



## Early selection in a population of the mangaba (*Hancornia speciosa* Gomes)

### *Seleção precoce de uma população de mangaba (Hancornia speciosa Gomes)*

Maria Clideana Cabral Maia<sup>1\*</sup>, Lúcio Borges de Araújo<sup>2</sup>, Carlos Tadeu dos Santos Dias<sup>3</sup>, Luís Cláudio de Oliveira<sup>4</sup>, Lúcio Flavo Lopes Vasconcelos<sup>5</sup>, Paulo Sarmanho da Costa Lima<sup>5</sup>

**Abstract:** The mangaba is a fruit tree with great potential for the northeast of Brazil. Due to the scarcity of improved cultivars, and as it is a species that is still in the process of domestication, exploitation has been based on rational and sustainable extractivism by farmer-gatherers. The aim of this study was to analyse the correlations between technological variables of the mangaba, and to carry out an early selection of the genotypes that make up the base population of the mangaba improvement program of Embrapa Meio-Norte. The variables fruit weight, pulp weight and percentage pulp show significant correlation with fruit length and fruit diameter, allowing indirect selection for the first set of variables by means of the latter, which are easily measured. Genotypes 4, 13, 16, 21, 25, 32, 35 and 49 show a positive association with fruit weight, fruit length, fruit diameter, pulp weight and percentage weight, and are therefore candidates for selection. Principal component 1 is associated with fruit weight, fruit length, fruit diameter, skin weight, number of seeds per fruit, seed weight, TTA and pulp weight. These variables can be selected with greater mathematical certainty, since this component concentrates more information regarding variability and is therefore more important.

**Key words:** Principal Components. Phenotypic Correlations. Genetic Improvement.

**Resumo:** A mangabeira é uma frutífera de grande potencial para a região nordeste do Brasil. Devido à escassez de oferta de cultivares melhorados, sua exploração tem se fundamentado no extrativismo racional e sustentável por agricultores-coletores, sendo uma espécie que ainda se encontra em processo de domesticação. Assim, objetivou-se com este trabalho analisar as correlações entre variáveis tecnológicas da mangaba e realizar uma seleção precoce dos genótipos que compõem a população base do programa de melhoramento da mangaba da Embrapa Meio-Norte. As variáveis massa do fruto, massa da polpa e % da polpa apresentam correlações significativas com as variáveis comprimento do fruto e diâmetro do fruto, permitindo a seleção indireta para as primeiras por meio destas variáveis, por serem de fácil mensuração. Os genótipos 4, 13, 16, 21, 25, 32, 35 e 49 apresentam associações positivas com os atributos massa do fruto, comprimento do fruto, diâmetro do fruto, massa da polpa e porcentagem da massa sendo, portanto, candidatos à seleção. O componente principal 1 está associado às variáveis massa do fruto, comprimento do fruto, diâmetro do fruto, massa da casca, número semente/fruto, massa da semente, ATT e massa da polpa. A seleção dessas variáveis pode ser feita com maior certeza matemática uma vez que esse componente é o que concentra maior informação de variabilidade, sendo, portanto, mais importante.

**Palavras-chave:** Componentes Principais. Correlações Fenotípicas. Melhoramento Genético.

\*Corresponding author

Submitted for publication on 04/12/2019, approved on 06/03/2020 and published on 14/04/2020

<sup>1</sup>Eng. Agrônoma, Doutora, Embrapa Agroindústria Tropical, Brasil. Endereço: Rua Dra. Sara Mesquita, nº 2.270, Bairro Planalto do Pici, 60511-110, Fortaleza, CE. E-mail: [clideana.maia@embrapa.br](mailto:clideana.maia@embrapa.br)

<sup>2</sup>Matemático/Estatístico, Doutor, Universidade Federal de Uberlândia/UFU, Departamento de Ciências Exatas, Brasil. Endereço: Av. João Naves de Ávila, 2121 - Santa Mônica, Uberlândia - MG, 38408-100. E-mail: [lucio@famat.ufu.br](mailto:lucio@famat.ufu.br)

<sup>3</sup>Eng. Agrônomo, Doutor, Universidade Federal do Ceará / UFC Departamento de Ciências do Solo, Brasil, Campus do Pici, BL. Endereço: Av. Mister Hull, s/n - Pici, Fortaleza - CE. 60455-760. E-mail: [ctsdias@ufc.br](mailto:ctsdias@ufc.br)

<sup>4</sup>Eng. Florestal, Mestre, Embrapa Acre, Brasil. Endereço: Rodovia BR-364, Km 14, Rio Branco - AC, 69900-970. E-mail: [luis.oliveira@embrapa.br](mailto:luis.oliveira@embrapa.br)

<sup>5</sup>Eng. Agrônomo. Doutores. Embrapa Meio-Norte, Brasil. Endereço: Av. Duque de Caxias, 5650 - Buenos Aires, Teresina - PI, 64008-780. E-mails: [lucio.vasconcelos@embrapa.br](mailto:lucio.vasconcelos@embrapa.br), [Paulo.costa-lima@embrapa.br](mailto:Paulo.costa-lima@embrapa.br)

## INTRODUCTION

The mangaba is a lactiferous fruit species native to Brazil, of family Apocynaceae, having social, economic and cultural importance in the areas where it occurs, with the state of Sergipe being the largest producer (SILVA JÚNIOR; LÉDO, 2006). The mangaba is widely dispersed geographically, and occurs in various ecosystems, such as the Amazon, Atlantic Forest and Cerrado (DONADIO *et al.*, 2002). Despite the great potential shown by the mangaba both as a fruit tree and for producing rubber, and in the absence of significant rational and technified plantations, extractivism is currently the only form of exploitation, but constitutes a barrier to realising the full potential of the plant.

Consumption of the mangaba offers health benefits due to the high supply of natural antioxidants (ALMEIDA *et al.*, 2011). The plant produces aromatic, delicate, tasty and nutritious fruit, with a protein content of 1.3 to 3.0%, and higher levels of vitamins and minerals than most fruit species, and enjoys excellent acceptance in the market (FERREIRA; MARINHO, 2007). The pulp has a distinctive flavour, a low lipid and calorie content, and is a source of fibre and minerals (SILVA *et al.*, 2008; MARIN *et al.*, 2009). The fruit, in addition to being consumed *in natura*, is used in the production of sweets, syrups, jams, wine, vinegar, liqueur, jelly and especially juice and ice cream (VIERA NETO, 1994; AGUIAR FILHO, 1998). As well as the fruit, the mangaba tree can be used for firewood (LORENZI, 1998).

The species has not been domesticated, and therefore the characteristics to be considered for genetic improvement are still not well established. However, for commercial cultivation to be viable, it is important that selected genotypes have characteristics which are superior to those of plants in their natural state, especially with regard to productivity, production stability and fruit quality. Studies of natural variability are fundamental for designing conservation strategies, since there has been a drastic and worrying reduction in the areas of natural occurrence (GANGA, 2010).

In systematising a program of genetic improvement, an analysis of the distribution of genetic variability, estimates of genetic parameters and a prediction of gains are essential, all of which depend on the different environments in which the genotypes are found. Under the environmental conditions of the state of Piauí, where Embrapa is starting a genetic improvement program of the mangaba, this type of information is scarce, but essential for establishing strategies of conservation and germplasm selection (PINHEIRO *et al.*, 2019). Therefore, the aim of this study was to carry out an early selection of the genotypes that make up the base population of the mangaba breeding program of Embrapa Meio-Norte.

## INTRODUÇÃO

A mangabeira é uma espécie frutífera e laticífera nativa do Brasil, da família Apocynaceae, que apresenta importância social, econômica e cultural nas áreas em que ocorre, sendo o estado de Sergipe o maior produtor (SILVA JÚNIOR; LÉDO, 2006). A mangabeira apresenta ampla dispersão geográfica, ocorrendo em diversos ecossistemas, como Amazônico, Mata Atlântica e Cerrado (DONADIO *et al.*, 2002). Apesar do grande potencial apresentado pela mangabeira como planta frutífera e até mesmo como produtora de borracha, e face à inexistência de plantios racionais e tecnificados significativos, o extrativismo apresenta-se, atualmente, como única forma de exploração, constituindo-se, assim, numa grande barreira ao aproveitamento de todas as suas potencialidades.

