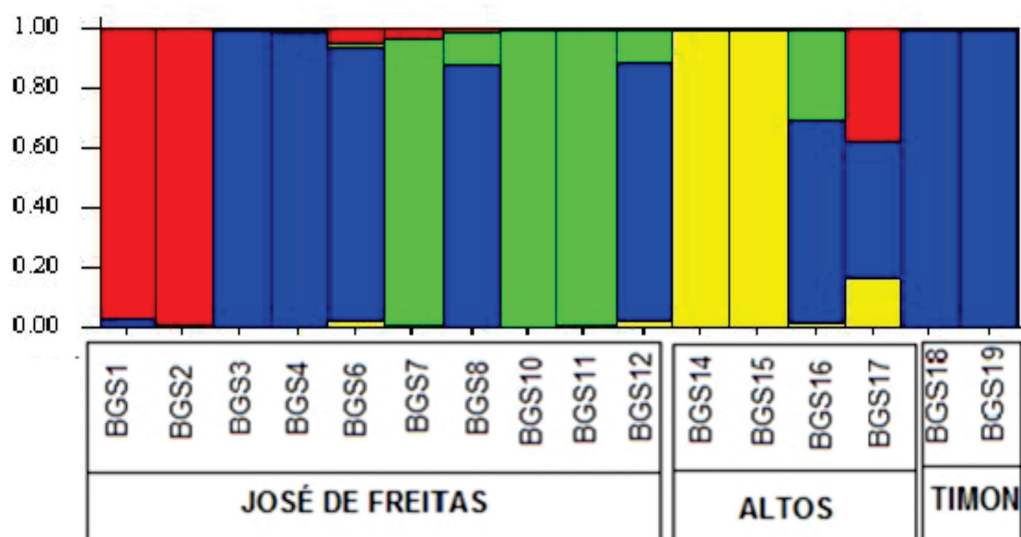


Figure 2. Attribution test for the sapucaia accessions evaluated with K = 4. The vertical colored bars represent the accessions shown. The same color in different population indicates they belong to the same group. Different colors in the same bar indicate the percentage of the genome shared with each group. The values of Q are in the Y axis: participation coefficient of one individual in a certain group, and the populations are in axis X.

According to the migrant test, which was run on the Structure program, all individuals showed a high probability of having pure ascendency in the population previously informed.

Genetic similarity among the accessions

Genetic similarity, evaluated based on the Jaccard coefficient for all 17 accessions of sapucaia, had values ranging from 0.28 to 0.82. The average genetic similarity among the pairs was 0.48. Maximum similarity (0.8276) was observed between accessions BGS 14 and BGS 15, and the lowest similarity (0.2836) was observed between accessions BGS 2 and BGS 4.

DISCUSSION

In this study, ISSR markers revealed that the germplasm collection of sapucaia, kept at Embrapa Meio Norte, possesses a high level of polymorphism (94.79%). This shows there is a high degree of genetic diversity in this species, since the proportion of polymorphic loci is used widely to quantify genetic diversity in plant populations (Borém and Caixeta, 2009). It is likely that the low level of polymorphism found in the Timon occurrence area, MA (20.83%), was related to the sample size, since only two accessions representing this locality were sampled.

The results obtained by Nei's diversity index ($h = 0.31$) and by Shannon's index ($I = 0.47$) confirmed there was high genetic diversity in the sapucaia germplasm. This suggests that the sapucaia AGB from Embrapa Meio-Norte can contribute to the maintenance of genetic variability in this species. In addition, this existing variability may be used in breeding programs. In the sapucaia germplasm, the lowest values for diversity indexes were found in Timon, Maranhão ($h = 0.08$; $I = 0.12$). Genetic studies in living individuals of *Castanea pumila* (Dane et al., 1999) revealed that, although this species of American chestnut was devastated by a severe fungal disease, it possesses a higher level of genetic diversity ($H_E = 0.151$) than observed for Timon sapucaias. However, this is not the true genetic variability of the individuals in this region due to the small number of accessions sampled. This finding highlights the need to obtain new collections so that this population is well represented in the sapucaia AGB.

