Universidade Federal Rural do Semi-Árido
Pró-Reitoria de Pesquisa e Pós-Graduação
https://periodicos.ufersa.edu.br/index.php/caatinga
ISSN 1983-2125 (online)

http://dx.doi.org/10.1590/1983-21252025v3811873rc

Scientific Article

Genetic variability of wild yellow plum (*Ximenia americana* L.) based on rapd markers

Variabilidade genética de ameixa selvagem (*Ximenia americana* L.) baseada em marcadores rapd

Elizanilda R. do Rêgo[1]*, Laertty G. de S. Cabral[2], Angela M. dos S. Pessoa[1], João E. Moreira Filho[3], Fabiane R. da C. Batista[4], Mailson M. do Rêgo[4]

[1]Universidade Federal da Paraíba, Areia, PB, Brazil. [2]Department of Experimental Pathophysiology, Universidade de São Paulo, São Paulo, SP, Brazil. [3]Centro Universitário Vale do Salgado, Icó, CE, Brazil. [4]Instituto Nacional do Semiárido, Campina Grande, PB, Brazil.

ABSTRACT - Several ethnobotanical studies highlighted the importance of yellow plum as a useful plant species in the Brazilian Northeast region. In order to domesticate and incorporate a native species in production systems it is necessary to obtain information on its development and genetic variation. This work aimed to analyze the genetic variability of individuals of *X. americana*, accessed by Random Amplified Polymorphic DNA (RAPD) molecular markers. 26 genotypes of *X. americana*, randomly collected at the private reserve Campo das Emas Farm located in the Caturité City, Paraíba State, Brazil, were used in the study. Through the RAPD technique it was possible to assess the genetic variations among native genotypes within a wild plum population. Based on Sokal and Rohlf dissimilarity coefficient it was possible to form four distinct groups using the UPGMA method. The Cophenetic Correlation Coefficient was 0.83, revealing variability in the consistency of the cluster pattern. The genotypes of *X. americana* have genetic diversity detected by RAPD markers, which have shown to be efficient in characterizing individuals. Genotypes 4, 8, 9, 11, 13, 14, 15, 16, 18, 21, 24, and 25 are the most divergent and should be used in conservation strategies and breeding programs.

Keywords: Molecular characterization. Diversity. Genetic resources.

RESUMO - Vários estudos etnobotânicos destacaram a importância da ameixa amarela como espécie vegetal útil no Nordeste brasileiro. Para domesticar e incorporar uma espécie nativa em sistemas produtivos é necessário obter informações sobre seu desenvolvimento e variação genética. Este trabalho teve como objetivo analisar a variabilidade genética de indivíduos de *X. americana*, acessada por marcadores moleculares amplificados ao acaso (RAPD). Foram utilizados 26 genótipos de *X. americana*, coletados na reserva particular Fazenda Canto das Emas, na cidade de Caturité, estado da Paraíba, Brasil. Por meio da técnica RAPD, foi possível avaliar as diferenças genéticas entre genótipos nativos em uma população de ameixa silvestre. Baseado no coeficiente de dissimilaridade de Sokas e Rohlf, foi possível formar quatro grupos distintos pelo método UPGMA. O coeficiente de correlação cofenética foi de 0,83, revelando variabilidade na consistência do padrão de agrupamento. Os genótipos de *X. americana* apresentam diversidade genética detectada por marcadores RAPD, que têm se mostrado eficientes na caracterização de indivíduos. Os genótipos 4, 8, 9, 11, 13, 14, 15, 16, 18, 21, 24, e 25 são os mais divergentes e devem ser utilizados em estratégias de conservação e programas de melhoramento.

Palavras-chave: Caracterização molecular. Diversidade. Recursos genéticos.

Conflict of interest: The authors declare no conflict of interest related to the publication of this manuscript.

Received for publication in: March 24, 2023.
Accepted in: June 27, 2024.

***Corresponding author:**
<elizanilda@cca.ufpb.br>

INTRODUCTION

The Brazilian territory has a diversity of biomes, composed of rich and diverse flora. The Caatinga is one of the main biomes and the main plant formation existing in the Brazilian northeast region. It is considered a unique ecosystem due to its heterogeneity, showing a significant number of rare and/or endemic taxa. All this diversity is represented by several species with different potentials for use, such as native fruits (QUEIROZ et al., 2017).

