Genetic Divergence Among a Breeding Population of *Hancornia Speciosa* Gomes (Mangabeira) as Determined by Multivariate Statistical Methods

Maria Clideana Cabral Maia (DSc)
Embrapa Tropical Agroindustry, Brazil
Mirian Fernandes Carvalho Araújo(DSc)
Faculty of Mathematics, Federal University of Uberlândia. Brazil
Lucio Borges de Araújo(DSc),
Faculty of Mathematics, Federal University of Uberlândia. Brazil
Carlos Tadeu dos Santos Dias(DSc),
School of Agriculture "Luiz de Queiroz" / University of São Paulo, Brazil
Luís Cláudio de Oliveira(MSc),
Embrapa Acre, Brazil
Cosme Damião Cruz(DSc),
Department of General Biology, Federal University of Viçosa - UFV, Brazil
Lúcio Flavio Lopes Vasconcelos(DSc),
Embrapa Mid-North, Brazil
Luciano Medina Macedo(DSc),
Department of Genetics and Plant Breeding, Federal Technological University of Paraná – UFPR, Brazil
Gilberto Ken Iti Yokomizo(DSc),
Embrapa Amapá, Brazil
Paulo Sarmanho da Costa Lima(DSc).
Embrapa Mid-North, Brazil

Abstract
        The mangabeira its figure out among the mains native fruit tree explored by extractivism in Brasil. The objective evaluate the genetic divergence of landraces in orientation of crosses using multivariate techinics. The complete random blocks experimental design with four repetitions was used to evaluate twelve quantitative characteristics from twelve genotypes elite of mangabeiras concerning to divergence genetic using the software R (2012). Three groups genetically divergent were composed by biplot graphic and stored by UPGMA cluster analysis (Unweighted Pair-Group Method

using Arithmetic Average / Weighted Clustering Method not using the Arithmetic Mean) showing genetic diversity and variability among 12 mangabeira accesses. Forty-four possible crosses are planned among genotypes of genetically dissimilar three groups and six among individuals in group III. Multivariate techniques were appropriate in the study of genetic divergence.

Keywords: Apocynaceae, analysis of grouping, principal component analysis, biometrics

Introduction

The species *Hancornia speciosa* Gomes is of social, economic and cultural importance to collector farmers (extractive farmers) as well as being a direct nutritional source for wild animals inhabiting the biomes where natural populations of the plant occur. The flowers supply nectar to natural pollinating agents responsible for increasing genetic variability, the contribution of which is necessary for the selection and production of improved genetic material required for the formation of orchards affording significant gains in productivity relative to the original populations.

Mangabeira is a tropical fruit tree belonging to the family Apocynaceae, and is found more frequently as natural populations in the Cerratinga (cerrado and caatinga) biomes of the central-west and northeastern states of Brazil. Despite the importance of mangabeira as a source of income and nutrition (being rich in essential vitamins, iron, calcium and protein) for populations of collector farmers (sustainable extractivism) in these regions, there is an imminent loss of entire natural populations that are, literally, in the process of diminishing their effective intra- and inter-population genetic sizes. These populations are subject to conditions of genetic vulnerability that pose an evident risk of extinction caused mainly by the effects of deforestation carried out in support of agro-productive activities based on homogeneous cultures. Although such practices are theoretically more profitable, they are bio-environmentally unsustainable in the long-term and will reduce the natural occurrence of this important native fruit tree.

There are few records of economical orchards and no records of orchards formed with the aim of providing genetic longevity, prolongation of harvesting and good yields, simultaneously. Considering that mangaba is a perennial fruiting species, definitive agricultural advantages can only be achieved if the individuals that form the orchards exhibit considerable genetic diversity.

Among the northeastern states, Sergipe is responsible for 55% of the production, followed by Bahia with 19.7% and Paraiba with 14%, along with smaller contributions from Rio Grande do Norte and Alagoas. There is also a

record in CONAB of the participation of the states of Minas Gerais and Maranhão in the production of mangaba (Vieira Neto et al. 2009; Macedo 2011). Commercialization occurs for the most part at free fairs, in public markets and alongside the highways of the producing regions (Gonçalves et al. 2013).

