

## Path analysis under multicollinearity for papaya production of the Solo and Formosa groups

Sarah Ola Moreira<sup>1</sup>, Karin Tesch Kuhlcamp<sup>2</sup>, Fabíola Lacerda de Souza Barros<sup>2</sup>,  
Sabrina Batista de Oliveira<sup>3</sup>, Roberto dos Santos Trindade<sup>4</sup>

**Abstract** – Correlations and path analysis allow the understanding of the interrelations between characteristics of interest to the plant breeding. However, in order for their results to be reliable, the undesirable effect of multicollinearity must be excluded. The objective of this study was to estimate the correlations and their partitioning in direct and indirect effects, by path analysis, on the fruit production per plant (FP) of papaya from the heterotic Solo and Formosa groups, using two different strategies to circumvent multicollinearity. Eleven agronomic variables were evaluated in twelve papaya genotypes from the Solo group and nine from the Formosa group. Path analysis was obtained with the FP as the basic variable and to eliminate multicollinearity were used the discard of variables and the ridge path analysis. For the Solo group, fruit length and pulp thickness had greater direct effects on FP. In the Formosa group, the number of commercial fruits had a direct and indirect effect on FP. The two methodologies used to circumvent multicollinearity had high coefficient of determination, with better values for the ridge path analysis. The results indicated that the interrelation between the study characters was different in the Solo and Formosa groups. Thus, indirect selection strategies should be specific for each heterotic group.

**Index terms:** *Carica papaya* L., ridge path analysis, indirect selection, correlations.

## Análise de trilha sob multicolinearidade para produção de mamoeiro dos grupos Solo e Formosa

**Resumo** – As correlações e a análise de trilha permitem o entendimento das inter-relações entre características de interesse ao melhoramento de plantas. Porém, para que os resultados dessas análises sejam confiáveis, é preciso excluir o efeito indesejado da multicolinearidade. O objetivo desse trabalho foi estimar as correlações e seus desdobramentos em efeitos diretos e indiretos, pela análise de trilha, sobre a produção de frutos por planta (FP) de mamoeiro dos grupos heteróticos Solo e Formosa, usando duas diferentes estratégias para contornar a multicolinearidade. Foram avaliadas 11 variáveis agronômicas em 12 genótipos de mamoeiro do grupo Solo e nove do grupo Formosa. A análise de trilha foi obtida tendo o FP como variável básica e para eliminar a multicolinearidade foi utilizado o descarte de variáveis e a análise de trilha em crista. Para o grupo Solo, o comprimento do fruto e a espessura da polpa tiveram maiores efeitos diretos sobre o FP. No grupo Formosa, o número de frutos comerciais teve grande efeito direto e indireto sobre FP. As duas metodologias utilizadas para contornar a multicolinearidade tiveram alto coeficiente de determinação, com melhores valores para a análise de trilha em crista. Os resultados indicaram que a inter-relação entre os caracteres em estudo foi diferente nos grupos Solo e Formosa. Assim, as estratégias de seleção indireta devem ser específicas para cada grupo heterótico.

**Termos para indexação:** *Carica papaya* L., análise de trilha em crista, seleção indireta, correlações.

**Corresponding author:**  
sarah.ola@gmail.com

**Received:** April, 27 2017.  
**Accepted:** December 11, 2017.

**Copyright:** All the contents of this journal, except where otherwise noted, is licensed under a Creative Commons Attribution License.



<sup>1</sup>D.Sc. Researcher at Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural. E-mail: sarah.ola@gmail.com

<sup>2</sup>M.Sc. Researcher at Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural. E-mails: fabiola.barros@incaper.es.gov.br karin.kuhlcamp@incaper.es.gov.br

<sup>3</sup>Forestry Engineer. Scholarship holder at Researcher at Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural. E-mail: sabrina.bo@live.com

<sup>4</sup>D.Sc. Researcher at Embrapa Milho e Sorgo. E-mail: roberto.trindade@embrapa.br

## Introduction

Due to the low number of papaya cultivars (*Carica papaya* L.) available to the producer, this crop becomes more susceptible to damage caused by pests, diseases and climatic adversities. Therefore, it is necessary to develop new papaya cultivars, ensuring their sustainability (OLIVEIRA et al., 2010; SILVA et al., 2016). The search for new genotypes is facilitated when the relationships between characters used in selection are known, especially when the variable of interest is polygenic, has its phenotype expressed late or of medium or low heritability, such as fruit production.

Correlations indicate the direction and degree of association between a pair of characters, and if the interrelationship is due to genetic factors - pleiotropy or gene linkage - or environmental factors, increasing the possibility of genetic gains. However, the correlation does not evaluate the cause and effect relationships, that is, they do not indicate the direct and indirect effects of other variables on the pair of characters under study (TEIXEIRA et al., 2012).

To circumvent this limitation, Wright (1923) proposed the path analysis, which consists in quantifying the direct and indirect effects of explanatory characteristics on a basic variable. Estimates of this analysis are obtained by regression equations, where the variables are previously standardized (WRIGHT, 1923). In a review on path analysis, Olivoto et al. (2016) have shown that it is used in different areas of knowledge, such as plant and animal breeding, and environmental and social sciences.

