Path analysis of yield components in tropical pumpkin[1]

Miguel Espitia-Camacho[2], Franco A. Vallejo-Cabrera[3], Diosdado Baena-Garcia[3] and Linda Wessel-Beaver[4]

ABSTRACT

Path coefficient analysis was carried out in tropical pumpkin (*Cucurbita moschata* Duchesne) for yield per plant as a function of number of fruit per plant and weight per fruit, using phenotypic and genetic correlations from two diallels: one using five varieties (variety diallel) and a second using five S1 lines derived from the varieties (line diallel). A randomized complete block design with five replications was used to evaluate 15 genotypes from each diallel, not including reciprocals. Direct effects were 300 to 500% greater than indirect effects in the line diallel. In contrast, direct effects were only 17 to 68% greater than indirect effects in the variety diallel. Effects (direct and indirect) were greater (17 to 500%) when genetic correlations rather than phenotype correlations (42 to 440%) were used in the path analysis. Fruit weight showed a greater effect on yield than did number of fruit, with values between 3.2 to 4.9 times the residual effects. Fruit weight and number of fruit can be used as selection criteria to increase yield in tropical pumpkin.

Key words: *Cucurbita moschata*, cucurbits, yield components

RESUMEN

Análisis de sendero de componentes de rendimiento en la calabaza

El presente estudio tuvo como objetivo realizar un análisis de sendero para producción por planta en función del número de frutas por planta y el peso por fruta en dos dialélicos de calabaza (*Cucurbita moschata* Duchesne): uno entre cinco variedades y otro entre cinco lineas S1, originadas de las anteriores. Se utilizaron correlaciones fenotípicas tanto como correlaciones genéticas. Se utilizó un diseño en bloques completos al azar con cinco repeticiones para evaluar los 15 genotipos de cada dialelo (no se incluyeron los recíprocos). Los efectos directos resultaron mayores que los efectos indirectos en el dialélico de líneas (300% a 500%) que en el de variedades (17% a 68%). Los efectos (directos e indirectos) fueron superiores cuando se usaron las correlaciones genéticas (17% a 500%), que cuando se usaron las correlaciones fenotípicas (42% a 440%) en el análisis de sendero. El peso por fruta mostró mayor efecto directo sobre la producción que el número de frutas, con valores entre 3.2 a 4.9 veces el efecto residual. Tanto

el peso por fruto como el número de frutas pueden usarse como criterios de selección para incrementar la producción en *C. moschata*.

Palabras clave: *Cucurbita moschata*, **cucúrbitas, componentes de rendimiento**

INTRODUCTION

Worldwide, the area planted to tropical pumpkin (*Cucurbita moschata* Duchesne) grew 94% during the period from 1994 to 2001 (Le Buanec, 2002). Tropical pumpkin is an important vegetable crop in countries of the Caribbean Basin. Our study was conducted in Colombia, where tropical pumpkin is planted primarily by small farmers (Toro, 2001) and is recognized as important in terms of its export potential and area planted (FAO, 2002; CORPOICA, 2001; Hernández, 2001), as well as for its nutritional, decorative and artisan value (Vallejo and Mosquera, 1998). Little research has been conducted in Colombia on the crop with the exception of work carried out within the "Program for Genetic Improvement, Agronomy and Seed Production of Vegetables" at the National University of Colombia, Palmira Campus (CEUNP, for its initials in Spanish) over the past 17 years (Vallejo and Mosquera, 1998).

A principal focus of the CEUNP program is to develop higher yielding tropical pumpkin cultivars. One way to achieve this result for a quantitative low-heritability trait, such as yield, is via indirect selection for one or more traits correlated with yield. It is also important to consider that increasing yield per se may not be the only goal in a vegetable crop breeding program. Yield components, such as number of fruit and fruit size (including traits such as weight, length and width), can be important characteristics, particularly when marketing vegetable crops. Thus, the relationships between yield and its components, as well as the relationships among yield components, are important considerations in a breeding program.

