

Path analysis in multicollinearity for fruit traits of pepper

Análisis de ruta con multicolinealidad de las características de la fruta de la pimienta

Anderson Rodrigo da Silva^{1*}, Moysés Nascimento¹, Paulo Roberto Cecon¹,
Moryb JLC Sapucay², Elizanilda Ramalho do Rêgo³, Lidiane Aparecida Barbosa³

ABSTRACT

This study aimed to use path analysis in multicollinearity to assess the genotypic correlation coefficients and their partitions in direct and indirect effects of morphological traits of fruits on the dry matter content of pepper. The experimental plot was a randomized block design with 2 replications and 10 fruits per plot. The data were obtained from the characterization of 10 accessions (varieties) of four pepper species of the genus *Capsicum*. We assessed the peduncle length, fruit length, largest fruit diameter, pericarp thickness, average fruit weight and dry matter content. We carried out the multicollinearity diagnosis and the ridge path analysis to partition the genotypic correlation coefficients into direct and indirect effects, considering dry matter content as the basic variable. The largest fruit diameter showed negative genotypic correlation with the basic variable; however, it is the morphological trait that showed, in isolation, the greatest importance to explain the variation of dry matter content in fruits, and may therefore be used as an auxiliary criterion in processes of indirect selection.

Key words: *Capsicum* spp., correlation; genetic breeding; pepper; indirect selection.

RESUMEN

Este estudio tuvo como objetivo utilizar el análisis de ruta bajo multicolinealidad para evaluar los coeficientes de correlación genotípica y su evolución en los efectos directos e indirectos de las características morfológicas de la fruta en el contenido de materia seca de frutos de pimienta. El experimento se realizó en un diseño de bloques al azar con dos repeticiones de 10 frutos por parcela. Los datos utilizados fueron de las caracterizaciones de la fruta de diez accesiones (variedades) de cuatro especies de pimienta. Las características fueron: longitud del pedúnculo, longitud del fruto, mayor diámetro de fruto, espesor del pericarpio, peso promedio del fruto y el contenido de materia seca. Se realizó el diagnóstico de multicolinealidad y el análisis de la ruta cresta a la partición de los coeficientes de correlación genotípica para los efectos directos e indirectos, teniendo en cuenta el contenido de materia seca. El diámetro del fruto más grande mostró correlación genotípica negativa con la variable de base; por otra parte, era el carácter que mostró aislamiento morfológico, mayor importancia para explicar la variación en el contenido de materia seca de las frutas que se pueden utilizar como criterio auxiliar en los procesos de selección indirecta.

Palabras clave: *Capsicum* spp., correlación, mejora genética, pimienta, selección indirecta.

Introduction

Fruit traits are essential for pepper fruits used in the industry or cooking, either *in natura* or as paprika, or paste or dehydrated pepper or in conserves (Pinto *et al.*, 2007). For Sousa & Maluf (2003), Rêgo *et al.* (2006), Rêgo *et al.* (2010),

the dry matter content, as well as the variables of chemical quality, total soluble solid contents and color pungency, are important traits in pepper used in the industry of pepper powder.

To obtain greater yields in traits of industrial importance, such as the dry matter content of the fruit, it is essential to develop new pepper varieties through

¹ Universidade Federal de Viçosa (UFV), Departamento de Estatística, Av. P.H. Rolfs, s/n, Campus Universitário, 36570-000, Viçosa, MG, Brasil.

² UFV, Departamento de Fitotecnia.

³ Universidade Federal da Paraíba, Laboratório de Biotecnologia Vegetal, Caixa Postal 02, 58397-000, Areia, PB, Brasil.

* Corresponding author: anderson.rodrigo@ufv.br

genetic breeding programs. In these programs, the correlation between variables is important when we aim to carry out simultaneous selection among the traits or when the trait of interest shows low heritability and/or difficulties to be measured or identified.

