GMR

Path analysis and canonical correlations for indirect selection of *Jatropha* genotypes with higher oil yield

L.A. Silva[1], L.A. Peixoto[1], P.E. Teodoro[1], E.V. Rodrigues[2], B.G. Laviola[2] and L.L. Bhering[1]

[1]Universidade Federal de Viçosa, Viçosa, MG, Brasil
[2]Embrapa Agroenergia, Parque Estação Biológica, Brasília, DF, Brasil

Corresponding author: L.L. Bhering
E-mail: leonardo.bhering@ufv.br

Genet. Mol. Res. 16 (1): gmr16019562
Received December 2, 2016
Accepted January 19, 2017
Published March 22, 2017
DOI http://dx.doi.org/10.4238/gmr16019562

ABSTRACT. *Jatropha* is a species with great potential for biodiesel production, and the knowledge on how the main agronomic traits are correlated will contribute to its improvement. Therefore, the objectives of this study were to estimate the genetic parameters of the traits: plant height at 12 and 40 months, canopy projection on the row at 12 and 40 months, canopy projection between the row at 12 and 40 months, number of branches at 40 months, grain yield, and oil yield; to verify the existence of phenotypic correlation between these traits; to verify the influence of the morphological traits on oil yield by means of path analysis; and to evaluate the relationship between the productive traits in *Jatropha* and the morphological traits measured at different ages. Sixty-seven half-sib families were evaluated using a completely randomized block design with two replications and five plants per plot. Analysis of variance was used to estimate the genetic value. Phenotypic correlations were given by the Pearson correlation between traits. For the canonical correlation analysis, two groups of traits were established: group I, consisting of traits of economic importance for the culture, and group II, consisting of morphological traits. Path analysis was carried out considering oil yield as the main dependent variable. Genetic variability was observed among *Jatropha* families. Productive traits can be indirectly selected via morphological traits due to the correlation between these two groups of traits. Therefore, canonical correlations and path analysis are two strategies that may be useful in *Jatropha*-breeding program when the objective is to select productive traits via morphological traits.

Key words: *Jatropha curcas*; Biometrics; Breeding; Quantitative genetics

INTRODUCTION

Jatropha curcas L. is a perennial species, and it can be used as hedges, or for phytoremediation and medicinal purposes (Sharma et al., 2012). The grains are the most appreciated part of the plant for presenting high oil content (Montes and Melchinger, 2016), reaching 1500 kg/ha from the fourth year of cultivation (Laviola et al., 2014), and thus, it can be used for biofuel production. Despite the enormous energetic potential, *Jatropha* is still under domestication in Brazil. Thus, efforts have been made to develop *Jatropha* cultivars with high oil yield, contributing to the insertion of this crop in the world's bioenergy scenario.

One of the main objectives of *Jatropha*-breeding programs is the selection of genotypes with high yield potential (Bhering et al., 2013; Peixoto et al., 2016; Silva Junqueira et al., 2016). To this end, knowing the association between traits is necessary, especially if selecting one of them is difficult, owing to the low heritability and/or difficulty in measurement and identification. Estimates of correlation coefficient make it possible to evaluate the magnitude and direction of the relationship between two traits, and consequently the possibility of obtaining gains for one of them by means of indirect selection via the other trait. In some cases, indirect selection based on the correlated response can be more effective and faster than direct selection of the desired trait (Cruz et al., 2012).

Despite the great usefulness of the correlation coefficients in breeding programs, they only estimate the association between traits, and do not allow conclusions on cause and effect relationships. Path analysis studies the unfolding of the correlation coefficient developed by Wright (1923). This analysis consists of the study on the direct and indirect effects of traits on a main variable, whose estimates are obtained by means of regression equations, in which the variables are previously standardized (Vencovsky and Barriga, 1992; Cruz et al., 2012).

Studies on the association between productive traits in *Jatropha* are scarce, and they consider grain yield as the main dependent variable (Laviola et al., 2011; Teodoro et al., 2016). Therefore, the objectives of this study were to verify the influence of the morphological traits on oil yield by means of path analysis, and to evaluate the relationship between morphological traits measured at different ages and the productive traits in *Jatropha* via canonical correlations.

MATERIAL AND METHODS

Experimental design

The experiment was installed in the experimental area of Embrapa Cerrados, in Planaltina, DF, Brazil (latitude 15°35'30"S, longitude 47°42'30"W, at 1007 m asl) in November 2008. The climate is tropical with dry winter and rainy summer (Aw), according to the Köppen classification, with average annual temperature of 22°C, relative humidity of 73%, and average annual rainfall of 1100 mm. The predominant soil in the site was classified as Red Latosol, with high clay content.

