# Morphological and molecular characterization of Umbu-Cajá tree in the northeast region of Brazil

Euriann Lopes Marques Yamamoto[1], Ioná Santos Araújo Holanda[2], Patrícia Lígia Dantas de Morais[3], Moaciria de Souza Lemos[4], Emanuela de Oliveira Alves Araújo[4]

[1]Faculdade de Ciências Agrárias, Universidade Federal da Grande Dourados. Rodovia Dourados/Itahum, km 12 - Unidade II. Dourados. Mato Grosso do Sul. Brasil. Caixa Postal 364.
e-mail: marques.agronomia@gmail.com

[2]Departamento de Ciências Vegetais, Universidade Federal Rural do Semi-Árido. Avenida Francisco Mota, 572 - Costa e Silva, Mossoró. Rio Grande do Norte. Brazil. Caixa Postal 474.

[3]Departamento de Biotecnologia. Universidade Federal do Ceará. Fortaleza. Av. da Universidade, 2853 - Benfica Ceará. Brazil. Caixa Postal 6001.

[4]Departamento de Biotecnologia. Universidade Federal do Rio Grande do Norte. Campus Universitário Lagoa Nova, CEP 59078-97. Rio Grande do Norte. Brazil. Caixa Postal 1524.

**ABSTRACT**

The umbu-cajá (*Spondias* sp.) is a fruit specie native of the brazilian semiarid region with high potential for cultivation. The juicy pulp is used in the production of juices and nectars. The aim of this study was to characterize 18 individuals of umbu-cajá tree at three sites located in the Brazilian northeast municipalities of Mossoró/RN and Apodi/RN to identify individuals with superior attributes for fruit processing and/or *in natura* consumption The morphological, physical and physicochemical characterization of 20 fruits of each individuals were performed. Leaves of trees were used for molecular analyses with Inter Simple Sequence Repeats (ISSR) and Random Amplified Polymorphic DNA (RAPD) markers. It was possible to verify that there are phenotypic variability among the fruits of umbu-cajá trees.  The individuals of the same region were grouped in the same group. JZ5, JZ6 and MA6 individuals can be used for pulp processing, while JZ1, JZ2, JZ3 and JZ6 are adequate for *in natura* consumption. The molecular analyses showed polymorphism for the two molecular markers used. The ISSR marker generated the greatest polymorphism among the individuals. However, the RAPD marker made it possible to achieve a more consistent cluster, which is in agreement with the geographical distribution of these individuals. The use of morphological markers together with molecular markers constitutes an important tool for characterizing umbu-cajá individuals.

Keywords: genetic resources, ISSR, RAPD, *Spondias* sp.

## Caracterización morfológica y molecular de árboles de Umbu-cajá en la región nordeste del Brasil

**RESUMEN**

Umbu-cajá (*Spondias* sp.) es una especie de fruto nativa de la región semiárida brasileña con alto potencial de cultivo. La pulpa jugosa se utiliza en la producción de jugos y néctares. El objetivo de este estudio fue caracterizar a 18 individuos de umbu-cajá en tres sitios ubicados en los municipios noreste de Mossoró / RN y Apodi / RN para identificar individuos con atributos superiores para procesamiento de frutas y / o consumo *in natura*. Se realizó la caracterización morfológica, física y fisicoquímica de 20 frutos de cada individuo. Las hojas de los árboles se utilizaron para los análisis moleculares mediante Secuencias Intergénicas Repetidas Simples (ISSR) y la Amplificación aleatoria de ADN polimórfico (RAPD). Se pudo verificar la existencia de variabilidad fenotípica entre los frutos de los árboles umbu-cajá. Los individuos de la misma región se agruparon en el mismo grupo. Los individuos JZ5, JZ6 y MA6 pueden usarse para el procesamiento de pulpa, mientras que JZ1, JZ2, JZ3 y JZ6 son adecuados para el consumo *in natura*. Los análisis moleculares mostraron polimorfismo para los dos marcadores moleculares utilizados. El ISSR generó el mayor polimorfismo entre los individuos. Sin embargo, el marcador RAPD permitió lograr un grupo más consistente, lo que está de acuerdo con la distribución geográfica de estos individuos. El uso de marcadores morfológicos junto con marcadores moleculares constituye una herramienta importante para caracterizar individuos umbu-cajá.

Palabras clave: ISSR, RAPD, recursos genéticos, *Spondias* sp.

## INTRODUCTION

Brazil produces the most fresh fruit in the world in terms of tropical variety, however, none of these fruits are native (Reetz *et al.*, 2015). The umbu-cajá (*Spondias* sp.) is a perennial fruit tree that is propagated asexually, and belonging to the genus *Spondias* (*Anacardiaceae*). It has a succulent fruit and sweet-acidulous pulp (average pulp yield of 55% to 65% and soluble solids around 10 °Brix), which can be used for *in natura* consumption or processed in juices, ice cream and popsicle (Santos *et al.*, 2016).

The fruits of umbu-cajá tree have increased contribution in the Northeast regions agribusiness, mainly in commerce. Nevertheless, there being no commercial plantations and an irregular seasonal fruit production. Despite this, the marketing as a fruit pulp is assured since there is a consumer market that uses the product in the form of juices and ice creams (Brito *et al.*, 2009). Therefore, protecting this tree against the risk of genetic erosion, which can threaten unexploited native species, is a matter of urgency.

Different factors affects the physical and physicochemical characteristics of fruits. Among then stand out genetic constitution, soil and climatic conditions, cultivation and post-harvest treatment are mentioned (Rodrigues *et al.*, 2012). The use of techniques for the characterization of plant material are essential for breeding and cultivar protection programs. Morphological markers are still widely used, but may present problems such us low efficiency in the detection of genetic variability. On the other hand, molecular markers are accurate and detect differences in DNA, as well as distinguish polymorphism between different individuals (Samal *et al.*, 2011).

For these reasons, morphological, physicochemical or molecular characterization are necessary to learn how to maintain the genetic variability that exists among plants of the same species (Santana *et al.*, 2011; Santos *et al.*, 2016). Studies on the morphological characterization of umbu-cajá, although incipient ones are found in the literature (Gondin *et al.*, 2013; Freitas *et al.*, 2015). However, with respect to molecular markers, the scenario is more critical, with rare researches that use molecular markers for the characterization of fruits of umbu-cajá (Santana *et al.*, 2011).

The molecular markers ISSR and RAPD are dominant and based on Polymerase Chain Reaction (PCR) amplifying DNA segments. The advantage of using two molecular markers for analysis is because different regions of the genome of the species will be sampled and their application does not need any prior information about the target sequences on the genome (Samal *et al.*, 2011).

The aim of this study is to characterize, morphological, physicochemical and molecularly, umbu-cajá individual trees that have superior agronomic attributes for processing the fruits and/or *in natura* consumption.

## MATERIAL AND METHODS

### *Plant material*

Fruits of umbu-cajá individual trees were used for morphological characterization and leaves were employed for molecular studies.

