GROUPING OF CATINGUEIRA GENOTYPES BASED ON MORPHOLOGICAL CHARACTERISTICS[1]

Leonardo Silva Souza[2*], Ricardo Franco Cunha Moreira[3], Teresa Aparecida Soares de Freitas[3], Andrea Vita Reis Mendonça[3], Sandra Domingos João Afonso[2] and Carlos Alberto da Silva Ledo[4]

[1] Received on 10.02.2015 accepted for publication on 24.05.2016.

[2] Universidade Federal do Recôncavo da Bahia, Programa de Pós-Graduação em Ciências Agrárias-CCAAB, Cruz das Almas, BA - Brasil. E-mail: <leouenf@hotmail.com> and <sandra.afonso3@gmail.com>.
[3] Universidade Federal do Recôncavo da Bahia, Centro de Ciências Agrárias Ambientais e Biológicas (CCAAB), Cruz das Almas, BA - Brasil. E-mail: <ricardofcm@ufrb.edu.br>, <teresa@ufrb.edu.br> and <andrea@ufrb.edu.br>.
[4] Embrapa Mandioca e Fruticultura, Cruz das Almas, BA - Brasil. E-mail: <carlos.ledo@embrapa.br>.
*Corresponding author.

ABSTRACT – The objective of this study was to obtain information about the characteristics of *Poincianella pyramidalis* (Tul.) through morphological characters. We have analyzed eleven genotypes of *P. pyramidalis*, at least 100 meters far apart among themselves, located in the municipalities of Santa Teresinha and Castro Alves, in the State of Bahia, Brazil. The genotypes were identified with aluminum plaques, georeferenced, evaluated according to their height and diameterat ground level, number of tillers per plant and fruit production. One hundred fruits for each genotype were collected and measured relative to their length, width (in the proximal, medial and distal positions), thickness and number of loci.Analyses have been performed: descriptive, Pearson correlation between the variables in addition to the cluster analysis for qualitative variables and subsequent clustering to define the genetic diversity among the species genotypes using the Euclidean average linkage and a cutoff point as an "NbClust" package. The results showed variability among genotypes, with the formation of eleven major groups of genetic dissimilarity. The largest genetic distances are among the G7ST and the G1ST genotypes located in the municipality of and the G1CA and the G8ST genotypes in the municipalities of Castro Alves and Santa Teresinha. The variable with the greatest contribution to divergence was the one with the number of tillers.

Keywords: Caatinga; Genetic divergence; Variability.

FORMAS DE AGRUPAMENTO DE GENÓTIPOS DE CATINGUEIRA COM BASE NAS CARACTERÍSTICAS MORFOLÓGICAS

RESUMO – O objetivo deste trabalho foi obter informações sobre as características de ***Poincianella pyramidalis*** *(Tul.) através de caracteres morfológicos. Foram analisados 11 genótipos de* ***P. pyramidalis*** *distantes no mínimo 100 metros entre si, localizados nos municípios de Santa Teresinha e Castro Alves - BA. Os genótipos foram identificados com placas de alumínio, georreferenciados e avaliados quanto altura, diâmetro do colo, número de perfilhos, produção de fruto. Para cada genótipo, foram coletados 100 frutos e medidos em relação ao seu comprimento, largura (nas posições proximal, mediana e distal), espessura e número de lóculos. Foram realizadas as análises: descritiva, correlação de Pearson entre as variáveis além da análise de agrupamento para variáveis qualitativas, e posterior agrupamento para definição da diversidade genética entre genótipos utilizando a Distância Euclidiana Média e como ponto de corte o pacote "NbClust". Os resultados revelaram variabilidade entre os genótipos, com a formação de onze grupos principais de dissimilaridade genética. As maiores distâncias genéticas foram encontradas entre os genótipos G7ST e G1ST localizados no município de Santa Teresinha e os genótipos G1CA e G8ST localizados nos municípios de Castro Alves e Santa Teresinha. A variável com maior contribuição para divergência foi a número de perfilhos.*

Palavra chave: Caatinga; Divergência genética; Variabilidade.