Sixty-seven half-sib families were evaluated using a randomized block design with two replications and five plants per plot, spaced 4 x 2 m apart. Management practices were based on Dias et al. (2007), adapted according to the results of studies on *Jatropha* in Brazil and in the world.

Evaluated traits

The evaluated traits were plant height at 12 (PH12) and 40 (PH40) months after planting; canopy projection on the row at 12 (CPR12) and 40 (CPR40) months after planting; canopy projection between the row at 12 (CPB12) and 40 (CPB40) months after planting; number of branches at 40 (NB40) months after planting; grain yield (GY) and oil yield (OY) in 2014, which corresponds to the sixth year after planting. Plant height, canopy projection on the row, and canopy projection between the row were measured in m, while grain yield and oil yield were estimated in kg/ha and L/ha, respectively.

Statistical analyses

Initially, analysis of variance (ANOVA) was carried out for each trait, according to the statistical model described in Equation 1:

$$Y_{ij} = \mu + B_j + G_i + e_{ij} \qquad \text{(Equation 1)}$$

in which Y_{ij} is the observation in the j-th block, evaluated in the i-th genotype; μ is the overall mean of the experiment; B_j is the fixed effect of the j-th block; G_i is the random effect of the i-th genotype; e_{ij} is the random error associated with the Y_{ij} observation.

Phenotypic correlations (r_F) between pairs of traits were estimated according to Equation 2:

$$r_F = \frac{COV_{F(xy)}}{\sqrt{\hat{\sigma}^2_{Fx} \ x \ \hat{\sigma}^2_{Fy}}} \qquad \text{(Equation 2)}$$

in which $COV_{F(xy)}$ is the phenotypic covariance between traits X and Y; $\hat{\sigma}^2_{Fx}$ is the phenotypic variance of trait X; $\hat{\sigma}^2_{Fy}$ is the phenotypic variance of trait Y. Phenotypic correlations were tested by the *t*-test with n-2 degrees of freedom.

Phenotypic correlation network was used to graphically express the functional relationship between the estimates of the coefficients of phenotypic correlation between traits, in which the proximity between the nodes (traces) is proportional to the absolute value of the correlation between these nodes. The thickness of the edges was controlled by a cut-off value of 0.60, which means that only $|r_{ij}| \geq 0.60$ has their edges highlighted. Finally, positive correlations were represented in green, whereas negative correlations were represented in red.

Subsequently, multicollinearity analysis of the X'X correlation matrix was carried out, which revealed severe multicollinearity (condition number >1000), according to the classification of Montgomery and Peck (2001). Therefore, a constant k = 0.10 was added to the diagonal of the X'X matrix for canonical correlation analysis, and a constant k = 0.05 was added to path analysis, which resulted in weak multicollinearity (condition number <100).

For the canonical correlation analysis, two groups of traits were established: group I consisting of traits of economic importance for the culture (GY and OY); and group II, consisting of the other traits evaluated (PH12, CPR12, CPB12, NB40, PH40, CPR40 and CPB40). The first canonical correlation (r_1) between the linear combination of traits of groups I and II was estimated by Equation 3:

$$r_1 = \sqrt{\lambda_1} \qquad \text{(Equation 3)}$$

$$H = \begin{bmatrix} 1.10 & r_{\mathrm{GY\,x\,OY}} \\ r_{\mathrm{OY\,x\,GY}} & 1.10 \end{bmatrix}^{-1}_{2\,\mathrm{x}\,2} \mathrm{x} \begin{bmatrix} 1.10 & \cdots & r_{\mathrm{PH12\,x\,GY}} \\ \vdots & \ddots & \vdots \\ r_{\mathrm{GY\,x\,PH12}} & \cdots & 1.10 \end{bmatrix}_{2x7} x \begin{bmatrix} 1.10 & \cdots & r_{PH12xCPB40} \\ \vdots & \ddots & \vdots \\ r_{CPB40xPH12} & \cdots & 1.10 \end{bmatrix}^{-1}_{7x7} x \begin{bmatrix} 1.10 & \cdots & r_{PH12xGY} \\ \vdots & \ddots & \vdots \\ r_{GYxPH12} & \cdots & 1.10 \end{bmatrix}_{2x7} \qquad \text{(Equation 4)}$$

Equations 5 and 6 were used to obtain the first canonical pair, composed of X_I and Y_I, which are linear combinations between traits of groups I and II:

$$X_i = a'X \qquad \text{(Equation 5)}$$

$$Y_i = b'Y \qquad \text{(Equation 6)}$$

in which a' and b' are eigenvectors associated with the first eigenvalue of the matrix H; X and Y are the vector of phenotypic values of the traits that make up groups I and II, respectively. The other three canonical correlations and canonical pairs were estimated using the eigenvectors associated with eigenvalues in descending order following the equations previously described.

