CHARACTERIZATION AND GENETIC DIVERSITY OF THE CASHEW NATURAL POPULATION THROUGH MORPHOLOGICAL DESCRIPTORS

Marcones Ferreira Costa[1], Luana Souza Silva[1], Luana Alves Carneiro[1], Maria Fernanda da Costa Gomes[2], Michelli Ferreira dos Santos[3], Sérgio Emílio dos Santos Valente[4]

ABSTRACT – The characterization and the evaluation of genetic relationships in natural populations are extremely important for the management and conservation of genetic diversity. From this perspective, this work aimed to characterize and estimate the genetic diversity of 34 cashew trees from a native and natural population of the rural area of Floriano-Piauí, Brazil, through seven morphological variables related to leaf and inflorescence. Data were subjected to multivariate analysis of main and cluster components. The descriptors inflorescence length and inflorescence width were the major contributors to the genetic variation observed in the population. The biplot graph obtained by the two-dimensional projection of the first two main components showed dispersion of accesses within the population, which was confirmed by cluster analysis by the UPGMA method and by the Tocher optimization method. As for genetic diversity, the results show that accesses are organized into two distinct groups, indicating that there is interpopulation phenotypic differentiation. This information is very important for the better management and conservation of natural populations and may support future pre-breeding programs.

Keywords: cerrado region, genetic parameters, multivariate analisys, phenotypic differentiation.

CARACTERIZAÇÃO E DIVERSIDADE GENÉTICA DE POPULAÇÃO NATURAL DE CAJUEIRO POR MEIO DE DESCRITORES MORFOLÓGICOS

RESUMO – A caracterização e avaliação das relações genéticas em populações naturais são extremamente importantes para o manejo e conservação da diversidade genética. Nesta perspectiva, este trabalho teve por objetivo caracterizar e estimar a diversidade genética de 34 cajueiros de uma população nativa e natural da zona rural da cidade de Floriano-Piauí, Brasil, por meio de sete variáveis morfológicas, relacionadas à folha e inflorescência. Os dados foram submetidos a análises multivariadas de componentes principais e de agrupamento. Os descritores comprimento da inflorescência e largura da inflorescência foram os que mais contribuíram para a variação genética observada na população. O gráfico biplot obtido pela projeção bidimensional dos dois primeiros componentes principais demonstrou dispersão dos acessos dentro da população, o que foi confirmado pelas análises de agrupamento pelo método UPGMA e pelo método de otimização de Tocher. Quanto à diversidade genética, os resultados demonstram que os acessos estão organizados em dois grupos distintos, indicando que há diferenciação fenotípica interpopulacional. *Estas informações serão muito importantes para o melhor manejo e conservação de populações naturais, podendo vir a subsidiar futuros programas de pré-melhoramento.*

Palavras chave: análise multivariadas, cerrado, diversidade fenotípica, parâmetros genéticos.

---

[1] Universidade Federal do Piauí, Campus Amílcar Ferreira Sobral Campus, Brazil. E-mail: marconescosta@ufpi.edu.br
[2] Universidade Estadual do Piauí, Campus Ariston Dias Lima, Brazil.
[3] Universidade Federal do Piauí, Campus Senador Helvídio Nunes de Barros, Brazil.
[4] Universidade Federal do Piauí, Campus Ministro Petrônio Portela, Brazil.

INTRODUCTION

Cashew (*Anacardium occidentale* L.) is a tropical plant native to northeastern Brazil. It is a diploid species (2n = 42), which was introduced in West Africa and India in the 16th and early 17th centuries by the Portuguese, where it spread spontaneously forming wild and domesticated populations (Archak et al., 2009; Aliyu 2012; Adeigbe et al., 2015). Its origin extends from Mexico to northeast Brazil (Kouakou et al., 2018).