Among the species with these potentials, there is the wild plum (*Ximenia americana* L.), considered one of the main shrub-tree extracts of the Caatinga, belonging to the Olacaceae family and popularly known as yellow plum and sea lemon (SILVA et al., 2008). It is a cosmopolitan species established in Africa, Central America, India, New Zealand and South America (MOHAMED; FEYISSA, 2020). This species is widely distributed in Brazil, growing in all Brazilian states, in areas of crystalline and sedimentary Caatinga, as well as in Atlantic Forest and Cerrado enclaves. Because of its history of use in traditional communities, *X. americana* is the most widely known species of Olacaceae (LUCENA et al., 2021).

Yellow plum fruits have been analyzed for their nutritive composition (MAIKAI; KOBO; MAIKAI, 2010), phytochemicals, medicinal values

(FERNANDES; BIZERRA, 2020), and bioactive components and fatty acid profiles (BAZEZEW et al., 2021). Additionally, it yields acidic, berry-like wild edible fruits, although not commonly commercialized (MOHAMED; FEYISSA, 2020).

Ethnobotanical studies highlighted the importance of yellow plum as a useful plant species in the Brazilian Northeast (OLIVEIRA et al., 2021; SILVA et al., 2024). In order to domesticate and incorporate a native species in production systems it is necessary to obtain information on its development and genetic variation (COSTA et al., 2011). According to Mora et al. (2009), the characterization of under -utilized fruit species is important for beginning or improving artificial selection processes aimed at developing better quality of edible fruits, that is, increase the species' exploitation, and raising the genetic variability.

Furthermore, few reports are available on the germplasm resources of yellow plum, causing ineffective germplasm collection, conservation, and seedling breeding. As far as our knowledge is concerned, only one study of wild plum genetic diversity is available (MORA et al., 2009). These authors worked with fifteen individuals belonging to a population located at State of Puebla, México. Additionally, varieties proper for edible and medicinal purposes have not yet been identified and developed.

Random Amplified Polymorphic DNA (RAPD) molecular markers is the method that involves the detection of randomly amplified DNA polymorphism, used in the characterization of genotypes in various plant species (SOUZA, 2015). It has great advantages over other methods, due to its simplicity, speed in obtaining data and relatively low cost, in addition to immediate applicability in any type of organism (SILVA et al., 2019).

Thus, this work aimed to analyze the genetic variability of individuals of wild plum (*X. americana*), based on RAPD molecular markers.

MATERIAL AND METHODS

The work was developed at the Laboratory of Biotechnology and Plant Breeding at the Center for Agricultural Sciences, Federal University of Paraíba (CCA-UFPB), Paraíba State, Brazil.

The plant material was constituted of 26 genotypes of wild plum, collected randomly at the private reserve Campo das Emas Farm located in the Caturité City (latitude 7º 23' 16" S and longitude 36º 03' 53.6" W) in Paraíba State, Brazil, with altitude of 405 m and warm semi-arid climate according to the Köppen-Geiger classification: BSh.

Young leaves of the genotypes were collected in the field and transferred to the laboratory on ice for further analysis.

For molecular analyses, 200 mg of young leaf tissue from each accession was subjected to DNA extraction, according to the protocol described by Doyle and Doyle (1990).

The amount of DNA was analyzed on 0.8% agarose gel, aliquots of each DNA sample were applied to the gel wells, and the concentration of the samples was estimated by visually comparing the fluorescence intensity of the DNA bands with those of the known standard. The run was performed in TAE 1X buffer (0.04 M Tris-acetate and 1 mM EDTA) at 80 V, and the gel, ethidium bromide, was photographed under UV light in a Gel Logic 112® molecular imaging camera.

For DNA purification, the samples were incubated in a bath at 37 ºC, with DNA in a proportion of 1:1/2 RNAse (40 ng/mL; v:v), for 12 min. After that, 1:10 5 M NaCl was added, followed by 2/3 of the volume of cold isopropanol, and the samples were kept at -20 °C for 2 h. After the incubation period, the samples were centrifuged for 10 min at 14,000 rpm. The supernatant was removed and the micro tubes were washed twice with 70% and once with 95% ethanol and centrifuged at 14,000 rpm for 2 min per wash. Then, the supernatant was carefully discarded and the micro tubes were kept at room temperature for total ethanol evaporation; the precipitate was suspended in 40 µL of TE buffer.

To evaluate the genetic diversity between the genotypes, the samples were subjected to PCR (Polymerase Chain Reaction), using eighteen RAPD primers (Table 1).