In order for the results of the path analysis to be reliable, it is necessary that the assumptions of the model are met, among them, the lack of multicollinearity. Multicollinearity occurs when the basic variable is linked to a large number of explanatory variables, or when the correlation between the basic and explanatory characters is very high (TOEBE; CARGNELUTTI FILHO, 2013a, b; CRUZ et al., 2014). If path analysis is performed under multicollinearity, estimates of direct and indirect effects may be biased, leading to erroneous conclusions. To eliminate it, it has been recommended to remove from the analysis the highly correlated variables (TOEBE; CARGNELUTTI FILHO, 2013a, b). However, when character deletion is not interesting to the breeder, we can alternatively perform ridge path analysis, where a constant  $k$  is added to the elements of the diagonal of the correlation matrix (CRUZ et al., 2014).

For papaya, studies that seek to understand the correlations and their direct and indirect effects through path analysis for the production, number and mass of fruits are available in the literature (OLIVEIRA et al., 2010; SILVA et al., 2016). However, these analyzes do not consider the different heterotic groups of papaya, divided in Solo and Formosa, that have phenotypic patterns, such

as size and fruit mass and different plant architecture. This may cause misunderstandings in the interpretation of the results since the characteristics considered ideal are not necessarily the same for the two groups. In addition, these studies did not consider the different alternatives to circumvent multicollinearity.

The objective of this research was to study the phenotypic, genotypic and environmental correlations and their partitioning in direct and indirect effects, by path analysis, on the production of papaya fruits of the Solo and Formosa groups, and to evaluate different methods to overcome the effect of multicollinearity.

## Material and methods

The experiment was conducted in the municipality of Pinheiros, ES, (18°30'59" S, 40°17'38" W) in a private area with commercial papaya planting. The climate is hot humid and monsoon (Am), with average annual air temperature of 23.6°C and average annual rainfall of 1,308 mm (ALVARES et al., 2013).

Twenty-one genetic materials of papaya from the heterotic Solo and Formosa groups were evaluated. Within the heterotic Solo group, 12 genotypes were evaluated: H10-60, H26-60, H36-45, L06-08, L47-P8, L54-08, L78-08 (Embrapa Mandioca e Fruticultura); UC13, UC14, UC15 and UC16 (Universidade Estadual do Norte Fluminense Darcy Ribeiro - Uenf, in partnership with Caliman Agrícola S/A); and the cultivar Golden. For the heterotic Formosa group, nine materials were evaluated: UC03, UC10, UC11, UC12 (Uenf/Caliman), EW Sinta and EW 2747 (East West Seed); L10-08 (Embrapa Mandioca e Fruticultura), Tainung 01 (Takii do Brasil Ltda) and Rubi Incaper 511 (Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural - Incaper). Barros et al. (2017) described the agronomic characterization of these genotypes. The experimental design was a randomized block with five replications and six plants per plot. The commercial planting of papaya trees around the experiment was used as a border, and therefore all plants were considered useful.

The seedlings were produced in commercial substrate and transplanted to a previously prepared area. In the transplant, three seedlings per hole were used and, after the emergence of the flower buds, sexing was carried out, maintaining one hermaphrodite plant per hole. The spacing was 3.50 x 1.80 m, and irrigation was carried out with a central pivot. The planting area was monitored for elimination of plants with virus symptoms. Other cultural treatments were carried out according to the recommendations for culture (MARTINS; COSTA, 2003).

Were evaluated the following characteristics: fruit yield per plant (FP), in kg plant<sup>-1</sup>; number of commercial fruits (NCF); number of deformed fruits (NDF); number of nodes without fruit (NWF); fruit mass (FM), in g fruit<sup>-1</sup>;

fruit length (FL), in cm; fruit diameter (FD), in cm; pulp thickness (PT), in cm; soluble solids content (SS) in °Brix; plant height (PH) in m, measured from the soil level to the insertion of the youngest leaf; and, stem diameter (SD), in cm, measuring 20 cm from the soil. For the NCF, NDF, NWF, PH and SD evaluations, were sampled three plants per plot. For the other variables, a sample of five fruits per plot harvested at maturation stage 2 (25% of its yellow surface).

For each heterotic group, the original data were submitted to analysis of variance. Genotypic ( $r_g$ ), phenotypic ( $r_p$ ) and environmental ( $r_e$ ) correlations were estimated for all combinations of characters. The phenotypic correlations were partitioned on direct and indirect effects through path analysis (WRIGHT, 1923), considering FP as the basic variable and the other agronomic characters as explanatory variables.

The multicollinearity diagnosis was performed based on the condition number (CN), which represents the ratio between the highest and the lowest self-value of the correlation matrix  $X'X$  (MONTGOMERY et al., 2012). According to Toebe and Cargnelutti Filho (2013a), when the CN is greater than or equal to 100 there is multicollinearity. In this experiment, after verification of multicollinearity, two strategies were adopted to circumvent its effects: the discarding of variables that contributed to its appearance, that is, those that have the highest correlation among them; and the methodology of ridge path analysis, where a constant ( $k$ ) is introduced to the diagonal of the  $X'X$  matrix (CRUZ et al., 2014). For the data analysis, was used the Genes Program (CRUZ, 2013).