The grouping obtained by Nei's genetic distance (UPGMA) revealed a higher level of genetic similarity (87.49%) among the populations of José de Freitas and Altos, which were grouped with 81.1% consistency. This may be related to the high geographic proximity between the two areas. Although the Timon region is geographically more distant from the other two, and contains the Parnaíba river as a geographic barrier, the genetic distances found among the Timon population and the other ones was not high. Isolation through geographic distance can be used to explain population subdivisions in plants (Wright, 1965). In this study, there was a correlation between genetic similarity and geographic distance between the sapucaia populations. However, sapucaia is known to be a long-life tree species, and is preferably allogamous, its pollen is spread by bees and its seeds are spread by animals and humans (Mori and Prance, 1990). The results of some studies have shown that genetic distance does not always correlate with geographic distance (Rivas et al., 2013).

The dendrogram of the accessions shows that individuals were grouped regardless of their geographic origin. This is shown by the fact that more than 80% of the accessions belonged to a single group, which contained individuals from all populations. This may be related to the genetic similarities among the populations. However, with 56% genetic similarity it is possible to define several subgroups, showing that there is high genetic differentiation within the populations. Some groupings observed in this study are consistent with those reported in a study of phenotypic characteristics that used the same AGB as used here (Souza et al., 2008). In that study, four groups were formed and there was phenotypic variability among the accessions of sapucaia for most of the characteristics analyzed. However, phenotypic differences are not always representative of genetic differences (Sujii et al., 2013).

Variability found in the sapucaia germplasm can serve as a basis for future breeding. Since the most divergent accessions were BGS 2 and BGS 4, which were both from José de Freitas population, these may be good genitors. Breeders can select individuals that are genetically distant as genitors in order to generate hybrid combinations and maximize the expression of desirable characteristics, which contributes to the amplification of genetic

variance in segregating populations (Bered, 1999).

Indexes of genetic differentiation, together with AMOVA, when revealing that the greatest differentiation occurs within the populations, are in accordance with the description by Hamrick and Godt (1996). Those authors suggested that the reproductive system is important for the formation of genetic structure within a population. In general, species that are mainly allogamous possess a high level of differentiation within populations and a lower level of differentiation among populations. This can explain the results found for the sapucaia populations, because this species is mainly allogamous and bees are a main pollinator (Mori and Prance, 1990).

According to Hamrick (2012), gene flow between individuals within populations and among populations can have large impacts on the distribution of genetic variation. This is because when introducing new variations, gene flow reduces the genetic differentiation among the populations. One or more migrants among populations are sufficient to prevent substantial differentiation. Generally, if $N_m > 1$, then there will be little differentiation among the populations (McDermott and McDonald, 1993). Therefore, a small amount of migration is necessary to prevent significant differentiation among the populations when measured by the fixation index, F_{ST} . However, among the populations, differentiation is not insignificant, even if F_{ST} is very low (Wright, 1965). Furthermore, gene flow as a function of the F_{ST} is considered as historical flow, that is, it does not definitively indicate whether there was gene flow for a given certain reproductive event (Levin, 1984; Slatkin, 1987). Therefore, the significant genetic differentiation of 10.66% (Table 7) found among the sapucaia populations, even having been found what was considered gene flow ($N_m = 1.10$), may be related to random processes in allelic transmission between generations or by differences in the allele frequency among the initial founders of the populations (Hartl and Clark, 2010). According to Levin (1984), colonization of empty habitats by the spread of seeds alters the spatial structure of the species, leading to some genetic differentiation among the populations. This would be true for the populations in the present study, since sapucaia seeds are usually spread by anthropic action.

Consistent with the UPGMA grouping, grouping analysis by the Structure program identified three geographic groups of sapucaia form four independent groups, since some individuals presented a high probability of having originated from a different location. This could be related to the possibility of some individuals being immigrants or descendants of recent immigrants. However, the migrant text strongly grouped all individuals in accordance with their known population, and the probability of individuals from one population having ancestors in another population was low. Thus, the genetic differentiation among individuals within the populations can be explained by the success of the pollen flow.