The DNA amplification reaction was carried out with a final volume of 25 µL, with 23 µL of Master Mix (Buffer 10x 2.5 µL + $MgCl_2$ 1.5 µL + dNTP 0.5 µL + primer 2.5 µL + Taq DNA polymerase 0.2 µL + H_2O DDA 15.8 µL) and 2 µL of genomic DNA from the sample, in a concentration of 5 ng/µL. The amplification was performed in a thermocycler (Techne TC-Plus Bibby Scientific Ltd®), with the following programming: 94 °C for 5 min; 30 cycles, being: 94 °C for 5 min; 94 °C for 1 min; 36 °C for 1 min; 72 °C for 2 min, finally the last cycle with 72 °C for 5 min. After amplification, the samples were stored at -20 °C until used.

The amplification products were separated by electrophoresis in a horizontal chamber containing 0.5 TB of TBE buffer for 1h, in 1.5% agarose gel, at 80 V.

For each primer, the number of amplified bands, number of polymorphic loci, number of monomorphic loci, polymorphic information content (PIC), marker index (MI), resolving power (Rp) and expected heterozygosity (He) were estimated. PIC was calculated using the formula described by Serrote et al. (2020): PIC = 1 - (p2 + q2), where PIC is the polymorphic information content of the primer, p is the frequency of the bands present and q is the frequency of the absent bands. For dominant markers, the maximum value for PIC is 0.5 for p = 0.5 (SERROTE et al., 2020). MI and Rp were estimated as proposed by Gilbert et al. (1999), MI = PIC x polymorphic bands and Rp = ∑Ib, where Ib represents the band information, estimated as follows: Ib = 1 - (2*|0.5 - p|), where p is the proportion of accessions containing band I. The percentage of polymorphic loci (PLP) was also estimated. He was estimated considering the Hardy-Weinberg equilibrium (2pq), where p is the frequency of present bands and q is the frequency of absent bands. Allelic diversity analyses were performed using Microsoft® Excel for Mac software, version 15.25.

By reading the bands on the gel, a binary matrix was built (0.1) in which (0) indicated absence and (1), the presence of bands; from which genetic distances between individuals were estimated, generating a dissimilarity matrix, using the Sokal and Rohlf dissimilarity coefficient (1962). The UPGMA cluster analyses were performed using the Genes program (CRUZ, 2016). The principal component analysis (PCA) was done using RStudio 2023.12.1+402 version (ALLAIRE, 2012) and the Devtools (WICKHAM et al., 2022), Factoextra and FactoMineR (KASSAMBARA, 2017a, b). All graphs were created using ggplot2 (WICKHAM et al.,

2016).

RESULTS AND DISCUSSION

The 18 RAPD markers generated a total of 83 amplified bands, 70 polymorphic loci, and 13 monomorphic loci, with an average of 84.37% polymorphic loci, highlighting the genetic diversity among the plants of the *Ximenia americana* L. population (Table 1). The polymorphic information content (PIC) ranged from 0.08 (UB-068) to 0.49 (UB-02). SERROTE et al. (2020) reported that the PIC estimate is equivalent to the observed heterozygosity, when the molecular marker is dominant (RAPD), as used in this study. According to the formula for estimating PIC = 1 - (p2 + q2), heterozygosity will be maximum when it reaches 0.5. Then, PIC values close to 0.5 means the primer is highly informative (SHETE et al., 2000).

Estimates of primer resolving power (Rp) ranged from 0.52 (UB-06) to 6.65 (UB-14) (Table 2). According to Prevost and Wilkinson (1999), Rp has a high correlation with genotype diagnosis. The UB-14 primer was considered the most efficient for evaluating genetic diversity, as it had the highest Rp rate.

The marker index (MI), which is a function of PIC and the number of polymorphic loci, ranged from 0.14 (UB-10) to 4.32 (UB-14) (Table 1). This index (MI) was used by Milbourne et al. (1997) to compare techniques that access genetic diversity; however, the function has not been evaluated as a measure of the ability of the primer to discriminate genotypes, although it is cited in works for this purpose (VALADEZ-MOCTEZUMA et al., 2015). On the other hand, Prevost and Wilkinson (1999) reported that there is little or no correlation between MI and the ability of the primer to distinguish genotypes and that this function is of little practicality and limited value for this purpose.

The expected heterozygosis (He) ranged from 0.09 (UB-06) to 0.50 (UB-02 and UB-14). According to SERROTE et al. (2020), He values are higher than PIC values. The least informative primers could be discarded in future studies. Field visits and research revealed that most of the *X. americana* germplasms were extremely similar in morphology, making it difficult to distinguish between different germplasms. In this concern, DNA molecular markers can quickly identify and describe germplasms that cannot be accurately distinguished phenotypically.

Table 1. RAPD markers sequences, polymorphic loci number (PLN), monomorphic loci number (MLN), polymorphic loci percentage (PLP), polymorphic information content (PIC), resolving power (R_p), marker index (MI) and expected heterozygosity (He) in *Ximenia americana* L.