## Results and Discussion

There were significant differences ( $p < 0.05$ ) for all traits studied, except for SD in the Solo group and FL for the Formosa group, indicating that there is variability among genetic materials within the same heterotic group. The cultivars of the Solo group had, on average, FM of 591.7 g and SS of 13.7 °Brix. Within the heterotic Formosa group, the FM was 1,298.8 g and the SS was 13.0 °Brix (Table 1). These values corroborate with Costa et al. (2013) and Luz et al. (2015), which indicated FM for the Solo group between 450 and 900 g and FM for the Formosa group from 1,100 to 2,700 g. For the soluble solids content, the values obtained were higher than 12.0 °Brix, a value considered optimal for the papaya, independently of the heterotic group (LUZ et al., 2015).

For the FP, the cultivars of the Solo group ranged from 7.3 to 32.0 kg plant<sup>-1</sup>. In the Formosa group, the FP was between 9.5 and 43.4 kg plant<sup>-1</sup>. These values indicate the wide genetic variability available for the character, justifying the need to identify the direct and indirect effects of different variables on FP and that can be used in the

selection of more productive genotypes.

The coefficients of experimental variation (CV) ranged from 4.2 to 34.8% for the Solo group and from 4.9 to 39.7% for the Formosa group. For the two heterotic groups, the lowest CV was obtained for SS and the highest for NDF (Table 1). The CV value for FP was also considered high, from 28.1% for Solo and 36.1% for Formosa. This result evidences that this character, being the result of the expression of multiple variables, is complex and influenced by environmental variations. However, these high CV did not prevent the identification of significant differences for the study variables.

For the Solo group, the FP had a significant phenotypic correlation with NDF, FM, FL, FD and PT, reinforcing its complexity. With NDF this correlation was negative (-0.58), which was expected because the deformed fruits are not considered in the evaluation of the production per plant. The significant correlations between FP and the other variables were positive and ranged from 0.66 to 0.81 for FD and PT, respectively. The characteristics FM, FL, FD and PT also had phenotypic correlations with each other and higher than 76% (Table 2). These results are in agreement with Oliveira et al. (2012), indicating the strong interrelation between these variables and, because they have high magnitude and same sign, indicate that the selection practiced for one of them will favor the others.

For the heterotic Solo group, significant phenotypic correlations ( $p < 0.05$ ) were identified between FM x NCF, NDF, NWF, FD, with values of 0.70 (FM x NDF) at 0.95 (FM x FD). Then, the variable with the greatest number of significant interactions was NCF, which correlated with NWF, FM, FD and FP (Table 3). Contrary to that observed for the Solo group, for the Formosa group the FP had a significant correlation only with NWF (0.83). This variation observed between the correlations in the two heterotic groups was possibly due to differences in the fruit pattern of each group. In the Solo group, the fruits are smaller and more numerous, thus, small variations in FD, FL and PT can alter FM and FP. In the Formosa group, which has a pattern of larger and fewer fruits, NCF seems to have a greater influence on FP when compared to the morphological characteristics of the fruit.

In the Solo group, in 83.64% of the relationships studied, the genotypic correlations surpassed the phenotypic correlations. For the Formosa group, this value was 74.55%. This indicates the overlap of the genetic effects on the environmental in the manifestation of the evaluated characters and that the sampling errors were less expressive (CRUZ et al., 2014).

**Table 1.** Average and coefficient of variation (CV, in %) of eleven agronomic characters in papaya from the heterotic Solo and Formosa groups.

	NCF	NDF	NWF	FM	FL	FD	PT	SS	PH	SD	FP
Solo Group											
Average	27.1*	9.7*	9.6*	591.7*	15.7*	33.8*	2.3*	13.7*	2.3*	13.1 <sup>ns</sup>	16.0*
CV	22.9	34.8	33.9	16.0	7.1	5.9	5.3	4.2	11.1	9.9	28.1
Formosa Group											
Average	16.1*	5.2*	13.8*	1,298.8*	22.3 <sup>ns</sup>	11.4*	2.9*	13.0*	2.3*	13.5*	19.7*
CV	23.5	39.7	23.8	22.17	11.5	7.6	7.4	4.9	7.3	8.6	36.1

\*significant at 5% probability by the F test; <sup>ns</sup> not significant; NCF: number of commercial fruits; NDF: number of deformed fruits; NWF: number of nodes without fruit; FM: fruit mass; FL: fruit length; FD: fruit diameter; PT: pulp thickness; SS: soluble solids content; PH: plant height; SD: stem diameter; FP: production of fruits per plant.

**Table 2.** Estimating the coefficients of phenotypic ( $r_p$ ), genotypic ( $r_g$ ) and environmental ( $r_e$ ) correlation between eleven agronomic characters in papaya from the heterotic Solo group.