However, the high correlation between two traits may result from the effect of a third trait on them, or from a group of traits (Cruz & Regazzi, 1994). Furthermore, simple correlation does not provide information on the direct and indirect effects of a group of traits regarding a given trait considered as a trait of greater interest (Cruz, 2001).

The partition of correlations between traits into direct and indirect effects on a basic variable of greater interest, a methodology known as path analysis (Wright, 1921) has been used in several crops, such as peanut (Santos *et al.*, 2000), bean (Furtado *et al.*, 2002), pepper (Rêgo *et al.* 2001), rice (Marchezan *et al.*, 2005) and tomato (Sobreira *et al.*, 2009), among others.

Path analysis may be understood as a multiple regression analysis, and as such it is usual to observe multicollinearity between the explicative variables, which may greatly affect estimates for path coefficients and therefore lead to misinterpretations. Variances of coefficient estimates and predictors may increase significantly, masking the significance of variables used in the model, limiting its applicability (Souza, 1998).

The adverse effects of multicollinearity may be overcome by eliminating the variables from the regression model or by carrying out ridge path analysis as proposed by Carvalho (1995), as an alternative to the estimation method of minimum squares to estimate parameters (Carvalho *et al.*, 1999). This method was also used by Carvalho *et al.* (1999) in bell pepper.

The objective of this study was to assess, using path analysis, the coefficients of genotypic correlation and their partition into direct and indirect effects of morphological fruit traits on the dry matter content of pepper using multicollinearity, for indirect selection of the variable of dry matter content of the fruit.

Materials and Methods

The experiment was conducted in an experimental area at the Center for Agricultural Sciences of the Universidade Federal da Paraíba (CCA-UFPB), in the municipality of Areia, Paraíba State, Brazil,

during the year 2009. We used data collected from morphological characteristics of fruits of ten accessions (varieties) of four pepper species from the genus *Capsicum*, which are: *Capsicum chinense* (6 accessions), *C. annuum* (1 accession), *C. baccatum* (1 accession) and *C. frutescens* (2 accessions). The species belonged to the Vegetable Germplasm Bank (VGB) of CCA-UFPB. The plants were planted in expanded polystyrene trays with 128 cells containing commercial substrate. After shooting of six true leaves, the plants were transplanted to the field in a randomized block design with 2 replications and 10 plants per plot, and each replication corresponded to the arithmetic mean of 30 fruits. A total of 20 data were used in the analysis. Crop care was applied in accordance with Filgueira (2000). The fruits were collected at a mature stage and immediately characterized according to the *Capsicum* descriptors (IPGRI, 1995). We evaluated the traits of peduncle length (PL), fruit length (FL), largest fruit diameter (LFD), pericarp thickness (PT) in centimeters; average fruit weight (AFW) in grams and dry matter content (DMC) in percentage ($100 \times$ ratio dry matter/fresh matter).

To investigate the existence of genetic variability among the accessions, estimates of genotypic correlation and the coefficient of genotypic determination were obtained according to Mode & Robinson (1959) and Vencovsky & Barriga (1992), respectively.

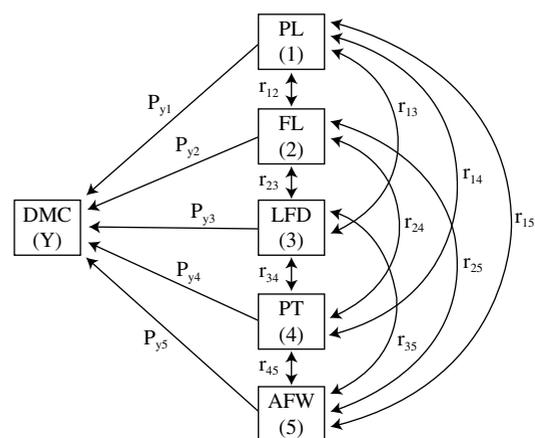


Figure 1. Previously established causal diagram, considering the DMC as the basic variable. The unidirectional arrow indicates the direct effect of each explicative variable on the basic variable. The bidirectional arrow represents the interdependence of two explicative variables, whose magnitude is quantified in the genotypic correlation.