O consumo de mangaba oferece benefícios à saúde através do fornecimento elevado de antioxidantes naturais (ALMEIDA *et al.*, 2011). A mangabeira produz frutos aromáticos, delicados, saborosos e nutritivos, com teor de proteína de 1,3 a 3,0%, com teores de vitaminas e sais minerais superiores aos da maioria das espécies frutíferas, tendo excelente aceitação no mercado (FERREIRA; MARINHO, 2007). A polpa possui sabor peculiar, tem baixo conteúdo lipídico e calórico e é fonte de fibras e de minerais (SILVA *et al.*, 2008; MARIN *et al.*, 2009). O fruto, além do consumo *in natura*, é utilizado na produção de doces, xaropes, compotas, vinho, vinagre, licor, geléia e principalmente suco e sorvete (VIERA NETO, 1994; AGUIAR FILHO, 1998). Além do fruto, a mangabeira pode ser usada como lenha (LORENZI, 1998).

A espécie não passou por processo de domesticação e, por isso, os caracteres a serem levados em consideração no melhoramento genético ainda não estão bem estabelecidos. Entretanto, para que o cultivo comercial seja viável, é importante que genótipos selecionados reúnam características superiores às plantas em estado natural, principalmente, com relação à produtividade, estabilidade na produção e qualidade de frutos. Estudos sobre a variabilidade natural são fundamentais no delineamento de estratégias de conservação, uma vez que a redução de seus locais de ocorrência natural tem acontecido de forma drástica e preocupante (GANGA, 2010).

Na sistematização de um programa de melhoramento genético, são imperiosas as análises de distribuição da variabilidade genética, estimativas de parâmetros genéticos e predição de ganhos, os quais dependem dos diferentes ambientes em que se encontram os genótipos. Nas condições ambientais do estado do Piauí, onde a Embrapa está iniciando o programa de melhoramento genético da mangabeira, este tipo de informação é escasso, porém imprescindível ao estabelecimento de estratégias de conservação e de seleção de germoplasma (PINHEIRO *et al.*, 2019). Logo, objetivou-se com este trabalho realizar uma seleção precoce dos genótipos que compõem a população base do programa de melhoramento da mangaba da Embrapa Meio-Norte.

## MATERIAL AND METHODS

The working mangaba collection of Embrapa Meio-Norte comprises 30 half-sibling progeny, planted in a randomised block design. Sixty-nine genotypes of these families were evaluated in this study.

The variables under evaluation were coded as follows: V1: Fruit weight (g); V2: Fruit length (mm); V3: Fruit diameter (mm); V4: Skin weight (g); V5: Number of seeds per fruit; V6: Seed weight (g); V7: Total Soluble Solids - TSS in °Brix; V8: pH; V9: Total Titratable Acidity - TTA; V10: TSS/TTA ratio; V11: Pulp Weight (g); V12: Percentage Pulp.

The genotypes were coded as follows: G1: F10P12 E-ASM; G2: F10P6 E-IAA; G3: F10P8 E-EGF; G4: F10P9 E-EGF; G5: F13P14 E-ASM; G6: F1P10 E-IAA2; G7: F1P12 E-IAA3; G8: F1P14 E-SPA; G9: F1P2 E-FCPS; G10: F1P3 E-FCPS; G11: F2P12 E-IAA2; G12: F2P15 E-SPA; G13: F2P4 E-JB; G14: F2P6 E-JB; G15: F2P7 E-IAA; G16: F2P7 E-IAA2; G17: F2P8 E-IAA2; G18: F3P10 E-TBS; G19: F3P12 E-MBN; G20: F3P13 E-FCPS; G21: F3P14 E-ASM; G22: F3P15 E-IBM; G23: F3P2 E-RUB; G24: F3P6 E-IAA; G25: F3P7 E-JB; G26: F3P8 E-IAA3; G27: F4P10 E-TBS; G28: F4P11 E-EGF; G29: F4P12 E-MBN; G30: F4P4 E-SPA; G31: F4P5 E-IAA2; G32: F4P6 E-IAA; G33: F4P7 E-JB; G34: F4P83 E-FCPS; G35: F4P9 E-IBM; G36: F5P1 E-KTS; G37: F5P15 E-IBM; G38: F5P2 E-IAA2; G39: F5P4 E-EGF; G40: F5P8 E-KTS; G41: F6P10 E-FCPS; G42: F6P11 E-ASM; G43: F6P4 E-RUB; G44: F6P5 E-MBN; G45: F6P6 E-IBM; G46: F6P7 E-IAA; G47: F6P8 E-KTS; G48: F7P1 E-KTS; G49: F7P10 E-IAA2; G50: F7P11 E-JB; G51: F7P13 E-IAA; G52: F7P15 E-ASM; G53: F7P3 E-ASM; G54: F7P4 E-IAA2; G55: F7P5 E-SPA; G56: F8P1 E-KTS; G57: F8P10 E-IAA2; G58: F8P11 E-JB; G59: F8P12 E-IBM; G60: F8P13 E-IAA; G61: F8P4 E-IAA2; G62: F8P6 E-RUB; G63: F8P7 E-FCPS; G64: F9P1 E-TBS; G65: F9P10 E-ASM; G66: F9P11 E-ASM; G67: F9P12 E-ASM; G68: F9P2 E-TBS; G69: F9P8 E-EGF.

Initially, an exploratory analysis of the data was carried out based on Pearson's correlation coefficient, to verify the existence or not of a correlation between the variables under study.

This initial analysis was followed by principal component analysis (PCA). PCA is a technique of multivariate analysis that consists in explaining a complicated structure of variance and covariance of a set of variables using few linear combinations of these variables, in order to reduce the dimensionality of the set of variables, and facilitate interpretation of the interdependence between them (JOHNSON; WICHERN, 2007).

## MATERIAL E MÉTODOS

A coleção de trabalho da mangaba da Embrapa Meio-Norte está constituída por 30 progênies de meios irmãos implantados em delineamento em blocos ao acaso e 69 genótipos dessas famílias foram avaliados nesse estudo.

As variáveis avaliadas foram codificadas da seguinte forma: V1: Massa do Fruto (g); V2: Comprimento do Fruto (mm); V3: Diâmetro do fruto (mm); V4: Massa da casca (g); V5: N° semente/fruto; V6: Massa da semente (g); V7: Sólidos Solúveis Totais - SST em °Brix; V8: pH; V9: Acidez Total Titulável - ATT; V10: Relação SST/ATT; V11: Massa da Polpa (g); V12: % Polpa.