Marker	Sequence (5'-3')	PLN	MLN	PLP (%)	PIC*	R_p	MI	He
OPAT-01	CAGTGGTTCC	3	1	75	0.25	2.95	0.75	0.26
OPAT-03	GACTGGGAGG	5	0	100	0.28	5.74	1.40	0.36
OPAT-05	ACACCTGCCA	4	1	80	0.41	4.12	1.64	0.42
UB-01	AGACGGCTCC	3	0	100	0.36	3.74	1.08	0.38
UB-02	GTTCGGAACC	7	0	100	0.49	6.06	3.43	0.50
UB-03	GGGCGACTAC	4	1	80	0.37	4.97	2.8	0.39
UB-04	GTGCGCAATG	3	0	100	0.33	3.66	0.99	0.35
UB-05	TCGCATCCAG	4	0	100	0.45	3.96	1.80	0.46
UB-06	CAGAAGCGGA	2	6	25	0.08	0.52	0.16	0.09
UB-07	CACAGCGACA	3	0	100	0.39	5.00	1.17	0.40
UB-08	CAAAGCGCTC	6	1	85.7	0.48	6.26	2.88	0.48
UB-09	TCCCCATCAC	3	0	100	0.24	3.81	0.72	0.25
UB-10	TGCGGGTCCT	1	1	50	0.14	5.23	0.14	0.15
UB-11	CAGGATTCCC	3	0	100	0.38	3.76	1.14	0.39
UB-12	GTGGAGTCAG	5	1	83.3	0.47	4.74	2.35	0.48
UB-13	AAGTCCGCTC	4	0	100	0.42	3.94	1.68	0.43
UB-14	CAGCACTGAC	9	1	90	0.48	6.65	4.32	0.50
UB-15	GACAGGAGGT	1	0	100	0.16	1.26	0.16	0.18
Total		70	13					
Average		3.89	0.72	84.37	0.32	4.24	1.59	0.36

*PIC observed heterozygosity.

The estimated values of genetic dissimilarity ranged from 0.21 to 0.89 (Table 2). The lowest dissimilarity value was observed between the genotypes 25 and 26 (0.21) and the highest values were found between the genotypes 24 and 25 (0.89), 21 and 24 (0.87), 8 and 24 (0.86), 26 and 25 (0.84), 14 and 22 (0.83), and 9 and 24 (0.83) (Table 2). They are also represented in the heat map, where deep red color means high similarity between genotypes (Figure 1). Accession 24 is the most divergent among all the studied accessions, showing deep purple color (Figure 1). Mora et al. (2009), when working with 15 individuals of a wild plum population based on fruit quality traits, found that the greatest distance between

individuals was 0.23 (23%). In a wild population of Passiflora, the distances based on RAPD markers ranged from 0.096 to 0.324. The genetic distances between individuals of another Olacaceae species (*Erythropalum scandens*), based on ISSR markers, ranged from 0.0447 to 0.3420 (YANG et al., 2023). These ranges were smaller than the ones presented in this work. Despite this, the cited authors considered those values as indicative of high diversity between genotypes, which is highly desirable in the selection process for different purposes, such as conservation and indication as parents to form a base population to be used in possible breeding programs.