The diagnosis of multicollinearity of the $X'X$ matrix was verified according to Montgomery & Peck (1981), which is based on the condition number (CN) that consists of the ratio between the largest and smallest eigenvalue of the $X'X$ matrix, where: $CN < 100$ – weak multicollinearity; $100 < CN < 1000$ – moderate to severe multicollinearity; $CN > 1000$ – severe multicollinearity. Moreover, Montgomery & Runger (2008) report that the presence of multicollinearity may also be easily observed by the magnitude of variance inflation factors (VIF), which are the diagonal elements of the $X'X$ matrix, meaning that if any VIF exceeds 10, the multicollinearity will be a problem.

After determining the presence of multicollinearity, we used the method proposed by Carvalho (1995) to obtain path coefficients, which consists of adding a constant k to the $X'X$ matrix diagonal terms, slightly altering the normal equation system. Thus, the path coefficients were obtained by solving the system: $(X'X - kI)\theta^* = X'Y$, where $X'X$ is the matrix of genotypic correlations between the explicative variables of the model; k is a small amount subtracted from the elements of the main diagonal axis of $X'X$ matrix; I is the identity matrix; θ^* is the vector of estimators of path coefficients (P_{yi}) and $X'Y$ is the matrix of genotypic correlations between the basic variable and each explicative variable. The value of k was established in the graphic analysis by plotting the estimates of path coefficients against the value of k in the interval $0 < k < 1$, and by obtaining the ridge trace according to the regression method for ridges proposed by Hoerl & Kennard (1970). We used the smallest value of k capable of stabilizing most estimates of path coefficients.

The results of the path analysis were interpreted according to Vencovsky & Barriga (1992), considering that high correlation coefficients and direct effects

(path coefficients) indicate that these independent variables explain part of the change of the basic variable and that positive or negative correlation coefficients, but with direct effect of different or insignificant coefficient signal, indicate that variables with greater indirect effects should be considered simultaneously to explain the change in the basic variable.

All analyses were carried out using the Genes program, version 2009.7.0 (Cruz, 2006).

Results and Discussion

Genetic variability among the accessions for all fruit traits assessed was observed. All traits showed high heritability indexes (coefficient of genotypic determination) in the broad sense, ranging from 87.33% (PL) to 98.58% (DMC), which is useful for genetic breeding through indirect selection. High values (above 80%) of heritability in the broad sense were found by Rêgo *et al.* (2010) studying the phenotypic diversity, correlation and importance of traits in *Capsicum baccatum*. The coefficients of variation (CV%) of the experiment in relation to the morphological traits ranged from 6.02% (PL) to 23.57% (AFW), which are considered low to moderate according to Silva *et al.* (2011).

The basic variable (DMC) correlated negatively with all other variables (Table 1). Sapucay *et al.* (2009) studied four pepper species and found similar results, indicating a negative phenotypic correlation between DMC and LFD, AFW and PT, and insignificant correlation with variables PL and FL. Rêgo *et al.* (2010) have shown that in *C. baccatum*, DMC correlated negatively with the fruit traits evaluated, namely pericarp thickness, larger and smaller fruit diameter of fruits and fruit weight, which are in line with the results found in our study.

Table 1. Estimates of genotypic correlation coefficients (r_g) between traits of pepper fruits: peduncle length (PL), fruit length (FL), largest fruit diameter (LFD), pericarp thickness (PT), average fruit weight (AFW) and dry matter content (DMC).

Traits	PL (1)	FL (2)	LFD (3)	PT (4)	AFW (5)	DMC (Y)
PL (1)	1					
FL (2)	0.1723 ^{NS}	1				
LFD (3)	0.6210 ^{**}	0.0013 ^{NS}	1			
PT (4)	0.3492 ^{NS}	0.0817 ^{NS}	0.6791 ^{**}	1		
AFW (5)	0.7106 ^{**}	0.1470 ^{NS}	0.9289 ^{**}	0.7766 ^{**}	1	
DMC (Y)	-0.4563 [*]	-0.4806 [*]	-0.7435 ^{**}	-0.7089 ^{**}	-0.7195 ^{**}	1

^{NS}, * and **: non-significant, significant at 5% and at 1% probability by the t test, respectively.