Os genótipos foram codificados da seguinte forma: G1: F10P12 E-ASM; G2: F10P6 E-IAA; G3: F10P8 E-EGF; G4: F10P9 E-EGF; G5: F13P14 E-ASM; G6: F1P10 E-IAA2; G7: F1P12 E-IAA3; G8: F1P14 E-SPA; G9: F1P2 E-FCPS; G10: F1P3 E-FCPS; G11: F2P12 E-IAA2; G12: F2P15 E-SPA; G13: F2P4 E-JB; G14: F2P6 E-JB; G15: F2P7 E-IAA; G16: F2P7 E-IAA2; G17: F2P8 E-IAA2; G18: F3P10 E-TBS; G19: F3P12 E-MBN; G20: F3P13 E-FCPS; G21: F3P14 E-ASM; G22: F3P15 E-IBM; G23: F3P2 E-RUB; G24: F3P6 E-IAA; G25: F3P7 E-JB; G26: F3P8 E-IAA3; G27: F4P10 E-TBS; G28: F4P11 E-EGF; G29: F4P12 E-MBN; G30: F4P4 E-SPA; G31: F4P5 E-IAA2; G32: F4P6 E-IAA; G33: F4P7 E-JB; G34: F4P83 E-FCPS; G35: F4P9 E-IBM; G36: F5P1 E-KTS; G37: F5P15 E-IBM; G38: F5P2 E-IAA2; G39: F5P4 E-EGF; G40: F5P8 E-KTS; G41: F6P10 E-FCPS; G42: F6P11 E-ASM; G43: F6P4 E-RUB; G44: F6P5 E-MBN; G45: F6P6 E-IBM; G46: F6P7 E-IAA; G47: F6P8 E-KTS; G48: F7P1 E-KTS; G49: F7P10 E-IAA2; G50: F7P11 E-JB; G51: F7P13 E-IAA; G52: F7P15 E-ASM; G53: F7P3 E-ASM; G54: F7P4 E-IAA2; G55: F7P5 E-SPA; G56: F8P1 E-KTS; G57: F8P10 E-IAA2; G58: F8P11 E-JB; G59: F8P12 E-IBM; G60: F8P13 E-IAA; G61: F8P4 E-IAA2; G62: F8P6 E-RUB; G63: F8P7 E-FCPS; G64: F9P1 E-TBS; G65: F9P10 E-ASM; G66: F9P11 E-ASM; G67: F9P12 E-ASM; G68: F9P2 E-TBS; G69: F9P8 E-EGF.

Inicialmente, realizou-se uma análise exploratória dos dados, com base no coeficiente de correlação de Pearson, para verificar a existência ou não de correlação entre as variáveis estudadas.

Depois desta análise inicial, procedeu-se com a análise de componentes principais (ACP). A ACP é uma técnica de análise multivariada que consiste em explicar uma estrutura de variâncias e covariâncias complicada de conjunto de variáveis através de poucas combinações lineares dessas próprias variáveis, com o objetivo de reduzir a dimensionalidade do conjunto de variáveis e facilitar a interpretação da interdependência entre elas (JOHNSON; WICHERN, 2007).

For a vector of  $p$  variables  $\mathbf{x}^T = (X_1, X_2, \dots, X_p)$  and with a covariance matrix  $\text{cov}(\mathbf{x}) = \Sigma$ , it is possible to obtain the following linear combinations:

$$\begin{aligned} Y_1 &= \mathbf{I}_1^T \mathbf{x} = l_1 X_1 + l_2 X_2 + \dots + l_{1p} X_p \\ Y_2 &= \mathbf{I}_2^T \mathbf{x} = l_2 X_1 + l_2 X_2 + \dots + l_{2p} X_p \\ &\vdots \\ Y_p &= \mathbf{I}_p^T \mathbf{x} = l_{p1} X_1 + l_{p2} X_2 + \dots + l_p X_p \end{aligned}$$

These linear combinations are called the principal components. The components are independent, and the variances are the largest possible (BARROSO; ARTES, 2003; JOHNSON; WICHERN, 2007). Thus, the  $i$ -th component is a linear combination,  $Y_i = \mathbf{I}_i^T \mathbf{x}$ , which maximises the variability of  $Y_i$ , with the following restrictions:  $\mathbf{I}_i^T \mathbf{I}_i = 1$  e  $\text{Cov}(Y_i, Y_j) = 0$ , for any  $i \neq j$ .

Since  $\Sigma$  is the covariance matrix associated with the vector of random variables  $\mathbf{x}$ , and  $(\lambda_1, \boldsymbol{\alpha}_1), (\lambda_2, \boldsymbol{\alpha}_2), \dots, (\lambda_p, \boldsymbol{\alpha}_p)$  are the eigenvalues and standardised orthogonal eigenvectors associated with  $\Sigma$ , ordered so that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ , the  $i$ -th component is given by:

$$Y_i = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_{1p} X_p \quad (1)$$

where  $\text{var}(Y_i) = \lambda_i$  and  $\text{Cov}(Y_i, Y_j) = 0$ , i.e. the principal components are uncorrelated, and have variances equal to the eigenvalues of  $\Sigma$ . Expression (1) shows that the value of  $\alpha_{ij}$  measures the importance of the  $j$ -th variable for the  $i$ -th principal component.

Based on the PCA, it is possible to create a *biplot* (GABRIEL, 1971) that represents, on a two-dimensional chart, the variables (physical or chemical) and individuals (genotypes). The *biplot* originates with the PCA, and to build the chart, an approximation for the data matrix is sought, based on decomposition into a singular value. For this, consider a matrix  $\mathbf{X}_{(n \times q)}$  with  $n$  rows,  $q$  columns and rank  $p$ , at  $p = \min(n, q)$ . A matrix  $\mathbf{Y}_{(n \times q)}$  is wanted that approximates to  $\mathbf{X}_{(n \times q)}$ , but with a rank equal to 2. This approximation is obtained by the singular value decomposition of  $\mathbf{X}$ :

$$\mathbf{X}_{(n \times q)} = \mathbf{U}_{(n \times p)} \boldsymbol{\Lambda}_{(p \times p)} \mathbf{V}'_{(p \times q)}$$

with:

$\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ ,

$\mathbf{U}$ : orthogonal eigenvector matrix of  $\mathbf{X}\mathbf{X}'$ ;

$\mathbf{V}$ : orthogonal eigenvector matrix of  $\mathbf{X}'\mathbf{X}$ ;

Para um vetor de  $p$  variáveis  $\mathbf{x}^T = (X_1, X_2, \dots, X_p)$  e com matriz de covariâncias  $\text{cov}(\mathbf{x}) = \Sigma$ , é possível obter as seguintes combinações lineares,

Essas combinações lineares são chamadas de componentes principais. Tem-se que os componentes são independentes e as variâncias são as maiores possíveis (BARROSO; ARTES, 2003; JOHNSON; WICHERN, 2007). Assim a  $i$ -ésima componente é uma combinação linear  $Y_i = \mathbf{I}_i^T \mathbf{x}$ , que maximiza a variabilidade de  $Y_i$ , com as seguintes restrições:  $\mathbf{I}_i^T \mathbf{I}_i = 1$  e  $\text{Cov}(Y_i, Y_j) = 0$ , para qualquer  $i \neq j$ .

Sendo  $\Sigma$  a matriz de covariância associado ao vetor de variáveis aleatórias  $\mathbf{x}$  e  $(\lambda_1, \boldsymbol{\alpha}_1), (\lambda_2, \boldsymbol{\alpha}_2), \dots, (\lambda_p, \boldsymbol{\alpha}_p)$  os autovalores e autovetores ortogonais padronizados associados a  $\Sigma$ , ordenados de modo que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ , tem-se que o  $i$ -ésimo componente é dada por:

em que  $\text{var}(Y_i) = \lambda_i$  e  $\text{Cov}(Y_i, Y_j) = 0$ , ou seja, os componentes principais são não correlacionados e têm variâncias iguais aos autovalores de  $\Sigma$ . Pela expressão (1) observa-se que o valor de  $\alpha_{ij}$  mede a importância da  $j$ -ésima variável para o  $i$ -ésimo componente principal.