There are several reasons (such as history, evolution and conservation of the species) that justify studies on genetic divergence including breeding of selected and genetically different parents and guided crosses to obtain segregating and transgressive progenies (Cruz et al, 2011).

Considering the obvious social, economic and cultural potential of mangaba, it is essential to carry out studies with this species in order to generate information for the advancement of genetic improvement and theoretical support for its conservation. According to Falconer and Mackay (1996), knowledge of the degree of genetic variability obtained from divergence studies is fundamental to the process of identifying new source of genes of interest.

Thus, the objective of the present work was to evaluate the relationship between traits and genetic divergence between accessions of mangabeira in guided crossings using multivariate techniques.

Materials and Methods
Twelve family of half sib of *H. speciosa* (coded 1, 3-10, and 12-14) were selected among those originating from Empresa de Pesquisa Agropecuária da Paraíba (EMEPA) and maintained at the AGB of Embrapa Meio-Norte. Genotype selection was based on superior fruit production recorded during the 2011/2012 and 2012/2013 harvestings. The experiment was of randomized block design with four blocks, two genotypes per plot and the mean values of variables determined for each accession were used in the multivariate statistical analysis. Twenty fruits/plant were randomly sampled at the stage of complete physiological maturity and values of fruit mass (FM; g), fruit length (FL; mm), fruit diameter (FD; mm), peel-bagasse mass (PBM; g), number of seeds per fruit (NS), seed mass (SM; g), total soluble solids (TSS; °Bx), pH, total titratable acidity (TTA; % of citric acid), TSS/TTA ratio, pulp mass (PM; g) and PM per pulp yield (PMPY; %) were recorded.

An exploratory analysis of the data was performed using Pearson correlation coefficients to identify possible correlations between the variables and to detect discrepant data points. Subsequently unweighted-pair-group-method-with-arithmetic-mean-analysis (UPGMA) and principal component analysis (PCA) were performed to visualize genetic distances and construct a phylogenetic tree, and to establish common origins between the populations.

The UPGMA algorithm is based on Euclidean distances and generates a dendrogram that reflects the structure present in a pairwise similarity or

dissimilarity matrix. The nearest two clusters are combined into a higher-level cluster and the distance between any two clusters is taken to be the mean distance between elements of each cluster, thereby avoiding the characterization of dissimilarities based on the extreme values of two individuals. This method tends to generate higher values of the cophenetic correlation coefficient (CCC), determined as the Pearson linear correlation between the elements of the dissimilarity matrix and the elements of the cophenetic matrix, and produces less distortion regarding the similarity of the genotypes in the dendrogram. Besides the Euclidian distances were calculated for standardizes data. The CCC was used to evaluate the quality of groupings obtained from UPGMA, with values > 0.70 indicating good representation of clusters (Cruz and Carneiro 2006), and the significance of the coefficient was assessed by the Mantel randomization test.

Hierarchical methods are effective for visualizing the structure of the genomic data but do not provide clusters automatically. The ideal number of significant clusters may be established based on research experience, by visual inspection of the dendrogram considering significant changes in branching, or through application of statistical tools such as those suggested by Mojena (1977). The method employed in this study involved selection of the number of groups in step j that primarily satisfied the inequality $\alpha_j > \theta_k$ in which α_j is the distance of the fusion level corresponding to step j ($j = 1, 2,.., n$), θ_k is the cut-off point given by $\theta_k = \bar{\alpha} + k\hat{\sigma}_\alpha$, $\bar{\alpha}$ is the mean of the values of α, $\hat{\sigma}_\alpha$ is the standard deviation of the values of α, and k is a constant, the value of which was selected to define the optimal number of clusters ($k = 1.25$; Milligan and Cooper 1985).

