Cabral Maia, Maria Clideana; Borges de Araújo, Lúcio; dos Santos Dias, Carlos Tadeu; de
Oliveira, Luís Cláudio; Flavo Lopes Vasconcelos, Lúcio; Vasconcelos de Carvalho Júnior,
José Eduardo; Simeão, Marcelo; Gagarin Muniz Bastos, Yuri
Selection of mango rosa genotypes in a breeding population using the multivariate-biplot
method

Selection of mango rosa genotypes in a breeding population using the multivariate-biplot method

Seleção de genótipos em uma população de melhoramento de manga rosa usando o método multivariado *biplot*

Maria Clideana Cabral Maia[I*] Lúcio Borges de Araújo[II] Carlos Tadeu dos Santos Dias[III] Luís Cláudio de Oliveira[IV] Lúcio Flavo Lopes Vasconcelos[I] José Eduardo Vasconcelos de Carvalho Júnior[I] Marcelo Simeão[I] Yuri Gagarin Muniz Bastos[I]

ABSTRACT

Mango (***Mangifera indica*** L.) *trees stand out among the main fruit trees cultivated in Brazil. The mango rosa fruit is a very popular local variety (landrace), especially because of their superior technological characteristics such as high contents of Vitamin C and soluble solids (SS), as well as attractive taste and color. The objective of this study was to select a breeding population of mango rosa (polyclonal variety; ≥5 individuals) that can simultaneously meet the fresh and processed fruit markets, using the multivariate method of principal components and the biplot graphic. The principal components, biplot graphic, and phenotype correlations were obtained using the R (2012) software. Pulp percentage and the pulp, skin, and seed mass variables can be indirectly selected using the smallest fruit diameter, which allowed an easier measurement. The P23R AREA3, P30R AREA3, and P32R AREA3 genotypes are selection candidates due to the presence of alleles, which are important agro-technological traits for mango breeding. This study showed that the biplot analysis is a valuable tool for decision making and visualization of interrelationships between variables and genotypes, facilitating the mango selection process.*

Key words: *Mangifera indica* L.*, breeding, biometry.*

RESUMO

*Dentre as principais fruteiras cultivadas no Brasil, destaca-se a mangueira (**Mangifera indica L.**). A manga rosa é uma variedade local bastante apreciada, especialmente devido a suas características tecnológicas superiores, tais como alto teor de Vitamina C, conteúdo de sólidos solúveis totais (SS), bem como sabor e cor atraentes. O objetivo deste estudo foi selecionar uma população de melhoramento de manga rosa (variedade policlonal; ≥5 indivíduos) que atenda simultaneamente aos mercados de frutas frescas e processadas, empregando o método multivariado de componentes principais e o gráfico biplot. Os componentes principais, o gráfico biplot e as correlações fenotípicas foram obtidos usando o software R (2012). As variáveis porcentagem de polpa e as massas de polpa, casca e semente podem ser selecionadas indiretamente por meio do diâmetro menor do fruto, que permite uma mensuração mais fácil. Os genótipos P23R ÁREA3, P30R ÁREA3, e P32R ÁREA3 são candidatos à seleção devido à presença de alelos para caracteres agrotecnológicos, que são importantes para o melhoramento da manga. Este estudo mostrou que a análise biplot é uma ferramenta valiosa para a tomada de decisão e visualização das inter-relações entre variáveis e genótipos, facilitando o processo seletivo da manga.*

Palavras-chave: *Mangifera indica* L., *melhoramento genético, biometria.*

INTRODUCTION

Mango (***Mangifera indica*** L.) trees stand out among the main fruit trees grown in Brazil, which is the seventh largest mango producer in the world, after India, China, Thailand, Indonesia, Pakistan, and Mexico (FAOSTAT - FAO, 2013). Mango fruits are sources of fiber and vitamins such as vitamin C and provitamin A (beta-carotene). However, these fruits are little commercialized in the world, and most of this production is consumed in domestic markets (ITO et al., 2012).

This fruit is a naturalized variety widely adapted to the oscillating conditions and agro-biological fluctuations of the biome that represents the

[I]Embrapa Meio-Norte, 64006-245, Teresina, PI, Brasil. E-mail: clideana.maia@gembrapa.br. *Corresponding author.
[II]Universidade Federal de Uberlândia (UFUB), Uberlândia, MG, Brasil.
[III]Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ), Universidade de São Paulo (USP), Piracibaba, SP, Brasil.
[IV]Embrapa Acre, Rio Branco, AC, Brasil.