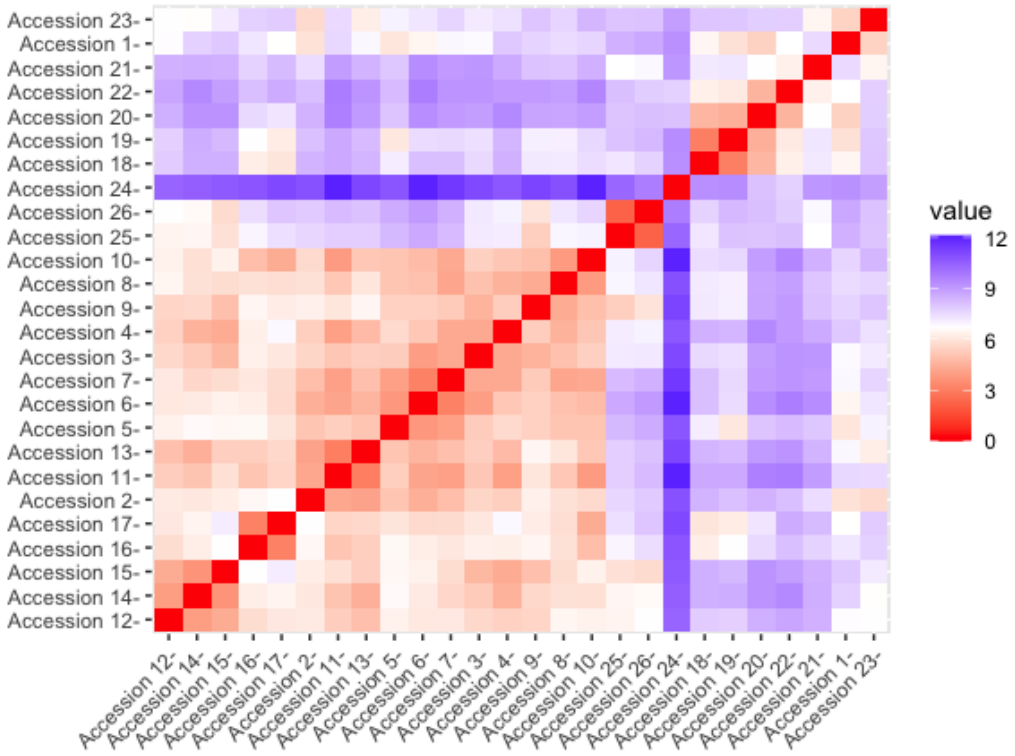

Figure 1. Graphical representation of the dissimilarity matrix between the 26 accessions of *Ximenia americana* L. The color level is proportional to the dissimilarity value between the accessions. Deep red: when the distance between xi and xj = 0. Deep purple: when the distance between xi and xj = 1.

Table 2. Matrix of genetic distance between genotypes of *Ximenia americana* L., based on RAPD markers.

GE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	0.47	0																							
3	0.53	0.50	0																						
4	0.67	0.50	0.39	0																					
5	0.50	0.47	0.49	0.52	0																				
6	0.45	0.34	0.32	0.43	0.31	0																			
7	0.53	0.43	0.40	0.37	0.36	0.29	0																		
8	0.62	0.54	0.48	0.45	0.49	0.41	0.36	0																	
9	0.63	0.55	0.41	0.52	0.46	0.42	0.45	0.41	0																
10	0.59	0.47	0.49	0.43	0.41	0.42	0.38	0.31	0.42	0															
11	0.58	0.33	0.46	0.32	0.46	0.36	0.36	0.45	0.53	0.32	0														
12	0.53	0.57	0.54	0.53	0.56	0.49	0.54	0.64	0.49	0.56	0.43	0													
13	0.56	0.37	0.50	0.47	0.43	0.40	0.47	0.60	0.61	0.44	0.26	0.47	0												
14	0.64	0.57	0.49	0.47	0.60	0.50	0.46	0.58	0.53	0.47	0.44	0.39	0.41	0											
15	0.66	0.56	0.41	0.43	0.56	0.52	0.48	0.54	0.46	0.55	0.52	0.42	0.47	0.37	0										
16	0.62	0.58	0.55	0.56	0.61	0.49	0.51	0.55	0.60	0.41	0.38	0.51	0.46	0.58	0.60	0									
17	0.54	0.58	0.51	0.62	0.53	0.44	0.47	0.51	0.53	0.34	0.45	0.51	0.46	0.52	0.61	0.27	0								
18	0.55	0.73	0.56	0.72	0.59	0.58	0.64	0.61	0.58	0.53	0.64	0.68	0.70	0.74	0.75	0.52	0.46	0							
19	0.50	0.70	0.58	0.70	0.44	0.50	0.57	0.58	0.55	0.59	0.68	0.65	0.68	0.74	0.67	0.58	0.48	0.29	0						
20	0.40	0.65	0.69	0.79	0.64	0.67	0.64	0.74	0.71	0.63	0.71	0.71	0.70	0.74	0.75	0.61	0.52	0.44	0.39	0					
21	0.63	0.57	0.76	0.73	0.61	0.68	0.68	0.67	0.64	0.59	0.65	0.75	0.68	0.74	0.73	0.63	0.64	0.68	0.64	0.62	0				
22	0.54	0.59	0.71	0.71	0.63	0.69	0.63	0.76	0.73	0.69	0.70	0.73	0.73	0.83	0.71	0.65	0.67	0.59	0.55	0.45	0.61	0			
23	0.43	0.47	0.57	0.67	0.60	0.50	0.61	0.69	0.70	0.67	0.58	0.61	0.51	0.58	0.63	0.69	0.63	0.72	0.73	0.72	0.57	0.75	0		
24	0.62	0.78	0.74	0.80	0.75	0.76	0.72	0.86	0.83	0.79	0.81	0.79	0.82	0.81	0.82	0.82	0.74	0.81	0.76	0.68	0.87	0.73	0.79	0	
25	0.69	0.61	0.58	0.64	0.71	0.66	0.67	0.62	0.41	0.49	0.58	0.55	0.66	0.58	0.51	0.59	0.59	0.61	0.71	0.66	0.58	0.72	0.73	0.89	0
26	0.72	0.64	0.55	0.59	0.73	0.68	0.66	0.64	0.49	0.56	0.63	0.60	0.68	0.58	0.48	0.64	0.63	0.69	0.73	0.69	0.62	0.68	0.75	0.84	0.21

GE = Genotypes.