Revista Árvore, Viçosa-MG, v.40, n.3, p.427-434, 2016
http://dx.doi.org/10.1590/0100-67622016000300006

1. INTRODUCTION

The Caatingais an exclusively Brazilian biome and, according to the Brazilian Institute of Geography and Statistics (IBGE, 2013), it presents itself as the fourth most extensive biome in the country, with an area of 826,411 square kilometers extending over the entire State of Ceará (100%), over half of the State of Bahia (54%), Paraíba (92%), Pernambuco (83%), Piauí (63%) and Rio Grande do Norte (95%), almost half of the State of Alagoas (48%) and Sergipe (49%), besides, it also covers small portions of Minas Gerais (2%) and Maranhão (1%).

The *Poincianella pyramidalis* (Tul.) L. P. Queiroz, comb. Nov. widely known as catingueira, Pig's Caatinga and Rat stick, belongs to the Fabaceae family, of an endemic species and one of the most representative of the Caatinga biome (GIULIETTI et al., 2004).

The *P. pyramidalis*, a midsized tree with arboreal habitwithout thorns, normally4 to 6 m tall but can reach up to 12 m, open and irregular crown, and, when adult,its bark is light gray, sometimes brown with spots of a yellowish, greenish and white color, and it releases the superficial layer in slightly elongated blades (MAIA, 2004). According to the same author the *P. pyramidalis* presents a promising potential in veterinary use, reforestation, agroforestry and fodder for sheep and cattle feeding.

Its wood is used for poles, fence posts; building of mud houses as well as for firewood and charcoal, and because of its hardiness it is suitable for reforestation (LORENZI; MATOS, 2008). Besides it is considered a very rough species because it adapts itself very well to most soils and climates and also to be quite tolerant to droughts (BARROS et al., 1997). The species has been scientifically proven to have medicinal properties (anti-inflammatory, healing and antimicrobial ones) (SALVAT et al., 2004).

The success of a breeding program, conservation and sustainable exploitation lies in the existence of genetic variability (ZUCCHI, 2009). So it is recommended for the formation of the base populationcrossbreeding between upper and divergent genotypes. This divergency can be evaluated from agronomic, morphological and molecular characteristics, among others. Multiple information of each genotype is expressed in dissimilarity measures which represent the diversity that exists in the studied biological system.

The determination of the genetic dissimilarity, through simultaneous assessment of several characters, can be a powerful tool for the identification of superior genotypes, making possible the concentration of efforts on the most promising combinations (MOURA et al., 1999). Thus, the multivariate techniques have proven useful for assessing the individual in several aspects and provide a holistic view of each access (CRUZ et al., 2004).

Cluster analysis has the purpose of gathering, according to some classification criteria, a set of sampling units (treatments, parents, individuals, etc.) in groups by evaluating variables, so that there is homogeneity in each group and heterogeneity between groups (RIBEIRO JUNIOR, 2001).

According to Scheeren et al. (2000) the clustering criteria used in most algorithms, is a measurement of similarity or distance between elements of a data array.

To Rossmann (2001) phenotypic variability can be known by means of phenotypic, genotypic and environmental correlation coefficients of genotypic and phenotypic fluctuations, and other genetic parameters, which reflect the nature of the genetic material and environmental action, allowing the prediction of gains resulting from the selection and definition of breeding strategies that could be adopted.

This study aimed to study the ways of clustering *P. pyramidalis* genotypes based on morphological characteristics.

2. MATERIALAND METHODS

To carry out this work, eleven *P. pyramidalis* genotypes, selected from eight genotypes of the municipality of Santa Teresinha-BA and three genotypes of Castro Alves-Bahia, were used. The genotypes were georeferenced with GPS assistance (Table 1) and identified with aluminum plaques complying with the minimum distance of 100 meters apart from themselves. The two municipalities are in region of the Paraguaçu river valley which is part of the Drought Polygon, being the period of greatest rainfall only in the months from May to July with rainfall reaching an annual average of 865 mm, ranging from 1,496 mm (maximum) to 329 mm (minimum) (SILVA, 1993). The vegetation consists of deciduous forest, seasonal caatinga-forest and semi-deciduous forest (QUEIROZ, 2009).