Mid-North portion of the Northeast region in Brazil. Their fruits show attractive size, skin and pulp color and smell, thin skin, high sugar content (fructose), intermediate pulp fiber content, and low occurrence of internal collapse (pulp deterioration due to nutritional imbalance between calcium and nitrogen, relative to the Tommy Atkins mango). This variety has a significant participation in the total mango fruit commercialization in this region. Despite the genetic variability accumulated over decades of evolution in these multi-environments, some genotypes exhibit a great uniformity for these characters. In general, a single genotype does not carry all characters searched for in the breeders. In addition, it presents a possible genetic vulnerability along its perennial cycle.

The simple and partial correlation coefficients and path analysis are among the statistics used to measure the association between characters. The first of them provides a single estimate of the association between two characteristics. The second allows to estimate the partial correlations between two characters, eliminating other linear relationships between the two other characters. The third allows a more detailed analysis, unfolding the 'characters' direct and indirect effects on a basic variable (CRUZ & REGAZZI, 1997). Although these analyzes have a wide applicability and usefulness in breeding programs, they do not provide a global evaluation of data patterns. Additionally, they do not allow variable-genotype interrelations to be inferred.

YAN & RAJCAN (2002) proposed an alternative to circumvent these limitations. It is based on a biplot graphical analysis generated from the principal component matrix, where lines are the genotypes and columns are the study variables. The biplot graphical analysis allows to identify the best-performing genotypes for the characteristics assessed (SILVA FILHO et al., 2009).

In an improved cultivar, several desirable characters are simultaneously gathered in an elite genetic material before being recommended to producers. In this context, it becomes clear the importance of studies on correlation and genotype-character interrelations (selection target) in which statistical procedures are used. In such procedures, several variables and the study genotypes are simultaneously considered in an accurate, robust, and integrated (PCA-biplot) analysis. In breeding programs, this analysis is an important tool to identify and rank the superior genotypes.

Given the importance of quantitative characters for breeding programs, studies have been conducted with multivariate techniques for various species such as vine (BORGES et al., 2010; BATISTA et al., 2015), mango (SILVA et al., 2012), banana (ALVES et al., 2012), strawberry (MORALES et al.; 2011), papaya (RODRIGUES et al., 2010), and palm oil (OKOYE et al., 2007) trees.

The aim of this study was to apply the phenotype correlations and genotype-variable interrelations using the multivariate biplot technique to select a mango rosa breeding population (polyclonal variety) that can meet simultaneously the fresh and processed fruit markets.

MATERIALS AND METHODS

The Embrapa Meio-Norte breeding program for mango rosa uses the recurrent selection method presented by RESENDE (2007) to develop varieties combining alleles that dominate the favorable characters of mango rosa and desirable characteristics of commercial varieties involved in crossings.

A biparental cross scheme was implemented in mango rosa varieties as female parents, surrounded by commercial varieties as probable male parents. Due to the proximity effect (Duarte; personal communication), plants that are closer to each other have a higher probability of crossing between themselves. Thus, we can only infer the probable male parents based on the erratic and unsystematic (stochastic) "foraging" by pollinating insects. Molecular analyzes will be required to identify the male parents in the study population.

The genetic material was obtained by open pollination between mango rosa matrices and commercial varieties. The male parents were the following commercial cultivars. Their main characteristics are described as follows (PINTO et al., 2002):

- Irwin: small open-canopy tree; oval and elongated fruits (average: 340-450g); orange-yellow color with deep red hints; yellow soft pulp (little fiber); monoembrionic seed.

- Keitt: sprawling-canopy tree (high productivity); late ripening; large oval fruit (550-740g); yellowish-green color, with red-rosy hints; fiber only around the seed; monoembrionic seed; pulp/fruit ratio about 0.7.

- Palmer: open canopy tree; dark red fruits; thin skin; yellow firm pulp (little or no fiber); monoembrionic seed; pulp/fruit ratio: 0.7.

- Tommy Atkins: full and dense tree; oval medium-size fruits (450g); orange-yellow, covered with intense red and purple color; thick

skin; firm juicy pulp (average fiber content); Brix: medium (17%).