Cashew has a significant agronomic role worldwide, especially regarding the use of chestnuts (fruit) and of the pseudofruit (peduncle). The pseudofruits can be consumed fresh or used to make candy or pulp for juices and other beverages, and the process waste is used as a component of animal food (Santos et al., 2019).

The chestnut and the pseudofruit have different compounds, such as flavonoids, carotenoids and vitamins (C and A) and minerals such as calcium, iron and phosphorus. In addition, the extract has been used as an alternative medicine due to its anesthetic and bactericidal properties (Kulis et al., 2012; Silva et al., 2013).

Cashew cultivation is an important socioeconomic activity in northeastern Brazil, especially in the states of Ceará, Piauí, and Rio Grande do Norte, since its fruiting period is during the off-season of traditional crops, for example, rice and beans. Cashew planting is an important alternative source of income and food for rural communities.

The cashew tree shows great potential as a source of genes for the genetic improvement of cultivated species belonging to the genus (Maia et al., 2019). The use of morphological descriptors stands out as one of the main instruments for the characterization and quantification of genetic diversity, the initial step for breeding studies. This characterization has been performed through a set of botanic, morphological, and agronomic descriptors.

The availability of genetic resources necessarily involves the characterization and evaluation of germplasm, which aim to describe accesses through characteristics of interest such as productivity, fruit mass, pulp thickness, and number of seeds per fruit. Thus, from these data, and with the use of genetic-statistical methodologies, it is possible to estimate the genetic diversity of a population through morphological or phenotypic descriptors and quantify their potential use in conservation and pre-breeding programs (Marim et al., 2010).

To this end, morphological descriptors become useful tools, providing information to quantify the genetic diversity among individuals through the use of multivariate techniques, such as cluster and main components analysis, making it possible to determine the importance of the variables studied for the knowledge of genetic variability. There are several studies that aimed to compare the genetic diversity of cashew germplasm in different regions using morphological markers (Aliyu and Awopetu, 2007; Chipojola, 2009; Vieira et al., 2014; Borges, 2015; Carneiro et al., 2019; Andrade et al., 2019).

Understanding the genetic basis of morphological characteristics is relevant to the conservation, characterization and the use of genetic resource programs. Thus, the characterization and study of genetic diversity of a natural cashew population in a rural area of Floriano-Piauí, Brazil, was conducted, based on multi-categorical descriptors of leaves and flowers, in order to evaluate the relation and importance of these diferente descriptors for characterization and utilization of genetic resources programs in cashew trees.

MATERIAL AND METHODS

For this study, the genetic divergence of 34 cashew tree accessions were characterized and estimated, named G-1 to G-34 according to the order of collection in a native and natural population of the rural area of Floriano-Piauí, Brazil (S 06 ° 46 ‘24 “South and W43 ° 00’ 43”), taking into consideration a minimum distance of about 3 km between the plants. The climate of the region is tropical semi-arid, hot and humid, with an average temperature of 28.1ºC. Its vegetation is typical of a transition area with the presence of species from Caatinga and mainly from Cerrado (Lopes et al., 2016).

Seven quantitative descriptors (categorical) were evaluated in each plant: Inflorescence length (CIN), taken randomly in the most intense flowering period, with the aid of a scale graduated in centimeters; Maximum inflorescence width (LIN), maximum distance between the inflorescence branches, with the aid of a scale graduated in centimeters; Total number of inflorescence branches (NRIN), number of inflorescences in relation to the main axis; Petiole length (PC), with the aid of a scale graduated in centimeters; Lamina length (CL) or leaf blade length, with the aid of a scale graduated in centimeters; Lamina Width (LL) or leaf blade width, with the aid of a scale graduated in centimeters, and Lamina Index (IF), the ratio between length and width of the leaf.

These characteristics are important, as they vary according to the spatial and temporal distribution of the species, light availability and plant age. The samples were evaluated under field conditions and the definition of the phenotypic class of each descriptor was based on the evaluation of 10 leaves and flowers of each plant. The collection points were established randomly, considering the places where the species could be found.