Table 1 – Location data of eleven *P. pyramidalis* genotypes, georeferenced and identified in municipalities in the semi-arid region of the Caatinga ,Bahia State, Brazil.
Tabela 1 – *Dados de localização de onze genótipos de* ***P. pyramidalis*** *georeferenciados e identificados em municípios da região semiárida da Caatinga do Estado da Bahia, Brasil.*

	Geographical Coordinates		
Genotype	Latitude(S)	Longitude(W)	Origin
G1ST	12°45'15"	39° 31'01"	Santa Teresinha
G2ST	12°45'13"	39° 31'02"	Santa Teresinha
G3ST	12°45'09.7"	39°31'01"	Santa Teresinha
G5ST	12°45'10"	39°30'56"	Santa Teresinha
G6ST	12°45'09"	39°30'52"	Santa Teresinha
G7ST	12°45'06"	39°30'49"	Santa Teresinha
G8ST	12°45'01.1"	39°30'48"	Santa Teresinha
G9ST	12°45'01.2"	39°30'43"	Santa Teresinha
G1CA	12°44'48"	39°26'39"	Castro Alves
G6CA	12°44'47"	39°27'02"	Castro Alves
G7CA	12°44'48.4"	39°26'39.6"	Castro Alves

The genotypes of the municipality of Castro Alves are situated in the midst of pasture, while the ones from Santa Teresinha are composing a gallery forest of the river called "Laziness".

The height of the trees was measured through the use of Hypsometer. The fruits were collected in two locations between the months of September and October. To help with the fruit gathering, trimmer, pruning shears, plastic bags and plastic canvas were used.

The fruits were sent to the Seed Laboratory of the Federal University of Bahia Reconcavo (UFRB), then they wereplaced to dry in the shade under ventilation, and their cut opening, to obtain the seeds, was performed manually.

The following aspects were considered for the characterization of fruit: dimensions (length, width and thickness), texture and number of loci.

These evaluations were made through a random sampling of 100 fruits of each matrix. The fruits were measured with the aid of a 0.1 mm precision caliper. The length was measured from the base up to its apex, the width was measured in three positions (proximal, intermediate and distal of the fruit), and the thickness was measured in the midline of the fruit.

For the description and evaluation of the genotypes, eight variables were analyzed: height (m), diameter at ground level (cm), number of tillers (unit), fruit production (kg), fruit length (cm), width of the fruit (cm), fruit thickness (cm) and number of loci (unit).

The Pearson correlation coefficients of the data were calculated and tested by Student's T test with a 5% and 10% of significance.

The multicollinearity diagnostic for the linear correlation matrix at the discretion of Montgomery and Peck (1981) has also been performed.

Then, based on the average of quantitative traits, the Euclidean distance average linkage was estimated as a measure of dissimilarity, by analyzing the correlation between the traits, according to Cruz et al. (2004).

Hierarchical clustering of analyzes from the genetic distance matrix was obtained by UPGMA - Unweighted Pair Group Method with Arithmetic Mean (SNEATH; SOKAL, 1973). For the relative contribution of each evaluated characteristic of quantitative data grouping, it was performed a variance analysis between groups using the T test and it was obtained the relative contribution of quantitative variables for divergence between genotypes according to Singh (1981). To obtain the genetic distance matrix and the calculation of the cophenetic correlation coefficients, the Genes program has been used (CRUZ, 2008). The dendrogram was obtained using the STATISTICA program (STATSOFT, 2005). The variance analysis has been performed using SAS (SAS, 2006).

