CrossMark

ORIGINAL ARTICLE

Path analysis in fruits of the buriti (*Mauritia flexuosa* L. f) in populations from the Brazilian Middle North Region

Fabíola dos Santos Dias[1] · Francisco Charles dos Santos Silva[1] · Ronaldo Silva Gomes[1] · Débora Cristina da Silva[2] · Deyvison de Asevedo Soares[3] · Daniele Piano Rosa[1] · Amilton Ferreira da Silva[4]

Abstract The *Mauritia flexuosa* L.f. tree is of immense socioeconomic significance in the Brazilian Middle North Region for its manifold benefits, but mainly for its fruit. However, the potential of this species has still not been extensively studied. The objectives of this work were to study the phenotypic and genotypic correlations between the characteristics of the *M. flexuosa* fruits and to establish the direct and indirect effects of the secondary characteristics of the fruit on the pulp yield. Samples of ten fruits per genotype were gathered from four natural populations, from 240 different genotypes. These samples were assessed in terms of the fruit, almond and hull weights, equatorial and polar fruit diameters, and the fruit, almond and pulp, volumes, as well as the pulp yield. The genotypic correlations showed greater magnitude, higher than the phenotypic correlations, and both were significant and revealed equal signs. The indirect selection of the *M. flexuosa* genotypes with greater fruit weight and almond volume favored the increase in the pulp yield in this fruit species.

✉ Ronaldo Silva Gomes
ronaldo.s.gomes@ufv.br

[1] UFV, Viçosa, MG, Brazil

[2] UFPE, Serra Talhada, PE, Brazil

[3] Unesp, Ilha Solteira, SP, Brazil

[4] Department of Agronomy, UFJS, Sete Lagoas, MG, Brazil

Introduction

Mauritia flexuosa L.f. is a palm fruit, native to the Central and North regions of South America (Passos and Mendonça 2006). The pulp extracted from the fruits of this palm contains a rich supply of provitamin A carotenoids, vitamin B, and vitamin C, besides minerals such as Ca and Fe (Albuquerque et al. 2003). The fruits of this palm are also abundant in oil, which finds use in a variety of ways like cooking (Lognay et al. 1987), as an additive in polymer manufacturing (Durães et al. 2006) and as a fuel (Goulding and Smith 2007).

The versatility of *M. flexuosa* has escalated this species to a significant socioeconomic status, and its fruits, the chief produce of this species, are extensively consumed in the Brazilian Middle North Region. However, the exploitation of this tree continues to remain merely extractive, highlighting the need for the selection of superior genotypes in order to derive high-quality germplasm.

Besides focusing on the development of the main characteristics, the selection process performed during the breeding programs should simultaneously guarantee the maintenance or improvement of a series of secondary characteristics (Lopes et al. 2002). In this instance, knowledge regarding the relationships between the characteristics, such as those assessed through correlations, are highly relevant in plant breeding, supplying pertinent information to the breeders during the selection process. The correlation estimates facilitate the indirect selection practice for a main characteristic with low heritability and/or of difficult evaluation, thus resulting

in faster genetic gains in relation to direct selection (Cruz et al. 2004).

Estimating the correlations between the main and secondary traits is crucial, for example, in understanding the inheritance of a quantitative trait, such as production. However, the correlations by themselves do not determine the relative importance of the direct and indirect influences of the secondary characteristics on pulp yield. The quantification and interpretation of the coefficients of correlation between two characteristics may cause misconceptions during selection because a high correlation may be due to the effect of a third characteristic or even a group of characteristics (Cruz and Carneiro 2003).

To more clearly understand the associations existing among the different traits, Wright (1921) proposed a method based on the unfolding of the estimated correlations of a set of secondary variables on a basic one, in terms of direct and indirect effects. This method is termed path analysis.

The success in path analysis lies primarily in estimating the cause and effect relationships between the variables. In addition, the unfolding of the correlations depends on the set of traits studied, normally established from prior studies, their importance, and possible interrelationships expressed in the path diagrams (Cruz and Regazzi 1997).

This work aimed at evaluating the phenotypic and genotypic correlations existing between the characteristics present in the *M. flexuosa* fruits and to determine the direct and indirect effects of the secondary characteristics on the pulp yield, in natural populations of this species found in the Brazilian Middle North Region.