### Morphological characterization

Eighteen umbu-cajá individual trees that have superior agronomic attributes for processing the fruits and/or *in natura* consumption were collected in different geographical areas of northeast region of Brazil.

In Apodi and Mossoró sites located in the Oeste Potiguar microregion of Rio Grande do Norte (RN) the trees were selected. Its named as follows: ML1, ML2, ML3, ML4, ML5 and ML6 from the Moacir Lucena settlement, Apodi/RN; JZ1, JZ2, JZ3, JZ4, JZ5 and JZ6 from the Juazeiro II Rural Community, Apodi/RN and MA1, MA2, MA3, MA4, MA5, MA6 from the Coqueiro rural community, Mossoró/RN.

According to the classification of Thornthwaite (1948), the area has a type «E»- arid climate, with average annual rainfall of 590 mm, and average annual temperature of 31 °C.

Twenty ripe fruits were harvested from each individual. For physical analyses, it were evaluated: longitudinal diameter (mm) and transverse diameter (mm). Besides, it was

calculated the longitudinal/transverse diameter ratio and determined the fruit mass (g), pulp mass (g), bark mass (g) and seed mass (g).

For physicochemical analyses, the 20 fruits pulp were processed to get a single sample for determining: pH, total soluble solids (TSS) (%) (directly determined from the juice with a digital refractometer, PR-100 Pallete model, Atago brand) and total titratable acidity (TTA) as recommended by AOAC (2005) (% in citric acid). The TSS/TTA ratio was calculated. The content of vitamin C was determined using Tillmans method according to Strohecker and Henning (1967) in ascorbic acid 100 $g^{-1}$ pulp. The pulp yield (%) was calculated by subtracting bark and seed weight from the total fruit weight and technological index (%), by pulp percentage and TSS percentage ratio.

The averages, standard deviation and coefficient of variation of the evaluated characteristics were estimated using BioEstat 5.0 software. The Unweighted Pair Group Method with Arithmetic Mean (UPGMA) hierarchical clustering model was used to assemble the dendrogram, which makes it possible to verify the degree of similarity between individuals or similar groups. Cluster analyses were performed by Genes software (Cruz, 2013).

**Molecular characterization**

Leaf samples of each umbu-cajá individuals were collected and stored in labeled plastic bags and packed in thermal boxes. DNA was extracted from young and healthy umbu-cajá leaves in order to perform the molecular analyses. Plant DNA extractions were performed immediately after the leaves were collected from the field.

The protocol described by Yamamoto *et al.* (2013) was used to extract the DNA from the umbu-cajá tree leaves.

*ISSR analysis*

The protocol and 25 primers proposed for Santana *et al.* (2011) was used for ISSR analysis.

The amplification reactions were performed in a reaction volume of 12 µl. Each reaction contained 10 ng of template DNA, 10 mM (pH 8.3) Tris-HCl, 50 mM KCl, 2.5 µM $MgCl_2$, 2 mM of each dNTPs, 5 mM RAPD primer, 1 unit *Taq* DNA polymerase and sterile ultra purified water to complete the reaction volume.

The amplification program included pre-denaturation step of 4 min at 94°C followed by 35 cycles of denaturation at 94°C for 40 s, differents temperatures for each primer (48°C, 50°C or 52°C for 40 s) and extension at 72°C for 1 min. The final extension was made at 72°C for 7 min. About 25 primers was selected based on clarity and scorability of banding patters. Amplified products were electrophoresed in a 2.5% (w/v) agarose-1x Tris-borate-EDTA gel stained with ethidium bromide. The gel was then exposed to ultraviolet light and photographed using a digital camera Canon.

*RAPD analysis*

For umbu-cajá molecular characterization 40 arbitrary RAPD primers were used from Operon Technologies Inc. (Alameda, CA, USA). The amplification reactions were performed in a reaction volume of 12 µl. Each reaction contained 10 ng of template DNA, 10 mM (pH 8.3) Tris-HCl, 50 mM KCl, 2.5 µM $MgCl_2$, 2 mM of each dNTPs, 5 mM RAPD primer, 1 unit *Taq* DNA polymerase and sterile ultra purified water to complete the reaction volume.

The amplification program included pre-denaturation step of 5 min at 94°C, amplification reactions were cycled 40 times at 94°C for 30 s, 35°C for 1 min, and 72°C for 1 min in a thermocycler. A final extension was allowed for 7 min for 72°C. About 40 primers was selected based on clarity and scorability of banding patters. Amplified products were electrophoresed in a 1.0% (w/v) agarose-1x Tris-borate-EDTA gel stained with ethidium bromide. The gel was then exposed to ultraviolet light and photographed using a digital camera Canon.

*Data analysis*

A binary matrix was developed based on visual scoring of the bands in the cases of both RAPD and ISSR. The Jaccard similarity coefficient was used to interpret the molecular analyses, which generated an estimated genetic similarity among the individuals, using Genes software (Cruz, 2013).

## RESULTS AND DISCUSSION

### Morphological characterization

A phenotypic variation was observed among the individuals according to the analyses. The MA4 and MA5 had the largest fruits (IBPGR, 1980), whereas, the JZ4 and JZ2, MA2 had the smallest fruit sizes (Table 1). The fruits analyzed were classified as small if <35 mm. The values found for the LD and TD variables are consistent with the values found in the literature (Carvalho *et al.*, 2008; Gondim *et al.*, 2013).

The fruits from the different individuals had an average rating for the LD/TD ratio, indicated that these fruits have a rounded shape (Table 1). As a result, the variability for this characteristic was low. More rounded fruits are the best for pulp processing industries, since cleaning and pulping are simpler tasks (Chitarra and Chitarra, 2005).

There were variations regarding to AFM for all the individuals analyzed (Table 1). Some were classified as small (JZ1, JZ2, JZ3, JZ4, JZ5, JZ6, ML1, ML2, ML3, ML4, ML5, ML6, MA2, MA1 and MA6), medium (MA3) and large (MA4 and MA5), according to Bosco *et al.* (1999). The average found (10.7 g) was lower than that obtained by Carvalho *et al.* (2008) (19.8 g) and Freitas *et al.* (2015) (13.1 g) originating from Rio Grande do Norte.