Regarding the data analysis, the normality of the data was previously verified by the Shapiro-Wilk test, where it was observed that the sample presented a non-normal behavior. A Spearman correlation with 5% significance was used between the estimates of genetic dissimilarities obtained based on all morphological descriptors.

Genetic dissimilarities among the 34 cashew tree accessions were calculated based on all morphological descriptors. Estimates were based on the average Euclidean distance matrix. Based on the genetic dissimilarity matrices, dendrogram accession grouping analyzes were performed using the Unweighted Pair-Group Method using Arithmetic Averages (UPGMA) criterion and the Tocher optimization grouping method. Graphic dispersion was based on multidimensional scales through main component analysis (PCA). Statistical analyses were performed with the aid of R (R Development Core Team, 2018) and Genes (Cruz, 2013) programs.

RESULTS AND DISCUSSION

The morphological characterization based on the variables selected for this study allowed the distinction of population accessions, allowing them to be grouped according to the level of genetic similarity, so that there is greater similarities between accessions of the same group and greater dissimilarities between accessions of distinct groups.

The cashew trees were classified according to specification provided by the Ministry of Agriculture, Livestock and Supply for the characterization of mango (MAPA, 2011), a plant belonging to the same taxonomic family of cashew, the Anacardiaceae. The variables petiole length, Lamina length and Lamina width varied from 4.3 to 9.7 cm, 10.3 to 15.2 cm and 7.1 to 9.1 cm, respectively.

According to the MAPA criterion, the leaf petiole length can be classified as short (≤ 4.0 cm), medium (4.0 <x ≤ 8.0 cm) and long (> 8.0 cm). 65.7% of the accessions of the population had medium petiole leave length, 23.5% with large petioles and 10.8% with short petioles. For the lamina length, all samples presented short lengths (x ≤ 20.0 cm). No cashew trees in the population had narrow leaves, as the samples had lamina width values greater than 4.0 cm. All leaves were classified as broad (> 5.0 cm).

The lamina index represents the length and width relation, all cashew trees evaluated showed lamina index values ranging from 1.0 to 2.0. According to Costa et al. (2016), plant species with lamina index values greater than one (IF> 1.0) invest more in the longitudinal growth of leaves than in width, in response to suitability to environmental conditions, since this characteristic is strongly related to maximizing the use of light energy, essential for the photosynthesis process.

The coefficients of variation of the morphological descriptors of the sampled cashew population reveal that the variables lamina width, length and index are more dispersed around the averages (CV>25%). While the lamina width, inflorescence length, inflorescence width, and total number of inflorescence branches had a low coefficient (CV<10%), which indicates a low dispersion of morphological variables around the mean for all individuals (Table 1).

Table 1 - Mean, standard deviation and coefficient of variation (CV%) of the morphological variables for the cashew trees of the natural population analyzed.

Variable	Mean	Standard deviation	CV%
Inflorescence Length (cm)	14.7	0.06	4.49
Inflorescence Maximum Width (cm)	14.4	0.07	4.45
Total number of inflorescence branches	10	3.15	3.13
Petiole Length (cm)	7.12	0.013	18.3
Lamina length (cm)	9.5	0.242	25.4
Lamina width (cm)	7.7	0.07	25.4
Lamina Index (cm)	1.68	0.042	26.28

All cashew trees in the study had short inflorescences with their length ranging from 14 to 22 cm. Regarding the maximum inflorescence width all accessions were classified as small inflorescences (15.0 < x ≤ 20.0 cm). The total number of inflorescence branches was low, since the accessions presented less than 15 primary branches per inflorescence.

The Spearman's correlation was used to assess the association between the variables. Most results showed a weak correlation, which implies that these variables are not highly correlated (Table 2). The same happened between the inflorescence length and maximum width, that presented one of the highest correlation values (r = 0.682), but without significance (P>0.05). The correlation value between lamina width and index was strong, significant and positive (r = 0.607), indicating that the widest the leaf the larger the index.