PCA is an eigenvector-based multivariate analysis technique that explains the internal structure of the data (variance and covariance of a set of variables) by means of linear combinations, referred to as principal components (PCs). The aim of PCA is to reduce the dimensionality of the data and facilitate interpretation of the interdependence between the variables (Johnson and Wichern 1998; Barroso and Artes 2003). For a vector of p-variables $\mathbf{x}^T = (X_1, X_2,\ldots, X_p)$ and covariance matrix cov $(\mathbf{x}) = \sum$, it is possible to calculate the eigenvalues and their respective standardized orthogonal eigenvectors $(\lambda_1, \boldsymbol{\alpha}_1)$, $(\lambda_2, \boldsymbol{\alpha}_2),\ldots,(\lambda_p, \boldsymbol{\alpha}_p)$ where $\lambda_1 \geq \lambda_2 \geq,\ldots,\geq \lambda_p \geq 0$. The i-th PC is then given by $Y_i = \alpha_{11}X_1 + \alpha_{12}X_2 + \ldots + \alpha_{1p}X_p$ in which var $(Y_i) = \lambda_i$, cov $(Y_i, Y_j) = 0$ and α_{ij} measures the importance of the j-th variable in the i-th PC. The PCA biplot (Gabriel 1971) is a graphical representation of PCA in which the physicochemical variables and the individual genotypes are presented in a two-dimensional display (Cartesian coordinate system). In the present study, the biplot was constructed using a low-rank approximation (matrix Y) to a transformed version of the data matrix X data based on singular value

decomposition (DVS). The correlation matrix was used to obtain the PCs (Barroso and Artes 2003) so that the results were not influenced by the magnitude of the variable units. The correlation matrix, PCs and PCA biplot were obtained using R software (R Core Team 2013).

Results and Discussion

The genetic or phenotypic characteristics of a population of a plant species are intrinsic and non-transferable to other populations. For this reason, comparisons between studies involving genetic divergence are not helpful even when similar mathematical approaches have been employed. Furthermore, it is not possible to extrapolate or juxtapose the results from one study to another because the characteristics of a population are the product of the manifestation of particular genes and are determined by the differential influence of the culture environment depending on the degree of heritability of the variables. For these reasons, we have not compared our results with those derived from previous studies involving other mangabeira genotypes, but have limited citations to theoretical inferences. It should be mentioned, however, that exceptions to the above could be justified when considering monogenic variables.

Descriptive statistics and phenotypic correlations relating to physicochemical variables of the 12 mangabeira genotypes studied (Table 1) reveal significant variation regarding, in order of decreasing CVp, the traits TSS/TTA, NS, SM, PM, FM, PBM and TTA. This implies that part of the phenotypic variability is of genetic nature, representing the results of selection processes. In this sense, selection elicits effective genetic gains that are transmitted to future generations so that progenies produce fruits with improved characteristics for the final consumer. According to Ganga et al. (2010), phenotypic variation in native plants from the Cerrado is strongly influenced by uncontrolled environmental factors such as *anthropization*, soil, climate and plant age as well as genetic differences between individuals. ***Table 1.*** Descriptive statistics of 12 variables of 12 genotypes of *Hancornia speciosa* (mangabeira) maintained at the active germplasm bank of Embrapa Meio-Norte (Teresina, PI, Brazil), and mean estimates of phenotypic correlations (below the diagonal) and their corresponding *P* values (above the diagonal) as determined using unweighted-pair-group method with arithmetic mean analysis (UPGMA).

	Descriptive statistics											
	F	FL	FD	PB	NS	SM	TS	pH	TT	TSS/	PM	PM
Mean ($\bar{x}$)	19.	33.	31.	3.0	14.	3.9	17.	3.5	1.6	12.3	13.	63.
Standard	8.0	4.8	4.8	1.1	7.4	2.0	2.9	0.2	0.5	6.46	5.6	8.2
CVp (%)	40.	14.	15.	39.	51.	51.	16.	8.2	35.	52.1	42.	13.
Group I	13.	29.	27,	2.4	11.	2.7	17.	3.6	1.3	14.2	9.2	44.