Table 1. Physical characteristics of fruits collected from 18 individuals of umbu-cajá tree (*Spondias* sp.) in three locations of northeast of Brazil.

| IND. | LD | TD | LD/TD | AFM | APM | ABM | ASEM |
|------|------|------|-------|-------|-------|-------|-------|
| JZ1 | 26.03 | 23.02 | 1.14 | 8.85 | 6.00 | 1.50 | 1.50 |
| JZ2 | 23.17 | 20.59 | 1.14 | 6.75 | 4.50 | 1.00 | 1.00 |
| JZ3 | 27.23 | 24.69 | 1.11 | 10.25 | 7.00 | 2.00 | 1.50 |
| JZ4 | 24.08 | 21.83 | 1.14 | 7.50 | 5.00 | 1.50 | 1.00 |
| JZ5 | 26.61 | 23.65 | 1.10 | 9.25 | 7.00 | 1.50 | 1.50 |
| JZ6 | 28.15 | 25.19 | 1.12 | 10.75 | 8.00 | 1.00 | 1.50 |
| ML1 | 27.80 | 24.82 | 1.11 | 11.50 | 7.00 | 2.50 | 2.00 |
| ML2 | 25.62 | 24.28 | 1.07 | 9.45 | 6.00 | 2.00 | 1.50 |
| ML3 | 26.96 | 24.36 | 1.08 | 10.50 | 6.50 | 2.50 | 1.50 |
| ML4 | 27.40 | 25.81 | 1.06 | 11.00 | 7.00 | 2.50 | 1.50 |
| ML5 | 28.46 | 26.10 | 1.09 | 12.00 | 7.00 | 3.00 | 2.00 |
| ML6 | 27.82 | 26.05 | 1.09 | 11.50 | 7.00 | 2.50 | 2.00 |
| MA1 | 25.07 | 23.53 | 1.05 | 9.00 | 5.00 | 2.00 | 1.50 |
| MA2 | 24.85 | 21.60 | 1.15 | 7.50 | 4.00 | 1.50 | 1.50 |
| MA3 | 28.48 | 26.03 | 1.10 | 13.00 | 9.00 | 2.00 | 2.00 |
| MA4 | 31.05 | 28.83 | 1.09 | 16.00 | 10.50 | 2.50 | 3.00 |
| MA5 | 31.09 | 28.69 | 1.10 | 16.00 | 11.00 | 2.50 | 3.00 |
| MA6 | 27.11 | 26.43 | 1.03 | 12.00 | 9.00 | 1.50 | 2.00 |
| Min. | 23.20 | 20.6 | 1.00 | 6.80 | 4.00 | 1.00 | 1.00 |
| Max. | 31.10 | 28.8 | 1.20 | 16.00 | 11.00 | 3.00 | 3.00 |
| Mean | 27.05 | 24.75 | 1.10 | 10.71 | 7.03 | 1.97 | 1.75 |
| CV(%) | 7.75% | 8.94% | 3.00% | 24.02% | 27.33% | 29.45% | 31.38% |

*Legend: IND individuals, LD longitudinal diameter (mm), TD transversal diameter (mm), AFM average fruit mass (g), APM average pulp mass (g), ABM average bark mass (g) ASEM average seed mass (g)*

Most of the umbu-cajá fruits used were small. However, it is worth considering that, in addition to the AFM variable being greatly influenced by the environment, this characteristic is inherent to its own individual. This is flexible within the typical limits of each individual, when the fruits reach full physiological development (Chitarra and Chitarra, 2005).

For the APM variable, the range obtained was between 4.0 (JZ1) and 11 g (MA5) (Table 1). In these sense, Carvalho *et al.* (2008) found a variation between 6.5 and 19.1 g in umbu-cajá individuals. The fruits of the individuals evaluated have a lesser amount of pulp, may be because these umbu-cajá trees spontaneously occur in the Oeste Potiguar microregion and are not subjected to appropriate agronomical practice.

With respect to the average bark mass (ABM), the values obtained in the analyzed trees showed an average of 1.97 g. The JZ2 had the lowest value (1.0 g) while the ML5 had the highest (3.0 g) (Table 1). Some fruit pulping industries consider the bark as an integral part of the pulp. However, it is important to consider the need to select individuals with lower ABM values, because part of the pulp is connected to the bark during the pulping process, which may increases the volume of waste generated in the process.

Regarding the average of seed mass (ASEM), the variable ranged from 1.0 to 3.0 g and the average was 1.75 g. The JZ4 and JZ2 had the lowest ASEM values, while the MA4 and MA5 had the highest (Table 1). Carvalho *et al.* (2008) found values ranging between 1.3 and 9.3 g, with an average of 6.3 g, which is higher than the value found during our study. However, Santos *et al.* (2010) reported an average value of 1.63 g. Gondin *et al.* (2013) informed that this variable is related to the mass of the fruits. The heavier the fruit, the greater the seed percentage. This statement can be proven by looking at the JZ4, JZ5, JZ6 and MA5 results. The seed mass characteristic is directly influences the pulp yield, which is related to product quality in industrial processing (Gondim *et al.*, 2013).

The values of TTA ranged 0.79% to 1.6%, which indicates a high variance between the evaluated individuals (Table 2). According to the Brazilian Quality and Identity Standard (PIQ) for yellow mombin (*Spondias mombin* L.) - a specie of the same genus that is widely found throughout Brazil (Brazil, 2000) - the minimum value for this variable must be 0.90%. With the exception of JZ1 (0.79%) and JZ2 (0.80%), all the other individuals had TTA values above those recommended for yellow mombin pulp, meaning fruit from the umbu-cajá tree has great potential for its industrialization. Santos *et al.* (2010) found similar TTA values.

The average pH of the analyzed fruits was 2.91, ranged between 2.51 (ML2) and 3.27 (JZ6), where the minimum was 2.2 (Table 2). High pH values (low acid) are preferred for *in natura* consumption (Brazil, 2000). However, this fact highlights a problem that industries face stemming from enzyme activities and the development of microorganisms being encouraged (Santos *et al.*, 2010). At this point, genetically improving the umbu-cajá tree is justified, with the need to select individuals that are suitable for both pulp processing, producing, and marketing *in natura* fruits.

The total soluble solids (TSS) content ranged between 10.1 and 16.0 °Brix (Table 2). All the individuals evaluated reached values above 9.5 °Brix (minimum value required) (Brazil, 2000). The overall TSS average is above that found by other authors such as Lira *et al.* (2005) who reported mean values of 10.1 °Brix and Tiburski *et al.* (2011) that mentions TSS content of 14.9 °Brix. Therefore, there was reasonable variation for this characteristic, with the TSS content variability being a notable aspect among the individuals tested.

Both, the fresh fruit consumer market and the processed fruit market prefer fruits with a high TSS/TTA ratio. These variable ranged between 8.32 (ML5) and 19.61 (JZ2) (Table 2). Analyzing ripe fruits from umbu-cajá trees, different authors reported variations between 3.7 and 10.6 (Carvalho *et al.*, 2008), 9.97 and 13.75 (Gondim *et al.*, 2013), 11.77 and 15.64 (Vieira, 2013). According to the results, JZ2 could be selected due to the higher TSS/TTA ratio. However, fruits from this individual are smaller and lighter, and, despite of low acidity, it cannot be used for processing. This means that the JZ3 and JZ6 are the candidates selected for the process.