The Cophenetic Correlation Coefficient (CCC) was 0.83, highly significant by the t-test ($p < 0.01$), which reveals variability in the consistency of the cluster pattern. With the degree of distortion, it was 1.41%, confirming the cophenetic correlation coefficient as a good representation of similarity matrices in the form of a dendrogram. When the CCC values are greater than 0.80%, they are considered efficient and correlate the distance matrix and the grouping matrix (SOKAL; ROHLF, 1962).

Stress is a statistical representation that determines the accuracy of graphic projection adjustment, based on the projection of the similarity matrix in the dendrogram, and measures the distortion between the original matrix and that obtained after the construction of the dendrogram (ALBUQUERQUE et al., 2016). In this research, it was considered good, with a value of 11.87%, allowing comparisons to be made between similarity coefficients.

Cluster analysis grouped 26 accessions of *X. americana* L. into four distinct clusters based on Mojena's (1977) criteria (Figure 2). Group I was formed by only one individual (24). This genotype was the most dissimilar among all the evaluated individuals (Figure 2), corroborating the data from the dissimilarity matrix (Table 2). Group II included plants 21 and 23. The individuals 22, 18, 19, 1 and 20 formed Group III. The other 18 individuals formed Group IV. The results of cluster analysis performed by Yang et al. (2023) revealed that the 34 germplasms of *E. scandens*, another Olacaceae species, were classified into three clusters. Likewise, Mora et al. (2009) grouped 15 individuals of a wild plum population analyzed into two groups. Despite the small number of groups formed in these studies, the researchers considered the genetic diversity analyses efficient in discriminating the germplasm and facilitating the rapid advance in breeding.

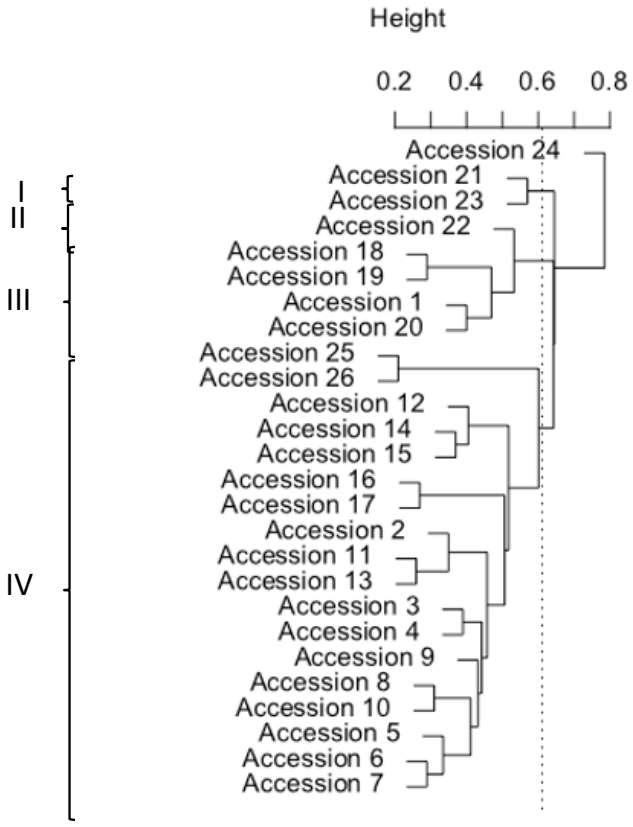

Figure 2. Dendrogram formed from the dissimilarity matrix of 26 accessions of *Ximenia americana* L., through UPGMA cluster analysis. The dotted line represents the cutoff point established by the criteria of Mojena (1977), forming four groups.

The first two principal components explained 50.34% of all genetic variation in the analyzed population and the first six components together explained 81.38% of this variation (Table 3). Similar results were reported by Mora et al. (2009), when working with wild plum., where the first four principal components explained 60% of the total variation. On the other hand, the first two principal components explained only 19.32% of variation for other Olacaceae species (YANG et al., 2023).