The criteria used for the groups formation and for determining the cutoff point was based on the "NbClust" package (CHARRAD et al., 2012) belonging to the computer program R (R DEVELOPMENT CORE TEAM, 2010), in which fifteen indices for the ideal number of clusters were determined. These indices are proposed by the following authors: Friedman and Rubin (1967), Halkidi and Vazirgiannis (2001), Frey and van Groenewoud (1972), Duda and Hart (1973), Calinski and Harabasz (1974), Baker and Hubert (1975) Hartigan (1975), Hubert and Levin (1976), Davies and Bouldin (1979), Dunn (1974), Sarle (1983) Milligan and Cooper (1985), Krzanowski and Lai (1988), Walesiak and Dudek (2011) and Tibshirani et al. (2001), where the choice of the best grouping is generally based on taking into account the maximum and minimum values and the difference between the indices. The choice of cutoff point is made according to the proportion that the number of groups repeats itself in larger quantities.

3. RESULTS

For the results found it is appropriate to point out that the plants are in different natural locations and conditions, being the genotypes of Castro Alves in the middle of a pasture and the genotypes of Santa Teresinha in a gallery forest. Thus, according to Pires (2011) the performance of a genotype in relation to another can change according to the environment, so that genotypes that are superior in an environment may not be in another.

The linear Pearson correlation coefficients tested by Student's t test were significant and positive (Table 2) for correlations between EF and LF variables (0.66 *) with a 5% significance, indicating that plants with a higher thickness in the fruits, theses will tend to be wider.

In the sequence, the variables that showed a significant positive correlation to 10% significance were DAS and ALT (0,54+), LF and CF (0,47+), EF and CF (0,54+), NL and PF (0,55+), and the negative correlation was identified between the variables NP and DAS (-0,51+), for the remainder of the variables were not significant at the 5% level and a 10% probability.

The multicollinearity diagnostic revealed that the number of condition (NC) is 95.07 for the used variables, being defined as weak collinearity according to Montgomery and Peck rating table (1981). The cluster analysis performed with the matrix of the Euclidean distance average linkage, obtained among genotypes from variables, plant height (ALT), diameter at ground level (DAS), number of tillers (NP), fruit production (PF), fruit length (FL), fruit width (LF), fruit thickness (EF), number of loci (NL) indicates that the genotypes of less distant catingueira were the G3ST and the G6CA with value of 0.16, from the municipalities of Santa Teresinha and Castro Alves, respectively. These two genotypes show morphological characteristics such as height, diameter at ground level, length and width of fruit and number of loci very closeas noted by the descriptive results. The largest genetic distance of 0.66 (Table 3) was observed between G7ST and G1ST genotypes (Santa Teresinha) as well as in the G1CA (Castro Alves) and G8ST (Santa Teresinha) genotypes.

The cophenetic correlation coefficient (r = 0.83). The variables that contributed the most to the divergence among the genotypes were the number of tillers (NP) and fruit production (PF), accounting for 52.40% and 20.55%, respectively (Table 4).

According to the dendrogram of the genetic dissimilarity (Figure 1) obtained based on the eight morphological characteristics of the *P. pyramidalis*,

Table 2 – The linear Pearson correlation coefficientsamong the morphological characteristics evaluated from eleven *P. pyramidalis* genotypes from the municipalities of Castro Alves and Santa Teresinha in the state of Bahia.
Tabela 2 – *Correlação linear de Pearson entre as características morfológicas avaliadas de onze genótipos de* ***P. pyramidalis*** *provenientes dos municípios de Castro Alves e Santa Teresinha no Estado da Bahia.*

	ALT	DAS	NP	PF	CF	LF	EF	NL
ALT	1.00	0.54^{+}	-0.11^{ns}	0.42^{ns}	0.02^{ns}	0.20^{ns}	0.08^{ns}	0.18^{ns}
DAS		1.00	-0.51^{+}	0.29^{ns}	-0.35^{ns}	-0.13^{ns}	0.013^{ns}	-0.09^{ns}
NP			1.00	0.33^{ns}	0.10^{ns}	0.06^{ns}	0.10^{ns}	0.25^{ns}
PF				1.00	0.05^{ns}	-0.36^{ns}	0.13^{ns}	0.55^{+}
CF					1.00	0.47^{+}	0.54^{+}	0.32^{ns}
LF						1.00	0.66*	-0.42^{ns}
EF							1.00	0.02^{ns}
NL								1.00