Table 2 - Spearman correlation coefficients between descriptors Inflorescence length (CIN), Maximum inflorescence width (LIN), Total number of inflorescence branches (NRIN), Petiole length (PC), Lamina length (CL), Lamina Width (LL) and Lamina Index (IF).

	CIN	LIN	NRIN	CP	CL	LL	IF
CIN		0.682^{ns}	-0.101^{*}	-0.278*	-0.090^{*}	-0.060*	-0.375^{ns}
LIN			0.0407^{*}	0.037*	-0.016^{*}	-0.272^{*}	-0.202^{*}
NRIN				-0.036*	-0.257^{*}	-0.072^{*}	0.0879^{*}
CP					0.441^{ns}	0.032*	0.254*
CL						0.091^{*}	0.100^{*}
LL							0.607^{*}
IF							

* = P <0.05; ns = not significant.

By establishing the section point in the dendrogram it is possible to observe the formation of two groups, the same amount observed in the Tocher optimization method (Table 3). The Tocher optimization method is concomitantly used with the hierarchical UPGMA method, providing correspondence for element allocation in clusters (Cantelli et al., 2016).

From the estimates obtained by the Euclidean distance matrix it can be observed that there is genetic variability among the population accessions. The minimum distance was 0.18 between the G-20 and G-22 cashew trees, considered as the most similar, however the estimated maximum distance was 2.83 between the G-22 and G-3 cashew trees.

The Tocher grouping organized the cashew trees into two groups considering their estimated characteristics (Table 3). Group I had 97.11% of the plants and group II gathered only the G-03 cashew tree. Groups formed by a single individual suggest that these are the most divergent from the others (Vieira et al., 2005). The distance between the representatives of group I was 1.26, while the distance between group I and II was 2.37. This indicates that the genetic pool for this study was small, which contributed to the low genetic diversity observed, with relatively homogeneous distribution among the cashew trees of this population.

The UPGMA cluster analysis for cashew plants is shown in Figure 1. The performance of cluster analysis, based on morphological descriptors, was revealed through the cophenetic correlation coefficient (ccc). According to Rohlf (1970), cophenetic correlation values lower than 70% indicate inadequacy of the clustering method. In the present study, the dendrogram cophenetic correlation was 87.2%; indicating that the UPGMA method was adequate for the grouping of accessions of this population.

Table 3 - Clustering of the 34 cashew plants obtained from the Tocher optimization method through the analyzed morphological variables

Groups	Cashew trees evaluated
I	G-24, G-25, G-5, G-31, G-33, G-34, G-15, G-17, G-28, G-18, G-26, G-12, G-13, G-22, G-6, G-27, G-1, G-20, G-4, G-23, G-16, G-19, G-21, G-10, G-29, G-2, G-11, G-8, G-9, G-14, G-7, G-30, G-32
II	G-3

Tocher optimization method, along with other clustering techniques can be used to compare different types of multivariate analysis. The choice of the method to be used depends on the desired precision, complexity of analysis and data collection process and there is no defined parameter of choice for the study of genetic divergence for a group of accessions or genotypes (Cruz et al., 2012; Hamawaki et al., 2012).