- Amrapali: precocious fruit tree; dwarf-size tree (regular productivity); medium-size fruits; sweet pulp (no fiber).

- Sensation: productive tree (large open canopy); attractive fruits; reddish color; good-quality pulp without fiber (MARANCA, 1985).

After the random samples (five fruits per genotype) reached full physiological maturation, they were evaluated in the hybrid population (19 plants), which was obtained from biparental crosses. The population targeted for the selection process was installed in an experimental field, and experimental design was not used.

The variables evaluated in this study were coded as follows: V1: fruit mass (g); V2: fruit length (mm); V3: largest fruit diameter (base; mm); V4: smallest fruit diameter (apex; mm); V5: skin mass (g); V6: seed mass (g); V7: pulp mass (g); V8: pulp percentage (%); V9: pulp firmness (N); V10: total soluble solid content (SS; °Brix); V11: pH, and V12: SS/acidity ratio.

The *biplot* graphic (GABRIEL, 1971), which represents the physical or chemical variables and individuals (genotypes) in a Cartesian plane (or in the three-dimensional space) is originated from the PCA. To build this graphic, an approximation (matrix Y) is sought for the X data matrix based on decomposition in singular values. Thus, matrix Y can be factored into a product of matrices GH', in which matrix G rows are genotype markers (X line markers), and columns H' represent variable markers (X column markers). With these markers, both

position of an observation relative to another and importance of each variable for each genotype can be verified. Verifying how genotypes and variables are grouped is also possible. More details can be found in GOWER & HAND (1996).

The principal components, biplot graphic, and phenotype correlations were obtained using the R (2012) software and princomp and biplot functions available in the stats package.

RESULTS AND DISCUSSION

Genetic associations between the main target characteristics of mango selection improvement program shown in table 1 and discussed below.

All fruit physical variables (fruit mass and length, smallest and largest fruit diameters, and skin, seed, and pulp masses) except pulp percentage and firmness, showed correlation with each other. However, pulp percentage is associated with the smallest fruit diameter and pulp mass. Thus, fruit, pulp, skin, seed, and pulp masses can be studied as a function of fruit length and diameters, as evaluation of these variables is easy in the experimental field.

Due to the significant association between fruit length and pH, measurements of the first variable can extrapolate the values for the pH variable, whose measurement is more expensive. Similarly, the pulp percentage variable, which requires fruit depulping, can be studied as a function of the total soluble solids content and pulp firmness variables greatly facilitating the selection process.

It can also be noted that a significant phenotypic correlation between pulp percentage

Table 1 - Phenotype correlation matrix for the physical and chemical variables of mango rosa genotypes. Teresina-PI, 2013.

Variables	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
V1	1.00	0.8700*	0.8929*	0.9139*	0.8592*	0.5858*	0.9869*	0.3671	0.1603	-0.3910	-0.4122
V2		1.00	0.7247*	0.7042*	0.8017*	0.4578*	0.8553*	0.2411	0.1482	-0.3542	-0.4786*
V3			1.00	0.9221*	0.7634*	0.7372*	0.8542*	0.2152	0.1309	-0.3408	-0.2582
V4				1.00	0.7087*	0.5511*	0.9140*	0.4663*	0.2003	-0.4497	-0.3558
V5					1.00	0.6349*	0.7828*	0.0034	0.0202	-0.3103	-0.5167*
V6						1.00	0.4690*	-0.3820	-0.1240	-0.0159	-0.0459
V7							1.00	0.4968*	0.2079	-0.4181	-0.4023
V8								1.00	0.5477*	-0.4778*	-0.3069
V9									1.00	-0.5188*	-0.4083
V10										1.00	0.3608

V1: fruit mass; V2: fruit length; V3: largest fruit diameter; V4: smallest fruit diameter; V5: skin mass; V6: seed mass; V7: pulp mass; V8: pulp percentage; V9: pulp firmness; V10: total soluble solids (TSS; °Brix); V11: pH.
*Significant phenotype correlation at 5% level (P<0.05).

exists with the smallest fruit diameter and pulp mass variables. The pulp percentage and mass are the primary components of production. Besides these components, fruit firmness and total soluble solids content, are technological characters that reflect the shelf life and fruit quality, respectively. Therefore, they are target characteristics in mango breeding programs. Thus, pulp percentage, mass, and firmness and total soluble solids content can be studied indirectly by means of the smallest fruit diameter. This scan for association between variables (even if it is classified as moderate) is important to ensure application of multivariate techniques (JOHNSON & WICHERN, 1998).