The observation that there is no significant genetic differentiation among the populations from both states is consistent with the grouping obtained by Jaccard coefficient and with the Bayesian analysis through the Structure program, where the accessions of Maranhão were grouped together with those from Piauí. Thus, elevated gene flow could be limiting the genetic divergence on this hierarchy level. However, the number of individuals sampled in the Timon population was very small, and is insufficient to reveal true genetic differentiation among the two groups.

In conclusion, the results of the present study showed that ISSR markers represent an important tool for the study of genetic diversity in sapucaia. The accessions BGS 2 and BGS 4 constitute a good combination for breeding programs, since they are the more divergent and present significant phenotypic variability for several characteristics.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We are thankful to CAPES (agency for graduate training, Brasília, DF, Brazil) and EMBRAPA, Teresina, PI, Brazil, for the financial support, and for providing the space to perform the experiments.

REFERENCES

- Bered F (1999). Variabilidade genética: ponto de partida para o melhoramento de plantas. In: Genética, para que te quero? (Sacchet AMOF, eds.). UFRGS, Porto Alegre.
- Borém A and Caixeta ET (2009). Marcadores Moleculares. 2nd edn. Universidade Federal de Viçosa, Viçosa.
- Cruz CD and Carneiro PCS (2012). Modelos biométricos aplicados ao melhoramento genético. 4th ed. Universidade Federal de Viçosa, Viçosa.
- Dane F, Hawkins LK and Huang H (1999). Genetic Variation and Population Structure of *Castanea pumila* var. ozarkensis. *J. Am. Soc. Hortic. Sci.* 124: 666-670.
- Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620. <http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x>
- Excoffier L, Laval G and Schneider S (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1: 47-50.
- Faleiro FG (2007). Marcadores genético-moleculares aplicados a programas de conservação e uso de recursos genéticos, Emprapa, Cerrados, Planaltina, DF.
- Falush D, Stephens M and Pritchard JK (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.
- Falush D, Stephens M and Pritchard JK (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7: 574-578. <http://dx.doi.org/10.1111/j.1471-8286.2007.01758.x>
- Freitas FO, Moretzsohn MC and Valls JFM (2007). Genetic variability of Brazilian Indian landraces of *Arachis hypogaea* L. *Genet. Mol. Res.* 6: 675-684.
- Gupta M, Chyi YS, Romero-Severson J and Owen JL (1994). Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. *Theor. Appl. Genet.* 89: 998-1006.
- Hammer O, Harper DAT and Ryan PD (2001). PAST: paleontological statistics software package for education and data analysis. *Paleontol. Electron.* 4: 1-9.
- Hamrick JL (2012). Gene movement in tropical tree populations tropical breeding systems: one and done? *Heredity* (Edinb.) 109: 330-331. <http://dx.doi.org/10.1038/hdy.2012.47>
- Hamrick JL and Godt MJW (1996). Conservation genetics of endemic plant species. In: Conservation genetics, case histories from nature, (Avice JC and Hamrick JL, eds.). Chapman and Hall, New York.
- Hartl DL and Clark AG (2010). Princípios de genética de populações. 4th ed. Artmed, Porto Alegre.
- Levin DA (1984). Immigration in plants: an exercise in the subjunctive. In: Perspectives on plant population ecology, (Dirzo R and Sarvkhan J, eds.). Sunderland, Sinaue.
- Lynch M and Milligan BG (1994). Analysis of population genetic structure with RAPD markers. *Mol. Ecol.* 3: 91-99. <http://dx.doi.org/10.1111/j.1365-294X.1994.tb00109.x>
- Madesis P, Abraham EM, Kalivas A, Ganopoulos I, et al. (2014). Genetic diversity and structure of natural *Dactylis glomerata* L. populations revealed by morphological and microsatellite-based (SSR/ISSR) markers. *Genet. Mol. Res.* 13: 4226-4240. <http://dx.doi.org/10.4238/2014.June.9.8>
- McDermott JM and McDonald BA (1993). Gene flow in plant pathosystems. *Annu. Rev. Phytopathol.* 31: 353-373. <http://dx.doi.org/10.1146/annurev.py.31.090193.002033>
- Miller MP (1997). Tools for Population Genetic Analysis (TFPGA). Version 1.3. Available at [<http://www.marksgeneticssoftware.net/tfpga.htm>]. Accessed February 10, 2014.
- Mori SA and Prance GT (1990). Taxonomy, ecology and economic botany of the Brazil nut (*Bertholletia excelsa* Humb. & Bonpl.: Lecythidaceae). *Adv. Econ. Bot.* 8: 130-150.