Com base na ACP, é possível construir gráfico *biplot* (GABRIEL, 1971) que representa, em um gráfico de duas dimensões, as variáveis (físicas ou químicas) e indivíduos (genótipos). O *biplot* origina-se na ACP e para construir este gráfico busca-se uma aproximação para a matriz de dados, com base na decomposição em valor singular. Para isto, considere uma matriz  $\mathbf{X}_{(n \times q)}$  com  $n$  linhas,  $q$  colunas e posto  $p$ , em  $p = \min(n, q)$ . Deseja-se encontrar uma matriz  $\mathbf{Y}_{(n \times q)}$  que se aproxima de  $\mathbf{X}_{(n \times q)}$ , mas que tenha posto igual 2. Essa aproximação é obtida pela decomposição em valor singular de  $\mathbf{X}$ :

com:

$\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ ,

$\mathbf{U}$ : matriz ortogonal de autovetores de  $\mathbf{X}\mathbf{X}'$ ;

$\mathbf{V}$ : matriz ortogonal de autovetores de  $\mathbf{X}'\mathbf{X}$ ;

So, the matrix can be written as:

$$\mathbf{X} = \lambda_1 \mathbf{u}_1 \mathbf{v}_1' + \lambda_2 \mathbf{u}_2 \mathbf{v}_2' + \dots + \lambda_p \mathbf{u}_p \mathbf{v}_p'$$

where:

$\lambda_i$ : is the  $i$ -th nonzero singular value of matrix  $\mathbf{X}\mathbf{X}'$  (or  $\mathbf{X}'\mathbf{X}$ );

$\mathbf{u}_i$ : is the  $i$ -th eigenvalue of  $\mathbf{X}\mathbf{X}'$  associated with eigenvalue  $\lambda_i$ ;

$\mathbf{v}_i$ : is the  $i$ -th eigenvalue of  $\mathbf{X}'\mathbf{X}$  associated with eigenvalue  $\lambda_i$ ;

Thus, matrix  $\mathbf{Y}$  with a value equal to 2 (GOWER; HAND, 1996; GABRIEL, 1971), which approximates to  $\mathbf{X}$  is given by:

$$\mathbf{Y} = \lambda_1 \mathbf{u}_1 \mathbf{v}_1' + \lambda_2 \mathbf{u}_2 \mathbf{v}_2'$$

Consequently,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1' \\ \mathbf{v}_2' \end{bmatrix}$$

To obtain the *biplot* it is necessary to write matrix  $\mathbf{Y}$  as the product of two matrices  $\mathbf{G}\mathbf{H}'$ , where  $\mathbf{G}$  is a matrix of dimension ( $n \times 2$ ) and  $\mathbf{H}$  is a matrix of dimension ( $q \times 2$ ). Thus,  $\mathbf{Y}$  can be written as:

$$\mathbf{Y} = \begin{bmatrix} u_1 \sqrt{\lambda_1} & u_2 \sqrt{\lambda_2} \\ u_2 \sqrt{\lambda_1} & u_2 \sqrt{\lambda_2} \\ \vdots & \vdots \\ u_{1n} \sqrt{\lambda_1} & u_{2n} \sqrt{\lambda_2} \end{bmatrix} \begin{bmatrix} v_1 \sqrt{\lambda_1} & v_2 \sqrt{\lambda_1} & \dots & v_{1n} \sqrt{\lambda_1} \\ v_2 \sqrt{\lambda_2} & v_2 \sqrt{\lambda_2} & \dots & v_{2n} \sqrt{\lambda_2} \end{bmatrix} = \mathbf{G} \mathbf{H}'$$

Therefore, the lines of matrix  $\mathbf{G}$  are the markers for the genotypes ( $\mathbf{X}$  line markers) and the  $\mathbf{H}'$  columns represent the markers for the variables ( $\mathbf{X}$  column markers). With these markers, it is possible to visually check the position of one observation relative to another and the importance of each variable to each genotype. It is also possible to see how the genotypes and variables are grouped. Further details can be found in Gower and Hand (1996).

In order for the results not to be influenced by the magnitude of the units of each variable, the correlation matrix was used to obtain the principal components (BARRSO; ARTES, 2003).

The principal components, the *biplot* and the correlations were obtained using the R software (2012).

Assim a matriz pode ser escrita como:

em que:

$\lambda_i$ : é  $i$ -ésimo valor singular não nulo da matrix  $\mathbf{X}\mathbf{X}'$  (ou de  $\mathbf{X}'\mathbf{X}$ );

$\mathbf{u}_i$ : é o  $i$ -ésimo autovalor de  $\mathbf{X}\mathbf{X}'$  associado ao autovalor  $\lambda_i$ ;

$\mathbf{v}_i$ : é o  $i$ -ésimo autovalor de  $\mathbf{X}'\mathbf{X}$  associado ao autovalor  $\lambda_i$ ;

Assim, a matriz  $\mathbf{Y}$  com igual a 2 (GOWER; HAND, 1996; GABRIEL, 1971), que aproxima  $\mathbf{X}$  é dada por:

Consequentemente,

Para obter o *biplot* é necessário escrever a matriz  $\mathbf{Y}$  como o produto de duas matrizes  $\mathbf{G}\mathbf{H}'$ , onde  $\mathbf{G}$  é uma matriz de dimensão ( $n \times 2$ ) e  $\mathbf{H}$  é uma matriz de dimensão ( $q \times 2$ ). Assim  $\mathbf{Y}$  é escrito como:

Assim, as linhas da matriz  $\mathbf{G}$  são os marcadores para os genótipos (marcadores das linhas de  $\mathbf{X}$ ) e as colunas  $\mathbf{H}'$  representa os marcadores para as variáveis (marcadores das colunas de  $\mathbf{X}$ ). Com esses marcadores, é possível verificar visualmente a posição de uma observação em relação a outra e a importância de cada variável para cada genótipo. Também é possível verificar como os genótipos e as variáveis se agrupam. Mais detalhes podem ser encontrados em Gower e Hand (1996).

Para que os resultados não fossem influenciados pela magnitude das unidades das variáveis, utilizou-se a matriz de correlação para obtenção dos componentes principais (BARRSO; ARTES, 2003).

Os componentes principais, o gráfico *biplot* e as correlações foram obtidos utilizando o software R (2012).



## RESULTS AND DISCUSSION

The genetic or phenotypic characterisation of a genetic population of any species is unique and not transferable to another population. Therefore, comparing the results of studies of a genetic nature, even when employing the same mathematical principle, is irrelevant and misleading. Such results are not susceptible to extrapolation or juxtaposition, since they are the final product of the manifestation of particular genes and, depending on the degree of heritability of the variables, determined by different levels of intrinsic influence from the growing environment. In this work therefore, no comparisons with the results obtained from other genetic populations are reported; one exception can be justified, when the analysis is based on monogenic variables (MAIA *et al.*, 2018; MAIA *et al.*, 2019; PINHEIRO *et al.*, 2019). It was therefore necessary to restrict this study to citing information, and to inferences of a theoretical nature.

The associations between characteristics are due to genetic linkage and/or pleiotropy. In the first case, they are temporary and result from linkage disequilibrium in the population; in the second case, they are permanent, since the same genes are involved in the genetic control of the characteristic (VENCOVSKY; BARRIGA, 1992).

It can be seen from Table 1 that there is a significant correlation between some variables. Fruit weight is positively correlated with fruit length, fruit diameter, number of seeds per fruit and pulp weight; fruit length is positively and significantly correlated with fruit diameter, number of seeds per fruit, seed weight and pulp weight; fruit diameter has a positive correlation with the same variables that correlated with fruit length, it is also strongly related to percentage pulp.

Skin weight is negatively correlated with TSS. Therefore, despite the fact that fruits with thinner skin are sweeter and have higher yields, they may have a shorter shelf life. For Maia *et al.* (2014), firmness is another characteristic that defines the *shelf life* of the fruit.

The number of seeds per fruit, in addition to the correlations already mentioned, also correlates positively with fruit weight, seed weight and pulp weight. Seed weight is positively correlated with the variables fruit weight, fruit length, fruit diameter, number of seeds per fruit and pulp weight, and as expected, negatively with percentage pulp. Pulp weight, in addition to the correlations mentioned above, also has a positive correlation with percentage pulp.

## RESULTADOS E DISCUSSÃO

A caracterização genética ou fenotípica de populações genéticas de quaisquer espécies são próprias e intransferíveis para outras populações, desta forma, as comparações de resultados entre estudos de natureza genética mesmo com emprego de abordagens de mesmo princípio matemático tornam-se impertinentes e equivocadas não sendo também passíveis de extrapolações ou justaposições uma vez que são produto final da manifestação dos genes particulares e, adicionalmente, determinados por níveis diferenciados de influência intrínseca do ambiente de cultivo a depender do grau de herdabilidade de suas variáveis, portanto, neste trabalho não se reportou, para efeito de comparações, a resultados obtidos a partir de outras populações genéticas. Uma ressalva pode ser justificada quando a análise recai sobre variáveis monogênicas (MAIA *et al.*, 2018; MAIA *et al.*, 2019; PINHEIRO *et al.*, 2019). Passando-se assim, necessariamente, neste trabalho, a se restringir a citações de informações e inferências de referencial teórico.