Table 2. Physicochemical characteristics of fruits collected from 18 individuals of umbu-cajá tree (*Spondias* sp.) in three locations of northeast of Brazil.

| IND. | TTA | pH | TSS | VITC | TSS/TTA | TI | PY |
|---|---|---|---|---|---|---|---|
| JZ1 | 0.79 | 3.21 | 14.45 | 9.26 | 18.40 | 9.88 | 67.80 |
| JZ2 | 0.80 | 3.15 | 15.75 | 9.26 | 19.61 | 10.62 | 66.70 |
| JZ3 | 0.93 | 3.17 | 15.40 | 9.26 | 16.52 | 10.52 | 68.30 |
| JZ4 | 0.96 | 3.07 | 16.00 | 9.26 | 16.75 | 10.99 | 66.70 |
| JZ5 | 1.12 | 3.12 | 16.00 | 9.26 | 17.97 | 11.68 | 75.70 |
| JZ6 | 0.95 | 3.27 | 16.00 | 9.26 | 16.87 | 11.53 | 74.40 |
| ML1 | 1.50 | 2.65 | 13.65 | 9.50 | 9.08 | 8.25 | 60.90 |
| ML2 | 1.48 | 2.51 | 13.25 | 9.50 | 8.94 | 8.13 | 63.50 |
| ML3 | 1.50 | 2.61 | 12.55 | 7.65 | 8.37 | 7.71 | 61.90 |
| ML4 | 1.24 | 2.85 | 11.60 | 9.50 | 9.34 | 7.17 | 63.60 |
| ML5 | 1.60 | 2.60 | 13.30 | 9.50 | 8.32 | 7.76 | 58.30 |
| ML6 | 1.48 | 2.62 | 13.95 | 7.65 | 9.45 | 8.36 | 60.90 |
| MA1 | 0.96 | 2.62 | 12.50 | 9.50 | 13.05 | 7.15 | 55.60 |
| MA2 | 1.08 | 2.85 | 10.10 | 9.50 | 9.35 | 5.64 | 53.30 |
| MA3 | 1.14 | 2.80 | 15.00 | 9.50 | 13.20 | 10.21 | 69.20 |
| MA4 | 1.27 | 2.95 | 13.45 | 9.50 | 10.58 | 8.91 | 65.60 |
| MA5 | 1.12 | 3.23 | 13.10 | 9.26 | 11.67 | 9.04 | 68.80 |
| MA6 | 1.04 | 3.10 | 13.85 | 9.26 | 13.31 | 10.10 | 75.00 |
| Min. | 0.80 | 2.50 | 10.10 | 7.70 | 8.30 | 5.60 | 53.30 |
| Max. | 1.60 | 3.30 | 16.00 | 9.50 | 19.60 | 11.70 | 75.70 |
| Mean | 1.16 | 2.91 | 13.88 | 9.19 | 12.82 | 9.09 | 65.34 |
| CV(%) | 22.00% | 8.96% | 11.80% | 6.22% | 30.54% | 18.63% | 9.64% |

*Legend: IND individuals, TTA total titratable acidity (in % citric acid), TSS total soluble solids (°Brix), pH, TSS/TTA ratio, VITC vitamin C (in mg/100 g of pulp), TI technology index (%), PY pulp yield (%)*

The VITC levels had a range between 7.65 mg/100 g (ML3 and ML6) and 9.5 mg/100 g of ascorbic acid (ML1, ML2, ML4, ML5, MA1, MA2, MA3 and MA4) (Table 2). The values are consistent with the range found by Carvalho *et al*. (2008) between 3.8 and 16.4 mg/100g and Santos *et al*. (2010), with an average of 8 mg/100 mg. Consuming a 300 ml glass of juice daily would supply around 82% of the requirement of this vitamin for a balanced diet (Brazil, 2005).

The average percentage of the TI ranged from 5.64% for MA2, until 11.68%, for JZ5 (Table 2). Santos *et al*. (2010) reported an average of 6.97%, furthermore, Vieira (2013) presented an average higher than 11.16%. The agricultural industry prefers fruits that have the highest industrial yield indices, because these fruits are most likely to have high levels of soluble solids (Sacramento *et al*., 2007). JZ5 (11.68) and JZ6 (11.53) stood out in terms of their technology index.

Concerning PY, the JZ5 (75.7%), JZ6 (74.4%), and MA6 (75%) had the highest values, which indicates that these individuals have potential for selection (Table 2). The PY results were higher than those informed by Santos *et al*. (2010), with 69.7% and Gondim *et al*. (2013), with an average of 68.53%.

According to the PIQ (Brazil, 2000), except for JZ2, all individuals met the basic requirements set out for the quality of fruit for both fresh consumption and for processing

pulp (Table 3). A minimum pH 2.2, titratable acidity $\geq$ 0.90 g/100 g of citric acid, soluble solids $\geq$ 9.5 and Brix soluble solids/titrable acidity greater than 10.0, indicating the possibility of selecting individuals for most features.

Figure 1 shows the dendrogram of similarity created for the 18 umbu-cajá trees under study based on the 14 analyzed variables. Given a 0.33 cut-off point of similarity in the dendrogram, formations of four groups subdivided into other subgroups were observed.

From the morphological analyzes it was possible to verify that there is variability between fruits of umbu-cajá, with a relative grouping of individuals of the same region in the same group. Group I was characterized by including all the individuals from the Juazeiro II rural community. It contained those that had the highest TI and PY values. Group II contained all the individuals from the Moacir Lucena Settlement and MA1, which had the smallest TSS/TTA ratio according to the physical and chemical characterization of the fruits. It is possible that these individuals were grouped together because it had the same climatic conditions, showing similar values for most of the analyzed variables.

Table 3. Reference to fruit quality of umbu-cajá (*Spondias* sp.) for fresh consumption and pulp processing (Brazil, 2000).

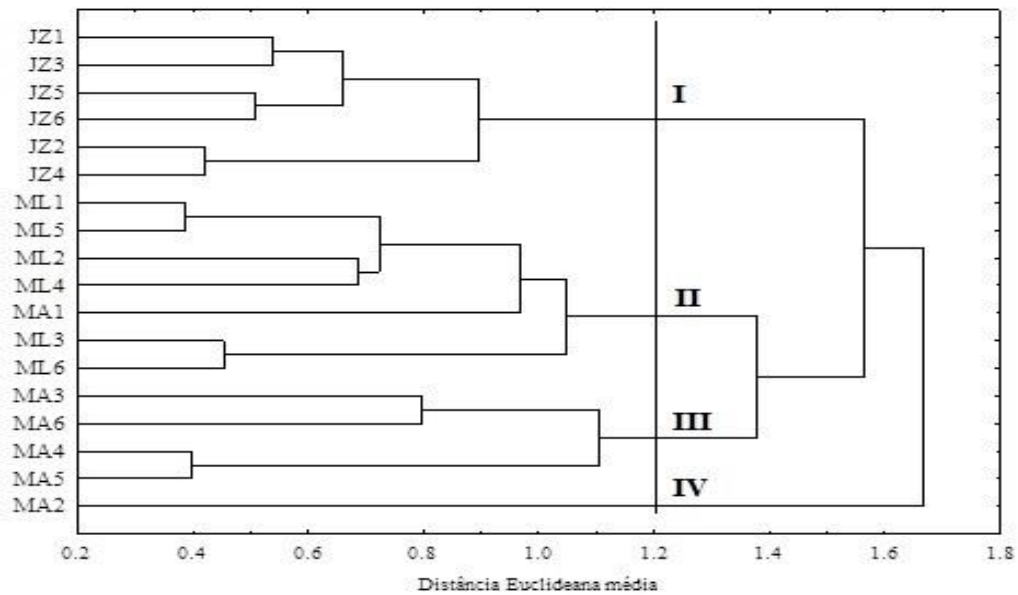|  | Quality Fruit | |
| --- | --- | --- |
|  | Fresh consumption | Pulp processing |
| pH | high | between 2.2 $\geq$ x $\leq$ 3.5 |
| TTA | low | $\geq$ 0.90 g/100g citric acid |
| TSS | high | $\geq$ 9.5 ºBrix |
| TSS/TTA | high | $\geq$ 10 |
| PY | - | $\geq$ 50% |
| LD/TD | Round fruits | Round fruits |