Variable <sup>1</sup>	$r$	NDF	NWF	FM	FL	FD	PT	SS	PH	SD	FP
NCF	$r_p$	-0.13	-0.76*	-0.34	-0.31	-0.42	-0.05	-0.31	0.03	0.10	0.39
	$r_g$	-0.18	-0.85	-0.40	-0.37	-0.48	-0.09	-0.44	-0.06	-0.32	0.26
	$r_e$	0.01	-0.39	-0.03	0.03	-0.13	0.14	0.11	0.38	0.60	0.78
NDF	$r_p$		-0.07	-0.54	-0.56	-0.42	-0.51	-0.11	0.42	0.40	-0.58*
	$r_g$		-0.02	-0.62	-0.66	-0.50	-0.62	-0.11	0.42	0.61	-0.76
	$r_e$		-0.28	-0.10	0.01	-0.09	0.13	-0.13	0.42	0.28	0.00
NWF	$r_p$			0.38	0.51	0.32	0.37	0.05	-0.21	-0.21	-0.20
	$r_g$			0.44	0.59	0.37	0.45	0.04	-0.20	-0.31	-0.19
	$r_e$			0.02	-0.03	0.08	-0.20	0.08	-0.25	-0.16	-0.24
FM	$r_p$				0.95*	0.96*	0.87*	-0.09	-0.18	0.06	0.73*
	$r_g$				0.97	0.98	0.89	-0.05	-0.21	0.08	0.78
	$r_e$				0.71	0.83	0.69	-0.33	-0.04	0.08	0.54
FL	$r_p$					0.83*	0.92*	-0.07	-0.23	0.10	0.68*
	$r_g$					0.89	0.94	-0.06	-0.29	0.14	0.73
	$r_e$					0.33	0.60	-0.19	0.22	0.13	0.48
FD	$r_p$						0.76*	-0.10	-0.21	0.03	0.66*
	$r_g$						0.80	-0.05	-0.22	0.03	0.73
	$r_e$						0.46	-0.32	-0.17	0.04	0.37
PT	$r_p$							-0.35	-0.36	0.04	0.81*
	$r_g$							-0.39	-0.47	-0.01	0.89
	$r_e$							-0.16	0.31	-0.22	0.53
SS	$r_p$								0.32	0.29	-0.34
	$r_g$								0.37	0.57	-0.41
	$r_e$								0.14	0.06	-0.12
PH	$r_p$									0.61*	-0.20
	$r_g$									0.73	-0.36
	$r_e$									0.71	0.36
SD	$r_p$										0.13
	$r_g$										-0.25
	$r_e$										0.58

<sup>1</sup>NCF: number of commercial fruits; NDF: number of deformed fruits; NWF: number of nodes without fruit; FM: fruit mass; FL: fruit length; FD: fruit diameter; PT: pulp thickness; SS: soluble solids content; PH: plant height; SD: stem diameter; FP: production of fruits per plant. \*significant at 5% probability by test t.



**Table 3.** Estimates of phenotypic ( $r_p$ ), genotypic ( $r_g$ ) and environmental ( $r_e$ ) correlation coefficients among eleven agronomic characters in papaya from the heterotic Formosa group.

Variabel <sup>1</sup>	$r$	NDF	NWF	FM	FL	FD	PT	SS	PH	SD	FP
NCF	$r_p$	0.51	-0.84*	-0.73*	-0.08	-0.77*	-0.54	-0.52	0.54	0.02	0.83*
	$r_g$	0.53	-0.92	-0.85	-0.08	-0.85	-0.61	-0.59	0.56	-0.03	0.92
	$r_e$	0.26	-0.51	-0.04	-0.12	0.08	0.02	0.11	0.26	0.26	0.57
NDF	$r_p$		-0.45	-0.70*	-0.73*	-0.56	-0.66*	-0.12	0.13	0.07	0.03
	$r_g$		-0.51	-0.76	-0.92	-0.59	-0.72	-0.21	0.10	-0.02	0.07
	$r_e$		-0.17	-0.37	-0.46	-0.33	-0.25	0.62	0.42	0.42	-0.16
NWF	$r_p$			0.77*	0.15	0.76*	0.59	0.28	-0.42	-0.31	-0.66
	$r_g$			1.02	0.27	0.92	0.77	0.38	-0.47	-0.41	-0.77
	$r_e$			-0.10	-0.08	0.01	-0.18	-0.23	-0.17	-0.10	-0.36
FM	$r_p$				0.44	0.95*	0.89*	-0.09	-0.44	0.06	-0.29
	$r_g$				0.26	0.99	0.89	-0.04	-0.55	-0.06	-0.63
	$r_e$				0.87	0.81	0.90	-0.34	0.23	0.35	0.73
FL	$r_p$					0.20	0.45	-0.21	0.40	-0.13	0.40
	$r_g$					0.12	0.34	-0.20	0.55	-0.41	0.28
	$r_e$					0.51	0.83	0.34	0.09	0.22	0.61
FD	$r_p$						0.85*	0.01	-0.52	0.20	-0.41
	$r_g$						0.88	0.06	-0.59	0.19	-0.67
	$r_e$						0.65	-0.43	0.12	0.27	0.67
PT	$r_p$							-0.13	-0.37	0.34	-0.09
	$r_g$							-0.12	-0.46	0.32	-0.30
	$r_e$							-0.17	0.33	0.45	0.66
SS	$r_p$								-0.04	0.09	-0.64
	$r_g$								-0.06	0.09	-0.76
	$r_e$								0.24	0.10	-0.21
PH	$r_p$									-0.06	0.56
	$r_g$									-0.25	0.65
	$r_e$									0.69	0.24
DS	$r_p$										0.14
	$r_g$										0.02
	$r_e$										0.35

<sup>1</sup>NCF: number of commercial fruits; NDF: number of deformed fruits; NWF: number of nodes without fruit; FM: fruit mass; FL: fruit length; FD: fruit diameter; PT: pulp thickness; SS: soluble solids content; PH: plant height; SD: stem diameter; FP: production of fruits per plant. \*significant at 5% probability by t test.