A two dimensional representation of the relative positions of each individual result (Figure 3) was generally consistent with UPGMA clustering analysis (Figure 2). It is important highlight that there is no overlapping of plants in the plot, meaning each individual is unique and important to maintain the diversity of this population.

Table 3. Principal component analysis of wild plum (*Ximenia americana*).

Principal Components	Eigenvalues	Cumulative Variance (%)
PC1	8.0326	34.74
PC2	4.0568	50.34
PC3	3.1609	62.50
PC4	1.8816	69.74
PC5	1.6667	74.15
PC6	1.3612	81.38

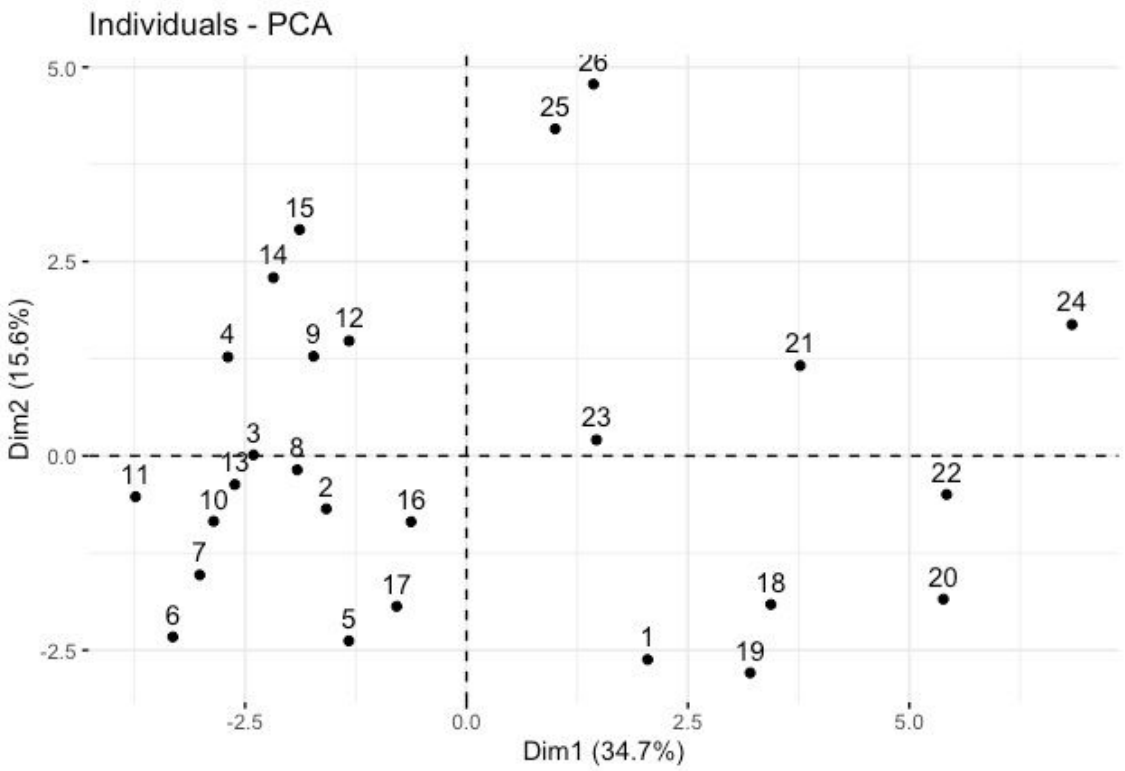

Figure 3. Principal component analyses based on genetic distance. 1 - 26 germplasm of *Ximenia americana.*

CONCLUSION

RAPD markers were efficient in characterizing individuals in this population. Therefore, the plants 4, 8, 9, 11, 13, 14, 15, 16, 18, 21, 24, and 25 are the most divergent and should be used in conservation strategies and in breeding programs.

REFERENCES

ALBUQUERQUE, M. A. et al. Comparação entre coeficientes similaridade uma aplicação em ciências florestais. **Revista Eletrônica Matemática e Estatística em foco**, 4: 102-114, 2016.

ALLAIRE, J. RStudio: integrated development environment for R. **Boston, MA**, 770: 165-171, 2012.

BAZEZEW, A. M. et al. Bioactive composition, free radical scavenging and fatty acid profile of Ximenia americana grown in Ethiopia. **Heliyon**, 7: 1-8 2021.

COSTA, T. S. et al. Diversidade genética de acessos do banco de germoplasma de mangaba em Sergipe. **Pesquisa Agropecuária Brasileira**, 46: 499-508, 2011.

CRUZ, C. D. Genes Software – extended and integrated with the R, Matlab and Selegen. **Acta Scientiarum Agronomy**, 38: 547-552, 2016.