Figure 1. Dendrogram of dissimilarity among 18 individuals of the umbu-cajá tree (*Spondias* sp.) based on 14 fruit variables by the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) hierarchical clustering model from the average standardized euclidean distance. Cophenetic value 0.80.

Group III included four of the six umbu-cajá tree individuals from the Coqueiro rural community, in which geographical locations are visually closer together than the other trees, allocated in groups II and IV, from the same region. Regarding to physical variables, this group contained individuals that had the highest values for LD, TD, AFM, APM, ASEM and TTA that were analyzed. Finally, group IV was made up only by the MA2, which have smaller AFM and lower TSS, TI and PY values. The clustering of single individual in a separate group may suggest that it has different phenotypic characteristics.

Umbu-cajá tree is propagated principally by asexual way because the 90% of fruits lack seeds (Carvalho *et al.*, 2008), which indicates that the individuals located in surrounding regions have a tendency to resemble each other. This fact possibly justifies what occurred with the trees from the Moacir Lucena Settlement as well as the Juazeiro II and Coqueiro rural communities. However, the same thing happened between the MA1 and MA2 individuals, which is not only due to environmental causes, but also likely due to a result of sexual reproduction or natural mutations (Souza *et al.*, 2006).

**Molecular characterization**

*ISSR analysis*

Nineteen of the 25 ISSR primers tested during the study, performed by Santana *et al.* (2011), showed suitable results in terms of producing consistent fragments, a satisfactory amplification profile and variable intensity. Table 4 presents the names of the selected primers, their respective sequences and the number of fragments produced during the molecular analyses with ISSR.

Table 4. Results with Inter Simple Sequence Repeats (ISSR) selected primers for amplifying the umbu-cajá tree (*Spondias* sp.) individuals of three location of northeast of Brazil.

| Primer (5'-3') | Primer sequence* | Tm (°C) | NAF | NPF | %P |
|---|---|---|---|---|---|
| DiCA5'CY | $CY(CA)_8$ | 50 | 8 | 3 | 37.5 |
| DiGA3'C | $(GA)_8C$ | 50 | 7 | 3 | 42.9 |
| DiGA3'RC | $(GA)_8RC$ | 50 | 8 | 6 | 75.0 |
| DiGA3'T | $(GA)_8T$ | 50 | 7 | 3 | 42.9 |
| TriCAC3'RC | $(CAC)_5RC$ | 50 | 6 | 1 | 16.7 |
| TriGTG | $(GTG)_5$ | 50 | 6 | 2 | 33.3 |
| TriGTG3'YC | $(GTG)_5YC$ | 50 | 8 | 6 | 75.0 |
| TriGTG5'CY | $CY(GTG)_5$ | 50 | 8 | 3 | 37.5 |
| TriTGT3'RC | $(TGT)_5RC$ | 50 | 8 | 5 | 62.5 |
| TriTGT5'CY | $CY(TGT)_5$ | 50 | 8 | 3 | 37.5 |
| TriAAC3'RC | $(AAT)_5RC$ | 50 | 9 | 3 | 33.3 |
| TriAAG3'RC | $(AAG)_5RC$ | 50 | 8 | 6 | 75.0 |
| TriACA3'RC | $(ACA)_5RC$ | 48 | 7 | 2 | 28.6 |
| TriAGG3'RC | $(AGG)_5RC$ | 50 | 6 | 1 | 16.7 |
| TriTCA3'RC | $(TCA)_5RC$ | 50 | 11 | 4 | 36.4 |
| TriTCT3'RC | $(TCT)_5RC$ | 50 | 6 | 1 | 16.7 |
| TriTGA3'RC | $(TGA)_5RC$ | 48 | 4 | 1 | 25.0 |
| TriCGA3'RC | $(CGA)_5RC$ | 50 | 9 | 5 | 55.6 |
| TriGAA3'RC | $(GAA)_5RC$ | 50 | 10 | 6 | 60.0 |
| | Total | - | 144 | 64 | |
| | Mean | - | 7.58 | 3.37 | 44.4 |

Legend: *Tm annealing temperature, NAF number of amplified of fragments, NPF number of polymorphic fragments, %P polymorphism percentage. *R=A, G; Y=C, T*

There was a low observed variability among the umbu-cajá tree individuals from the three locations under analysis, which indicates low genotypic variation. This fact was an expected result because the variability within a population is smaller than among populations (Reis and Grattapaglia, 2004).

In this study, we found 44.4% polymorphism and an average of 7.58 polymorphic fragments/primer, varying between 1 and 6 fragments/primer. The polymorphism percentage was lower than reported by Santana *et al.* (2011). The authors studying 25 ISSR primers in umbu-cajá trees and found an 80% polymorphism, which corresponds to an average of 10 fragments for primer, varying between 5 and 16 fragments/primer.

The association between the distances obtained by the Jaccard coefficient and the cophenetic matrix provided a value that indicated a high representativeness (92%), thereby inferring that the clusters made are consistent. The dissimilarity interval ranged from 0.010 to 0.208, with the shortest distance between the ML1 and ML5 and the greatest dissimilarity degree between the JZ1 and MA2 (Figure 2). Based on the cut-off point corresponding to the average genetic

dissimilarity (agd=0.09) (60%), the dendrogram classified the umbu-cajá tree individuals into three groups that were subdivided into several other groups.

The individuals were generally distributed among the different groups, which was probably because it come from the same region (Oeste Potiguar) and presumably the same genetic source, since the umbu-cajá tree only propagates asexually.