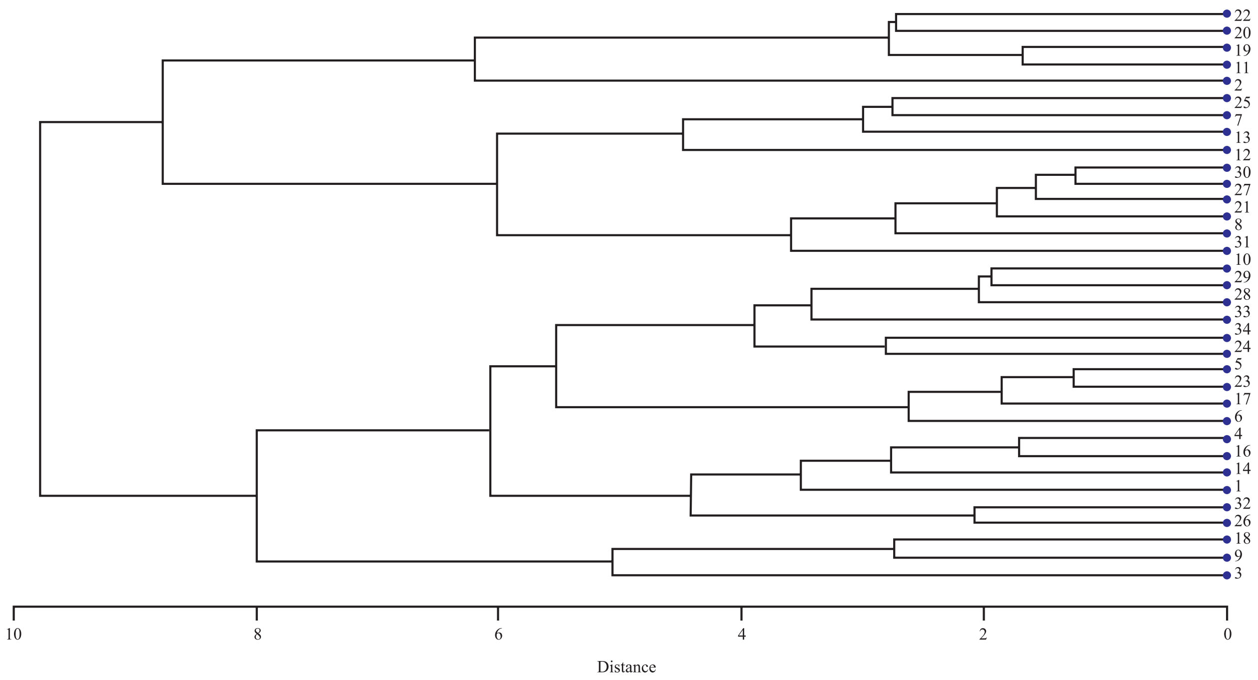

Figure 1 - Dendrogram of similarity between accessions of the natural cashew population obtained by the UPGMA method.

Group I of the dendrogram was formed by the cashew trees G-1, G-3, G-4, G-5, G-6, G-9, G-14, G-15, G-16, G-17, G -18, G-23, G-24, G-26, G-28, G-29, G-32, G-33 and G-34, the rest were grouped in group 2 (Figure 1). The cophenetic correlation coefficient (ccc) allowed ratifying the efficiency of the UPGMA method.

An analysis of the main components showed that the first three axes explained 70.33% of the total variation (Table 4). The first and second main components contributed respectively to 28.62% and 27.12% for the observed variability; while the third represented 14.59%. According to the Kaiser criteria (1960), the first three main components were selected, since their values were greater than 1.0. According to Cruz et al. (2011) the ideal is that the first two main components concentrate the largest amount of data variance so that there is divergence between genotype groups.

However, there are studies in which main component analysis have been shown to be effective in assessing genetic diversity, even when the 80% limit between the first two main components is not reached (ChabiSika et al., 2015; Kouakou et al., 2018). Vieira et al. (2014), using the main component method to reveal the main leaf shape variation trends of two species of the gender *Anacardium*, indicated that 11 descriptors are needed to express more than 80% of the variation and yet, the scores of these components made it possible to distinguish the groups from the leaf morphometric data.