Group II	19.	33.	31.	2.7	13.	3.7	16.	3.5	1.6	11.3	13.	64.
Group III	24.	35.	33.	3.7	17.	5.0	19.	3.5	1.7	14.3	14.	44.
	Phenotypic correlations											
	F	FL	FD	PB	NS	SM	TS	pH	TT	TSS/	PM	PM
FM	1	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	0.3	0.1	0.31	**0.0**	0.8
FL	0.9	1	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	0.3	0.1	0.72	**0.0**	0.9
FD	0.9	0.8	1	**0.0**	**0.0**	**0.0**	0.0	0.4	0.1	0.49	**0.0**	0.9
PBM	0.8	0.4	0.7	1	**0.0**	**0.0**	**0.0**	0.2	0.3	**0.04**	**0.0**	0.2
NS	0.7	0.7	0.7	0.6	1	**0.0**	**0.0**	0.8	0.2	0.69	0.0	0.3
SM	0.9	0.8	0.9	0.7	0.8	1	**0.0**	0.8	0.2	0.31	**0.0**	0.2
TSS	0.6	0.6	0.5	0.4	0.8	0.7	1	0.9	0.7	0.05	0.1	0.1
pH	-	-	-	-	0.0	-	-	1	0.2	0.71	0.3	0.1
TTA	0.4	0.4	0.4	0.3	0.3	0.4	0.1	-	1	0.27	0.2	0.7
TSS/TTA	0.3	0.1	0.2	0.5	0.1	0.3	0.5	-	-	1	0.2	0.1
PM	0.9	0.8	0.9	0.7	0.5	0.8	0.4	-	0.3	0.33	1	0.8
PMPY	-	0.0	0.0	-	-	-	-	-	0.1	-0.44	0.0	1

Abbreviations: CVp, coefficient of phenotypic variation; FM, fruit mass (g), FL, fruit length (mm); FD, fruit diameter (mm), PBM, peel-bagasse mass (g); NS, number of seeds per fruit; SM, seed mass (g); TSS, total soluble solids (°Bx), pH, TTA, total titratable acidity (units?); PM, pulp mass (g); PMPY, PM per pulp yield (%).
Significant correlations are shown in bold ($P < 0.05$).

The natural pollinating agents of mangabeira are mainly nocturnal insects of the order Lepidoptera, such as sphinx moths or sphingids (Sphingidae), as well as diurnal butterflies (Hesperiidae) and bees of the tribe Euglossini (Apidae) (Vasconcelos, personal communication). Thus, hybrid progenies of mangabeira are likely formed according to the following possibilities: (i) specific fertilization events - when the infrutescence is visited by Lepidoptera, it is expected that seed formation will derive from a crossing of specific parental genotypes since these insects use the proboscis to fertilize the ova; (ii) non-specific fertilization events - when bees visit flowers, pollen adheres to the dense pilose areas of the legs such that seed formation may derive, at least theoretically, from crossings between a greater number of parental phenotypes. In mangabeira, such mixed crossings were likely responsible for the large variability observed in the dimensional characteristics of fruits of both half- and full-sib progenies. Multiple fertilization events can favor high variability in NS and, thereby, determine directly the emergence of fruits with inconsistent dimensions in each of the individuals of the population. Considering that a portion of the phenotypic variation observed in the attributes of the studied genetic population derived from genetic natural effectiveness and genetic gains, it would be expected that the variability in fruit size would be a result of the selection process.