Phenotypic correlations between traits are calculated by using directly observed measurements. There are two causes of phenotypic correlation between traits: genetic and environmental. Statistical techniques are used to estimate these correlations from directly observed measurements (Cruz and Regazzi, 1997; Falconer and Mackay, 1996; Hallauer and Miranda, 1981). Genetic correlations are particularly important because they provide the plant breeder with information about heritable associations between traits. This information is useful in the development of appropriate strategies for breeding for multiple traits in a plant improvement program or for employing indirect selection for a particular trait such as yield. Selection for one trait will result in a response in genetically correlated traits.

Path coefficient analysis is a statistical tool that helps the plant breeder to better understand what types of relationships exist between

two variables. The relationship between two variables can be such that one variable directly causes an effect on the other variable. Alternatively, a variable can cause an indirect effect by way of (via the "path" of) a third, or more variables. In general, the path coefficient analysis consists of partitioning the correlation coefficient into the direct and indirect effects that trait X_i has on trait Y (X is assumed to be the independent, and Y the dependent, variable). A correlation coefficient measures the total linear association between two traits, say Y and X_1, and includes both the direct effect of X_1 on Y plus the indirect effects of other traits ($X_2, \ldots, X_n$) related to X_1. For example, the correlation between Y and X_1 can be partitioned into the direct effect of trait X_1 (quantified by the path coefficient, P_1, which measures the direct influence independent from other variables) and the indirect effects of X_1 via its correlation with other traits, X_2 to X_n. Path analysis can use phenotypic, genetic or environmental correlation coefficients (Cruz and Regazzi, 1997; Vencovsky and Barriga, 1992).

Path analysis has been used in a number of studies in *Cucurbita,* usually to study the components of yield (Mohanty, 2001; Gwanama et al., 1998; Kumaran et al., 1998; Vinasco et al., 1998; Singh et al., 1992; Pandita et al., 1989; Doijode and Sulladmath, 1986; Rana et al., 1985). Several general conclusions are found in these studies: (1) Correlations (both phenotypic and genetic) between yield and its principal components, number of fruit and weight per fruit, are generally significant and positive. (2) Weight per fruit normally presents the greatest direct effect on yield, followed by number of fruit. (3) Fruit length and width (size index) are the variables that generally have the greatest direct effects in path analysis of weight per fruit, followed by (not always in this order): seed weight per fruit, number of seeds per fruit, weight of 100 seeds, size of seed cavity and mesocarp thickness. Most authors of these studies suggest that breeding programs for *Cucurbita* species select for weight per fruit or number of fruit to increase yield. When the goal is to obtain heavy fruit, selection of plants with fruits with a high size index, large seed weight and thick mesocarp is recommended.

The objective of this study was to carry out a path analysis for yield per plant as a function of number of fruits per plant and weight per fruit in two diallels of *C. moschata*: one using five varieties and another using five S1 lines derived from the varieties.

MATERIALS AND METHODS

Field plantings were done at the Experimental Center of the CE-UNP located 3°24' latitude north and 76°26' longitude west, at 980 m above sea level. The location has an average annual temperature of 24.6° C, precipitation of 1,100 mm, and relative humidity of 69%.

Two diallels were used for this study. The variety diallel consisted of five open-pollinated varieties [Bolo Verde (BV)], Soler (SO), accession 88 (A88), accession 34 (A34) and accession 126 (A126)]. Bolo Verde was released from the CEUNP program; Soler was released from the Agricultural Experiment Station of the University of Puerto Rico, Mayagüez. The other three accessions are from the *C. moschata* collection at CEUNP. The S1 lines (lines produced from one generation of self pollination) diallel was made up of five S1 lines, one from each variety (LBV-S1, LSO-S1, LA88-S1, LA34-S1 and LA126-S1). Fifteen genotypes (five parents and 10 crosses) were evaluated from each diallel. No reciprocals were included. For each diallel the 15 genotypes were arranged in a randomized complete block design with five replications. Both within- and between-row spacing was 3 m. Plots consisted of a single row of eight plants. The sampling unit consisted of the middle six plants (an area of 54 m²). Fruit was harvested separately from each plant in a plot. Each fruit was weighed separately and the total number of fruit counted. Additional agronomic variables were measured but found not to be useful in the path analysis since they were not strongly correlated with yield.