As associações entre caracteres são devidas à ligação gênica e, ou, à pleiotropia. No primeiro caso são temporárias e decorrentes da falta de equilíbrio de ligação na população; no segundo caso, são permanentes, já que os mesmos genes estão envolvidos no controle genético do caráter (VENCOVSKY; BARRIGA, 1992).

Pela Tabela 1, verifica-se que existe correlação significativa entre algumas variáveis. A massa do fruto está correlacionada positivamente com comprimento do fruto, diâmetro do fruto, nº semente/fruto e massa da polpa; o comprimento do fruto esta correlacionado de forma positiva e significativa com diâmetro do fruto, número de semente por fruto, massa da semente e massa da polpa; o diâmetro do fruto esta correlacionado positivamente com as mesmas variáveis que se correlacionaram com o comprimento do fruto, adicionalmente, está fortemente relacionado com % da polpa.

A massa da casca se correlaciona de forma negativa com SST. Assim, a despeito de frutos com casca mais fina serem mais doces e apresentarem maior rendimento, eles podem ter uma vida de prateleira menor. Para Maia *et al.*, (2014) a firmeza é outra característica que define o tempo útil de prateleira (*shelf-life*) dos frutos.

O número sementes/fruto, além das correlações já citadas, ainda se correlaciona de forma positiva com massa do fruto, massa da semente e massa da polpa; a massa da semente está correlacionada de forma positiva com as variáveis: massa do fruto, comprimento do fruto, diâmetro do fruto, nº semente/fruto, massa da polpa; e, como esperado, de forma negativa com % de polpa; para a massa da polpa, além das correlações citadas anteriormente, ela apresenta, também, relação positiva com % de polpa.

Percentage pulp is correlated with the following variables: fruit diameter, seed weight and pulp weight. This variable can therefore be indirectly selected from fruit diameter (MAIA *et al.*, 2016), and can also be measured in the field.

The most important variables for selection are those that reflect the productivity and quality of the fruit of the genetic materials under evaluation. In this case, fruit weight, pulp weight and percentage pulp are the primary components of production, and therefore targets for selection. However, for Cruz (2010), variables of a complex genetic nature are influenced by the environment, and their selection is usually complicated due to the significant stochastic effects involved, however, these variables have significant associations with fruit length and fruit diameter, making their indirect selection possible (Table 1).

A porcentagem de polpa está correlacionada às seguintes variáveis: diâmetro do fruto, massa da semente e massa da polpa. Dessa forma, essa variável pode ser selecionada indiretamente a partir do diâmetro do fruto (MAIA *et al.*, 2016), podendo ser medido ainda em campo experimental.

As variáveis mais importantes para a seleção são aquelas que refletem a produtividade e qualidade dos frutos dos materiais genéticos avaliados. Nesse caso, massa do fruto, massa da polpa e % de polpa constituem os componentes primários da produção, portanto, alvos da seleção. No entanto, para Cruz (2010), variáveis de natureza genética complexa são influenciadas pelo ambiente sendo que a sua seleção, geralmente, é dificultada devido ao efeito estocástico significativo envolvido na sua manifestação, contudo, essas variáveis apresentam associações significativas com as variáveis comprimento do fruto e diâmetro do fruto, permitindo a sua seleção indireta (Tabela 1).

**Table 1** - Correlation between the physical and chemical variables below the main diagonal and *p* values of the respective correlation above the main diagonal

**Tabela 1** - Correlação entre as variáveis físicas e químicas abaixo da diagonal principal e *p* valores da respectiva correlação acima da diagonal principal

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
V1	1	<b>0.000</b>	<b>0.000</b>	0.040	<b>0.000</b>	<b>0.000</b>	0.790	0.090	0.490	0.440	<b>0.000</b>	0.730
V2	<b>0.760</b>	1	<b>0.000</b>	0.550	<b>0.000</b>	<b>0.000</b>	0.470	0.260	0.460	0.480	<b>0.000</b>	0.570
V3	<b>0.730</b>	0.560	1	0.320	<b>0.000</b>	<b>0.000</b>	0.430	0.480	0.530	0.680	<b>0.000</b>	<b>0.000</b>
V4	0.240	0.070	0.120	1	0.030	0.140	<b>0.010</b>	0.910	0.700	0.610	0.620	0.800
V5	<b>0.860</b>	<b>0.710</b>	<b>0.780</b>	0.260	1	<b>0.000</b>	0.460	0.490	0.500	0.370	<b>0.000</b>	0.730
V6	<b>0.890</b>	<b>0.720</b>	<b>0.620</b>	0.180	<b>0.860</b>	1	0.760	0.260	0.470	0.400	<b>0.000</b>	<b>0.010</b>
V7	0.030	0.090	-0.100	<b>-0.330</b>	-0.090	0.040	1	0.410	0.810	0.060	0.820	0.200
V8	-0.210	-0.140	-0.090	-0.010	-0.080	-0.140	-0.100	1	<b>0.030</b>	0.240	0.200	0.280
V9	-0.090	-0.090	0.080	-0.050	-0.080	-0.090	0.030	-0.260	1	<b>0.000</b>	0.500	0.100
V10	0.090	0.090	-0.050	-0.060	0.110	0.100	0.230	0.140	<b>-0.800</b>	1	0.820	0.070
V11	<b>0.770</b>	<b>0.470</b>	<b>0.740</b>	0.060	<b>0.660</b>	<b>0.600</b>	-0.030	-0.150	0.080	-0.030	1	<b>0.010</b>
V12	-0.040	-0.070	<b>0.440</b>	-0.030	0.040	<b>-0.300</b>	-0.160	0.130	0.200	-0.220	<b>0.330</b>	1

V1: Fruit Weight; V2: Fruit Length (mm); V3: Fruit diameter (mm); V4: Skin weight (g); V5: Number of seeds per fruit; V6: Seed weight (g); V7: Total Soluble Solids - TSS in °Brix; V8: pH; V9: Total Titratable Acidity - TTA; V10: TSS/TTA ratio; V11: Pulp Weight (g); V12: Percentage Pulp. Values in bold below the main diagonal: Significant at 5% and 10% probability.

V1: Massa do Fruto; V2: Comprimento do Fruto (mm); V3: Diâmetro do fruto (mm); V4: Massa da casca (g); V5: N° semente/fruto; V6: Massa da semente (g); V7: SST em °Brix; V8: pH; V9: ATT; V10: Relação SST/ATT; V11: Massa da Polpa (g); V12: % Polpa.

Valores em negrito abaixo da diagonal principal: Significativos ao nível de 5% e 10% de probabilidade.

The next step considers the principal component (PC) analysis of the correlation matrix of the interrelated variables. Principal component analysis of the correlation matrix allowed the dimensionality of the interrelated variables to be reduced. From Table 2 it can be seen that PC1 explains 39.01% of the total variability, and PC2 explains 17.59%, which together explain 57% of the total variation.

It can also be seen from Table 2 that PC1 is associated with the variables fruit weight, fruit length, fruit diameter, skin weight, number of seeds per fruit, seed weight, TTA and pulp weight, with the respective coefficients 7.93, 6.57, 6.94, 1.90, 7.71, 7.38, -1.51 and 6.59. The variables linked to PC1 are important as comparative parameters between the remainder, as they extracted most of the standard variance from the data, ultimately lending themselves to selection of the genotypes of this population (full-sib hybrids and half-sib hybrids) derived from the possible crossings (MAIA *et al.*, 2016).