The multicollinearity diagnosis considering all variables was  $CN = 96,978.88$  for the Solo group and  $CN = 311,784.92$  for the Formosa group, indicating severe collinearity. According to Toebe and Cargnelutti Filho (2013a; b), the high degree of multicollinearity can result in trail analysis with direct and indirect biased effects, with values higher than 1 or lower than -1 and, therefore, without biological sense. To eliminate the multicollinearity of the analysis, the first strategy used was the exclusion of variables that had high correlation between themselves. However, although this is the statistical criterion, it is necessary to consider the objectives of the breeding program, and, if it is prudent, do not discard variables of interest despite the correlation (CRUZ et al., 2014). In this

study, FM was considered the most interesting variable, thus, it was preserved in the analysis even though it had a high correlation.

For the Solo group,  $CN < 100$ , indicating weak multicollinearity, was obtained after the exclusion of FD, FL, PT and NCF ( $CN = 85.53$ ), in that order. In the Formosa group, FD, PT, NWF and NCF were excluded to obtain  $CN = 62.31$ . Thus, after the exclusion of variables, the new matrix of phenotypic correlations had weak multicollinearity, being possible to carry out the analysis of the trail without its harmful effects.

In the Solo group, considering the FP as main variable and NDF, NWF, FM, SS, PH and SD as explanatory variables, we observed that the greatest direct

effect was the FM (0.52), followed by the SD (0.38). The NDF and NWF variables had direct negative effects (-0.49 and -0.35, respectively) (Table 4). In plant breeding, variables with greater direct and positive effects are more favorable to selection. Thus, the correlated response through indirect selection will be efficient, facilitating the breeder work (TEODORO et al., 2016). In this sense, for the Solo group, the FM for having greater direct and total effect on the main variable should be prioritized in the selection, since they indicate the relation of cause and effect. In addition, the indirect effect of FM via NDF was 0.26, increasing the possibility of success due to the indirect selection of genetic materials of papaya from the Solo group with higher FP.

For the heterotic Formosa group, the explanatory variables used in the trail analysis were NDF, FM, FL, SS, PH and SD. The highest positive direct effect under FP was obtained by PH (0.36), which added to the indirect effects had a total effect of 0.56 (Table 5). Therefore, according to this methodology, the PH is the characteristic that must be prioritized to obtain greater production of papaya fruits of the Formosa group, for its direct and indirect effect and for making possible the early selection.

The direct negative effects of NDF (-0.56), FM (-0.60) and SS (-0.77) on FP were observed (Table 5). According to Cruz et al. (2014), the relation of cause and effect on the basic variable is given by high direct correlation and indirect effect in a favorable sense. Thus, characteristics with a large direct effect, but with an indirect effect in the opposite direction, indicate that the auxiliary character is not the main determinant of the changes in the basic variable, and others may provide greater impact in terms of selection gain. In this sense, both NDF and FM are not a good determinant of the FP variation for papaya in the Formosa group, because despite having a great direct effect, its indirect effect has an opposite meaning, indicating absence of cause and effect relationship. On the other hand, these results indicate that in order to obtain higher FP, it would be necessary to decrease the soluble solids content of the fruits. The negative correlation between SS and FP observed in this study and by Oliveira et al. (2010) indicate the negative direction in the degree of association between these variables. This correlation may be related to a dilution effect of soluble solids by the fruit, that is, the larger the fruit size, the greater the amount of carbohydrates needed to raise the °Brix.

In the trail analysis in which the elimination of variables was used to circumvent multicollinearity, the coefficient of determination ( $R^2$ ) was 94% for the Solo group (Table 4) and 89% for the Formosa group (Table 5). The residual effect, in turn, was 0.25 for the Solo group and 0.33 for the Formosa group. This indicates that, although FP is a complex characteristic, the variables maintained in the analyzes explained their variation. However, despite a low residual effect, some indirect effect may have

acted under FP (TEODORO et al., 2016). These results corroborate with those obtained by Oliveira et al. (2010) for path analysis for the number of papaya commercial fruits, whose  $R^2$  was 0.87 and the residual effect was 0.25, indicating that the model in explaining the genetic effects related to the variable under analysis was well adjusted.

Despite the consistent results, the exclusion of characteristics should be evaluated with criteria, since the removal of variables with high explanatory power can reduce  $R^2$  and increase the residue (OLIVOTO et al., 2016). On the other hand, Toebe and Cargnelutti Filho (2013b) observed that the elimination of correlated variables, besides providing better  $R^2$ , can represent labor savings, since fewer variables would need to be measured to explain the main characteristic. The results of this research indicate that it is possible to exclude four variables, one of them being destructive (PT) and others difficult to measure (NCF and NWF), reducing labor costs and loss of commercial fruits. On the other hand, the elimination of variables was different for the two heterotic groups of papaya, evaluated under the same environmental conditions, probably due to the differences in plant and fruit architecture patterns between groups. Therefore, it is necessary to evaluate the direct and indirect effects on the variable of interest for the expected phenotypic pattern.