The individuals from the Juazeiro rural community were fully clustered (group I), indicating that it have a high level of genetic similarity. The individuals from the Moacir Lucena settlement were distributed in groups I and II. Its shows their genetic similarity with individuals from other locations. The ISSR technique presented an average value 59.6% for the resampling analysis, with a consistency of above 60% for most groups formed. Although small, the genetic variability found can be useful for genetic improvement programs. One of the advantages of genetic improvement programs with perennial species that are vegetative propagate is that there is a possibility to transfer all the genetic variance that exists in a given individual, in view of its method of reproduction.
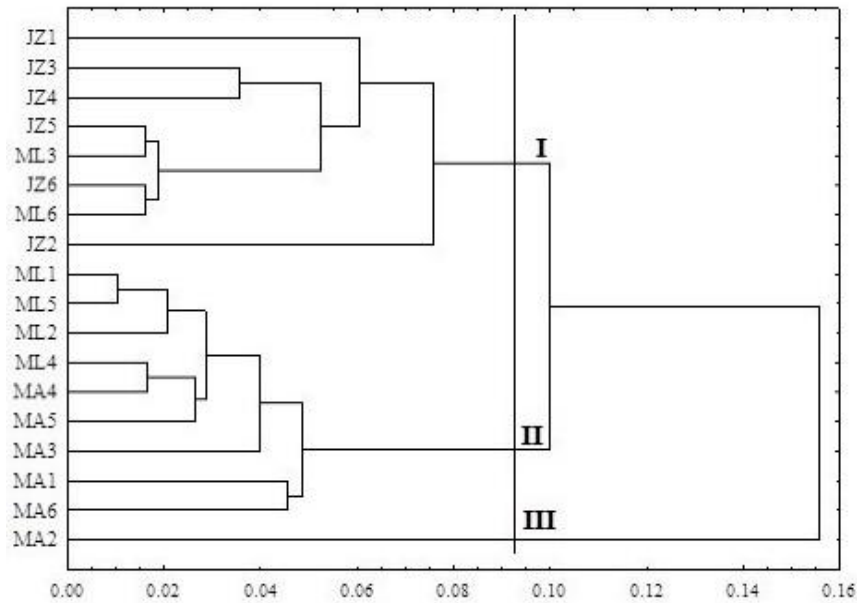


Figure 2. Cluster analysis of 18 umbu-cajá tree individuals (*Spondias* sp.), using the Jaccard coefficient arithmetical complement index to measure dissimilarity, based on 19 Inter Simple Sequence Repeats (ISSR) primers.

Almost all the individuals from the Coqueiro rural community were clustered in group II, with exception of MA2 (group III), which was separated from the others, showing a supposedly different genetic makeup from the others. It is worth mentioning that during the morphological analysis this same individual constituted a separate group, inferring that the analyzed variables could be related gene regions.

Generally, a high level of polymorphism occurs in tropical species when dominant molecular markers are used (Santana *et al.*, 2011). However, this polymorphism can be minimal depending on the population analyzed, which is due to the low genetic similarity that is also evidenced by the phenotypic similarity. Despite the sampled trees were collected from a large area, we can conclude that it were spread around the region from a few individuals. The low level of polymorphism found can be attributed to the low mutation rate among the repetitive sequences of the microsatellites and to the existence of similar regions of the genome (Reddy *et al.*, 2002).

The fact that a single individual was separated into a third group may indicate that its genetic nature is different from the others, suggesting that there was sexual reproduction or a manifestation of natural hybridizations among the umbu-cajá trees, as was affirmed by Souza *et al.* (2006). Then, we can consider the hypothesis that probable hybridizations suggest that there are plants in the natural environments that have intermediate characteristics, when compared to some species of the genus, indicating not only the possibility of natural cross breeding, but also that the incompatibility barriers within the *Spondias* genus can be weak.

*ISSR analysis*

From 40 RAPD primers that were screened, 17 of these presented fragments with greater distinctness and repeatability (Table 5). The low percentage of polymorphic markers shows reduced genetic variability found among the individuals analyzed (25.4%). Lima *et al.* (2011) found 145 fragments with 21 RAPD primers, which 115 where polymorphic in 32 cajá accessions. They found average number of fragments produced per primer of 6.8, 5.4 polymorphic fragments per primer

and 68.8% of genetic similarity. Moreira *et al.* (2007), while analyzing genetic diversity among umbu plants (*Spondias tuberosa* Arr. Câm.), detected low variability based on RAPD markers, with an average of one polymorphic fragment per primer. According to Ferreira and Grattapaglia (1998) an average number of 5 fragments per primer is considered satisfactory.

Based on a 60% cut-off point for genetic dissimilarity, there were two observed groups distributed into 4 other subgroups (Figure 3). On the visual analysis of the dendrogram, some individuals showed a 100% similarity to each other, in two groups, which was true even when using the Jaccard coefficient arithmetical complement, which is recommended for binary data from similar populations, thereby reinforcing the hypothesis forward by Carvalho *et al.* (2008).

Even though they are located in the same geographic region, the greatest dissimilarity among the individuals was ML2 and ML3, given that the ML3 individual was generally the most genetically distant among all the others, constituting a separate group (Figure 3). The results obtained from the RAPD technique suggests that there was a lower variability response within the communities or settlements than between themselves (Reis and Grattapaglia, 2004). Performing molecular analyses with RAPD markers in *Spondias* is a recent approach.

By means of the cophenetic correlation analysis, a 91% association between the distance matrixes, obtained by the Jaccard coefficient, and the cophenetic matrix (represented by the dendrogram) was verified, which indicates that there is a consistency between the clusters generated. Lima *et al.* (2011) demonstrated a cophenetic correlation of 92.4%, close to that found in this study.

For the comparison analysis of the clusters generated from the genetic dissimilarities by the Mantel test, a significance between the ISSR and RAPD molecular markers (r=0.47) and non-significance for the morphological characteristics was verified (data not show). However, as a result of its average magnitude, this indicates that they must be considered complementary measures, not allowing one technique to replace the other.

Table 5. Results with Random Amplified Polymorphic DNA (RAPD) selected primers for amplifying the umbu-cajá tree (*Spondias* sp.) individuals of three location of northeast of Brazil.

| Primer (5'-3') | Primer sequence | Tm (ºC) | NAF | NPF | %P |
|---|---|---|---|---|---|
| OPA-01 | CAGGCCCTTC | 40 | 10 | 5 | 50.0 |
| OPA-09 | GGGTAACGCC | 40 | 9 | 2 | 22.2 |
| OPA-13 | CAGCACCCAC | 40 | 11 | 1 | 9.1 |
| OPA-18 | AGGTGACCGT | 40 | 9 | 3 | 33.3 |
| OPAA-02 | GAGACCAGAC | 40 | 9 | 1 | 11.1 |
| OPAA-03 | TTAGCGCCCC | 40 | 8 | 1 | 12.5 |
| OPAA-04 | AGGACTGCTC | 40 | 8 | 2 | 25.0 |
| OPAA-07 | CTACGCTCAC | 40 | 11 | 2 | 18.2 |
| OPAA-08 | TCCGCAGTAG | 40 | 7 | 4 | 57.1 |
| OPAA-09 | AGATGGGCAG | 40 | 10 | 2 | 20.0 |
| OPAA-12 | GGACCTCTTG | 40 | 9 | 2 | 22.2 |
| OPAA-13 | GAGCGTCGCT | 40 | 5 | 1 | 20.0 |
| OPAA-14 | AACGGGCCAA | 40 | 7 | 3 | 42.9 |
| OPAA-15 | ACGGAAGCCC | 40 | 10 | 2 | 20.0 |
| OPD-02 | GGACCCAACC | 40 | 5 | 2 | 40.0 |
| OPD-05 | TGAGCGGACA | 40 | 6 | 1 | 16.7 |
| OPD-15 | CATCCGTGCT | 40 | 9 | 1 | 11.1 |
| Total | | - | 143 | 35 | |
| Mean | | - | 8.4 | 2.1 | 25.4 |