The physical variables capture a greater portion of the standard variation of the data, expanding the effectiveness and accuracy of the selection process as expected, due to being less-metric variables. It is clear that the selection of genetic materials based on these attributes is more efficient, due to the variability extracted from the total sum of squares showing less influence from stochastic effects, and capturing more of the standard variability of the data. These variables can therefore be selected with greater mathematical certainty. On the other hand, PC2 is linked to fruit length, fruit diameter, seed weight, TSS, TTA, pH, TSS/TTA ratio, pulp weight and percentage pulp, which had, respectively, the following coefficients: -1.16, 2.19, -1.59, -2.32, -1.75, 6.95, -7.17, 1.88, and 4.75. According to Johnson and Wichern (2007), PC2 is present in the variability shown by these variables, compromising the efficiency of selection based on these attributes.

For the genotypes, (Table 2), PC1 is represented by genotypes 4, 5, 6, 7, 8, 10, 12, 13, 14, 16, 18, 20, 21, 22, 25, 26, 27, 30, 32, 33, 34, 35, 36, 37, 43, 44, 50, 51, 52, 53, 54, 57, 59, 60, 62, 64, 65, 67, 68 and 69, since they have, respectively, the following coefficients: 3.67, 1.92, 1.43, 1.35, -1.58, -1.15, -1.05, 3.53, 3.36, 3.21, 1.44, -3.01, 3.16, 1.72, 4.36, -3.04, -2.11, -2.68, 4.7, -2.47, -1.72, 7.58, -1.76, -1.02, 4.05, -3.6, -1.09, -1.12, -1.61, -1.21, 1.8, -1.02, 1.61, -4.19, -3.36, -1.33, 1.33, -1.9, -1.58 and -1.15, which are the largest in absolute terms. For PC2, there are the following related genotypes: 4, 7, 11, 12, 13, 14, 15, 18, 19, 20, 22, 24, 26, 32, 40, 48, 54, 55, 58, 59, 61, 67, 68 and 69, which respectively have the following coefficients: 1.01, 1.21, -5.99, 1.28, -1.15, -2.96, 1.43, 1.10, 1.11, 1.77, 1.07, 1.01, -6.09, 1.39, 1.08, -1.45, -1.62, 1.14, -1.35, -2.87, 2.42, -1.06, -2.60 and 1.42.

A próxima etapa considera a análise de componentes principais (CP) da matriz de correlação das variáveis inter-relacionadas. A análise de componentes principais da matriz de correlação permitiu reduzir a dimensionalidade das variáveis inter-relacionadas. Assim, pela Tabela 2, verifica-se que o CP1 explica 39,01% da variabilidade total; e o PC2 17,59%, sendo que juntos explicam 57% da variação total.

Pode-se observar, também, na Tabela 2, que o CP1 está associado às variáveis massa do fruto, comprimento do fruto, diâmetro do fruto, massa da casca, nº semente/fruto, massa da semente, ATT, massa da polpa, com os respectivos coeficientes: 7,93; 6,57; 6,94; 1,90; 7,71; 7,38; -1,51; e 6,59. As variáveis ligadas ao CP1 são importantes como parâmetros comparativos entre as demais por terem extraído maior parte da variância padrão dos dados, se prestando, em última instância, para seleção dos genótipos desta população derivada (híbridos de irmão-completos e híbridos de meios-irmãos) dos cruzamentos possíveis (MAIA *et al.*, 2016).

As variáveis físicas estão captando maior porção da variação padrão dos dados, ampliando a efetividade e acurácia do processo seletivo como esperado por se tratar de variáveis menos métricas. Fica patente que a seleção dos materiais genéticos com base nesses atributos torna-se mais eficiente devido à variabilidade extraída da soma de quadrados total apresentar menos influência de efeitos estocásticos consequentemente maior captação da variabilidade padrão dos dados assim sendo a seleção dessas variáveis pode ser feita com maior certeza matemática. Já o CP2 está ligado ao comprimento do fruto, diâmetro do fruto, massa da semente, SST, ATT, pH, SST/ATT, massa da polpa e % de polpa, que apresentou, respectivamente, os seguintes coeficientes: -1,16; 2,19; -1,59; -2,32; -1,75; 6,95; -7,17; 1,88; e 4,75. De acordo com Johnson e Wichern (2007), o CP2 está presente na variabilidade apresentada por essas variáveis, comprometendo a eficiência da seleção com base nesses atributos.

Para os genótipos, (Tabela 2), tem-se que o CP1 é representado pelos genótipos 4, 5, 6, 7, 8, 10, 12, 13, 14, 16, 18, 20, 21, 22, 25, 26, 27, 30, 32, 33, 34, 35, 36, 37, 43, 44, 50, 51, 52, 53, 54, 57, 59, 60, 62, 64, 65, 67, 68 e 69, pois apresentam, respectivamente, os seguintes coeficientes: 3,67; 1,92; 1,43; 1,35; -1,58; -1,15; -1,05; 3,53; 3,36; 3,21; 1,44; -3,01; 3,16; 1,72; 4,36; -3,04; -2,11; -2,68; 4,7; -2,47; -1,72; 7,58; -1,76; -1,02; 4,05; -3,6; -1,09; -1,12; -1,61; -1,21; 1,8; -1,02; 1,61; -4,19; -3,36; -1,33; 1,33; -1,9; -1,58; e -1,15, que são os maiores em termos absolutos. Para o CP2, têm-se os seguintes genótipos relacionados: 4, 7, 11, 12, 13, 14, 15, 18, 19, 20, 22, 24, 26, 32, 40, 48, 54, 55, 58, 59, 61, 67, 68 e 69, que apresentam, respectivamente, os seguintes coeficientes: 1,01; 1,21; -5,99; 1,28; -1,15; -2,96; 1,43; 1,10; 1,11; 1,77; 1,07; 1,01; -6,09; 1,39; 1,08; -1,45; -1,62; 1,14; -1,35; -2,87; 2,42; -1,06; -2,60; 1,42.



**Table 2** – Principal components 1 (PC1) and 2 (PC2) for the genotypes, and the physical and chemical variables under evaluation*Tabela 2 - Componentes principal 1 (PC1) e 2 (PC2) para os genótipos e variáveis físicas e químicas avaliadas*

Genotype	PC1	PC2	Genotype	PC1	PC2	Genotype	PC1	PC2
G01	-0.2	0.25	G24	-0.95	1.01	G47	0.81	-0.64
G02	-2.3	0.55	G25	4.36	-0.29	G48	-0.36	-1.45
G03	-2.58	0.50	G26	-3.04	-6.09	G49	-0.29	0.91
G04	3.67	1.01	G27	-2.11	0.26	G50	-1.09	-0.21
G05	1.92	0.05	G28	0.37	0.12	G51	-1.12	-0.31
G06	1.43	-0.78	G29	-0.57	0.62	G52	-1.61	-0.16
G07	1.35	1.21	G30	-2.68	-0.19	G53	-1.21	0.53
G08	-1.58	0.06	G31	-0.21	0.36	G54	1.80	-1.62
G09	0.03	0.54	G32	4.7	1.39	G55	0.54	1.14
G10	-1.15	0.35	G33	-2.47	0.31	G56	0.07	0.65
G11	0.36	-5.99	G34	-1.72	0.87	G57	-1.02	0.57
G12	-1.05	1.28	G35	7.58	0.47	G58	-0.62	-1.35
G13	3.53	-1.15	G36	-1.76	-0.49	G59	1.61	-2.87
G14	3.36	-2.96	G37	-1.02	0.85	G60	-4.19	0.77
G15	-0.76	1.43	G38	-0.11	0.52	G61	0.19	2.42
G16	3.21	0.21	G39	-0.37	0.66	G62	-3.36	-0.28
G17	0.52	-0.89	G40	0.72	1.08	G63	0.11	0.16
G18	1.44	1.10	G41	0.29	0.73	G64	-1.33	0.12
G19	0.76	1.11	G42	-0.94	0.19	G65	1.33	-0.84
G20	-3.01	1.77	G43	4.05	-0.35	G66	-0.52	-0.06
G21	3.16	0.26	G44	-3.60	0.15	G67	-1.9	-1.06
G22	1.72	1.07	G45	-0.03	0.24	G68	-1.58	-2.60
G23	-0.05	0.89	G46	0.65	0.5	G69	-1.15	1.42
Variable			PC1	PC2				
Fruit Weight			7.93	-0.60				
Fruit Length			6.57	-1.16				
Fruit Diameter			6.94	2.19				
Skin Weight			1.90	0.47				
Number Seeds/Fruit			7.71	-0.40				
Seed Weight			7.38	-1.59				
Total Soluble Solids - TSS (°Brix)			-0.24	-2.32				
Total Titratable Acidity - TTA			-1.51	-1.75				
pH			-0.42	6.95				
SST/ATT Ratio			0.60	-7.17				
Pulp Weight			6.59	1.88				
Percentage Pulp			0.58	4.75				
<b>Eigenvalue</b>			<b>17.84</b>	<b>11.97</b>				
<b>Var (%)</b>			<b>39.01</b>	<b>17.59</b>				
<b>Accumulated value (%)</b>			<b>39.01</b>	<b>56.60</b>				