ALT = plant height (m); DAS= Diameter at breast height (cm); NP = Number of Tillers (unit); PF = Fruit Production (Kg); CF = Fruit Length (cm); LF = Width of the Fruit (cm); EF= Thickness of the Fruit (cm); NL = Number of locules (unit); ** , * , + significant at 1 %, 5% and 10, respectively, by the t-test and ns not significant.

Table 3 – Dissimilarity matrix among *P. pyramidalis* genotypes from the municipalities of Castro Alves and Santa Teresinha in Bahia State.
Tabela 3 – *Matriz de dissimilaridade entre os genótipos de* ***P. pyramidalis*** *nos municípios de Castro Alves e Santa Teresinha no estado da Bahia.*

Genótipo	G1ST	G2ST	G3ST	G5ST	G6ST	G7ST	G8ST	G9ST	G1CA	G6CA	G7CA
G1ST	0.00	0.46	0.54	0.38	0.51	0.66	0.56	0.52	0.48	0.57	0.48
G2ST		0.00	0.51	0.46	0.50	0.52	0.41	0.48	0.47	0.61	0.55
G3ST			0.00	0.52	0.44	0.28	0.39	0.23	0.56	0.16	0.51
G5ST				0.00	0.26	0.52	0.57	0.50	0.32	0.53	0.48
G6ST					0.00	0.43	0.60	0.50	0.38	0.43	0.41
G7ST						0.00	0.42	0.38	0.52	0.34	0.63
G8ST							0.00	0.24	0.66	0.49	0.57
G9ST								0.00	0.62	0.30	0.49
G1CA									0.00	0.57	0.58
G6CA										0.00	0.48
G7CA											0.00

Table 4 – Relative contribution of characters for divergence - Singh (1981) of eleven *P. pyramidalis* genotypes from the municipalities of Castro Alves and Santa Teresinha in Bahia State. Brazil.
Tabela 4 – *Contribuição relativa dos caracteres para divergência – Singh (1981) entre onze genótipos de* ***P. pyramidalis*** *provenientes dos municípios de Castro Alves e Santa Teresinha no Estado da Bahia.Brasil.*

Variable	Relative Contribution (%)
Number of tillers (Unit)	52.40
Production of Fruit (Kg)	20.55
Diameter in Ground height (cm)	19.28
Heights (m)	4.67
Number of locules (Unit)	1.30
Fruit Length (cm)	1.67
Fruit Width (cm)	0.11
Fruit thickness (cm)	0.011

collected in the municipalities of Santa Teresinha and Castro Alves in Bahia, shows the formation of eleven groups. The dendrogram cutoff point was done using the "Nbclust" package.

4. DISCUSSION

The knowledge of the correlations between the variables of interest is of great importance because it predicts the influence of the selection of a given character on the modification of the other's average (FARIAS NETO et al., 2005).

According to Lima (2012) the high degree of multicollinearity indicates interrelationships of the variables, making it difficult to interpret the set of variables used in the cluster for being more difficult to determine the individual effect of any variable. Regarding the analysis of the eigenvalues of the phenotypic correlation matrix, it was performed to identify the nature of the existing linear dependence among the characters, detecting those which contributed to the appearance of multicollinearity (BELSLEY et al., 1980).

The cophenetic correlation coefficient (r = 0.83), showed a good adjustment for the quantitative variables between the graphical representation of the distances and its original matrix. As suggested by Bussab et al. (1990),cluster analysis is acceptable if they produce a cophenetic correlation coefficient (CCC) from 0.80 onwards.