The accessions that showed greater fruit average also presented larger fruit diameter average ($r_g = 0.9289$). Furthermore, there were positive correlations between the explicative variables LFD and PL; LFD and PT; AFW and FL; AFW and PT, which are similar to the results observed in Rêgo *et al.* (2010).

The multicollinearity diagnosis showed CN = 384.72, indicating moderate to severe multicollinearity, with four VIF values greater than or equal to 10 in absolute values. In this case, in the path analysis with multicollinearity with the genotypic correlation matrix of explicative variables a value of $k = 0.04$ added, which was effective in stabilizing the estimates for path coefficients (Figure 2), with a determinant of the $X'X$ matrix = 0.045 and CN = 41.81, indicating weak multicollinearity. Nevertheless, the VIFs no longer exceeded the limit of 10.

The coefficient of determination of the model for the path analysis was high ($R^2 = 0.8384$), which shows that the variables explained a great part of the variation in the basic variable (Table 2). The traits PL and FL showed negative correlation with the basic variable (DMC) and low values of the path coefficients ($P_{y1} = -0.0394$ and $P_{y2} = -0.4850$), and the cause/effect relation was not observed. Therefore, we observe that these traits cannot be used to obtain satisfactory genetic gains in DMC.

The trait LFD showed the greatest direct effect ($P_{y3} = -0.8131$) on the basic variable, with significant

correlation and of the same sign, evidencing that this fruit trait can be used for the indirect selection of DMC, which means that by selecting fruits with lower LFD we are indirectly selecting fruits with greater DMC. According to Rêgo *et al.* (2009), this trait is determined by additive gene effects that make its selection effective in initial segregating generations of breeding programs based on hybridization methods. The authors also indicate the possibility to use backcrossing to insert the desirable trait due to the additive nature of genes.

Table 2. Path analysis of the basic variable DMC under the estimates of direct and indirect effects of explicative variables: peduncle length (PL), fruit length (FL), largest fruit diameter (LFD), pericarp thickness (PT), average fruit weight (AFW).

Variable	Effect	Estimate	VIF ¹	Estimator
PL (cm)	Direct	-0.0394	2.25	P_{y1}
	Indirect by FL	-0.0835	0.03	$P_{y2}r_{12}$
	Indirect by LFD	-0.5049	1.99	$P_{y3}r_{13}$
	Indirect by PT	-0.1555	0.30	$P_{y4}r_{14}$
	Indirect by AFW	0.3288	4.53	$P_{y5}r_{15}$
	Total (r_g)	-0.4563		
FL (cm)	Direct	-0.4850	1.16	P_{y2}
	Indirect by PL	-0.0067	0.06	$P_{y1}r_{12}$
	Indirect by LFD	-0.0010	0.00	$P_{y3}r_{23}$
	Indirect by PT	-0.0363	0.01	$P_{y4}r_{24}$
	Indirect by AFW	0.0680	0.19	$P_{y5}r_{25}$
	Total (r_g)	-0.4806		
LFD (cm)	Direct	-0.8131	5.60	P_{y3}
	Indirect by PL	-0.0244	0.80	$P_{y1}r_{13}$
	Indirect by FL	-0.0006	0.00	$P_{y2}r_{23}$
	Indirect by PT	-0.3024	1.15	$P_{y4}r_{34}$
	Indirect by AFW	0.4298	7.74	$P_{y5}r_{35}$
	Total (r_g)	-0.7435		
PT (cm)	Direct	-0.4453	2.71	P_{y4}
	Indirect by PL	-0.0137	0.25	$P_{y1}r_{14}$
	Indirect by FL	-0.0396	0.00	$P_{y2}r_{24}$
	Indirect by LFD	-0.5522	2.39	$P_{y3}r_{34}$
	Indirect by AFW	0.3593	5.41	$P_{y5}r_{45}$
	Total (r_g)	-0.7089		
AFW (g)	Direct	0.4627	9.70	P_{y5}
	Indirect by PL	-0.0280	1.05	$P_{y1}r_{15}$
	Indirect by FL	-0.0713	0.02	$P_{y2}r_{25}$
	Indirect by LFD	-0.7553	4.47	$P_{y3}r_{35}$
	Indirect by PT	-0.3458	1.51	$P_{y4}r_{45}$
	Total (r_g)	-0.7195		
Coefficient of determination		0.8384	R^2	
Effect of residual variable		0.4019	$(1 - R^2)^{0.5}$	
Value of k used in the analysis		0.0400		