The mangaba genetic improvement program of Embrapa Meio-Norte employs the method of recurrent selection in the support and continuous maintenance of alleles that are ideal for the selection process, as described by Resende (2001).

Figure 1 shows that initially the variables fruit weight, fruit length, fruit diameter, number of seeds per fruit, seed weight and pulp weight are related (Table 1). In addition, these variables show positive associations with genotypes 4, 13, 16, 21, 25, 32, 35, 49, i.e. the genotypes show high values for these variables, constituting the promising genotypes of this population, and are the genetic materials that are candidates for selection to form the base population of the mangaba improvement program of Embrapa Meio-Norte. Thus, these genotypes can be used as potential parents in the mangaba breeding program, as a source of genes for larger-sized fruit in the second cycle of recurrent selection.

O programa de melhoramento genético da mangabeira da Embrapa Meio-Norte emprega o método de seleção recorrente, na alimentação e manutenção contínua de alelos ideais ao processo seletivo, apresentado por Resende (2001).

Pela Figura 1, verifica-se que inicialmente as variáveis massa do fruto, comprimento do fruto, diâmetro do fruto, nº semente/fruto, massa da semente e massa da polpa estão relacionadas entre si (Tabela 1). Além disso, essas variáveis apresentam associações positivas com os genótipos 4, 13, 16, 21, 25, 32, 35, 49, ou seja, apresenta altos valores destas variáveis, constituindo-se nos genótipos promissores dessa população. Portanto, os materiais genéticos candidatos à seleção para formação da população base do programa de melhoramento da mangabeira da Embrapa Meio-Norte. Logo, esses genótipos podem ser utilizados como genitores potenciais no programa de melhoramento de mangabeira como fonte de genes para fruto de maior tamanho em segundo ciclo de seleção recorrente.

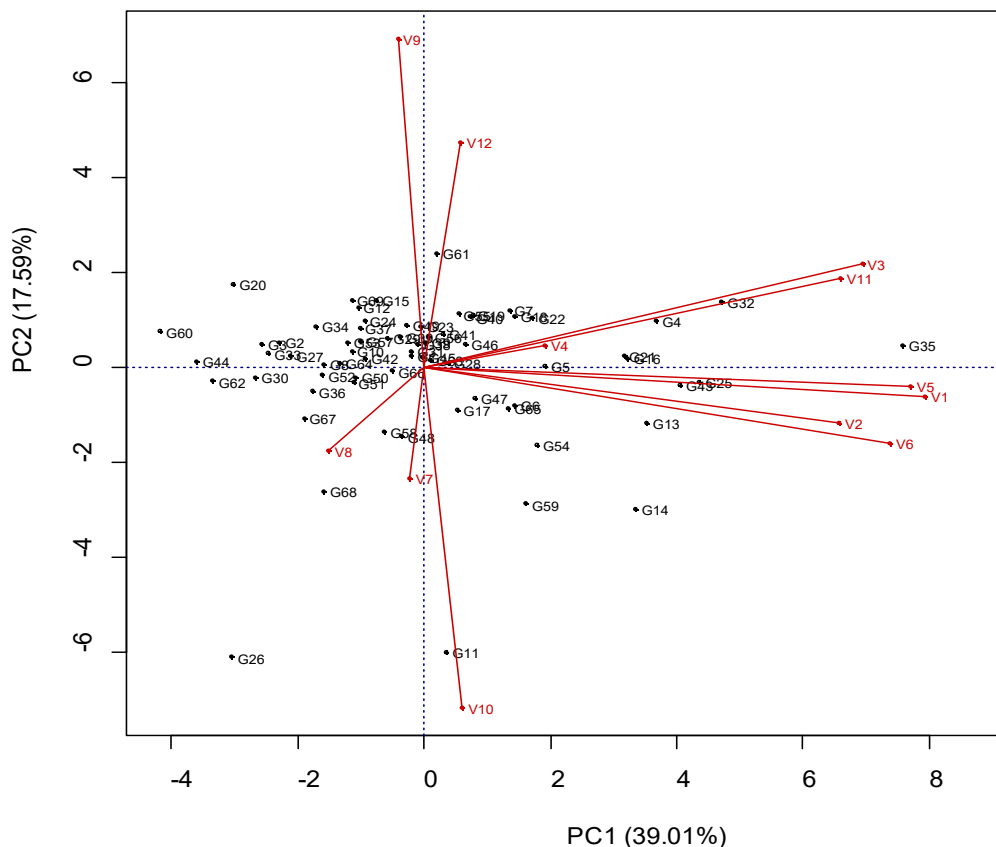


Figure 1 – *Biplot* with two components for the genotypes and the physical and chemical variables.

*Figura 1 - Biplot com dois componentes para os genótipos e as variáveis físicas e químicas.*

One advantage is selecting these genotypes, as they show important agrotechnological attributes for improving the mangaba, and can therefore be selected for future implantation of multi-environment experiments at the recommendation stage of a polyclonal variety representative of the environmental conditions of the mid-north region. (MAIA *et al.*, 2016). Genotypes 20, 44, 60 and 62 show low values for these variables, and can be eliminated from the improvement population of the program.

The variables TTA and percentage pulp are positively correlated and, in turn, show a negative correlation with the TSS/TTA ratio, with genotypes 11 and 26 showing high values for the TSS/TTA ratio and low values for TTA and percentage pulp. It can also be seen that the group formed by the variables TSS/TTA, TTA and percentage pulp shows no correlation with the group formed by the variables fruit weight, fruit length, fruit diameter, number of seeds per fruit, seed weight or pulp weight, since the angle formed by the vectors of these two groups is close to 90°, showing them to be independent (GABRIEL, 1971). All the other genotypes have intermediate amounts of the variables under evaluation.

## CONCLUSIONS

The variables fruit weight, pulp weight and percentage pulp present significant correlations with fruit length and diameter, allowing indirect selection by measuring these variables;

Genotypes 4, 13, 16, 21, 25, 32, 35 and 49 are candidates for selection as they show a positive correlation with the weight, length and diameter of the fruit and with pulp weight;

Principal component 1 is associated with the weight, length and diameter of the fruit, skin weight, number of seeds per fruit, seed weight, TTA and pulp weight. Selection based on these variables can be made with greater mathematical certainty.

Uma vantagem presente consiste na seleção desses genótipos por apresentarem atributos agrotecnológicos importantes para o melhoramento da mangabeira, sendo, portanto, selecionados para implantação futura de experimentos em multiambientes na etapa de recomendação de variedade poli-clonal para as condições representativas das condições ambientais da região Meio-Norte (MAIA *et al.*, 2016). Os genótipos 20, 44, 60 e 62 apresentam baixos valores destas variáveis, podendo ser eliminados da população de melhoramento do programa.