Material and methods

The *M. flexuosa* fruit samples were collected from four natural populations found in the municipalities of Pindaré-Mirim, Chapadinha, Brejo and Buriti, in the state of Maranhão, located in the Middle North Region of Brazil. A total of 200 genotypes were sampled. From September to January, the collections were done from plants of different ages during the harvesting period in the region. As a conservation method, the samples were placed in plastic bags containing silica gel until the time of evaluation.

A total of 10 randomly selected fruits were evaluated from each genotype. Evaluations were done in terms of the weights of the fruit (WF), almond (AW), hull (HW), and pulp (PW), by weighing the fruits on a 0.01-g precision electronic scale. Next, the volumes of fruit (VF), almond (AV) and pulp (PV), were assessed, by calculating the difference in volume after the immersion of a given variable in a 200-ml graduated glass. The fruits were also evaluated with respect to the dimensions of equatorial (EDF) and polar diameters (PDF), as well as to the pulp percentage (P%) in relation to the total fruit weight.

For the analysis, while each of the 240 genotypes was considered as a treatment, each of the ten fruits was regarded as a repetition. Initially, the coefficients of the genotypic and phenotypic correlations of the ten characteristics studied were tested by the bootstrap method, with 10,000 simulations.

The estimators of the phenotypic and genotypic correlation coefficients were obtained by the following expressions:

$$Phenotypic\ correlations\ (r\mathrm{P}) = \frac{APBG\ \mathrm{XY}}{\sqrt{ASBG\ \mathrm{X}.\ ASBG\ \mathrm{Y}}}$$

$$Genotypic\ correlations\ (r\mathrm{G}) = \frac{(APBG\ \mathrm{XY} - APBR\ \mathrm{XY})/\mathrm{r}}{\sqrt{\hat{\phi}g(X)\hat{\phi}g(Y)}}$$

$$= \frac{\hat{\phi}g(XY)}{\sqrt{\hat{\phi}g(X)\hat{\phi}g(Y)}}$$

where:

APBG XY	Average product between the genotypes for X and Y;
APBR XY	Average product between the residues for X and Y;
RAS X	Residue average square for X;
RAS Y	Residue average square for Y;
ASBG X	Average square between the genotypes for X;
ASBGY	Average square between genotypes for Y;
$\hat{\phi}g(XY)$	Estimator of genotypic covariance;

$$\hat{\phi}g(X) = \frac{ASBG\ \mathrm{X} - RAS\ \mathrm{X}}{r}\ \hat{\phi}g(Y)$$

$$= \frac{ASBG\ \mathrm{X} - RAS\ \mathrm{X}}{r}$$

where:

$\hat{\phi}g(X)$ and $\hat{\phi}g(Y)$ are the estimators of the quadratic components associated with the genotypic variability for the characters X and Y, respectively.

Path analysis was performed based on the phenotypic correlation indices, where the pulp yield was considered the basic variable. The multicollinearity was evaluated prior to the analysis, according to Cruz and Regazzi (1997).

The degree of multicollinearity of the correlation matrix between the independent variables in the model of regression was established, based on its number of conditions, given by the reason between the greatest and smallest eigenvalues of the correlation matrix. Thus, when the number of conditions is below 100, the multicollinearity is weak and does not imply any problem in the analysis; when it is between 100 and 1000, the multicollinearity is understood to be moderate to strong; and when it is above 1,00,000, the multicollinearity is regarded as severe (Montgomery et al. 1981).

The statistical analysis was done with the support of the Program Genes (Computer App in Genetic and Statistics) (Cruz 2013).

Results and discussion

The phenotypic and genotypic estimates of the correlations were predominantly high in magnitude, above 0.7, for all the pairs of combinations, barring the pair involving the volume and polar diameter of the fruits corresponding to (rP = 0.664) (Table 1) and (rG) = 0.679 (Tables 1 and 2), which were also significant.

The phenotypic correlation between the characteristics is a directly measurable correlation (Falconer and Mackay 1981), with genetic and environmental causes. However, only the causes of genetic origin include an inheritable association and, therefore, only this can be used in the breeding program orientations (Cruz and Regazzi 1997).

Based on this, no significant differences were observed, either in signal or contrasts of high magnitude between the phenotypic and genotypic correlation values. This revealed that the environment did not significantly exert any influence to the extent of overcoming the effect of the genotype.

Although the coefficients of linear correlation are valuable in quantifying the magnitude and direction of the factors that influence the determination of a complex character, they do not reveal the exact relative importance of the direct and indirect effects of these factors on this characteristic. In this case, path analysis is required to perform the unfolding of the correlation coefficients (Cruz and Regazzi 1997).