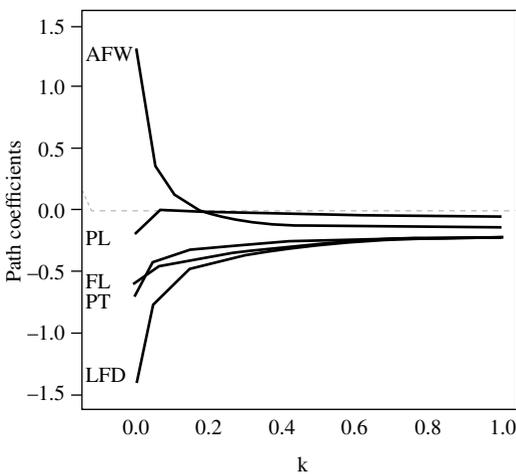


Figure 2. Estimates of path coefficients for k values obtained in the ridge path analysis, using dry matter content as the basic variable.

¹ Variance Inflation Factor.

For the PL trait we can observe that, despite the significant negative correlation ($r_g = -0.7089$), the direct effect on DMC is of the same sign and considered low ($P_{y4} = -0.4453$), given that, in absolute value it is equivalent to the residual variable effect (0.4019).

The AFW showed correlation and direct effect of different signs ($r_g = 0.4627$ and $P_{y5} = -0.7195$), which means that we should consider the LFD trait (greater indirect effect $P_{y3}r_{35} = -0.7553$) in the variation of the basic variable to be explained.

Given that the LFD measurements are simpler and less costly, because they do not require drying in an oven, which uses electricity, we should practice the indirect selection for the dry matter content through the largest fruit diameter trait.

The morphological trait of pepper fruits that presented, in isolation, the greatest importance in explaining the variation in fruit dry matter content was the largest fruit diameter (LFD), which can be used as a criterion to help in indirect selection.