Data were analyzed by following Gardner and Eberhart's analysis II model (1966) where genotypes are considered fixed. Analyses were conducted by using plot means of yield per plant, number of fruit per plant and weight per fruit. Phenotypic and genetic correlations, path coefficients (direct effects) and indirect effects were estimated with the program GENES (Windows version 2.1, 2004), developed by Cruz (2001) and available at www.ufv.br/dbg/genes/genes.htm (verified 8 April 2005). We carried out a total of four path analyses of the effect on yield of number of fruit and fruit weight, including analyses for phenotypic and genetic correlations in each of the two diallels. The GENES program automatically generates each correlation matrix (Table 1) from the analysis of variance. To estimate the direct effects in each of the analyses, this program partitions the correlation matrix (either phenotypic or genetic) into the matrices $\mathbf{P}, \mathbf{A^{-1}}$, and $\mathbf{R}$ where $\mathbf{P} = \mathbf{A^{-1}} \cdot \mathbf{R}$.

$\mathbf{A^{-1}}$ is the inverse of the matrix of correlations between causal variables (number of fruit and fruit weight in this study), $\mathbf{R}$ is the vector of the correlation coefficients between the causal variables and the effect variable (yield in this study); and $\mathbf{P}$ is the vector of path coefficients ($\mathbf{P_{NUM}}$ = the direct effect of number of fruit on yield; $\mathbf{P_{WT}}$ = the direct effect of fruit weight on yield). The correlation coefficients with yield are then partitioned as follows:

$$r_{NUM,YLD} = P_{NUM} + P_{WT}r_{NUM,WT}$$

$$r_{WT,YLD} = P_{WT} + P_{NUM}r_{NUM,WT}$$

TABLE 1. *Phenotypic (r_P), genetic (r_G) and environmental (r_E) correlations among the traits number of fruit per plant, weight per fruit and yield per plant in progenies from a diallel of five varieties (above diagonal) and a diallel of five S1 lines derived from the varieties (below diagonal) of tropical pumpkin* (Cucurbita moschata).

Trait	Type of correlation	Number of fruit per plant	Weight per fruit	Yield per plant
Number of	r_P		-0.65**	0.32
fruit per plant	r_G		-0.74**	0.18
	r_E		-0.26	0.64*
Weight per fruit	r_P	-0.22		0.47
	r_G	-0.20		0.49
	r_E	-0.31		0.50
Yield per plant	r_P	0.53*	0.67**	
	r_G	0.52*	0.70**	
	r_E	0.62*	0.43	

*, **Significant at $P \leq 0.05$ and 0.01, respectively.

As an example, in the first equation the correlation between number of fruit and yield ($\mathbf{r_{NUM,YLD}}$) is a function of the direct effect of number of fruit on yield (quantified by the path coefficient $\mathbf{P_{NUM}}$) plus the indirect effect of number of fruit on yield via fruit weight (quantified by $\mathbf{P_{WT}r_{NUM,WT}}$). The path coefficient due to residual effects or due to other variables not considered in this study is defined as $\mathbf{h} = [1 - (\mathbf{P_{NUM}r_{NUM,YLD}}) - \mathbf{P_{WT}r_{WT,YLD}})]^{\frac{1}{2}}$. There is no statistical test available to test the significance of path coefficients, but we considered values greater than 0.7 to indicate a strong effect of a variable.

RESULTS AND DISCUSSION

In addition to number of fruit and fruit weight, other agronomic variables (such as flowering) were considered in the model for path analysis of yield. The coefficients of determination (R^2) and residual effects (h) (data not shown) indicated that these variables did not improve the model. Thus, only number of fruit and fruit weight were used in the analyses presented.