Prior to performing the path analysis, the multicollinearity was assessed by the condition number of the matrices of phenotypic and genotypic correlations, showing a moderate multicollinearity problem. A system of regression analysis "*trilha em crista*" was adopted in order to reduce these effects (Carvalho and Cruz 1996).

Most studies involving path analysis have not taken into consideration the adverse effects of multicollinearity on the estimators of minimum squares adopted in the resolution of the equations systems. Thus, the results obtained in such conditions may not be reliable. One of the effects of high multicollinearity includes unstable estimates of the coefficients of regression and overestimations of the direct effect of explanatory variables on a main variable, which can lead to erroneous results (Cruz and Carneiro 2003).

The *M. flexuosa* fruit pulp comprises the most important characteristic in the breeding of this species, and this is the commercial product with highest market value. According to Barbosa et al. (2010) the fruit pulp percentage in *M. flexuosa* varies between 13.3% and 14.9%, a very low ratio, compared to most of the fruit trees domesticated till date. Additionally, the exact measurement of this characteristic is extremely

Table 1 Phenotypic correlations (rP) between the ten characteristics of the fruits evaluated in the 240 *M. flexuosa* genotypes from four natural populations in the Brazilian Middle North Region

Variables	EDF	PDF	FV	PY	AW	AV	HW
WF	0.914*	0.801*	0.874*	0.925*	0.958*	0.901*	0.803*
EDF		0.881*	0.751*	0.850*	0.890*	0.781*	0.841*
PDF			0.664*	0.796*	0.801*	0.707*	0.809*
FV				0.716*	0.842*	0.792*	0.737*
PY					0.853*	0.891*	0.805*
AW						0.846*	0.763*
AV							0.698*
HW							

*Significant at 1% by bootstrap method, with 10,000 simulations

WF (weight of fruit, g/fruit), *PY* (pulp yield, g/fruit), *EDF* (equatorial diameter of fruit, mm), *PDF* (polar diameter of fruit, mm), *VF* (volume of fruit, ml), *AW* (almond weight, g), *AV* (almond volume, ml), and *HW* (hull weight, g)

difficult, as the pasty pulp of the fruit is strongly adhered to the fruit hull and almond, which justifies the investigation for correlated characteristics that result in higher efficiency in the indirect selection for this character.

The fruit weight showed positive phenotypic and genotypic correlations with pulp yield (Tables 1 and 2). Almost all of the phenotypic correlation values observed occurred because of the direct effect of this variable. The fruit weight also positively affected the pulp yield through all the variables analyzed, suggesting that this is the most important characteristic in the selection for pulp yield in *M. flexuosa* (Table 3).

A high positive and significant phenotypic correlation was evident between the equatorial diameter of fruit and pulp yield (Table 1). Although an expressive indirect effect of this characteristic through the fruit weight and almond volume on pulp

Table 2 Genotypic correlations (rG) between the characteristics of the fruits evaluated in the 240 *M. flexuosa* genotypes from four natural populations in the Brazilian Middle North Region

Variables	EDF	PDF	FV	PY	AW	AV	HW
FW	0.924*	0.812*	0.884*	0.938*	0.985*	0.930*	0.819*
EDF		0.895*	0.766*	0.865*	0.924*	0.817*	0.863*
PDF			0.679*	0.813*	0.834*	0.742*	0.829*
FV				0.736*	0.881*	0.822*	0.756*
PY					0.874*	0.916*	0.819*
AW						0.865*	0.777*
AV							0.722*
HW							

*Significant at 1% by bootstrap method, with ten thousand simulations

WF (weight of fruit, g/fruit), *PY* (pulp yield g/fruit), *EDF* (equatorial diameter of fruit, mm), *PDF* (polar diameter of fruit, mm), *VF* (volume of fruit, ml), *AW* (almond weight, g), *AV* (almond volume, ml) and *HW* (hull weight, g)

Table 3 Unfolding of the phenotypic correlations through the direct and indirect effects, involving the main dependent variable *PY* (pulp yield, in g/fruit) and the independent explanatory variables *EDF* (equatorial diameter of fruit, mm), *PDF* (polar diameter of fruit, mm), *VF* (volume of fruit, ml), *AW* (almond weight, g), *AV* (almond volume, mm), and *HW* (hull weight, g) in 240 genotypes of *M. flexuosa*