Legend: *Tm annealing temperature, NAF number of amplified of fragments, NPF number of polymorphic fragments, %P polymorphism percentage*
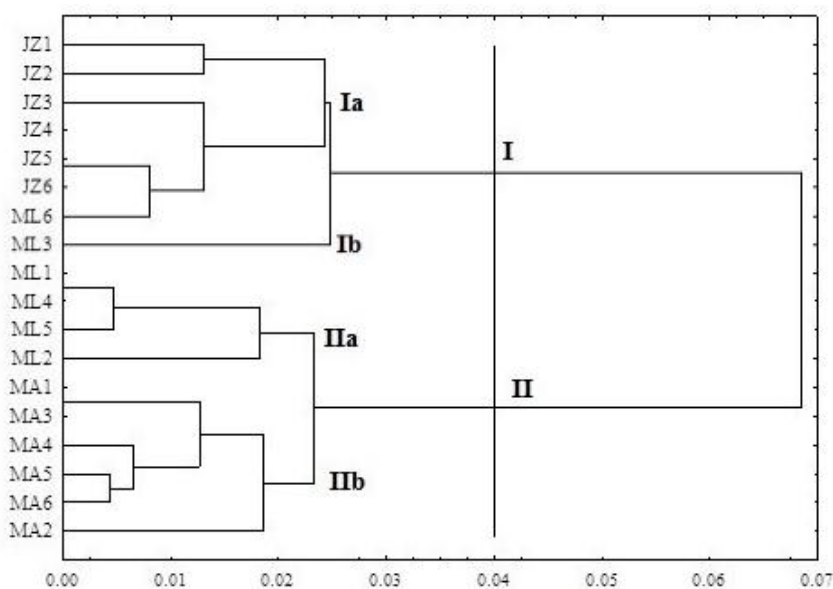


Figure 3. Cluster analysis of 18 umbu-cajá tree individuals (*Spondias* sp.), using the Jaccard coefficient arithmetical complement index to measure dissimilarity, based on 17 Random Amplified Polymorphic DNA (RAPD) primers.

Regarding to polymorphism, both markers had a low percentage, ISSR (45%) and RAPD (25.4%). However, the ISSR technique had a greater polymorphism than the RAPD, thereby confirming its superiority. There was a certain tendency for individuals from same location to cluster together in both molecular markers. A priori, the umbu-cajá tree is allogamous by nature, given the fact that its self-fertilization is hampered due to the stigma above the stamens (Nadia *et al.*, 2007).

Mishra *et al.* (2014) enumerate a number of interfering factors in the association between molecular markers, including reproducibility and quantity of fragments. The low concordance between the genetic matrices may have occurred due to the previously cited reason, since the comparison between the datasets was 64 and 35 polymorphic fragments of ISSR and RAPD, respectively.

Sampling different parts of the genome from the umbu-cajá tree between the molecular markers may have an influence on the correlations found among the genetic matrices. Furthermore, ISSR use longer primers (15-30 mers) as compared to RAPD primers (10 mers), which allows sequential use of high annealing temperature leading to higher stringency (Mishra *et al.,* 2014).

Comparing three dendrograms (morphological, ISSR and RAPD) we can see a certain similarity between the different groups, for example, all the individuals of the Juazeiro community were grouped into a single group in the three dendrogramas. Regarding the other individuals, there was a variation in the positioning of the same individuals, maintaining, however, some correspondence. Thus, the results show a greater discriminant power of the molecular markers in groupings containing individuals of another region in a single group, probably due to the possibility of greater coverage of the genome when compared to the morphological characterization.

By the way, the results indicated the existence of genetic variability among the different individuals. In this sense, the use of morphological markers together with molecular markers constitutes an important tool to the characterization of umbu-cajá for breeding or propagation purposes.

Researches involving in the characterization of umbu-cajá individuals in the state of Rio Grande do Norte should be encouraged so that the genetic variability of the specie is maintained throughout the generations, promoting the genetic improvement with the conservation of genetic resources.

## CONCLUSIONS

Based on a general analysis of the physical, chemical and physicochemical characterization of the 18 individuals, it is possible to infer that the JZ5, JZ6 and MA6 are suitable for industrial pulp processing, while the JZ1, JZ2, JZ3 and JZ6 are suitable for *in natura* consumption and can be part of future researches in genetic breeding of umbu-cajá plants.

The genetic dissimilarity generated by the molecular markers, in addition to the morphological dissimilarity, can provide a guide in the collection of divergent individuals and gene banks can be created to perform genetic improvement studies on umbu-cajá trees.

The ISSR marker generated the greatest polymorphism among the individuals. However, the RAPD marker made it possible to achieve a more consistent cluster, which is in agreement with the geographical distribution of these individuals.

The ISSR and RAPD molecular markers indicated that the interpopulational variability is smaller than the intrapopulational variability.

## REFERENCES

AOAC (2005) Official methods of analysis of the association of official analytical chemists. Arlington, Maryland; ISBN: 0-935584-77-3

Bosco J, Aguiar Filho SPD, Barros RV (1999) Banco ativo de germoplasma de cajá no Estado da Paraíba. In: Embrapa Recursos Genéticos e Biotecnologia (ed). Workshop para Curadores de Banco de Germoplasma de Espécies Frutíferas, pp. 80-85. Embrapa, Brasília

Brazil (2000) Instrução Normativa Nº 01 de 7 de janeiro de 2000. Diário Oficial da República Federativa do Brasil, Brasília DF

Brazil (2005) Resolução RDC nº 269, de 22 de setembro de 2005. Diário Oficial da União, Brasília DF

Brito CH, Costa NP, Batista JL, Silva AB, Lima AN (2009) Tratamento térmico de frutos da cajazeira utilizando vapor d'água, visando ao controle de *Ceratitis capitata* e à qualidade do fruto. Ciência Rural 39 (2):407-411; doi: 10.1590/S0103-84782009000200014

Carvalho PCL, Ritzinger R, Soares Filho WS, Ledo CAS (2008) Características morfológicas, físicas e químicas de frutos de populações de umbu-cajazeira no estado da Bahia. Revista Brasileira de Fruticultura 30 (1):140-147; doi: 10.1590/S0100-29452008000100026

Chitarra MIF, Chitarra AB (2005) Pós-colheita de frutos e hortaliças: fisiologia e manuseio. UFLA, Lavras; ISBN: 85-87692-27-5