These findings establish, in principle, the complexity of samples that represent genetic variability in *H. speciosa* accessions maintained in the AGB. They also reveal that conservation *in situ* or *in the field* is more viable and effectual since continuous cultivation and management of populations by farmers in agroecosystems in which the crop has evolved should ensure the evolutionary process and survival of the species. Actually, it is probable that the difficulty of domesticating *H. speciosa* has been an important factor in preventing its extinction.
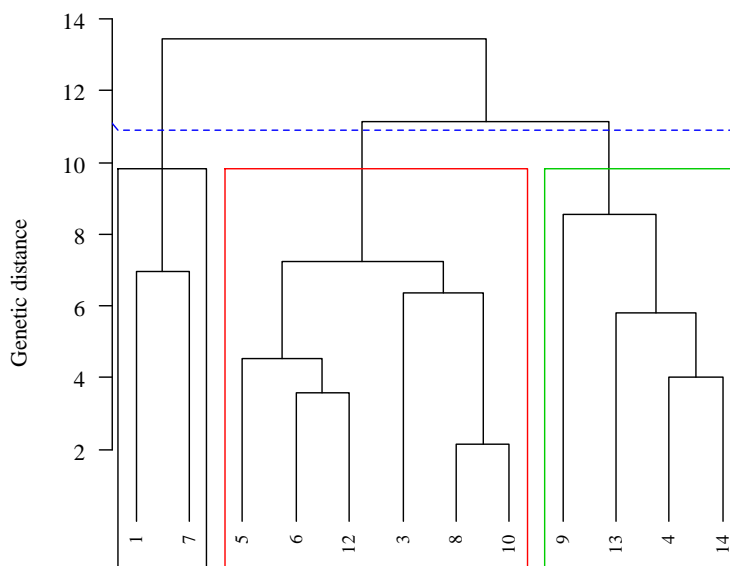
The mean values of the study variables (Table 1) show that the mangabeira population is potentially suitable for selection of genotypes with desirable fruit dimensions for future genetic improvement. The study population consisted of genotypes in the AGB collection that had been pre-selected for the production of fruits that were large at the time of harvesting but not with regard to chemical variables. Nonetheless, the fruits presented high mean TSS, TTA and TSS/TTA values, and these are important traits for the artisanal processing of jams, juices, sweets, ice creams and other products. This type of non-industrial agribusiness is more cost-effective when smaller quantities of artificial additives (sucrose, acidulants, etc) are required to preserve flavor or enhance taste and appearance.

The selection of genotypes among a genetically improved population can be very effective when various traits are combined in one or few homogeneous genotypes (polyclonal variety; ≥ 5 genotypes) from the original population. In such cases, the superior genotypes are expected to exhibit significant positive association with the target traits of the selection process, i.e. the superior genotypes will be considered as a single selection unit. Our results revealed positive and significant phenotypic associations between FM and the dimensional variables FL (0.92), FD (0.97), PBM (0.80), NS (0.76), SM (0.94), TSS (0.65) and PM (0.94). It is noteworthy that the variables FM, Fl. FD, PBM, NS, SM and PM presented positive and significant correlations between themselves except for the moderate positive correlation between PM and NS (0.57), which was not statistically significant ($P = 0.0541$). In addition, TSS was positively correlated with FM (0.65), FL (0.61), FD (0.55), PBM (0.44), NS 0.80), SM (0.72) and, of particular interest, with PM (0.48). Furthermore, the TSS/TTA presented a positive and significant correlation with PBM (0.58). On the other hand, moderate negative correlations were observed between PMPY and PBM (-034), TSS (-0.47), pH (-0.42), and TSS/TTA (-0.44).

Considering that the dimensional variables of the fruits were, in the main, positively and significantly correlated with each other and with TSS, it is possible to state that larger and heavier fruits present higher TSS content. Hence, it is possible to perform an indirect selection of genotypes that produce

sweeter fruit with higher PM based on field measurements of FM and FL, an approach that would facilitate the selection process.

Based on the UPGMA method with Euclidean distances applied to the 12 genotypes of *H. speciosa* (Figure 1), it was possible to identify a cut-off point of 10.88 using one of the procedures suggested by Mojena (1977). According to this cut-off point, three distinct genetically divergent clusters were present, namely: cluster I comprising genotypes 1 and 7, cluster II containing genotypes 5, 6, 12, 3, 8 and 10, and cluster III with genotypes 9, 13, 4 and 14. The calculated CCC was 0.78, a value that was highly significant ($P < 0.0001$) according to Mantel's randomization test, indicating that the three clusters included genotypes presenting a high degree of genetic similarity. On this basis, crossings between individuals belonging to the genetically divergent clusters could be performed with the aim of obtaining heterogeneous and transgressive segregation in progenies expressing good productivity traits (i.e. mean values for fruit/plant mass, pulp/plant mass, pulp percentage and length and diameter of fruit) and generating elite genotypes with high quality fruits (i.e. high levels of TSS and TAA, pH, TSS/TTA and adaptive capacity) simultaneously. In this context, a total of 44 combinations between the genetically divergent individuals of the three clusters could be used in the genetic improvement program at Embrapa Meio-Norte.