Literature Cited

- Carvalho, C.G.P. *et al.*
1999. Análise de trilha sob multicolinearidade em pimentão. *Pesquisa Agropecuária Brasileira*. Brasília, v. 34, n. 4, pp. 603-613.
- Carvalho, S.P.
1995. Métodos alternativos de estimação de coeficientes de trilha e índices de seleção, sob multicolinearidade. Viçosa: UFV. 163 p.
- Cruz, C.D.
2006. Programa Genes: Biometria. Viçosa: Editora UFV. 382 p.
- Cruz, C.D.; Regazzi, A.J.
1994. Modelos biométricos aplicados ao melhoramento genético. Viçosa: UFV. 390 p.
- Filgueira, F.A.R.
2000. Novo Manual de Olericultura: Agrotecnologia moderna na produção e comercialização de hortaliças. Viçosa: UFV. 402 p.
- Furtado, M.R. *et al.*
2002. Análise de trilha do rendimento do feijoeiro e seus componentes primários em monocultivo e em consórcio com a cultura do milho. *Ciência Rural*, v. 32, n. 2, pp. 217-220.
- Hoerl, A.E.; Kennard, R.W.
1970. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, v. 12, n. 1, pp. 55-68.
- Ipgri, Avrdc; Catie.
1995. Descriptors for *Capsicum* (*Capsicum* spp.). International Plant Genetic Resources Institute, Rome, Italy; the Asian Vegetable Research and Development Center, Taipei, Taiwan, and the Centro Agronómico Tropical de Investigación y Enseñanza, Turrialba, Costa Rica. 110 p.
- Marchezan, E. *et al.*
2005. Análise de coeficiente de trilha para os componentes de produção em arroz. *Ciência Rural*, v. 35 n. 5, pp. 1027-1033.
- Mode, J.C.; Robinson, H.F.
1959. Pleiotropism and the genetic variance and covariance. *Biometrics*, v. 15, n. 4, pp. 518-537.
- Montgomery, D.C.; Peck, E.A.
1981. Introduction to linear regression analysis. New York: John Wiley & Sons. 504 p.
- Montgomery, D.C.; Runger, R.C.
2008. Estatística aplicada e probabilidade para engenheiros. (Trad. Verônica Calado). Rio de Janeiro: LTC. 463 p.
- Pinto, C.M.F. *et al.*
2007. Pimenta (*Capsicum* spp.). In: Paula Júnior, T.J; Venzon, M. (Coords). 101 Culturas: manual de tecnologias agrícolas. Belo Horizonte: EPAMIG, pp. 625-632.
- Rêgo, E.R. *et al.*
2001. Correlações entre caracteres morfoagronômicas e produção de *Capsicum baccatum*. Anais do 1º Congresso Brasileiro de Melhoramento de plantas, Centro de Cultura e Convenções de Goiânia, Goiânia-GO.
- Rêgo, E.R. *et al.*
2006. Caracterização, diversidade e estimação de parâmetros genéticos em pimenteiros (*Capsicum* spp.). Anais do II Encontro Nacional do Agronegócio Pimentas (*Capsicum* spp.).
- Rêgo, E.R. *et al.*
2009. A diallel study of yield components and fruit quality in chilli pepper (*Capsicum baccatum*). *Euphytica* (Wageningen) 168: 275-287.
- Rêgo, E.R. *et al.*
2010. Phenotypic diversity, correlation and importance of variables for fruit quality and yield traits in Brazilian peppers (*Capsicum baccatum*). *Genetic Resources and Crop Evolution*. DOI 10.1007/s10722-010-9628-7.
- Santos, R.C.; Carvalho, L.P.; Santos, V.F.
2000. Análise de coeficiente de trilha para os componentes de produção em amendoim. *Ciência e Agrotecnologia*, v. 24, n. 1, pp. 13-16.
- Sapucay, M.J.L.C. *et al.*
2009. Diversidade genética, importância relativa e correlação de caracteres quantitativos em pimenteiros. In: 49º Congresso Brasileiro de Olericultura, Águas de Lindóia. *Horticultura Brasileira*. Brasília: ABH, v. 27, S1161-S1168.
- Silva, A.R. *et al.*
2011. Avaliação do coeficiente de variação experimental para caracteres de frutos de pimenteiros. *Rev. Ceres*, v.58, n. 1, pp. 695-700.
- Sobreira, F.M. *et al.*
2009. Análise de trilha em pós-colheita de tomate tipo salada. *Rev. Fac. Nal. Agr. Medellín*, v. 62, n. 1, pp. 4983-4988.

Souza, G.S.

1998. Introdução aos modelos de regressão linear e não-linear. Brasília: Embrapa-SPI/Embrapa-SEA. 505 p.

Souza, J.A.; Maluf, W.R.

2003. Diallel analyses and estimation of genetic parameters of hot pepper (*Capsicum chinense* Jacq). *Scientia Agricola*. v. 60, pp. 105-113.

Vencovsky, R.; Barriga, P.

1992. Genética biométrica no fitomelhoramento. Ribeirão Preto: Sociedade Brasileira de Genética. 486 p.

Wright, S.

1921. Correlation and causation. *Journal of Agricultural Research*, v. 20, n. 3, pp. 557-585.