The principal component analysis of correlation matrix, allowed to reduce the dimensionality of interrelated variables. The first (PC1: 56%) and second (PC2: 20%) principal components explained the total variability. Together, they explained 76% of the total variation (Table 2). This standard variation (noise-free) extracted from the total variability is considered significant due to the complex genetic control, i.e., the quantitative variables considered in this study. Similar values for the total variation were detected by the first principal component and reported for other perennial species. In mango fruit, the chemical variables were explained by PC1 (45.7%), PC2 (20.9%), and PC3 (16.9%) (SILVA et al., 2012). In palm oil (*Elaeis guineensis Jacq*.), they were explained by PC1 (39.7%) and PC2 (21.7%) (OKOYE et al., 2007).

Regarding variables, PC1 is associated with the following variables: fruit mass and length, largest and smallest fruit diameters, and skin and pulp masses. These characters are important as parameters for comparison with others, because most of the standard variance was extracted from the data. Ultimately, the characters are suitable for selection of genotypes of this cross-derived (full- and half-brother hybrids) population.

The PC2 is associated with seed mass, pulp percentage and firmness, total soluble solids content, and pH. The pulp percentage and firmness and total soluble solids content characters extracted a smaller, but significant portion of total variation. Furthermore, there is a strong genetic correlation that qualify them for genotype selection. PINTO et al. (2003) used principal component analysis in the physical, physicochemical, and chemical characterization of true cajá genotypes. They aimed to conduct a breeding research and identify industrial-interest materials, and they observed that the skin and pulp mass, total soluble solids and sugars, and industrial yield characters contributed most to form the PC1 and PC2 components.

Table 2 - Principal Components (PC) 1 and 2 for physical and chemical variables and genotypes of mango rosa fruit. Teresina-PI, 2013.

Variables	IPCA1	IPCA2
V1	0.3940	0.0623
V2	0.3516	0.0513
V3	0.3665	0.1560
V4	0.3749	0.0018
V5	0.3469	0.1927
V6	0.2343	0.4481
V7	0.3868	-0.0173
V8	0.1566	-0.5305
V9	0.1111	-0.5008
V10	-0.2039	0.3716
V11	-0.2074	0.2462
Genotypes		
F21P3 P17C	-0.0703	0.0151
P76C F5P4	-0.2057	0.3702
F1P3 P172C	-0.0991	-0.0827
P47CF12P1	0.3945	0.0029
P16C F28P3	-0.1180	0.0819
F22P3	-0.0019	-0.0780
F1P1	-0.2509	-0.1172
F5P4	-0.2542	0.2395
F13P4	-0.1175	0.1919
F6P1 76C	-0.1637	0.2825
P2 4C	0.4440	-0.2358
P 109C	-0.1947	0.0294
P 62C	-0.2650	-0.5734
P 119C	0.1253	0.1001
P91C FS4	-0.1284	-0.1522
P23R	0.2060	-0.0956
P30R AREA 3	0.1256	-0.2041
P32R AREA 3	0.1377	-0.1707
P72R AREA 3	0.4365	0.3961
Eigenvalues	6.16	2.20
Variances (%)	56.01	20.00
Accumulated variances (%)	56.01	76.01

The PC1 component is dominated by the P76CF5P4, P47CF12P1, F1P1, F5P4, P24C, P109C, P62C, P23R, and P72R AREA3 genotypes, whereas CP2 is dominated by the P76C F5P4, F13P4, F6P1 76C, P24C, P62C, P91CFS4, P30R AREA3, P32R AREA3, and P72R AREA3 genotypes.