In some instances, the exclusion of features that generate multicollinearity cannot be performed by the breeder, due to, for example, a small number of explanatory variables or the importance of knowing their effects (OLIVOTO et al., 2016). In these cases, a second option is to perform the analysis with all explanatory characteristics, but with the addition of a constant ( $k$ ) to the diagonal elements of the  $X'X$  matrix, known as ridge path analysis. The value of  $k$  to be used should be the lowest value able to stabilize most of the path coefficient estimators (TOEBE; CARGNELUTTI FILHO, 2013b; CRUZ et al., 2014). Thus, values of  $k = 0.05$  were previously employed by Amorim et al. (2008) and Moreira et al. (2013), and  $k = 0.10$  by Toebe and Cargnelutti Filho (2013a), allowing path analysis with all available variables.

The coefficients of the ridge path analysis obtained for the Solo group indicated that the greatest direct effects under FP were via NCF (0.39), PT (0.36) and FD (0.29), and FD still had large indirect effect via PT. Other characteristics related to FP were FM and FL, however, despite the large overall effect of FM and FL under FP, this effect is not direct, but indirect, via FD and PT (Table 6). Both PT and FD were previously excluded in path analysis to circumvent the effects of multicollinearity (Table 4). However, in this ridge path analysis, it is observed that these variables, besides the direct effects, exert indirect effects on each other, being able to be used both for the direct selection and for the indirect selection of papaya materials of the Solo group with higher FP. The two analyzes, however, allowed determination coefficients

above 94%, allowing the breeder to decide which of these variables to use in the selection. This choice can be based on the easiness of measurement, the precocity or the greater heritability of the characteristic.

In the heterotic Formosa group, the greatest direct positive effects were obtained by NCF (0.77) and FL (0.18) and negative by NDF (-0.20) and SS (-0.17). However, the NCF besides the great direct and total effect (0.83), provided indirect effects for the selection of the other analyzed variables, except FL and SD (Table 7). Therefore, this characteristic should be prioritized to evaluate the production of papaya from the Formosa group. However, NCF is a variable measured only when the fruits are at harvest point, delaying the selection process. In a trail analysis for NCF in papaya, Oliveira et al. (2010), observed that leaf width, plant height and number of flowers per peduncle are determinant characteristics of

variations in NCF, are easy to measure and that manifest themselves before production, making possible the early selection.

The coefficient of determination of ridge path analysis for papaya from the Formosa group was 0.93 (Table 6) for the Solo group was 0.97 (Table 5). This result corroborates to those obtained by Moreira et al. (2013), who had  $R^2$  of 89.59% and 82.41% with the ridge path analysis and with the discard of variables, respectively. On the other hand, Toebe and Cargnelutti Filho (2013a, b) with use of  $k = 0.10$ , indicated that the elimination of variables is more adequate than the ridge path analysis. For both heterotic groups, this analysis explained better the FP variation than the exclusion of variables, allowing the analysis with all the evaluated characters and the better understanding of the relationships between them.

**Table 4.** Estimates of direct (diagonal in bold) and indirect (off diagonal) effects of six agronomic characters on fruit yield per plant in papaya from the heterotic Solo group obtained by path analysis.

Variables <sup>1</sup>	NDF	NWF	FM	SS	PH	SD
NDF	<b>-0.49</b>	0.03	0.26	0.05	-0.21	-0.19
NWF	0.02	<b>-0.35</b>	-0.13	-0.02	0.07	0.07
FM	-0.28	0.20	<b>0.52</b>	-0.05	-0.10	0.03
SS	0.05	-0.02	0.04	<b>-0.41</b>	-0.13	-0.12
PH	-0.03	0.02	0.01	-0.02	<b>-0.07</b>	-0.05
SD	0.15	-0.08	0.02	0.11	0.23	<b>0.38</b>
Total	-0.58	-0.20	0.73	-0.34	-0.20	0.13
Determination coefficient						0.94
Effect of the residual variable						0.25

<sup>1</sup> NDF: number of deformed fruits; NWF: number of nodes without fruit; FM: fruit mass; SS: soluble solids content; PH: plant height; SD: stem diameter.

**Table 5.** Estimates of the direct (diagonal in bold) and indirect (off diagonal) effects of six agronomic characters on the fruit yield per plant in papaya from the heterotic Formosa group obtained by the path analysis.

Variables <sup>1</sup>	NDF	FM	FL	SS	PH	SD
NDF	<b>-0.56</b>	0.39	0.41	0.07	-0.07	-0.04
FM	0.42	<b>-0.60</b>	-0.27	0.05	0.27	-0.03
FL	0.01	0.00	<b>-0.01</b>	0.00	0.00	0.00
SS	0.09	0.07	0.17	<b>-0.77</b>	0.03	-0.07
PH	0.05	-0.16	0.14	-0.01	<b>0.36</b>	-0.02
SD	0.02	0.02	-0.04	0.03	-0.02	<b>0.30</b>
Total	0.03	-0.29	0.40	-0.64	0.56	0.14
Determination coefficient						0.89
Effect of the residual variable						0.33

<sup>1</sup> NDF: number of deformed fruits; FM: fruit mass; FL: fruit length; SS: soluble solids content; PH: plant height; SD: stem diameter.