Within both the variety and the S1 diallels, phenotypic and genetic correlations were of the same sign and similar magnitude (Table 1). Between the two diallels, correlations were also of the same sign, although their magnitude varied. The phenotypic and genetic correlations between number of fruit and fruit weight were negative and significant for the variety diallel, whereas those same correlations were also negative, but not significant, in the S1 diallel. Both number of fruit

and fruit weight were positively and significantly correlated with yield in the S1 diallel, but not correlated with yield in the variety diallel.

The R^2 values in all four analyses were over 89% (Figure 1), thus lending support to the hypothesis that most variation in yield can be explained by a simple model that includes number of fruit and fruit weight. Models based on the S1 diallel showed a slightly greater power to predict yield ($R^2 = 0.934$ to 0.951) than did models based on the variety diallel. This finding is possibly due to fewer non-additive genetic effects in the S1 diallel than in the variety diallel. Environmental cor-
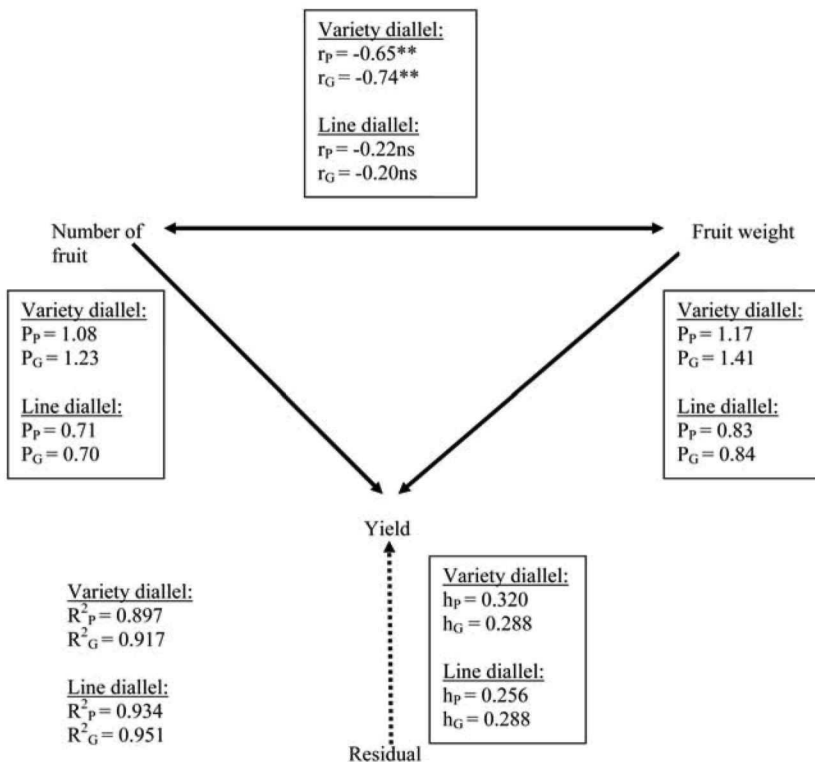

FIGURE 1. Path analysis of the effect of number of fruit and fruit weight on yield in *Cucurbita moschata* populations derived from diallels of five varieties (variety diallel) or five S1 lines (line diallel). Subscripts P and G refer to estimates made with phenotypic and genotypic correlations, respectively. The double-headed arrow represents the correlation (r) between number of fruit and fruit weight. Single-headed arrows represent direct effects of either number of fruit or fruit weight on yield. Direct effects are quantified by path coefficients (P). The dashed line represents the unexplained residual effects (h) in the model. The R^2 values indicate the fit of the model.

relations include both the correlations of environmental deviations and non-additive genetic deviations from breeding values (Falconer and Mackay, 1996). Thus, environmental correlations under- or over-estimate the real degree of association between the variables under study and make estimation of cause and effect relationships more difficult.