Figure 1 shows that the pulp firmness and percentage variables are positively related, whereas total soluble solids content and pH present an inverse relationship. In addition, a correlation is noted between the other variables (fruit mass and length, largest and smallest fruit diameter, and skin, seed, and pulp mass) as shown in table 1.
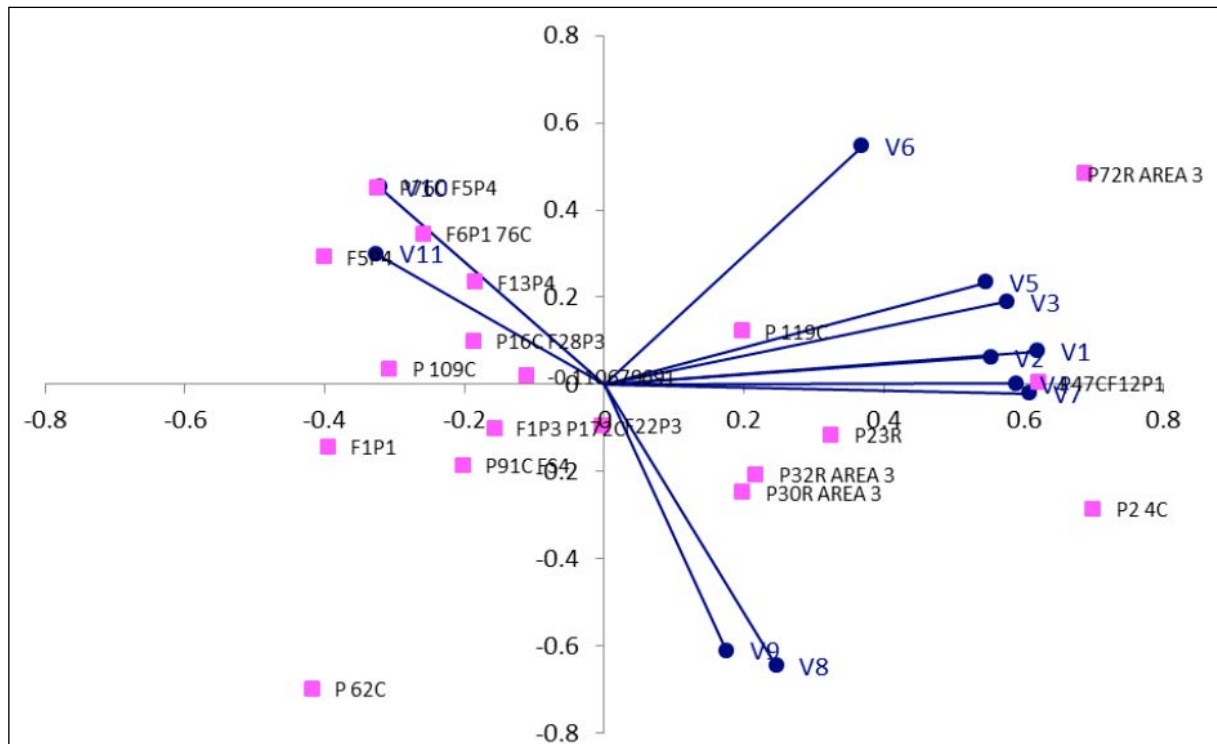
Figure 1 - *Biplot* graphic with two principal components for the physical and chemical variables of mango rosa genotypes, Teresina-PI, 2013.

The P72R AREA3, P47CF12P1, P2 4C, and P119C genotypes are similar among themselves, and have high values for fruit mass and length, largest and smallest fruit diameter, and skin, seed, and pulp mass. However, they have low values for total soluble solids content and pH. Therefore, these genotypes can be used as parents in the mango breeding program as a source of genes for a greater fruit size in a second cycle of recurrent selection. The P72R AREA3 hybrid presents desirable dimensional variables, a good TSS/pH relationship, and increased skin thickness (longer shelf life after harvest), and can; therefore, be both selected in the production of large fruits and utilized in the artisanal and industrial processing of pulp-derived products.

High values for pulp firmness and percentage were observed for the P23R, P30R AREA3, and P32R AREA3 genotypes. Selection of these genotypes is a further advantage, as they have important agrotechnological characters for mango fruit breeding. Therefore, they will be selected for experiments in the polyclonal variety recommendation phase in environmental conditions representing the Mid-North region. CARVALHO et al. (2004) used the principal components technique, identifying and recommending mango cultivars with desirable physicochemical characters for the edaphoclimatic conditions of Votuporanga (SP, Brazil).

The P 119C, F5P4, P76C F5P4, and F6P1 76C genotypes have high soluble solids content and pH values, and low indexes of fruit mass and length, largest and smallest fruit diameter, skin, seed, and pulp mass, and pulp firmness and percentage.

Regarding the F22P3 and F21P3 P17C genotypes, itshowed average values in all variables analyzed, may be involved in backcrossing with the superior genotypes of the studied population.