The G1ST and G9ST genotypes suggest that these are the most divergent of the total analyzed and can

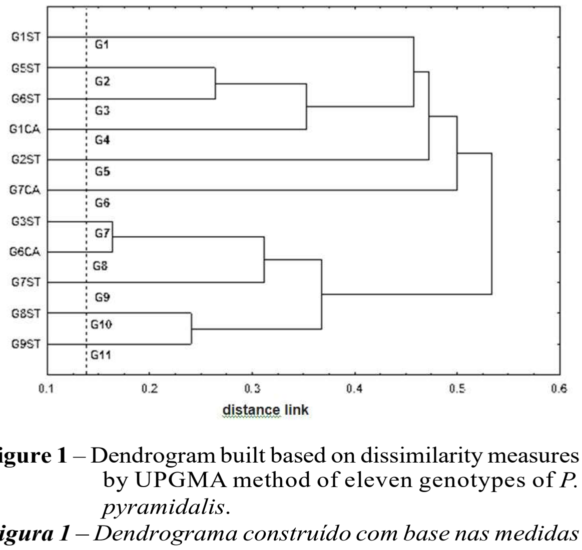

Figure 1 – Dendrogram built based on dissimilarity measures by UPGMA method of eleven genotypes of *P. pyramidalis*.
Figura 1 – *Dendrograma construído com base nas medidas de dissimilaridade pelo método de agrupamento UPGMA entre onze genótipos de* ***P. pyramidalis****.*

be used in breeding program of the species in relation to intraspecific crosses. The groups show formation with only one access each, featuring a good degree of divergency among them, corroborating the study done by Santos et al. (2012), in which was observed high genetic diversity among *P. pyramidalis* genotypes (Poincianellas pp) using RAPD markers. However, according to Sing et al. (1981) the best way of identifying divergence among genotypes is the combined use of molecular markers and agro-morphological descriptors for promoting complement to the results, proving the efficiency of the results with a simultaneous analysis.

Thus, future studies with a larger number of genotypes and in different locations should be considered as well as the use of molecular techniques in the analysis of data in order to compare and quantify the diversity of the species studied in this region and it can alsoassist in the management and conservation of the species. According to Santos et al. (2012) the practical use of information on the genetic diversity enables a better planning of seed collections for the improvement of tree species, as well as the genetic composition of future seed orchards.

5. CONCLUSIONS

1 - There is variability among *P. pyramidalis* genotypes, which can be exploited for the conservation and exploitation of the species.

2 - The variable with the greatest contribution to divergency was the one with the number of tillers.

6. ACKNOWLEDGMENT

The Bahia Research Support Foundation that has provided funding for this research. We thank Mr. Eduardo Nascimento for providing the area for this study.

7. REFERENCES

BARROS, N.N.; SOUSA, F.B.; ARRUDA, F.A.V. **Utilização de forrageiras e resíduos agroindustriais por caprinos e ovinos.** Sobral: Embrapa/CNPC, 1997. 28p. (Documentos, 26).

BAKER, F.B.; HUBERT, L.J. Measuring the power of hierarchical cluster analysis. **Journal of the American Statistical Association,** v.70, n.349, p.31-38, 1975.

BELSLEY, D.A.; KUH, E.; WELSCH, R.E. **Regression diagnostics:** identifying data and sources of colinearity. New York: J. Wiley, 1980. 292p.

BUSSAB, W.O.; MIAZAKI, E.S.; ANDRADE, D.F. Introdução à Análise de Agrupamentos. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 9., São Paulo. São Paulo: Associação Brasileira de Estatística, 1990. 105p.

CALINSKI, T.; HARABASZ, J. A dendrite method for cluster analysis. **Communications in Statistics -** theory and methods**.** 27 de janeiro 1974. Disponível em: http://dx.doi.org/10.1080/03610927408827101. Acesso em: 18 novembro 2013.

CHARRAD, M.; GHAZZALI, N.; BOITEAU, V.; NIKNAFS, A. NbClust: **An examination of indices for determining the number of clusters. R package version 1.4 2012.** Disponível em: http://cran.r-project.org/web/packages/NbClust/index.html

CRUZ, C.D. **Programa Genes (versão Windows):** Aplicativo computacional em genética e estatística. Viçosa, MG: Universidade Federal de Viçosa, 2008.