Similar results occur when comparing the path analysis model using genetic correlations versus phenotypic correlations. Coefficients of determination were $R^2 = 0.917$ and 0.951 for the former and $R^2 = 0.897$ and 0.934 for the latter (Figure 1). It appears that genetic correlations maximize the true relationship between the casual variables (number of fruit and fruit weight) and the effect variable (yield) in the model since they do not include environmental and non-additive genetic effects (which together make up the environmental correlation).

The results obtained for both diallels and both types of correlations were very similar in terms of the relative importance of direct effects over that of indirect effects (Table 2). Direct effects of number of fruit and fruit weight on yield were strong and positive, whereas the indirect effects were much smaller in magnitude and negative in sign. The negative indirect effect results from the negative correlation between number of fruit and fruit weight (Table 1). The absolute values (without taking sign into consideration) of direct effects were between 17% and 500% greater than the absolute values of indirect effects (Table 2). Both direct and indirect effects were smaller in the S1 diallel than in the variety diallel. However, in the S1 diallel, direct effects were greater than indirect effects by 300% to 500% compared to 17% to 68% in the variety diallel. This difference can be explained by the larger phenotypic and genetic correlations of yield with number of fruit and yield and fruit weight in the S1 diallel. It also suggests that it is easier to separate direct and indirect effects in more homogenous populations. Our results agree with a study of *C. maxima* by Vinasco et al. (1998). They detected large and significant correlations between yield and its components among F1 progenies but not in segregating generations.

In summary, in each of the four path analyses, direct effects explained much of the observed phenotypic and genetic correlation of yield with number of fruit and fruit weight. The direct effects were positive and larger than the indirect effects. The indirect effects were negative. Direct effects estimated by using the S1 diallel were greater than those estimated from the variety diallel. In the variety diallel, direct effects were greater when genetic correlations were used than when phenotypic correlations were used. In the S1 diallel there was very little difference between direct effects when using either genetic or phenotypic correlations. In all four path analyses, fruit weight had a somewhat greater direct effect on yield than did number of fruit. The

TABLE 2. *Summary of effects of number of fruit per plant and fruit weight on yield per plant as determined by path analysis in a tropical pumpkin (*Cucurbita moschata*) diallel of five varieties and a diallel of five S1 lines derived from those varieties, using phenotypic and genetic correlations.*

Diallel type	Correlation type	Correlation with yield	Effect on yield	
			Direct	Indirect
Number of fruit				
Variety	Phenotypic	0.32	1.08	-0.76
Variety	Genetic	0.18	1.23	-1.05
S1 line	Phenotypic	0.53*	0.71	-0.19
S1 line	Genetic	0.52*	0.70	-0.17
Fruit weight				
Variety	Phenotypic	0.47	1.17	-0.70
Variety	Genetic	0.49	1.41	-0.91
S1 line	Phenotypic	0.67**	0.83	-0.15
S1 line	Genetic	0.70**	0.84	-0.14

*,**Significant at $P \leq 0.05$ and 0.01, respectively; There are no statistical tests for direct and indirect effects, but effects greater than $|0.70|$ are considered to be large.

direct effects of fruit weight averaged about four times the size of the residual effects. This difference was more important in the S1 diallel than in the variety diallel, and more important when using genetic correlations than when using phenotypic correlations. The S1 diallel was better at producing a model that explained more of the variation (higher R^2 and fewer residual effects) and had greater sensitivity for detecting the importance of direct and indirect effects. Path analyses indicated that both fruit weight and number of fruit are strong candidates as indirect selection criteria for yield. Smaller fruits (<5 kg) are more attractive to smaller families. Supermarkets (perhaps in contrast to the *plaza de mercado*) also prefer smaller, easier to handle fruits. Thus, in order to obtain higher yields in *C. moschata*, emphasis should be on selection of plants with small heavy fruits in order to take advantage of the strong inverse correlation between number of fruit and fruit weight.