With a view to obtain segregating genotypes that simultaneously bring together desirable physical characteristics and fruit quality, selection of the P23R, P30R AREA3, P32R AREA3, and P72R AREA3 genotypes was a significant progress in the mango rosa breeding program to form a population according to the recurrent selection cycle of these individuals and compose a complete diallel crossing scheme.

CONCLUSION

The pulp percentage and pulp, skin, and seed mass variables may be indirectly selected by the smallest fruit diameter whose measurement is easier still in the experimental field.

The P23R AREA3, P30R AREA3, and P32R AREA3 genotypes are candidates for selection due to the presence of alleles for agrotechnological characters important for mango fruit breeding.

This study showed that biplot analysis is a valuable tool to make decisions and view variable-genotype interrelations, thus facilitating the mango selection process.

REFERENCES

ALVES, J.S. et al. Divergência genética entre genótipos de bananeira no estado do Rio de Janeiro. **Magistra**, v.24, p.116-122, 2012.

BATISTA, O.F. et al. Divergência genética entre variedades de videiras do Banco Ativo de Germoplasma da Embrapa Semiárido. **Revista Ciência Agronômica**, v.46, p.800-808, 2015.

BORGES, R.M.E. et al. Phenotypic divergence among wine grape accessions in the semi-arid region of Brazil. **Crop Breeding and Applied Biotechnology**, v.10, p.260-265, 2010.

CARVALHO, C.R.L. et al. Avaliação de cultivares de mangueira selecionadas pelo instituto agronômico de Campinas comparadas a outras de importância comercial. **Revista Brasileira de Fruticultura**, v.26, p.264-271, 2004.

FAO. **Organização das Nações Unidas para Alimentação e Agricultura**. Available from: <http://www.fas.fao.org/>. Accessed: Dec. 15, 2015.

GABRIEL, K.R. The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, v.58, p.453-467, 1971.

GOWER, J.C.; HAND, D.J. **Biplots**. London: Chapman and Hall, 1996. 277p.

ITO, A.P. et al. Efeito do processo de desidratação osmótica a pulso de vácuo na transferência de massa e nas propriedades reológicas e de cor de fatias de manga. **Ciência Tecnologia de Alimentos**, v.27, p.54-63, 2012.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. Madison: Prentice Hall International, 1998. 816p.

MARANCA, G. **Fruticultura comercial**: manga e abacate. São Paulo: Nobel, 1985. 138p.

MORALES, R.G.F. et al. Divergência genética em cultivares de morangueiro, baseada em caracteres morfoagronômicos. **Ceres**, v.58, p.323-329, 2011.

OKOYE, M. et al. Genotype by trait relations of oil yield in oil palm (***Elaeis guineensis*** Jacq.) based on GT biplot. **African Crop Science Conference Proceedings**, v.8. p.723-728, 2007.

PINTO, A.C.Q. et al. Principais variedades. In: GENÚ, P.J.C.; PINTO, A.C.Q. (Eds.). **A cultura da mangueira**. Brasília: Embrapa Informação Tecnológica, 2002. Cap.5, p.93-116.

PINTO, W.S. et al. Caracterização física, físico-química e química de frutos de genótipos de cajazeiras. **Pesquisa Agropecuária Brasileira**, v.38, p.1059-1066, 2003.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Austria, 2012. ISBN 3-900051-07-0. Available from: <http://www.R-project.org/>. Accessed: May 29, 2015.

RESENDE, M.D.V. de. Melhoramento de espécies perenes. In: NASS, L.L. **Recursos genéticos vegetais**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2007. p.63-119.

RODRIGUES, H.C de A. et al. Avaliação da diversidade genética entre acessos de mamoneira (***Ricinus communis*** L.) por meio de caracteres morfoagronômicos. **Ceres**, v.57, p.773-777, 2010.

SILVA, D.F.P. et al. Diversidade genética entre cultivares de mangueiras, baseada em caracteres de qualidade dos frutos. **Ceres**, v.59, p.225-232, 2012.

SILVA FILHO, J.L. da et al. Análise biplot genótipos x características em fases iniciais de melhoramento do algodoeiro, Campina Grande, PB, 2009. In: CONGRESSO BRASILEIRO DO ALGODÃO, 7., 2009, Foz do Iguaçu, PR. **Anais...** Foz do Iguaçu: Embrapa Algodão, 2009. p.1506-1511.