Characters			Phenotypic correlation	
Weight of fruit				
	Direct effect on PY	0.529	–	–
	Indirect effect through EDF	–	0.014	–
	Indirect effect through PDF	–	0.099	–
	Indirect effect through VF	–	−0.168	–
	Indirect effect through AW	–	−0.028	–
	Indirect effect through AV	–	0.336	–
	Indirect effect through HW	–	0.115	–
Total		–	–	0.925
Equatorial diameter of fruit				
	Direct effect on PY	0.015	–	–
	Indirect effect through WF	–	0.484	–
	Indirect effect through PDF	–	0.109	–
	Indirect effect through VF	–	−0.144	–
	Indirect effect through AW	–	−0.026	–
	Indirect effect through AV	–	0.291	–
	Indirect effect through HW	–	0.121	–
Total		–	–	0.850
Polar diameter of fruit				
	Direct effect on PY	0.124	–	–
	Indirect effect through WF	–	0.424	–
	Indirect effect through EDF	–	0.013	–
	Indirect effect through VF	–	−0.127	–
	Indirect effect through AW	–	−0.024	–
	Indirect effect through AV	–	0.264	–
	Indirect effect through HW	–	0.116	–
Total		–	–	0.796
Volume of fruit				
	Direct effect on PY	−0.192	–	–
	Indirect effect through WF	–	0.462	–
	Indirect effect through EDF	–	0.011	–
	Indirect effect through PDF	–	0.082	–
	Indirect effect through AW	–	−0.025	–
	Indirect effect through AV	–	0.296	–
	Indirect effect through HW	–	0.092	–
Total		–	–	0.7161
Almond weight				
	Direct effect on PY	−0.030	–	–
	Indirect effect through WF	–	0.507	–
	Indirect effect through EDF	–	0.013	–
	Indirect effect through PDF	–	0.099	–
	Indirect effect through VF	–	−0.162	–
	Indirect effect through AV	–	0.316	–
	Indirect effect through HW	–	0.110	–
Total		–	–	0.853
Almond volume				
	Direct effect on PY	0.373	–	–

Table 3 (continued)

Characters		Phenotypic correlation		
	Indirect effect through WF	–	0.477	–
	Indirect effect through EDF	–	0.012	–
	Indirect effect through PDF	–	0.088	–
	Indirect effect through VF	–	−0.152	–
	Indirect effect through AW	–	−0.025	–
	Indirect effect through HW	–	0.100	–
Total		–	–	0.891
Hull weight				
	Direct effect on PY	0.144	–	–
	Indirect effect through FW	–	0.425	–
	Indirect effect through EDF	–	0.013	–
	Indirect effect through PDF	–	0.100	–
	Indirect effect through VF	–	−0.122	–
	Indirect effect through AW	–	−0.023	–
	Indirect effect through HW	–	0.261	–
Total		–	–	0.805
Effect of the residual variable	0.337			
K value used in the analysis	0.051			
R^2	0.886			

yield had been noted in the analysis, the direct effect of the equatorial diameter of fruit on pulp yield was very low (Table 3). These effects are responsible for the high phenotypic correlation between the equatorial diameter of fruit and pulp yield (Table 1). The identical effect was also observed for polar diameter of fruit.

All the variables exerted a low and negative indirect effect on the pulp yield via the fruit volume and almond weight (Table 3). Although these variables revealed a high positive and significant phenotypic and genotypic correlation with pulp yield (Tables 1 and 2), their direct effects on this characteristic were low and negative. This shows that an intensified selection of fruit volume and almond weight may not give sufficient genetic gains in pulp yield in this species.

The almond volume showed significant phenotypic and genotypic correlations with pulp yield (Tables 1 and 2). The direct effect of this first characteristic on pulp yield was 0.3733, which was greater than the residual value (Table 3). Fruits presenting larger volumes of almonds tended to have a greater surface for pulp allocation. On the other hand, fruits with this characteristic were also likely to be lighter, which might be related to a higher reallocation of the photoassimilates during the pulp formation. This suggests that the selection of the more productive genotypes might be achieved by selecting the ones with larger-sized almonds.

Significant phenotypic and genotypic correlations were noted between the hull weight and pulp yield (Tables 1 and 2). However, the direct effect of this first characteristic on

pulp yield was insignificant (Table 3). This justifies the reason that a study based only on the correlation between these characteristics would not be efficient in the selection of characteristics for plant breeding. Gondim et al. (2008) observed an identical case between the dry biomass of the spike straw and grain yield in wheat.