Path analysis, considering OY as the main dependent variable, was carried out by the model described in Equation 7:

$$OY = \beta_1 PH12 + \beta_2 PH40 + \cdots + \beta_7 GY + \rho_{\varepsilon} \qquad \text{(Equation 7)}$$

in which: $\beta_1, \beta_2, \cdots, \beta_7$ are the estimators of direct effects of PH12, CPR12, CPB12, NB40, PH40, CPR40, and CPB40 on OY; p_{ε} is the residual effect. Thus, the system of equations was used to estimate the direct and indirect effects of each auxiliary variable on OY, according to Equation 8:

$$\begin{bmatrix} 1.05 & \cdots & r_{PH12xGY} \\ \vdots & \ddots & \vdots \\ r_{GYxPH12} & \cdots & 1.05 \end{bmatrix} x \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_7 \end{bmatrix} = \begin{bmatrix} r_{PH12xOY} \\ \vdots \\ r_{GYxOY} \end{bmatrix} \qquad \text{(Equation 8)}$$

The coefficient of determination (R^2) of the path analysis was obtained by Equation 9:

$$R^2 = \hat{\beta}_1 r_{PH12xOY} + \cdots + \hat{\beta}_7 r_{GYxOY} \qquad \text{(Equation 9)}$$

The residual effect ($\hat{p}_\varepsilon$) of the path analysis was obtained by Equation 10:

$$\hat{p}_\varepsilon = \sqrt{1 - R^2} \qquad \text{(Equation 10)}$$

All statistical analyses were carried out with the Genes software (Cruz, 2013), and followed the procedures recommended by Cruz et al. (2012).

RESULTS

ANOVA and estimates of genetic parameters

ANOVA confirmed significant genetic variance ($P < 0.05$) between the 67 *Jatropha* accessions for all traits evaluated (Table 1). The lowest coefficients of variation (CVe) were observed for plant height at 12 and 40 months (6.129 and 4.832%, respectively) and the highest values were obtained for OY (23.187%) and GY (22.473%).

Table 1. Analysis of variance for the traits plant height at 12 (PH12) and 40 (PH40) months after planting; canopy projection on the row at 12 (CPR12) and 40 (CPR40) months after planting; canopy projection between the row at 12 (CPB12) and 40 (CPB40) months after planting; number of branches at 40 (NB40) months after planting; grain yield (GY) and oil yield (OY) in 2014, in relation to the sixth year after planting, evaluated in 67 *Jatropha* accessions.

Traits	Mean squares		F calculated	CVe (%)
	Accessions	Residues		
PH12	0.053	0.015	3.462*	6.129
CPR12	0.065	0.038	1.701*	14.675
CPB12	0.065	0.038	1.714*	14.493
NB40	8.994	4.777	1.883*	18.207
PH40	0.055	0.018	3.081*	4.832
CPR40	0.046	0.020	2.240*	6.742
CPB40	0.067	0.043	1.557*	8.150
GY	284350.487	105673.688	2.691*	22.473
OY	53934.607	20616.703	2.616*	23.187

*Significant at 5% probability by the F-test; CVe: coefficient of experimental variation.

Table 2 shows that PH12, PH40, OY, and GY were little influenced by environmental factors, and they had most of the phenotypic variance ($\hat{\sigma}_f^2$) explained by the genetic variance ($\hat{\sigma}_g^2$), and presented the highest broad-sense heritability estimates ($\hat{h}^2$) (71.114; 67.540; 62.837, and 61.775%, respectively). CPR40 was the least affected by environmental variation, and therefore presented higher h^2 estimates, CVg and CVr, when compared with CPR12. On the other hand, CPB12, CPB40, and NB40 were highly influenced by the environment, and consequently presented lower h^2.

CVg is an important parameter for the evaluation of the genetic variability between accessions for a given trait. Table 2 shows that variables PH12, PH40, GY and OY presented values closer to or greater than 1 for the CVg/CVe ratio (1.110, 1.020, 0.920, and 0.899, respectively).