Table 4 - Estimation of the values associated with the main components and their percentage and cumulative variance of the morphological characters in cashew trees

Main component	AV	AV (%)	% Accumulated
MC1	2.57	28.62	28.62
MC2	2.44	27.12	55.74
MC3	1.31	14.59	70.33
MC4	0.87	9.62	79.95
MC5	0.75	8.37	88.32
MC6	0.75	8.43	96.75
MC7	0.30	3.25	100.0

The descriptors that most contributed to the variability of the first main component (MC1) were inflorescence length and width. For the second main component (MC2) the variables that contributed to the variation were total number of inflorescence branches and lamina length. Meanwhile, the total number of inflorescence branches and Lamina Index contributed to the variation observed in the third component (MC3).

The Biplot graph (Figure 2), with the relative positions of the accessions and variables, considering the first two components, shows that the cashew trees G-3, G-8, G-9, G-10, G-13, G-21, G-25 were grouped together because they presented the better data for inflorescence length and width, while G-4, G-6, G-17, G-23, G-28, G-32 and G -33 were grouped by presenting the highest values for total number of inflorescence branches, lamina index and width.

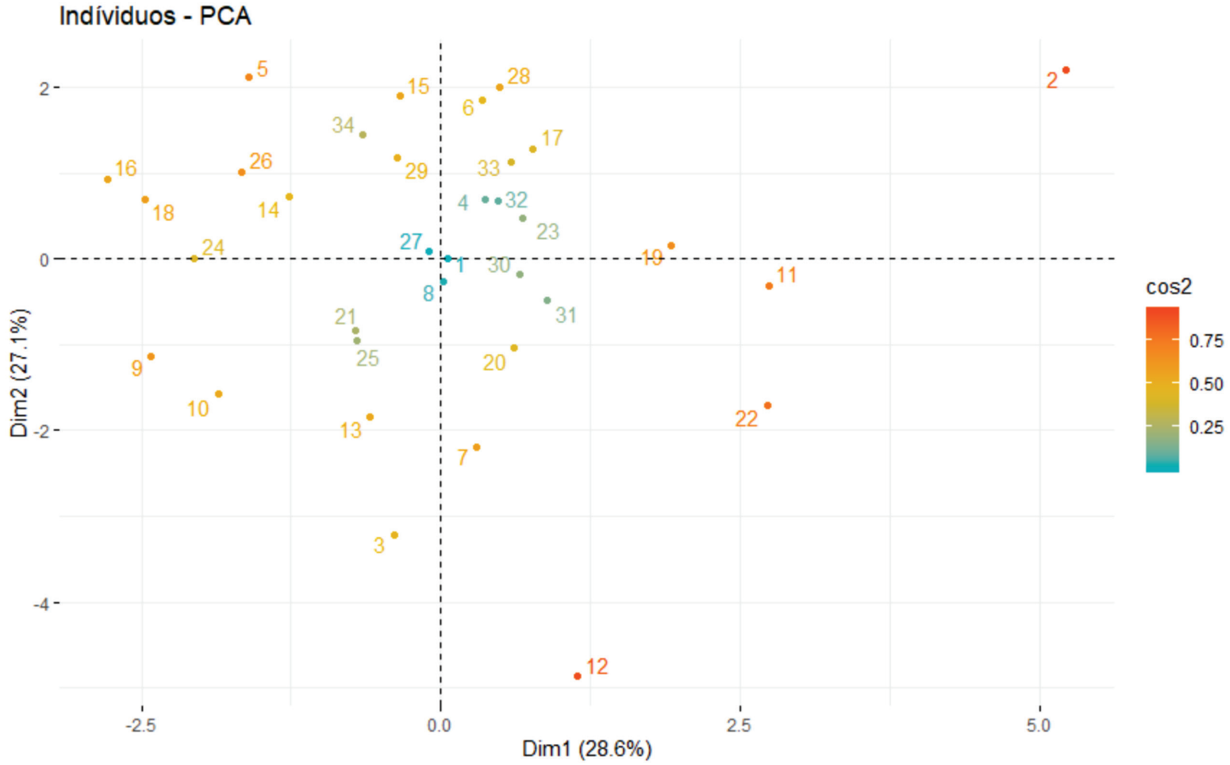

Figure 2 - Graphic dispersion via main component analysis – PCA, considering the first two components for variables in a natural cashew population in the municipality of Floriano-Piauí, Brazil.