Figure 1. Dendrogram based on unweighted pair-group method using arithmetic average (UPGMA) cluster analysis using data (12 physicochemical variables) obtained from 12 genotypes of *Hancornia speciosa* (mangabeira) maintained at the active germplasm bank of Embrapa Meio-Norte (Teresina, PI, Brazil).
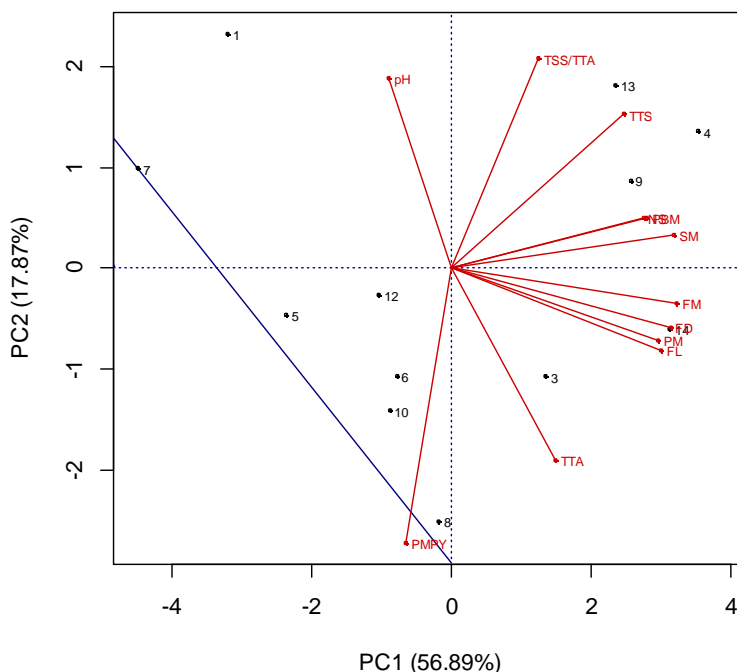
Even though sampling of the individuals maintained in the AGB was not representative, owing mainly to the limited sample size (12 genotypes) employed in the study and the differential reproduction performed *a priori*, the genetic divergence observed among the genotypes of the three clusters would allow successful selection and conservation management. In an earlier study, Silva et al. (2011) determined the genetic divergence in a population of mangabeira maintained in the AGB at Embrapa Tabuleiros Costeiros (Aracaju, SE, Brazil). Even though the sample size was small (only seven accessions), the authors reported separation of the population into two genetically divergent clusters that could be used for selection and genetic improvement, and concluded that future selection strategies should prioritize the most divergent accessions.

Since the mangabeira flowers are small, a crossing scheme could be implemented by making use of the apparent self-fertilization phenomenon, involving fertilization of the ovum by the male gamete of the same flower, and vicinism events in which close individuals tend to cross more frequently (Duarte, personal communication). Thus, each of the 44 planned crossings between the mangabeira genotypes would entail a scheme under field conditions in which each genotype of cluster I would be surrounded individually by each genotype from the other two clusters in order to allow the proximity of a greater number of different parental donors and pollen recipients at each planned crossing.