Table 2. Estimates of genetic parameters for the traits plant height at 12 (PH12) and 40 (PH40) months after planting; canopy projection on the row at 12 (CPR12) and 40 (CPR40) months after planting; canopy projection between the row at 12 (CPB12) and 40 (CPB40) months after planting; number of branches at 40 (NB40) months after planting; grain yield (GY) and oil yield (OY) evaluated in 67 *Jatropha* accessions.

Trait	$\hat{\sigma}_f^2$	$\hat{\sigma}^2$	$\hat{\sigma}_g^2$	h^2	CVg	CVr
PH12	0.026	0.008	0.019	71.114	6.800	1.110
CPR12	0.033	0.019	0.013	41.199	8.686	0.592
CPB12	0.032	0.019	0.013	41.647	8.657	0.597
NB40	4.497	2.388	2.109	46.889	12.096	0.664
PH40	0.027	0.009	0.018	67.540	4.929	1.020
CPR40	0.023	0.010	0.013	55.353	5.308	0.787
CPB40	0.034	0.022	0.012	35.786	4.302	0.528
GY	142,175.243	52,836.844	89,338.399	62.837	20.663	0.920
OY	26,967.303	10,308.351	16,658.952	61.775	20.843	0.899

$\hat{\sigma}_f^2$: phenotypic variance; $\hat{\sigma}^2$: environmental variance; $\hat{\sigma}_g^2$: genotypic variance; h^2: broad-sense heritability; CVg: coefficient of genetic variation; CVr: coefficient of variation, given by the ratio between the coefficient of genetic variation and the coefficient of environmental variation.

Phenotypic correlations

The highest estimated phenotypic correlation was observed between GY and OY (0.993), evidencing that the higher the grain yield of the accessions, the higher is the oil yield (Table 3). To facilitate the visualization and interpretation of the correlation coefficients, phenotypic correlation network was used (Figure 1), which is a graphical representation of the variables according to the magnitudes of the correlations, since the traits with high correlations are closer and connected by more expressive traces. Figure 1 shows that the smallest and most expressive trace is between GY and OY, due to the higher phenotypic correlation.

Correlations between PH, CPB and CPR evaluated at 12 and 40 months after planting were significant and positive, and the highest correlation was observed between PH12 and PH40 (0.718), followed by CPB12 and CPB40 (0.663), and finally by CPR12 and CPR40 (0.485). Correlations of OY and GY with plant height were higher at 40 months (PH40), when compared with the correlations with plants evaluated at 12 months (PH12), confirming that the higher the plant, the greater is the grain yield and the oil yield. Similar results were observed for correlations of OY and GY with CPB, which was higher when evaluated at 40 months (CPB40), and lower at 12 months (CPB12). On the other hand, correlations of OY and GY with the CPR were lower when evaluated at 40 months (Table 3).

Table 3 shows that all correlations between plant height and canopy projection are positive, both on the row and between the row, and at any age, i.e., the higher the plant, the greater is the canopy projection. Correlations between CPR and CPB are positive, proving the proportionality in plant growth, since the greater the canopy projection on the row, the greater is the projection between the row. These observations are also highlighted in Figure 1, evidencing that correlation between CPR and CPB is quite high when the plant is younger, at 12 months (0.908), and decreases when the plant reaches 40 months of age (0.666).

Figure 1 highlighted NB40 with thinner red traces, more distant from the other traits for they have the lowest magnitudes of correlations, mostly negative. Thus, for NB40, only the correlations with plant height, for both PH12 (-0.599) and PH40 (-0.407) are highlighted, justifying that the higher the plant, the smaller is the number of branches (Table 3).

Table 3. Phenotypic correlations between the traits plant height at 12 (PH12) and 40 (PH40) months after planting; canopy projection on the row at 12 (CPR12) and 40 (CPR40) months after planting; canopy projection between the row at 12 (CPB12) and 40 (CPB40) months after planting; number of branches at 40 (NB40) months after planting; grain yield (GY) and oil yield (OY) evaluated in 67 *Jatropha* accessions.

Trait	CPR12	CPB12	NB40	PH40	CPR40	CPB40	GY	OY
PH12	0.453	0.419	-0.599	0.718*	0.192	0.314	0.443	0.456
CPR12		0.908*	-0.078	0.477	0.485	0.651*	0.446	0.464
CPB12			-0.063	0.475	0.466	0.663*	0.462	0.477
NB40				-0.407	0.014	-0.037	-0.090	-0.094
PH40					0.138	0.289	0.463	0.476
CPR40						0.666*	0.318	0.321
CPB40							0.498	0.513*
GY								0.993*

*Significant at 5% probability by the *t*-test with 65 degrees of freedom.