In this regard, Okuyama et al. (2004) observed that the relationships between yield and its components determined through path analysis showed differences from those derived via the simple analysis of correlations. The apparent divergence is due to the analytical approach. While simple correlation identifies mutual associations between characteristics, the path analysis permits the determination of the relative magnitude of each effect (Cruz and Regazzi 1997).

The determinant coefficient of the path analysis model (R^2) demonstrated that 88.6% of the pulp yield variations in *M. flexuosa* could be explained by the effects of the secondary variables analyzed (Table 3). A few studies that address the linear relationships between the morphological characters and fruit production in this species are available in the literature, highlighting the significance of the findings from this study.

Conclusion

Despite finding positive and significant phenotypic and genotypic correlations with pulp yield, the variables fruit volume and almond weight exerted negative direct effects on the pulp

yield, suggesting that the indirect selection of the more productive *M. flexuosa* genotypes based on fruit volume and almond weight might not prove efficient.

The phenotypic and genotypic correlations and the path analysis indicated that the characteristics of fruit weight and almond volume can be used in the indirect selection of the *M. flexuosa* genotypes with higher pulp yield.

Data archiving statement   As recommended, the process of submitting the data in this work to "TreeGenes Database," https://dendrome.ucdavis.edu/index.php, has begun, as this seemed the most suitable. Once the submission process is completed, the TGGE editorial office will be notified.

References

Albuquerque MLS, Guedes I, Alcantra Júnior P et al (2003) Infrared absorption spectra of buriti (*Mauritia flexuosa* L.) oil. Vib Spectrosc 33:127–131. https://doi.org/10.1016/S0924-2031(03)00098-5

Barbosa RI, Lima AD, Mourão Júnior M (2010) Biometry of the buriti fruits (*Mauritia flexuosa* L. f. - *Arecaceae*): pulp and oil production in a savanna area of Roraima Amazônia. Ciência e Desenvolvimento 5:71–85

Carvalho SP, Cruz CD (1996) Diagnosis of multicollinearity: assessment of the condition of correlation matrices used in genetic studies. Braz J Genet 19:479–484

Cruz CD (2013) GENES - a software package for analysis in experimental statistics and quantitative genetics. Acta Scientiarum 35:271–276

Cruz CD, Carneiro PCS (2003) Modelos biométricos aplicados ao melhoramento de plantas Viçosa. UFV Press, Viçosa

Cruz CD, Regazzi AJ (1997) Modelos biométricos aplicados ao melhoramento genético. UFV Press, Viçosa

Cruz CD, Regazzi AJ, Carneiro PCS (2004) Modelos biométricos aplicados ao melhoramento genético. UFV Press, Viçosa

Durães JA, Drummond AL, Pimentel TAPF et al (2006) Absorption and photoluminescence of buriti oil/polystyrene and buriti oil/poly(methyl methacrylate) blends. Eur Polym J 42:3324–3332

Falconer DS, Mackay TFC (1981) Introduction to quantitative genetics. Longmans Green, Harlow

Gondim TCO, Rocha VS, Sediyama CS et al (2008) Path analysis for yield components and agronomic traits of wheat under defoliation. Pesq Agrop Brasileira 43:487–493

Goulding M, Smith N (2007) Palms: sentinels for Amazon conservation. Missouri Botanical Garden Press, Saint Louis

Lognay G, Trevejo E, Jordan E et al (1987) Investigaciones sobre el aceite de *Mauritia flexuosa* L. Consejo Superior de Investigaciones Científicas. Instituto de la Grasa, Sévilla

Lopes ACA, Vello NA, Pandini F et al (2002) Variability and correlations among traits in soybean crosses. Sci Agric 59:341–348. https://doi.org/10.1590/S0103-90162002000200021

Montgomery DC, Peck EA, Vining GG (1981) Introduction to linear regression analysis. John Wiley & Sons, New York

Okuyama LA, Federizzi LC, Barbosa Neto JF (2004) Correlation and path analysis of yield and its components and plant traits in wheat. Ciênc Rural 34:1701–1708. https://doi.org/10.1590/S0103-84782004000600006

Passos MAB, Mendonça MS (2006) Epidermis of leaf segments from *Mauritia flexuosa* L.f. (Arecaceae) in three phases of development. Actaamazonica 36:431–436. https://doi.org/10.1590/S0044-59672006000400005

Wright S (1921) Correlation and causation. J Agric Res 20:557–585