Knowledge of genetic diversity is an important step in the process of conservation of plant genetic resources. Therefore, knowing the population structure for the management of genetic resources is the first step in any selection process (Djekota *et al.* 2014; N'djolossè *et al.* 2019). The results of the present study show the existence of variability within the natural cashew population, especially in relation to inflorescence variables, which are important for taxonomy.

The observed variations can be explained by adaptations to ecological conditions. However, these results show that cashew quantitative parameters related to lamina width and index are not as discriminatory compared to the other descriptors analyzed. This means that the cashew trees of this population are not significantly different in terms of leaf area. These results can be explained by the fact that these trees are in the same geographical region, under the same selective pressures, being the result of adaptation to eco-geographical conditions, such as altitude, temperature, wind and water availability, all of which are similar among them.

CONCLUSIONS

The variation between accessions of the sampled natural population may be associated with genetic history, origin, and adaptation. The seven morphological descriptors evaluated were important in the characterization of cashew trees, since they presented important contributions in the discrimination of divergence, while the variables related to inflorescence were the ones that contributed the most to genetic diversity. The UPGMA and Tocher clustering method efficiently detected the genetic diversity and grouped the cashew trees effectively in two groups, through quantitative morphological descriptors. The results obtained in this study are relevant, as they will serve as a basis for future work involving pre-breeding, propagation and genetic conservation of the species. However, other descriptors such as those related to chestnut and peduncles are needed to study natural cashew populations, as they offer a broader view on the distribution and organization of genetic diversity.

LITERATURE CITED

ADEIGBE, O. O. et al. A review on cashew research and production in Nigeria in the last four decades. *Scientific Research and Essays*, v. 10, n. 5, p. 196-209, 2015.

ALIYU, O. M.; AWOPETU, J. A. Multivariate analysis of cashew (*Anacardium occidentale* L.) germplasm in Nigeria. *Silvae Genetica*, v. 56, n. 1-6, p. 170-179, 2007.

ALIYU, O. M. Genetic diversity of Nigerian cashew germplasm. *Genetic Diversity in Plants*, p. 163-184, 2012.

ARCHAK, S. et al. Comparative assessment of DNA fingerprinting techniques (RAPD, ISSR and AFLP) for genetic analysis of cashew (*Anacardium occidentale* L.) accessions of India. *Genome*, v. 46, n. 3, p. 362-369, 2003.

ANDRADE, I. M. et al. A morphometric study of the restinga ecotype of *Anacardium occidentale* (Anacardiaceae): wild coastal cashew populations from Piauí, Northeast Brazil. *Feddes Repertorium*, v. 130, n. 2, p. 89-116, 2019.

BORGES, A. N. C. *Caracterização genética em germoplasma de cajuí (Anacardium spp.) por meio de marcadores morfoagronômicos e moleculares ISSR*. 2015. Tese de Doutorado. Dissertação de Mestrado, Universidade Federal do Piauí, Brasil.

CANTELLI, D. A. V. et al. Analysis of the genetic divergence of soybean lines through hierarchical and Tocher optimization methods. *Genetics and molecular research*, v. 15, n. 4, 2016.

CARNEIRO, L. A. et al. Morphological characterization and genetic divergence of a cashew population in Floriano, Piauí, Brazil. *Genetics and molecular research*, v. 18, n. 71, 2019.

CHABI SIKA, K. et al. Morphological characterization and agronomic performances of cashew (*Anacardium occidentale* L.) accessions from Benin. *Journal of Agricultural and Crop Research*, v. 3, n. 2, p. 27-40, 2015.

CHIPOJOLA, F. M. et al. Morphological characterization of cashew (*Anacardium occidentale* L.) in four populations in Malawi. *African journal of Biotechnology*, v. 8, n. 20, 2009.