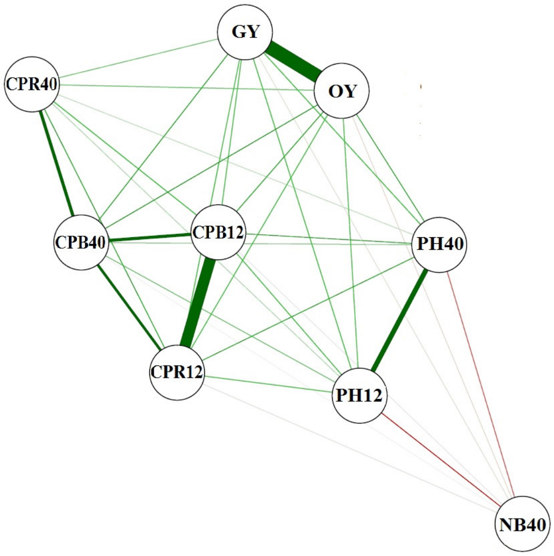

Figure 1. Phenotypic correlation network between the traits plant height at 12 (PH12) and 40 (PH40) months after planting; canopy projection on the row at 12 (CPR12) and 40 (CPR40) months after planting; canopy projection between the row at 12 (CPB12) and 40 (CPB40) months after planting; number of branches at 40 (NB40) months after planting; grain yield (GY) and oil yield (OY) evaluated in 67 *Jatropha* accessions.

Canonical correlation analysis

Canonical correlations were estimated between group 1, formed by the productive traits (GY and OY), and group 2, represented by the morphological traits evaluated at 12 and 40 months after planting (Table 4). Based on these results, the studied groups were not independent, since significant correlation was observed by the chi-square test ($P < 0.05$) for the first pair of canonical correlations ($r = 0.602$). The second pair of canonical correlations was not significant by the chi-square test ($P > 0.05$), and therefore the estimates of canonical coefficient were not expressed.

In relation to the first canonical pair, *Jatropha* accessions with higher grain yield and higher oil yield presented greater number of branches and, regardless of plant age (at 12 or 40 months), were higher and presented greater canopy projection between the row. CPR presented low estimates of canonical coefficients at both ages, which indicates absence of association with the productive traits (group 1).

Table 4. Estimates of canonical correlation and coefficients for the first canonical pair between group 1, formed by grain yield (GY) and oil yield (OY), and group 2 represented by plant height at 12 (PH12) and 40 (PH40) months after planting; canopy projection on the row at 12 (CPR12) and 40 (CPR40) months after planting; (CPB12) and 40 (CPB40) months after planting; and number of branches at 40 (NB40) months after planting.

Group 1	Productive traits
GY	0.354
OY	0.621
Group 2	Morphological traits
PH12	0.368
CPR12	-0.014
CPB12	0.131
NB40	0.219
PH40	0.358
CPR40	0.027
CPB40	0.470
R^2	0.602*
d.f.	14

*Significant at 5% probability by the chi-square test, with 14 degrees of freedom (d.f.). R^2 - coefficient of determination.

Path analysis

Table 5 shows the estimates of the direct and indirect effects of PH12, CPR12, CPB12, NB40, PH40, CPR40, CPB40, and GY on the main dependent variable OY. The estimate of the coefficient of determination was of high magnitude ($\hat{R}^2 = 0.944$), and the residual effect was of low magnitude (0.238) (Table 5). These results indicate that almost all variation in OY was explained by the auxiliary traits.

The highest direct effect on OY was obtained by GY (0.900), which is an estimate close to the phenotypic correlation (Table 5). Thus, GY is the main determinant in the variation of OY, and evidences the cause and effect relationship between these traits, i.e., the higher the grain yield, the higher is the oil yield of the plant, confirming the results obtained in the previous analyses.

On the other hand, Table 5 shows that all the other direct and indirect effects of the variables PH12, CPR12, CPB12, NB40, PH40, CPR40, and CPB40 on OY were low, and these values (in module) are lower than the effect of the residual variable, except for indirect effects via GY. All the traits had significant indirect correlation coefficients higher than the residual variable via GY, except for the NB40. Thus, the greatest indirect effects on OY via GY occurred for CPB40 (0.448), PH40 (0.417), CPB12 (0.416), and CPR12 (0.401).