It is important to note that cluster III presented the highest mean values of FM, FL, FD, PBM, NS, SM, TSS, TAA, TSS/TAA and PM, while clusters I and II exhibited the highest mean values of pH and PMPY, respectively. In contrast, the smallest mean values of FM, FL, FD, PBM, NS, SM, TAA, PM and PMPY were observed in cluster I, while the smallest mean values of TSS and TSS/TTA were observed in cluster II and the lowest pH was observed in cluster III (Table 1). In light of the above, six within-cluster crossings involving individuals of cluster III (9, 13, 4 and 14) could be planned since these genotypes exhibited the highest averages for the target variables relating to the desirable productivity traits that are essential for plant improvement. The genetic improvement program for mangabeira carried out at Embrapa Meio-Norte employs recurrent selection for the continuous maintenance of ideal alleles during the selection process (Resende 2001). Performance of 50 germane crossings of parental genotypes (i.e. 44 between-cluster and six within-cluster) with random sampling of 10 fruits containing 14 seeds/fruit on average in each of the crossings would generate a total of 7,000 (10 x 14 x 50) half- and full-sibs that could be used in the establishment of improved segregating populations.

As shown in the PCA biplot (Figure 2), principal components 1 and 2 explain, respectively, 56.89 and 17.87% of the variability found in the study

population and, taken together, account for 74.5% of the total variance. Thus, the first two components captured most of the cluster structure of the data and grouped the variables that were responsible for most of the phenotypic variance in the study population. The PCA biplot separated the study variables into three groups, namely group V1 comprising pH, TSS/TTA and TSS, group V2 containing PBM, NS, SM, FM, FD, PM and FL, and group V3 with TTA and PMPY. Interestingly, the V1 variables were negatively correlated with those of group V3. However, it should be noted that PCA describes the relationships between variables, genotypes and genotypes x variables without evaluating the statistical significance between these associations. Thus, although moderate positive and negative correlations are displayed in Figure 2, it was not possible to verify the significance of such correlations except for TTA x TSS because, in this case, the two variables were shown to be independent, i.e. the angles of the corresponding vectors were close to 90º.

Figure 2. Principal component analysis (PCA) biplot of 12 physicochemical variables determined for 12 genotypes of *Hancornia speciosa* (mangabeira) maintained at the active germplasm bank of Embrapa Meio-Norte (Teresina, PI, Brazil). Abbreviations: PC1, principal component 1; PC2, principal component 2; FM, fruit mass; FL, fruit length; FD, fruit diameter; PBM, peel-bagasse mass; NS, number of seeds per fruit; SM, seed mass; TSS, total soluble solids, TTA, total titratable acidity, PM, pulp mass; PMPY, PM per pulp yield.

430

Figure 2 shows that the physical variables (FM, FL, FD, PBM, NS, SM and PM) prevailed in component 1 while the chemical variables (TSS, pH, TTA, TSS/TTA and PMPY) prevailed in component 2. This means that, since the physical variables were responsible for most of the variance of the data, the effectiveness and accuracy of the selective process would increase because the measurements of these phenotypic variables are less complex.

Figure 2 also shows that, according to the sense of the vectors, genotypes were assembled into three genetically similar groups, namely group G1 (genotypes 1 and 7), group G2 (genotypes 3, 5, 6, 8, 10 and 12) and group G3 (genotypes 4, 9, 13 and 14). The G1 genotypes were negatively correlated with those of G3, whereas the G3 genotypes were negatively correlated with those of G2. Regarding the relationship between genotypes x variables, it may be observed that genotypes 1 and 7 (G1) were positively correlated with high pH (V1) but negatively correlated with the V2 and V3 variables. The genotypes 3, 6, 8 and 10 (G2) were positively correlated with high TTA and PMPY values but also with low pH, TSS/TTA and TSS values. The genotypes 4, 9, 13 and 14 (G3) were positively correlated with the physical V2 variables and negatively correlated with V3 variables (PMPY and TTA). The G3 genotypes were positively associated with the variables TSS/TTA, TSS, SM, FM, FD, PM and FL, while G1 genotypes were negatively associated with these variables, which is to say that the values of these variables were high in G3 and low in G1. Since the G3 genotypes presented favorable traits *per se* and permitted the best combinations, this heterotic group was pre-selected for within-group crossings.