COSTA, M. F. et al. Caracterização e divergência genética de populações de *Casearia grandiflora* no cerrado Piauiense. *Floresta e Ambiente*, v. 23, n. 3, p. 387-396, 2016.

CRUZ, C. D. Genes: a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum Agronomy*, v. 35, n. 3, p. 271-276, 2013.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. *Modelos biométricos aplicados ao melhoramento genético*. 3 ed. Viçosa: UFV, 2012.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. *Modelos biométricos aplicados ao melhoramento genético*. 4 ed. Viçosa: UFV, 2014.

DJEKOTA, C. et al. Morphological characterization of shea tree (*Vitellaria paradoxa* subsp. *paradoxa*) populations in the region of Mandoul in Chad. *International Journal of Biodiversity and Conservation*, v. 6, n. 2, p. 184-193, 2014.

HAMAWAKI, O. T. et al. Genetic parameters and variability in soybean genotypes. *Comunicata Scientiae*, v. 3, n. 2, p. 76-83, 2012.

KOUAKOU, C. K. et al. Agro-morphological characterization of cashew (*Anacardium occidentale* L.) ecotypes from North-Central of Côte d'Ivoire. *Fruits*, v. 73, n. 5, 2018.

KULIS, M. et al. Pepsinized cashew proteins are hypoallergenic and immunogenic and provide effective immunotherapy in mice with cashew allergy. *Journal of Allergy and Clinical Immunology*, v. 130, n. 3, p. 716-723, 2012.

LOPES, C. G. R. et al. Levantamento da flora apícola em área de cerrado no município de Floriano, estado do Piauí, Brasil. *Revista Brasileira de Biociências*, v. 14, n. 2, 2016.

MAIA, M. C. C. et al. Principal component and biplot analysis in the agro-industrial characteristics of *Anacardium spp*. *European Scientific Journal*, v.30, n. 15, p.1-11, 2019.

MARIM, B. G. et al. Variabilidade genética e importância relativa de caracteres em acessos de germoplasma de tomateiro. *Pesquisa Agropecuária Brasileira*, v. 44, n. 10, p. 1283-1290, 2010.

MAPA. (2011): *Instruções para execução dos ensaios de distinguibilidade, homogeneidade e estabilidade de cultivares de manga (Mangifera indica L.).* (available at http://www.agricultura.gov.br/assuntos/insumos-agropecuarios/insumos-agricolas/protecao-de cultivar/arquivos-frutiferas/mangueira_us.doc.).

N'DJOLOSSÈ, K. et al. Agro-morphological characterization of preselected cashew (Anacardium occidentale L.) mother trees in Benin farmer's plantations. *Journal of Agriculture and Environment for International Development (JAEID)*, v. 113, n. 1, p. 17-34, 2019.

ROHLF, F. J. Adaptive hierarchical clustering schemes. *Systematic Biology*, v. 19, n. 1, p. 58-82, 1970.

SANTOS, J. O. et al. Genetic diversity in wild populations of the restinga ecotype of the cashew (*Anacardium occidentale*) in coastal Piaui, Brazil. *Plant Systematics and Evolution*, v. 305, n. 10, p. 913-924, 2019.

SILVA, R. A. et al. Evaluation of antioxidant and mutagenic activities of honey-sweetened cashew apple nectar. *Food and chemical toxicology*, v. 62, p. 61-67, 2013.

TEAM, R. Core. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria, 2014.

VIEIRA, M.; MAYO, S. J.; DE ANDRADE, I. M. Geometric morphometrics of leaves of *Anacardium microcarpum* Ducke and *A. occidentale* L.(Anacardiaceae) from the coastal region of Piauí, Brazil. *Brazilian Journal of Botany*, v. 37, n. 3, p. 315-327, 2014.

Recebido para publicação em 18/03/2019, aprovado em 23/06/2020 e publicado 30/07/2020.