These results demonstrate that the auxiliary traits PH12, CPR12, CPB12, PH40, CPR40, and CPB40 indirectly influenced the main variable via GY. Thus, practicing direct selection on these traits to obtain gains in oil yield is not efficient, and indirect selection on GY should be prioritized in order to select the traits that present the highest gains in grain yield.

Table 5. Estimates of the direct (main diagonal) and indirect effects between the traits plant height at 12 (PH12) and 40 (PH40) months after planting; canopy projection on the row at 12 (CPR12) and 40 (CPR40) months after planting; canopy projection between the row at 12 (CPB12) and 40 (CPB40) months after planting; number of branches at 40 (NB40) months after planting; grain yield (GY) on the main productive variable, oil yield (OY), evaluated in 67 *Jatropha* accessions.

Traits	PH12	CPB12	CPB12	NB40	PH40	CPR40	CPB40	GY	Total
PH12	0.031	0.007	0.000	-0.011	0.018	-0.002	0.014	0.398	0.456
CPR12	0.014	0.014	0.000	-0.001	0.012	-0.005	0.028	0.401	0.464
CPB12	0.013	0.013	0.000	-0.001	0.012	-0.005	0.029	0.416	0.477
NB40	-0.018	-0.001	0.000	0.018	-0.010	0.000	-0.002	-0.081	-0.094
PG40	0.022	0.007	0.000	-0.007	0.025	-0.001	0.013	0.417	0.476
CPR40	0.006	0.007	0.000	0.000	0.004	-0.010	0.029	0.286	0.321
CPB40	0.010	0.009	0.000	-0.001	0.007	-0.007	0.044	0.448	0.513*
GY	0.014	0.006	0.000	-0.002	0.012	-0.003	0.022	0.900	0.993*
$\hat{R}^2$							0.944		
$\hat{p}_\varepsilon$							0.238		

*Significant at 5% probability, by the *t*-test, with 65 degrees of freedom.

DISCUSSION

ANOVA and estimates of genetic parameters

The population evaluated in this study has sufficient variability for selection of genotypes of superior agronomic performance. This fact is important for being and essential condition for the establishment of a breeding program. Cruz et al. (2012) report that experimental precision is high when CVe estimates are lower than 20%, as observed in this study for all morphological traits. Estimates of CVe higher than 20% for the productive traits possibly occurred due to the existence of variation within each accession, being similar to those reported in other studies with *Jatropha*, including this population (Laviola et al., 2010).

$\hat{h}^2$ estimates obtained for the traits PH12, PH40, CPR40, GY and OY are considered moderate (>50%), and are in accordance with the results reported by Laviola et al. (2010, 2012b). Peixoto et al. (2016), who evaluated productive traits in *Jatropha*, verified genetic variability between *Jatropha* accessions of the germplasm bank of EMBRAPA Agroenergy, and in this way, it is possible to obtain gain from selection on these traits.

CVg quantifies the proportion of genetic variability available for selection (Cruz et al., 2012), and the desirable values are high. The ratio between this parameter and CVe results in the CVr, which is an important variable for selection. According to Vencovsky et al. (1987), when CVr is greater than 1, gain with selection is possible. Thus, values obtained from CVg for plant height at all ages resulted in CVr superior to 1, which indicates favorable situation for the selection of genotypes for this trait. This result is significant, since selection of short *Jatropha* genotypes to facilitate fruitharvesting is one of the main objectives of the breeding programs of this crop (Dias et al., 2007).

Phenotypic correlations

Knowing the association between two traits is important for selection in breeding programs, as it allows indirect gains, and guides the establishment of selection indices (Cruz et al., 2012). The highest estimates of phenotypic correlations were obtained between the same traits evaluated at different ages, e.g., CPR12 x CPR40, CPB12 x CPB40, and between CPR12 x CPB40, CPR40 x CPB40, and GY x OY. These results are in accordance with those of Laviola et al. (2010, 2012a,b). Correlation between the canopy projection on the row and oil yield deserves attention, since it was moderate at both ages. The high correlation found between GY and OY is of great importance for *Jatropha*-breeding program, since the evaluation of OY is complex due to the process of oil extraction from the seeds, and to high-cost equipment and labor (Silva Junqueira et al., 2016). Thus, indirect selection of accessions with high OY value by selecting the best accessions for GY is recommended.