**Table 6.** Estimates of the direct (diagonal in bold) and indirect (off diagonal) effects of ten agronomic characters on fruit yield per plant (FP) in papaya from the heterotic Solo group obtained by the ridge path analysis.

Variables <sup>1</sup>	NCF	NDF	NWF	FM	FL	FD	PT	SS	PH	SD
NCF	<b>0.39</b>	-0.05	-0.30	-0.13	-0.12	-0.16	-0.02	-0.12	0.01	0.04
NDF	0.01	<b>-0.09</b>	0.01	0.05	0.05	0.04	0.05	0.01	-0.04	-0.04
NWF	0.18	0.02	<b>-0.24</b>	-0.09	-0.13	-0.08	-0.09	-0.01	0.05	0.05
FM	-0.07	-0.11	0.08	<b>0.21</b>	0.20	0.20	0.18	-0.02	-0.04	0.01
FL	-0.03	-0.06	0.06	0.10	<b>0.11</b>	0.09	0.10	-0.01	-0.02	0.01
FD	-0.12	-0.12	0.09	0.27	0.24	<b>0.29</b>	0.22	-0.03	-0.06	0.01
PT	-0.02	-0.18	0.13	0.31	0.33	0.27	<b>0.36</b>	-0.12	-0.13	0.02
SS	0.01	0.00	0.00	0.00	0.00	0.00	0.02	<b>-0.04</b>	-0.01	-0.01
PH	0.00	0.01	-0.01	0.00	-0.01	-0.01	-0.01	0.01	<b>0.02</b>	0.01
SD	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	<b>0.02</b>
Total	0.39	-0.58	-0.20	0.73	0.68	0.66	0.81	-0.34	-0.20	0.13
Determination coefficient										0.97
<i>k</i> value										0.05
Effect of the residual variable										0.18

<sup>1</sup> NCF: number of commercial fruits; NDF: number of deformed fruits; NWF: number of nodes without fruit; FM: fruit mass; FL: fruit length; FD: fruit diameter; PT: pulp thickness; SS: soluble solids content; PH: plant height; SD: stem diameter.

**Table 7.** Estimates of the direct (diagonal in bold) and indirect (off diagonal) effects of ten agronomic characters on the fruit yield per plant in papaya from the heterotic Formosa group obtained by the ridge path analysis.

Variables <sup>1</sup>	NCF	NDF	NWF	FM	FL	FD	PT	SS	PH	SD
NCF	<b>0.77</b>	0.39	-0.65	-0.56	-0.06	-0.59	-0.42	-0.40	0.42	0.02
NDF	-0.10	<b>-0.20</b>	0.09	0.14	0.14	0.11	0.13	0.02	-0.03	-0.01
NWF	0.09	0.05	<b>-0.11</b>	-0.08	-0.02	-0.08	-0.06	-0.03	0.05	0.03
FM	-0.10	-0.10	0.11	<b>0.14</b>	0.06	0.13	0.12	-0.01	-0.06	0.01
FL	-0.01	-0.13	0.03	0.08	<b>0.18</b>	0.04	0.08	-0.04	0.07	-0.02
FD	0.01	0.01	-0.01	-0.02	0.00	<b>-0.02</b>	-0.02	0.00	0.01	0.00
PT	-0.03	-0.03	0.03	0.05	0.02	0.04	<b>0.05</b>	-0.01	-0.02	0.02
SS	0.09	0.02	-0.05	0.01	0.04	0.00	0.02	<b>-0.17</b>	0.01	-0.01
PH	0.06	0.02	-0.05	-0.05	0.05	-0.06	-0.04	0.00	<b>0.12</b>	-0.01
SD	0.00	0.01	-0.04	0.01	-0.02	0.02	0.04	0.01	-0.01	<b>0.12</b>
Total	0.83	0.03	-0.66	-0.29	0.40	-0.41	-0.09	-0.64	0.56	0.14
Determination coefficient										0.093
<i>k</i> value										0.05
Effect of the residual variable										0.27

<sup>1</sup> NCF: number of commercial fruits; NDF: number of deformed fruits; NWF: number of nodes without fruit; FM: fruit mass; FL: fruit length; FD: fruit diameter; PT: pulp thickness; SS: soluble solids content; PH: plant height; SD: stem diameter.



## Conclusions

The correlations and effects of the variables on fruit weight by path analysis were distinct in the two heterotic groups of papaya, indicating that the relationships between the variables should be studied within each heterotic group.

The elimination of variables to circumvent multicollinearity allowed a good adjustment of the model to the main variable, which can generate savings of physical and financial resources. However, ridge path analysis resulted in a better determination of the variables influencing fruit production, both in the Solo and in the Formosa group.

## Acknowledgments

To Mr. Carlito Covre, for giving away the area and inputs to carry out this research, the Embrapa Mandioca e Fruticultura, the Universidade Estadual do Norte Fluminense Darcy Ribeiro and Caliman Agrícola S / A for making available the seeds of the evaluated genetic material.