In conclusion, there was considerable concordance between the results of UPGMA and PCA analysis. Although these statistical tools are based on different principles, both were very efficient in separating the study genotypes into three genetically divergent groups, each of which encompassed the same individuals, on the basis of 12 physicochemical traits. Furthermore, the two methods permitted the detection of positive and significant phenotypic associations between the physical variables (FM, FL, FD, PBM, NS, SM and PM), and of these variables with the chemical variable TSS. These results imply that the genetic divergence detected in the population of *H. speciosa* in the AGB at Embrapa Meio-Norte could be used to perform 50 germane crossings, generate a significant number ($n = 7,000$) of half- and full-sibs and, ultimately, produce improved segregating populations with desirable traits and high-quality fruits. Such genetic improvement program would make possible the establishment of heterogeneous, productive and long-living orchards that would benefit farmers, local people whose livelihoods depend on the artisanal agribusiness and the conservation of the species.

References:
1. Barroso, L. P. & Artes, R. (2003) Análise multivariada. Universidade Federal de Lavras, Lavras, 156p.
2. Cruz, C. D. & Carneiro P. C. S. (2006) Modelos biométricos aplicados ao melhoramento genético. 4th edn, Universidade Federal de Viçosa, Viçosa, 406p.
3. Cruz, C. D., Ferreira, F. M. & Pessoni, L. A. (2011) Biometria aplicada ao estudo da diversidade genética. Suprema, Visconde do Rio Branco, 620p.
4. Falconer, D. S. & Mackay, T. F. C. (1996) Introduction to quantitative genetics. Longman, London, 464p.
5. Gabriel, K. R. (1971) The biplot graphic display of matrices with application to principal component analysis. Biometrika 58: 453-467.
6. Ganga, R. M. D., Ferreira, G. A., Chaves, L. J., Naves, R. V. & Nascimento, J. L. (2010) Caracterização de frutos e árvores de populações naturais de *Hancornia speciosa* Gomes do cerrado. Revista Brasileira de Fruticultura 32: 101-113.
7. Gonçalves, L. G. V., Andrade, F. R., Marimon Júnor, B. H., Schossler, T. R., Lenza, & Marimon, B. S. (2013) Biometria de frutos e sementes de mangaba (*Hancornia speciosa* Gomes) em vegetação natural na região leste de Mato Grosso, Brasil. Revista de Ciências Agrárias 36: 31-40.
8. Johnson, R. A & Wichern, D. W. (1998) Applied multivariate statistical analysis. Prentice Hall International, Upper Saddle River, 816p.
9. Macedo, M. H. G. (2012) CONAB – Mangaba fruto: conjuntura março 2012. Available at <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/12_04_17_1 6_38_30_conjunturamangaba_marco_2012.pdf >. Accessed on October 25, 2017.
10. Milligan, G. W & Cooper, M. C. (1985) An examination of procedures for determining the number of clusters in a data set. Psychometrika 50: 159-179.
11. Mojena, R. (1977) Hierarchical grouping methods and stopping rules: an evaluation. The Computer Journal 20: 359-363.
12. R Core Team (2013), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
13. Resende, M. D. V. (2001) Melhoramento de espécies perenes. In Nas LL, Valois ACC, Melo IS and Valadares-Inglis MC (eds) Recursos genéticos e melhoramento - plantas. Rondonópolis: Fundação MT, p. 357- 421.
14. Scarpinati, E. A., Perecin, D., Paula, R. C., Bonine, C. A. V., Pavan,

B. E & Candido, L. S. (2009) Influência do modelo de análise estatística e da forma das parcelas experimentais na seleção de clones de *Eucalyptus* spp. Revista Árvore 33: 769-776.

15. Silva, A. V. C., Santos, A. R. F., Wickert, E., Silva Júnior, J. F & Costar, T. S. (2011) Divergência genética entre acessos de mangabeira (*Hancornia speciosa* Gomes). Revista Brasileira de Ciências Agrárias 6: 572-578.

16. Viera Neto, R. D., Melo, V.S & Dantas, J. O. (2009) Caracterização do sistema produtivo da mangabeira no município de Itaporanga D'Ajuda, Sergipe. Boletim de Pesquisa e Desenvolvimento 48. Embrapa Tabuleiros Costeiros, Aracaju, 19p.