To facilitate the interpretation of phenotypic correlations between morphological and productive traits, a phenotypic correlation network was used to detect complex phenotypic patterns, which is difficult to be obtained with other techniques (Figure 1). In the phenotypic correlation network, values of significant correlations by the *t*-test were highlighted with bold edges, negative correlations were represented by red-colored lines, and positive correlations were represented by green-colored lines. The efficiency of this innovative technique has already been reported by Ursem et al. (2008), DiLeo et al. (2011), and Silva et al. (2016). Phenotypic correlation networks facilitate the interpretation of the correlations between traits, and thus make it easier to verify which traits can be used for indirect selection. For instance, indirect selection of superior accessions for GY and OY via morphological traits will benefit *Jatropha*-breeding program due to early selection of accessions, since morphological traits can be measured a few months after planting. In addition, the cost with evaluation of the productive traits will decrease, since these traits require high-cost labor and equipment, and only begin to be measured 3 years after planting.

Canonical correlation analysis

By the interpretation of the first canonical pair, higher plants and greater canopy projection on the row are the ones with higher grain yield and oil yield. These results reveal the need to use these agronomic traits in the establishment of selection indices in order to obtain genetic gain on the productive traits. Thus, early selection of *Jatropha* accessions aiming at increasing oil yield could be carried out via morphological traits at the initial stages of growth.

As reported, the identification of short height genotypes is desirable, since this oilseed plant can exceed 5 m in height (Dias et al., 2007). However, since plant height correlated positively with oil yield, as confirmed by the canonical coefficients, some cycles of recombination between the most productive accessions are necessary in order to break this relation, and to enable the selection of genotypes with reduced size and high oil yield.

Path analysis

The main objective of the *Jatropha*-breeding program is the selection of accessions with high oil yield. However, the genetic variability for OY between *Jatropha* accessions is low, and this trait is greatly influenced by the environment (low heritability) (Silva Junqueira et al., 2016). Moreover, due to the high cost to evaluate oil yield, indirect selection for this trait, via morphological traits, is interesting. To this end, path analysis is a promising strategy for early selection.

In addition, besides being important, estimates of phenotypic correlations may lead to mistakes regarding the relationship between two traits, which may be the result of the effect that a third trait or a group of traits has on the pair of traits. Path analysis makes it possible to study the cause and effect relationship between traits, which allows the accurate establishment of the best criteria for indirect selection (Cruz et al., 2012). The estimate of R^2 for the established path model was superior to that of studies that employed this analysis in *Jatropha* (Spinelli et al., 2010; Reis et al., 2015; Teodoro et al., 2016).

The identification of traits that have high phenotypic correlation and high direct effect in the same direction on the main trait is desirable, since the correlated response by means of indirect selection can be effective. Therefore, indirect selection of genotypes with higher grain yield aiming to increase oil yield is a promising strategy due to the cause and effect relationship between these traits, as evidenced in this study.

Therefore, the selection of productive traits can be indirectly performed via morphological traits due to the correlation between these two groups of traits. Families with high performance for oil yield can be selected by means of grain production in order to reduce costs with selection for this trait, showing that canonical correlations and path analysis are two strategies that may be useful in *Jatropha*-breeding program, when the breeder intends to indirectly select productive traits via morphological traits.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We are thankful to CAPES (Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FAPEMIG (Fundação de Amparo à Pesquisa de Minas Gerais), Funarbe (Fundação Arthur Bernardes), and Universidade Federal de Viçosa for financial support. We also thank the Biometric Lab (Universidade Federal de Viçosa, Brazil) where all analyses were performed by remote access.

REFERENCES

Bhering LL, Barrera CF, Ortega D, Lavila BG, et al. (2013). Differential response of *Jatropha* genotypes to different selection methods indicates that combined selection is more suited than other methods for rapid improvement of the species. *Ind. Crops Prod.* 41: 260-265. http://dx.doi.org/10.1016/j.indcrop.2012.04.026

Cruz CD (2013). Genes - a software package for analysis in experimental statistics and quantitative genetics. *Acta Sci. Agron.* 35: 271-276. http://dx.doi.org/10.4025/actasciagron.v35i3.21251

Cruz CD, Regazzi AJ and Carneiro PCS (2012). Modelos biométricos aplicados ao melhoramento genético. 3 ed. Editora UFV, Viçosa.

Dias LAS, Leme LP, Laviola BG, Pallini Filho A, et al. (2007). Cultivo de pinhão manso (*Jatropha curcas L.*) para produção de óleo combustível. Universidade Federal de Viçosa. *Viçosa* 1: 1-40.