DOYLE, J. J.; DOYLE, J. L. Isolation of plant DNA from fresh tissue. **Focus**, 12: 13-15, 1990.

FERNANDES, P. R. D.; BIZERRA, A. M. C. Avaliação quantitativa de atividades antioxidantes das plantas nativas da Região do Alto Oeste Potiguar/RN. **Research, Society and Development**, 9: 1-20, 2020.

GILBERT, J. E. et al. Developing an appropriate strategy to

assess genetic variability in plant germplasm collections. **Theoretical and Applied Genetics**, 98: 1125-1131, 1999.

KASSAMBARA, A. **Practical guide to cluster analysis in R: Unsupervised machine learning**. London, UK: STHDA, 2017a. 187 p.

KASSAMBARA, A. **Practical guide to principal component methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra**. London, UK: STHDA, 2017b. 169 p.

LUCENA, D. D. S. et al. Olacaceae and Schoepfiaceae in eastern Northeast Brazil. **Rodriguésia**, 72: 1-15, 2021.

MAIKAI, V. A.; KOBO, P. I.; MAIKAI, B. V. O. Antioxidant properties of *Ximenia americana*. **African Journal of Biotechnology**, 9: 7744-7746, 2010.

MILBOURNE, D. et al. Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. **Molecular Breeding**, 3: 127-136, 1997.

MOHAMED, K.; FEYISSA, T. In vitro propagation of Ximenia americana L. from shoot tip explants: a multipurpose medicinal plant. *SINET:* **Ethiopian Journal of Science**, 43: 1 -10, 2020.

MOJENA, R. Hierarchical grouping methods and stopping rules: an evaluation. **The Computer Journal**, 20: 359-363, 1977.

MORA, V. H. F. et al. Characterization of wild plum (Ximenia americana L. var. americana; Olacaceae) fruit growing at Tepexi de Rodríguez, Puebla, Mexico. **Genetic Resources and Crop Evolution**, 56: 719-727, 2009.

OLIVEIRA, F. C. S. et al. The use and diversity of medicinal flora sold at the open market in the city of Oeiras, semiarid region of Piauí, Brazil. **Ethnobotany Research and Applications**, 22: 1-19, 2021.

PREVOST, A.; WILKINSON, M. J. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. **Theoretical and Applied Genetics**, 98: 107-112, 1999.

QUEIROZ, L. P. et al. Diversity and evolution of flowering plants of the Caatinga domain. In: SILVA, J. M. C. et al. (Eds.). **Caatinga**. Cham, Switzerland: Springer, 2017. cap. 2, p. 23-63.

SERROTE, C. M. L. et al. Determining the polymorphism information content of a molecular marker. **Gene**, 726: 144175, 2020.

SHETE, S. et al. On Estimating the Heterozygosity and Polymorphism Information Content Value. **Theoretical Population Biology**, 57: 265-271, 2000.

SILVA, C. N. et al. Variabilidade genética de progênies de maracujá maçã (*Passiflora maliformis* L.) usando marcadores RAPD e ISSR. **Agrotrópica**, 31: 45-52, 2019.

SILVA, G. G. et al. Caracterização do fruto de ameixa silvestre (*Ximenia americana* L.). **Revista Brasileira de Fruticultura**, 30: 311-314, 2008.

SILVA, T. et al. How does bitter taste guide the utilization of medicinal plants in the semiarid region of Brazil?. **Ethnobotany Research and Applications**, 28: 1-14, 2024.

SOKAL, R. R.; ROHLF, F. J. The comparison of dendrograms by objective methods. **Táxon**, 11: 33-40, 1962.

SOUZA, D. C. L. Técnicas moleculares para caracterização e conservação de plantas medicinais e aromáticas: uma revisão. **Revista Brasileira de Plantas Medicinais**, 17: 495-503, 2015.

VALADEZ-MOCTEZUMA, E. et al. Genetic diversity of Opuntia spp. varieties assessed by classical marker tools (RAPD and ISSR). **Plant Systematics and Evolution**, 301: 737-747, 2015.

WICKHAM, H. et al. ggplot2: Elegant Graphics forx' Data Analysis. 2. ed. Houston, Texas, USA: Springer, 2016. 260 p.

WICKHAM, H. et al. Package 'devtools'. **R Package, Version**, 2, 2022, 33 p.

YANG, T. et al. S. Evaluating the genetic diversity of Erythropalum scandens based on using inter-simple sequence repeat markers. **Genetic Resources and Crop Evolution**, 70: 2377-2390, 2023.