## References

- ALVARES, C.A.; STAPE, J.L.; SENTELHAS, P.C.; DE MORAES, G.; LEONARDO, J.; SPAROVEK, G. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, Berlin, v.22, n.6, p.711-728, 2013.
- AMORIM, E.P.; RAMOS, N.P.; UNGARO, M.R.G.; KIIHL, T.A. Correlações e análise de trilha em girassol. *Bragantia*, Campinas, v.67, n.2, p.307-316, 2008.
- BARROS, F.L.S.; KUHLCAMP, K.T.; ARANTES, S.D.; MOREIRA, S.O. Productivity and quality of papaya fruit Formosa and Solo in two harvest seasons. *Pesquisa Agropecuária Brasileira*, Brasília, DF, v.52, n.8, p.599-606, 2017.
- COSTA, A.F.S.; DANTAS, J.L.L.; PEREIRA, M.G.; CATTANEO, L.F.; COSTA, A.N.; MOREIRA, S.O. Botânica, melhoramento e variedades. *Informe Agropecuário*, Belo Horizonte, v.34, n.275, p.14-24, 2013.
- CRUZ, C.D. Genes - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum. Agronomy*, Maringá, v.35, n.3, p.271-276, 2013.
- CRUZ, C.D.; CARNEIRO, P.C.S.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético vegetal**. 3.ed.. Viçosa: Editora UFV, 2014. v.2, 668p.
- LUZ, L.N.; PEREIRA, M.G.; BARROS, F.R.; BARROS, G.B.; FERREGUETTI, G.A. Novos híbridos de mamoeiro avaliados nas condições de cultivo tradicional e no semiárido brasileiro. *Revista Brasileira de Fruticultura*, Jaboticabal, v.37, n.1, p.159-171, 2015.
- MARTINS, D.S.; COSTA, A.A.F. **A cultura do mamoeiro: tecnologias de produção**. Vitória: Incaper, 2003. 497 p. Disponível em: <<http://biblioteca.incaper.es.gov.br/digital/bitstream/item/924/1/Livro-A-Cultura-do-Mamoeiro-1.pdf>> Acesso em: 17 jan. 2017.
- MONTGOMERY, D.C.; PECK, E.A.; VINING, G.G. **Introduction to linear regression analysis**. 5<sup>th</sup> ed. New York: John Wiley & Sons, 2012. 672p.
- MOREIRA, S.O.; GONÇALVES, L.S.A.; RODRIGUES, R.; SUDRÉ, C.P.; AMARAL JÚNIOR, A.T.; MEDEIROS, A.M. Correlações e análise de trilha sob multicolinearidade em linhas recombinadas de pimenta (*Capsicum annuum* L.) *Revista Brasileira de Ciências Agrárias*, Recife, v.8, n.1, p.15-20, 2013.
- OLIVEIRA, E. J.; LIMA, D. S.; LUCENA, R. S.; MOTTA, T. B. N.; DANTAS, J. L. L. Correlações genéticas e análise de trilha para número de frutos comerciais por planta em mamoeiro. *Pesquisa Agropecuária Brasileira*, Brasília, DF, v.45, n.8, p.855-862, 2010.
- OLIVEIRA, E.J.; FRAIFE FILHO, G.A.; FREITAS, J.P.X.; DANTAS, J.L.L.; RESENDE, M.D.V. Plant selection in F<sub>2</sub> segregating populations of papaya from commercial hybrids. *Crop Breeding and Applied Biotechnology*, Londrina, v.12, n.3, p.191-198, 2012.
- OLIVOTO, T.; NARDINO, M.; CARVALHO, I.R.; FOLLMANN, D.N.; SZARESKI, V.J.; FERRARI, M.; PELEGRIN, A.J.; SOUZA, V.Q. Pearson correlation coefficients and accuracy of path analysis used in maize breeding: a critical review. *International Journal of Current Research*, Abingdon, v.8, n.9, p.37787-37795, 2016.
- SILVA, C.A.; SCHMILDT, E.R.; SCHMILDT, O.; ALEXANDRE, R.S.; CATTANEO, L.F.; FERREIRA, J.P.; NASCIMENTO, A.L. Correlações fenotípicas e análise de trilha em caracteres morfoagronômicos de mamoeiro. *Revista Agroambiente*, Roraima, v.10, n.3, p.217-227, 2016.

TEIXEIRA, D.H.L.; OLIVEIRA, M.S.P.; GONÇALVES, F.M.A.; NUNES, J.A.R. Correlações genéticas e análise de trilha para componentes da produção de frutos de açaizeiro. **Revista Brasileira de Fruticultura**, Jaboticabal, v.34, n.4, p.1135-1142, 2012.

TEODORO, P.E.; COSTA, R.D.; ROCHA, R.B.; LAVIOLA, B.G. Contribuição de caracteres agronômicos para a produtividade de grãos em pinhão-manso. **Bragantia**, Campinas, v.75, n.1, p.51-56, 2016.

TOEBE, M.; CARGNELUTTI FILHO, A. Multicollinearity in path analysis of maize (*Zea mays* L.). **Journal of Cereal Science**, London, v.57, n.3, p.453-462, 2013a.

TOEBE, M.; CARGNELUTTI FILHO, A. Não normalidade multivariada e multicolinearidade na análise de trilha em milho. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v.48, n.5, p.466-477, 2013b.

WRIGHT, S. The theory of path coefficients a reply to nile's criticism. **Genetics**, Austin, v.8, n.3, p.239-255, 1923.