As variáveis ATT e % de polpa apresentam correlação positiva entre si e, por sua vez, apresentam correlação negativa com SST/ATT, sendo que os genótipos 11 e 26 apresentam elevados valores de SST/ATT e baixos valores de ATT e % de polpa. Observa-se, também, que o grupo formado pelas variáveis SST/ATT, ATT e % de polpa não se correlaciona com o grupo formado pelas variáveis massa do fruto, comprimento do fruto, diâmetro do fruto, n° semente/fruto, massa da semente e massa da polpa, pois o ângulo formado pelos vetores destes dois grupos é próximo de 90°, indicando independência entre eles (GABRIEL, 1971). Todos os demais genótipos apresentam quantidades medianas das variáveis avaliadas.

## CONCLUSÕES

As variáveis massa do fruto, massa da polpa e % de polpa apresentam correlações significativas com comprimento e diâmetro do fruto, permitindo por meio da mensuração dessas variáveis a seleção indireta;

Os genótipos 4, 13, 16, 21, 25, 32, 35 e 49 são candidatos a seleção por apresentarem correlações positivas com massa, comprimento e diâmetro do fruto e massa da polpa;

O componente principal 1 está associado a massa, comprimento e diâmetro do fruto, massa da casca, n° semente/fruto, massa da semente, ATT e massa da polpa. A seleção com base nessas variáveis pode ser feita com maior certeza matemática.

## CITED SCIENTIFIC LITERATURE

- AGUIAR FILHO, S. P.; BOSCO, J.; ARAÚJO, I. A. **A mangabeira (*Hancornia speciosa*) domesticação e técnicas de cultivo**. João Pessoa: EMEPA-PB, 1998. 26p. (Documentos, 24).
- ALMEIDA, M. M. B.; SOUSA, P. H. M.; ARRIAGA, A. M. C.; PRADO, G. M.; MAGALHÃES, C. E. C.; MAIA, G. A.; LEMOS, T. L. G. Bioactive compounds and antioxidant activity of fresh exotic fruits from northeastern Brazil, **Food Research International**, v. 44, n. 7, p. 2155-2149, 2011.
- BARROSO, L. P.; ARTE, R. **Análise Multivariada**. Lavras: UFLA, 2003, 156p.
- DONADIO, L. C.; MÔRO, F. V.; SERVIDONE, A. A. **Frutas Brasileiras**. Jaboticabal: Novos Talentos, 2002. 288p.
- CRUZ, C. D. **Princípios de Genética Quantitativa**. Editora UFV. Viçosa: UFV, 2010, 394p.
- FERREIRA, E. G.; MARINHO, S. J. O. Produção de frutos da mangabeira para consumo *in natura* e industrialização. **Tecnologia e Ciência Agropecuária**, v. 1, n. 1, p. 9-14, 2007.
- GANGA, R. M. D.; FERREIRA, F. A.; CHAVES, L. J.; NAVES, R. V.; NASCIMENTO, J. L. Caracterização de frutos e árvores de populações naturais de *Hancornia speciosa* Gomes do cerrado. **Revista Brasileira de Fruticultura**, v. 32, n. 1, p. 101-113, 2010. DOI: <https://doi.org/10.1590/S0100-29452010005000019>
- GABRIEL, K. R. The *biplot* graphic display of matrices with application to principal component analysis. **Biometrika**, v. 58, n. 3, p. 453-467, 1971. DOI: <https://doi.org/10.1093/biomet/58.3.453>
- GOWER, J.C; HAND, D.J. **Biplot**. London: Chapman and Hall, 1996, 277p.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. Madison: Prentice Hall International, 2007, 808p.
- LORENZI, H. **Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil**. 2. ed. Nova Odessa: Plantarum, 1998, 352p.
- MAIA, M. C. C.; RESENDE, M. D. V.; OLIVEIRA, L. C.; VASCONCELOS, L. F. L.; LIMA NETO, J. F. P. Análise Genética em Genótipos de Manga Rosa Via REML/BLUP. **Revista Agrotecnologia**, v. 5, n. 1, p. 1-16, 2014. DOI: <https://doi.org/10.12971/2063>
- MAIA, M. C. C.; MACEDO, L. M.; VASCONCELOS, L. F. L.; AQUINO, J. P. A.; OLIVEIRA, L. C.; RESENDE, M. D. V. Estimates of Genetic Parameters Using Relm/Blup For Intra-Populational Genetic Breeding of *Platonia insignis* Mart.. **Revista Árvore**, v. 40, n. 3, p. 561-573, 2016. DOI: <https://doi.org/10.1590/0100-67622016000300020>
- MAIA, M. C. C.; ARAÚJO, L. B. DE; DIAS, C. T. DOS S; OLIVEIRA, L. C. DE; VASCONCELOS, L. F. L.; CARVALHO JÚNIOR, J. E. V. DE; SIMEÃO, M.; MUNIZ, Y. G. Selection of mango rosa genotypes in a breeding population using the multivariate-*biplot* method. **Ciência Rural**, v. 46, n. 10, p. 1689-1694, 2016. DOI: <https://doi.org/10.1590/0103-8478cr20130722>
- MAIA, M. C. C.; ARAUJO, M. F. C.; ARAÚJO, L. B. DE; DIAS, C. T. DOS S.; OLIVEIRA, L. C.; CRUZ, C. D.; VASCONCELOS, L. F. L.; MACEDO, L. M.; YOKOMIZO, G. K-I.; LIMA, P. S. C. Genetic Divergence Among a Breeding Population of *Hancornia Speciosa* Gomes (Mangabeira) as Determined by Multivariate Statistical Methods. **European Scientific Journal**, v. 14, n. 15, p. 421-433, 2018. DOI: <https://doi.org/10.19044/esj.2018.v14n15p421>
- MAIA, M. C. C.; ALMEIDA, A. S. ; ARAUJO, L. B. ; DIAS, C. T. S. ; OLIVEIRA, L. C. ; YOKOMIZO, G. K. I. ; ROSADO, R. D. S. ; CRUZ, C. D. ; VASCONCELOS, L. F. L. ; LIMA, P. S. C. ; MEDINA-MACEDO, L. Principal Component and *Biplot* Analysis in the Agro-industrial Characteristics of *Anacardium* spp. **European Scientific Journal**, v. 15, n. 30, p. 21-31, 2019. DOI: <http://dx.doi.org/10.19044/esj.2019.v15n30p21>
- MARIN, A. M. F.; SIQUEIRA, E. M. A.; ARRUDA, S. F. Minerals, phytic acid and tannin contents of 18 fruits from the Brazilian savanna. **International Journal of Food Sciences and Nutrition**, v. 60, p. 180-190, 2009. DOI: <http://dx.doi.org/10.1080/09637480902789342>

PINHEIRO, D. S.; MAIA, M. C. C.; PINTO M. F. E.; OLIVEIRA, L. C.; SILVA R., R. D.; YOKOMIZO, G. K. I. Repeatability estimation for mangaba selection using mixed models. **Agro@ambiente On-line**, v. 13, p. 243-255, 2019. DOI: <http://dx.doi.org/10.18227/1982-8470ragro.v13i0.5758>

R Development Core Team (2012), **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

RESENDE, M. D. V. de. Melhoramento de espécies perenes. In: NAS, L.L.; VALOIS, A. C. C.; MELO, I. S. de; VALADARES-INGLIS, M. C. (ed.) **Recursos genéticos e melhoramento - plantas**. Rondonópolis: Fundação MT, 2001. pg. 357-477.

SILVA, M. R.; LACERDA, D. B. C. L.; SANTOS, G. G. S.; MARTINS, D. M. de O. Caracterização química de frutos nativos do cerrado. **Ciência Rural**, v. 38, n. 6, p. 1790-1793, 2008. DOI: <https://doi.org/10.1590/S0103-84782008000600051>

SILVA JUNIOR, J. F. da; LÉDO, A. da S. (Ed.). Botânica. In: SILVA JUNIOR, J. F. da; LÉDO, A. da S. (Ed.). **A cultura da mangaba**. Aracaju: Embrapa Tabuleiros Costeiros, p. 25-33, 2006.

VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Revista Brasileira de Genética, 1992, 496p.