DiLeo MV, Strahan GD, den Bakker M and Hoekenga OA (2011). Weighted correlation network analysis (WGCNA) applied to the tomato fruit metabolome. *PLoS One* 6: e26683. http://dx.doi.org/10.1371/journal.pone.0026683

Laviola BG, Rosado TB, Bhering LL, Kobayashi AK, et al. (2010). Genetic parameters and variability in physic nut accessions during early developmental stages. *Pesqui. Agropecu. Bras.* 45: 1117-1123. http://dx.doi.org/10.1590/S0100-204X2010001000010

Laviola BG, Bhering LL, Mendonça S, Rosado TB, et al. (2011). Morpho-agronomic characterization of the germplasm bank of *Jatropha* young stage. *Biosci. J.* 27: 371-379.

Laviola BG, Alves AA, Gurgel FD, Rosado TB, et al. (2012a). Estimate of Genetic Parameters and Predicted Gains with Early Selection of Physic Nut Families. *Cienc. Agrotec.* 36: 163-170. http://dx.doi.org/10.1590/S1413-70542012000200004

Laviola BG, Alves AA, Gurgel Fd D, Rosado TB, et al. (2012b). Estimates of genetic parameters for physic nut traits based in the germplasm two years evaluation. *Cienc. Rural* 42: 429-435. http://dx.doi.org/10.1590/S0103-84782012000300008

Laviola BG, Anjos SD, Juhász ACP, Rocha RB, et al. (2014). Desempenho agronômico e ganho genético pela seleção de pinhão manso em três regiões do Brasil. *Pesqui. Agropecu. Bras.* 49: 356-363. http://dx.doi.org/10.1590/S0100-204X2014000500005

Montes JM and Melchinger AE (2016). Domestication and Breeding of *Jatropha curcas* L. *Trends Plant Sci.* 21: 1045-1057. http://dx.doi.org/10.1016/j.tplants.2016.08.008

Montgomery DC and Peck EA (2001). Introduction to linear regression analysis. 3 ed. John Wiley & Sons, New York.

Peixoto LA, Laviola BG, Bhering LL, Mendonça S, et al. (2016). Oil content increase and toxicity reduction in *Jatropha* seeds through family selection. *Ind. Crops Prod.* 80: 70-76. http://dx.doi.org/10.1016/j.indcrop.2015.10.034

Reis MVM, Damasceno Júnior PC, Campos TO, Diegues IP, et al. (2015). Variabilidade genética e associação entre caracteres em germoplasma de pinhão-manso (*Jatropha curcas* L.). *Rev. Cienc. Agron.* 46: 412-420.

Sharma S, Dhamija HK and Parashar B (2012). *Jatropha curcas*: a review. *Asian J. Pharm. Sci.* 2: 107-111.

Silva ARD, Rêgo ERD, Pessoa AMDS and Rêgo MMD (2016). Correlation network analysis between phenotypic and genotypic traits of chili pepper. *Pesqui. Agropecu. Bras.* 51: 372-377. http://dx.doi.org/10.1590/S0100-204X2016000400010

Silva Junqueira V, Azevedo Peixoto Ld, Galvêas Laviola B, Lopes Bhering L, et al. (2016). Bayesian Multi-Trait Analysis Reveals a Useful Tool to Increase Oil Concentration and to Decrease Toxicity in *Jatropha curcas* L. *PLoS One* 11: e0157038. http://dx.doi.org/10.1371/journal.pone.0157038

Spinelli VM, Rocha RB, Ramalho AR, Marcolan AL, et al. (2010). Primary and secondary yield components of the oil in physic nut (*Jatropha curcas* L.). *Cienc. Rural* 40: 1752-1758. http://dx.doi.org/10.1590/S0103-84782010005000129

Teodoro PE, Costa RH, Rocha RB and Laviola BG (2016). Contribuição de caracteres agronômicos para a produtividade de grãos em pinhão-manso. *Bragantia* 75: 51-56. http://dx.doi.org/10.1590/1678-4499.314

Ursem R, Tikunov Y, Bovy A, Van Berloo R, et al. (2008). A correlation network approach to metabolic data analysis for tomato fruits. *Euphytica* 161: 181-193. http://dx.doi.org/10.1007/s10681-008-9672-y

Vencovsky R and Barriga P (1992). Genética biométrica no fitomelhoramento. Sociedade Brasileira de Genética, Ribeirão Preto.

Vencovsky R, Paterniani E and Viegas GP (1987). Herança quantitativa. Melhoramento e produção de milho. 2. ed. Campinas.

Wright S (1923). The theory of path coefficients: a reply to Niles' criticism. *Genetics* 8: 239-255.