- Nei M (1987). Molecular evolutionary genetics. Columbia University Press, New York.
- Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
- Ramalho MAP, Santos JB and Pinto CABP (2012). Genética na agropecuária. 5th eds. Universidade Federal de Lavras, Lavras.
- Revilla J (2002). Plantas úteis da Bacia Amazônica. INPA/ SEBRAE, Manaus.
- Ribeiro RA and Rodrigues FM (2006). Genética da conservação em espécies vegetais do cerrado. *Rev. Cienc. Méd. Biol.* 5: 253-260.
- Rivas LH, Giustina LD, Luz LN, Karsburg IV, et al. (2013). Genetic diversity in natural populations of *Theobroma subincanum* Mart. in the Brazilian Amazon. *Genet. Mol. Res.* 12: 4998-5006. <http://dx.doi.org/10.4238/2013.October.24.12>
- Rohlf FJ (1992). Numerical taxonomy and multivariate analysis system. State University of New York, New York.
- Silva LL, Oliveira JPC, Gomes BS, Sousa-Neto, et al. (2012). Chemical constituents of *Lecythis pisonis* and cytotoxic activity. *Rev. Bras. Farmacogn.* 22: 1140-1144. <http://dx.doi.org/10.1590/S0102-695X2012005000053>
- Slatkin M (1987). Gene flow and the geographic structure of natural populations. *Science* 236: 787-792. <http://dx.doi.org/10.1126/science.3576198>
- Souza IGB, Souza VAB and Lima PSC (2012). Molecular characterization of *Platonia insignis* Mart. ("Bacurizeiro") using inter simple sequence repeat (ISSR) markers. *Mol. Biol. Rep.* 40: 3835-3845. <http://dx.doi.org/10.1007/s11033-012-2462-6>
- Souza VAB, Carvalho MG, Santos KS and Ferreira CS (2008). Características físicas de frutos e amêndoas e características químico-nutricionais de amêndoas de acessos de sapucaia. *Rev. Bras. Frutic.* 30: 946-952. <http://dx.doi.org/10.1590/S0100-29452008000400018>
- Sujii PS, Fernandes ET, Azevedo VCR, Ciampi AY, et al. (2013). Morphological and molecular characteristics do not confirm popular classification of the Brazil nut tree in Acre, Brazil. *Genet. Mol. Res.* 12: 4018-4027. <http://dx.doi.org/10.4238/2013.September.27.3>
- Vijayan K (2005). Inter simple sequence repeat (ISSR) polymorphism and its application in mulberry genome analysis. *Int. J. Indust. Entomol.* 10: 79-86.
- Weir BS and Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370. <http://dx.doi.org/10.2307/2408641>
- Wright S (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19: 395-420. <http://dx.doi.org/10.2307/2406450>
- Yeh FC, Yang RC and Boyle T (1999). POPGENE Version 1.31. Microsoft Window-based freeware for population genetic analysis. Quick user guide. University of Alberta, Centre for International Forestry Research.
- Yu J, Jing ZB and Cheng JM (2014). Genetic diversity and population structure of *Stipa bungeana*, an endemic species in Loess Plateau of China, revealed using combined ISSR and SRAP markers. *Genet. Mol. Res.* 13: 1097-1108. <http://dx.doi.org/10.4238/2014.February.20.11>
- Zietkiewicz E, Rafalski A and Labuda D (1994). Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20: 176-183. <http://dx.doi.org/10.1006/geno.1994.1151>