CRUZ, C.D.; REGAZZI, J.A.; CARNEIRO, P.C.S. (Ed.). **Modelos biométricos aplicados ao melhoramento genético.** Viçosa, MG: Universidade Federal de Viçosa, v.1, p.377–413, 2004.

DAVIES, D.L.; BOULDIN, D.W. A cluster separation measure. **IEEE Transactions on Pattern Analysis and Machine Intelligence,** v.1, p.2, 1979**.**

DUDA, R.O.; HART, P.E. **Pattern classification and scene analysis**. New York: John Wiley and Sons, 1973.

DUNN, J. Well separated clusters and optimal fuzzy partitions. **Journal Cybern,** v.1, p.9, 1974.

FARIAS NETO, J.T.; OLIVEIRA, M.S.P.; CRUZ NETO, A.J.; SANTOS, L.R. Avaliação de recursos genética em progênie jovens de açaizeiro. **Cernes**, v.11, n.4, p.336-341, 2005.

FRIEDMAN, H.P.; RUBIN, J. On some invariant criteria for grouping data. **Journal of the American Statistical Association,** v.1, p.1159-1178, 1967**.** http://cran.r-project.org/web/packages/NbClust/index.html

FREY, T.; van GROENEWOUD, H. A cluster analysis of the D-squared matrix of white spruce stands in Saskatchewan based on the maximum-minimum principle. **Journal of Ecology.** v.1, p.873-886, 1072.

GIULIETTI, A.M.; DU BOCAGE NETA, A.L.; CASTRO, A.A.J.F.; GAMARRA-ROJAS, C.F.L.; SAMPAIO, E.V.S.B.; VIRGÍNIO, J.F.; QUEIROZ, L.P.; FIGUEIREDO, M.A.; RODAL, M.J.N.; BARBOSA, M.R.V.; HARLEY, R.M. Diagnóstico da vegetação nativa do bioma Caatinga. In: SILVA, J.M.C.; TABARELLI, M.; FONSECA, M.T.; LINS, L.V. (Org.) **Biodiversidade da Caatinga: áreas e ações prioritárias para a conservação.** Brasília: Embrapa SemiÁrido, 2004.p.48-90..

HALKIDI, M.; BATISTAKIS, I.; VAZIRGIANNIS, M. On clustering validation techniques. **Journal of Intelligent Information Systems**, v.2 p.107-145, 2001.

HARTIGAN, J.A. **Clustering algorithms.** New York: John Wiley & Sons, 1975.

HUBERT, L.J.; LEVIN, J.R. A general statistical framework for assessing categorical clustering in free recall. **Psychological Bulletin**, v.2, p.193-218, 1976.

IBGE**. Mapa de biomas e de vegetação**. Rio de Janeiro: 2005. Disponível em: http://www.ibge.gov.br. Acesso em 29/11/2013.

KRZANOWSKI, W.J.; LAI, Y.T. A criterion for determining the number of groups in a data set using sum-of-squares clustering. **Biometrics,** v.44, n.1, p.23-34, 1988.

LIMA, M.S.S. **Seleção e propagação de genótipos de umbu-cajazeira (*Spondias sp.*) da Região semiárida da Bahia**. 2012. 47p. Dissertação (Mestrado em Recurso Genéticos Vegetais) - Universidade Federal do Recôncavo da Bahia, Cruz da Almas, BA, 2012.

LORENZI, H.; MATOS, F.J.A. **Plantas medicinais no Brasil**: Nativas e exóticas. 2ª ed. Nova Odessa: Instituto Plantarum, 2008. 544p.

MAIA, G.N. **Catingueira**. In: MAIA, G.N. Caatinga: árvores e arbustos e suas utilidades. São Paulo: Leitura e Arte, 2004. p.159-169.

MILLIGAN, G.; COOPER, M. An examination of procedures for determining the number of clusters in a data set. **Psychometrika,** v.2, p.159-179, 1985.

MONTGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis.** New York: John Wiley & Sons, 1981. 593p.

MOURA, W.M.; CASALI, V.W.D.; CRUZ, C.D. et al. Divergência genética em linhagens de pimentão em relação a eficiência nutricional de fósforo**. Pesquisa Agropecuária Brasileira,** v.34, n.2, p.217-224, 1999.

PIRES, I.E.; RESENDE, M.D.V.; SILVA, R.L.; JUNIOR, M.F.R.R. **Genética florestal.** Viçosa, MG: Arka, 2011. 318p.

QUEIROZ, L.P. **Leguminosas da Caatinga**. Feira de Santana: Universidade Estadual de Feira de Santana, 2009. 467p.

R DEVELOPMENT CORE TEAM. **A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2010.

RIBEIRO JÚNIOR, J.I. **Análises estatísticas no SAEG**. Viçosa, MG: Universidade Federal de Viçosa, 2001. 301p.

ROSSMANN, H. **Estimativas de parâmetros genéticos e fenotípicos de uma população de soja avaliada em quatro anos**. 2001. 80p. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2001.

WEE, A. K. S.; HONG LI, C.; HONG, Y. Genetic diversity in natural populations of Gmelina arborea : implications for breeding and conservation. **New Forests**, v.43, p.411-428, 2012.

SALVAT, A.; ANTONACCI, L.; FORTUNATO, R.H.; SUAREZ, E.Y.; GODO, H.M. Antimicrobial activity in methanolic extracts of several plant species from northern Argentina. **Phytomedicine**, v.11, p.230-234, 2004.

SANTOS, J.S.; MENDES, S.S.; CONDE, D.C.; DElMONDEl, R.C.; MANN, R.S.; THOMAZZl, S.M. Genetic diversity assessment of *Poincianella pyramidalis* (Tul.) l.P. Queiroz accessions using RAPD markers. **Scientia Plena**, v.8, n.3, p.1-8, 2012.

SARLE, W.S. SAS Technical Report. **Cubic clustering criterion.** Cary, N.C.: SAS Institute, 1983.

SAS Institute. SAS Technical Report. **SAS/STAT software: Changes and Enhancement**, Release 9.1, Cary NC: 2006.

SCHEEREN, L.W.; GEHRARDT, E.J.; FINGER, C.A.G. Agrupamento de unidades amostrais de *Araucária angustifólia* (Bert.) O. Ktze. Em função de variáveis do solo, da serrapilheira e das acículas, na região de Canela, RS. **Revista Ciência Florestal**, v.10, n.2, p.39-57, 2000.

SILVA, F.B.R. **Zoneamento agroecológico do Nordeste:** diagnóstico do quadro natural e agrossocioeconômico. Recife: Embrapa/CPATSA, 1993.

SINGH, D. The relative importance of characters affecting genetic divergence. **The Indian Journal of Genetics and Plant Breeding**, v.41, n.1, p.237-245, 1981.

SNEATH, P.H.; SOKAL, R.R. **Numerical taxonomy**: The principles and practice of numerical classification. San Francisco: W.H. Freeman, 1973. 573p.

STATSOFT, Inc. **Statistica for Windows** (data analysis software system), version 7.1. Tulsa: 2005.

TIBSHIRANI, R.; WALTHER, G.; HASTIE, T.Estimating the number of clusters in a data set via the gap statistic. **Journal of the Royal Statistical Society,** 2001. URL http://dx.doi.org/10.1111/1467-9868.00293.

WALESIAK, M.; DUDEK, A. **ClusterSim: Searching for optimal clustering procedure for a data set.** 2011. R package version 0.40-6, URL 2011. http://cran.r-project.org/web/packages/clusterSim.

ZUCCHI, M.I. **Diversidade genética em espécies medicinais.** 2009. Disponível:http://www.infobibos.com/Artigos/2009_4/DiversidadeGenetica/index.htm. Acesso em: 26/5/2013.