Cruz CD (2013) GENES - a software package for analysis in experimental statistics and quantitative genetics. Acta Scientiarum Agronomy 35 (3):271-276; doi: 10.4025/actasciagron.v35i3.21251

Ferreira ME, Grattapaglia D (1998) Introdução ao Uso de Marcadores Moleculares em Análise Genética. Embrapa-Cenargen, Brasília; doi: 10.13140/2.1.3757.8881

Freitas WES, Almeida MLB, Morais PLD, Silva JR (2015) Different 'umbu-cajá' clones fruit quality. Científica Jaboticabal 43 (3):236-240; doi: 10.15361/1984-5529.2015v43n3p236-240

Gondim PJS, Silva SM, Pereira WE, Dantas AL, Chaves Neto JR, Santos LF (2013) Qualidade de frutos de acessos de umbu-cajazeira (*Spondias* sp.). Revista Brasileira de Engenharia Agrícola e Ambiental 17 (11):1217-1221; doi: 10.1590/S1415-43662013001100013

IBPGR (1980) Tropical fruit descriptors. Available in: http://www.bioversityinternational.org/fileadmin/user_upload/online_library/publications/pdfs/289.pdf Accessed: 14/09/2016

Lima ATB, Souza VAB, Gomes RLF, Lima, PSC (2011) Molecular characterization of cajá (*Spondias mombin*) (Anacardiaceae) by RAPD markers. Genetics and Molecular Research 10 (4):2893-2904

Lira JJS, Musser RS, Melo EA, Maciel MIS, Lederman IE, Santos, VF (2005) Caracterização física e físico-química de frutos de cajá-umbu (*Spondias* spp.). Ciência e Tecnologia de Alimentos 25 (4):757-761; doi: 10.1590/S0101-20612005000400021

Mishra KK, Fougat RS, Ballani A, Thakur V, Jha Y, Bora M (2014) Potential and application of molecular markers techniques for plant genome analysis. International Journal of Pure & Applied Bioscience 2 (1): 169-188

Moreira PA, Pimenta MAS, Saturnino HM, Gonçalves NP, Oliveira DA (2007) Variabilidade genética de umbuzeiro na região Norte do Estado de Minas Gerais. Revista Brasileira de Biociências 5 (1):279-281

Nadia TL, Machado IC, Lopes VA (2007) Polinização de *Spondias tuberosa* Arruda (Anacardiaceae) e análise da partilha de polinizadores com *Ziziphus joazeiro* Mart. (Rhamnaceae), espécies frutíferas e endêmicas da caatinga. Revista Brasileira de Botânica 30 (1):89-100; doi: 10.1590/S0100-84042007000100009

Reddy MP, Sarla N, Siddig EA (2002) Inter-simple sequence repeat (ISSR) polymorphism and its application in plant breeding. Euphytica 129 (1):9-17; doi: 10.1023/A:1020691618797

Reetz ER, Kist BB, Santos CE, Carvalho C, Drum M (2015) Anuário Brasileiro da Fruticultura 2014. Editora Gazeta Santa Cruz do Sul, Santa Cruz do Sul

Reis AMM, Grattapaglia D (2004) RAPD variation in a germplasm collection of *Myracrodruon urundeuva* (Anacardiaceae), an endangered tropical tree: recommendations for conservation. Genetic Resources and Crop Evolution 51 (5):529-538; doi: 10.1023/B:GRES.0000024159.48644.e6

Rodrigues CNB, Souza PA, Coelho EL, Souza FX, Freitas RVS (2012) Qualidade de frutos de cajazeira em diferentes estádios de maturação provenientes de clones cultivados no Ceará. Revista Caatinga 25 (3):38-43

Samal KC, Jena RC, Swain SS, Das BK, Chand PK (2011) Evaluation of genetic diversity among commercial cultivars, hybrids and local mango (*Mangifera indica* L.) genotypes of India using cumulative RAPD and ISSR markers. Euphytica 76(5): 1-19; doi: 10.1007/s10681-011-0522-y

Sacramento CK, Matos CB, Souza CN, Barretto WS, Faria JC (2007) Características físicas, físico-químicas e químicas de cajás oriundos de diversos municípios da região sul da Bahia. Magistra 19 (8):283-289; doi: http://dx.doi.org/10.1590/1807-1929/agriambi.v18n08p856–860

Santana IBB, Oliveira EJ, Soares Filho WS, Ritzinger R, Amorim EP, Costa MAPC, Moreira RFC (2011) Variabilidade genética entre acessos de umbu-cajazeira mediante análise de marcadores ISSR. Revista Brasileira de Fruticultura 33 (3):868-876; doi: 10.1590/S0100-29452011005000090

Santos DC, Rocha APT, Gomes JP, Oliveira ENA, Albuquerque EMB, Araujo GT (2016) Storage of 'umbu-cajá' pulp powder produced by lyophilization. Revista Brasileira de Engenharia Agrícola e Ambiental 20 (12):1118-1123; doi: 10.1590/1807-1929/agriambi.v20n12p1118-1123

Santos CB, Cardoso RL, Fonseca AAO, Conceição MN (2010) Caracterização e qualidade de frutos de umbu-cajá (*Spondias tuberosa* x *S. mombin*) provenientes do recôncavo sul da Bahia. Revista Brasileira de Fruticultura 32 (4):1089-1097; doi: 10.1590/S0100-29452011005000015

Souza FX, Costa JTA, Lima RN, Crisostomo JR (2006) Crescimento e desenvolvimento de clones de cajazeira cultivados na chapada do Apodi, Ceará. Revista Brasileira de Fruticultura 28 (3): 414-420; doi: 10.1590/S0100-29452006000300017

Strohecker R, Henning HM (1967) Análisis de vitaminas: métodos comprobados. Editorial Paz Montalvo, Paz Montalvo

Tiburski JH, Rosenthal A, Deliza R, Godoy RLO, Pacheco S (2011) Nutritional properties of yellow mombin (*Spondias mombin* L.) pulp. Food Research International 44 (7): 2326–2331; doi: 10.1016/j.foodres.2011.03.037

Thornthwaite CW (1948) An approach towards a rational classification of climate. Geographical Review 38(1): 55-94; doi: 10.2307/210739

Vieira AR (2013) Propagação assexuada e qualidade de frutos de genótipos de umbu-cajazeira da mesorregião Centro-Sul do Ceará. Tese de Doutorado. Universidade Federal Rural do Semi-Árido, Mossoró, Brasil

Yamamoto ELM, Lemos MS, Maniçoba FE, Pinto KNC, Maia AKS, Holanda ISA, Antonio RP (2013) Otimização de protocolo para extração de DNA em umbu-cajazeira. In: Universidade Católica de Pernambuco (ed) 1ro Congresso Internacional de Ciências Biológica, Recife, 11-14/11/2013, pp.1-13. Universidade Católica de Pernambuco, Recife