LITERATURE CITED

Corporación Colombiana de Investigación Agropecuaria (CORPOICA), 2001. Plan de modernización de la horticultura Colombiana. URL:http://www.corpoica.org.co/html/planes/hortiuclutra/texto/horticultura.html

Cruz, C. D. and A.J. Regazzi, 1997. Modelos Biométricos Aplicados aõ Melhoramento Genético. Segunda edición. Editora UFV. Brazil. 390 p.

Cruz, C. D., 2001. Programa Genes. Versaõ Windows. Aplicativo Computacional em Genética e Estatística. Editora UFV. Universidade Federal de Vicosa. 648 pp.

Doijode, S. D. and U. V. Sulladmath, 1986. Genetic variability and correlation studies in pumpkin (*C. moschata* Poir). *Mysore J. Agric. Sci.* 20:1:59-61.

Falconer, D. S. and T. Mackay, 1996. Introduction to Quantitative Genetics. Fourth edition. Prentice Hall. 464 pp.

FAO, 2002. Estadísticas Agrícolas Mundiales. URL: http: //www.fao.org

Gardner, C. O. and S. A. Eberhart, 1966. Analysis and interpretation of the variety cross diallel and related populations. *Biometrics* 22:439-452.

Gwanama, C., M.S. Mwala and K. Nichterlein, 1998. Path analysis of fruit yield components of *Cucurbita moschata* Duch. *Tropical Agricultural Research and Extension* 1:19-22.

Hallauer, A. R. and J. B. Miranda, 1981. Quantitative Genetics in Maize Breeding. Iowa State University Press, Ames, Iowa. 468 pp.

Hernández, C. J. M., 2001. Programa nacional de recursos genéticos. URL:http://fwl.inifap.conacyt.mx/progs_a/agricola/prog_rec_fitoge.html

Kumaran, S. S., S. Natarajan and S. Thamburaj, 1998. Correlation and path analysis studies in pumpkin (*Cucurbita moschata* Poir). *South Indian Horticulture* 46:138-142.

Le Buanec, B., 2002. The world seed trade. *In:* 18 Seminario Panamericano de Semillas. Proceedings on CD. Santa Cruz-Bolivia. 1-3 of July 2002.

Mohanty, B. K., 2001. Studies on correlation and path analysis in pumpkin (*Cucurbita moschata*). *Haryana J. Hort. Sci.* 30(1-2):86-89.

Pandita, M. L., M. S. Dahiya and R. N. Vashistha, 1989. Studies on correlation and path analysis in summer squash (*Cucurbita pepo* L.): (a note). *Haryana J. Hort. Sci.* 18(3-4):295-298.

Rana, T.K., R.N. Vashistha and M.L. Pandita, 1985. Correlations and path coefficient studies in pumpkin (*Cucurbita moschata* Poir). *Haryana J. Hort. Sci.* 14(1-2):108-113.

Singh, J., J. C. Kumar and J. R. Sharma, 1992. Correlation and path coefficient analysis in pumpkin. *J. Res. Punjab Agric. Univ.* 29(2):207-212.

Toro, S. S., 2001. Heterosis y habilidad combinatoria entre poblaciones seleccionadas de *Cucurbita moschata* Poir. M.S. Thesis in Plant Breeding. Universidad Nacional de Colombia, Sede Palmira, Palmira, Colombia 39 pp.

Vallejo, F. A. and S. E. Mosquera, 1998. Transferencia del gen *Bu* a poblaciones de zapallo, *Cucurbita* sp. con crecimiento postrado. *Acta Agronómica* 48:7-18.

Vencovsky, R. Y. and P. Barriga, 1992. Genética Biométrica no Fitomelhoramento. Sociedad Brasileira de Genética. 496 pp.

Vinasco, L. E., G. D. Baena and M. García, 1998. Análisis genético de caracteres que afectan el hábito de crecimiento de zapallo *Cucurbita maxima* (Duch. Ex Lam). *Acta Agronómica* 48(3-4):12-18.