CORRELATIONS AND PATH ANALYSIS IN CHERRY TOMATO GENOTYPES

Felipe Barrera Sánchez¹, Larissa Pereira Ribeiro², Erina Vitório Rodrigues³, Leonardo Lopes Bhering², Paulo Eduardo Teodoro⁴

¹Departamento de Ciências Agrárias, Universidad Nacional de Colombia, Medellín, Colombia.
²Laboratório de Biometria, Universidade Federal de Viçosa (UFV), Viçosa, MG, 36570-900.
³Departamento de Ciências da Vida e da Terra, Universidade de Brasília (UnB), Planaltina, DF, 73345-010.
⁴Departamento de Ciências Agrárias, Universidade Federal de Mato Grosso do Sul (UFMS), Chapadão do Sul, MS, 79560-000.

Corresponding author: Leonardo Lopes Bhering (leonardo.bhering@ufv.br).

Abstract: Tomato is one of the most consumed vegetables in the world and in Brazil due to its nutritional and organoleptic quality and excellent flavor. This study aimed to evaluate the correlation among primary and secondary production components and their unfolding into direct and indirect effects on the total soluble solids content of cherry tomato genotypes. Five cherry tomato lines were crossed in complete diallel scheme of Griffing, Method 1, and parents and the 20 F₁ hybrids were evaluated for the traits: number of bunches per plant, number of flowers per bunch, number of fruits per bunch, number of fruits per plant, fruit weight, fruit yield and total soluble solids. Analysis of variance was performed considering information between and within plot. Subsequently, phenotypic and genotypic correlations were estimated, followed by path analysis. Genotypic correlations among the traits showed similar magnitude and same direction of their respective phenotypic correlations. Path analysis explained 92.2% of the variation on the total soluble solids among the evaluated genotypes. Based on the results obtained, it is possible to infer that the selection of yielding genotypes is the main strategy to increase the content of soluble solids in cherry tomatoes.

Keywords: Lycopersicon esculentum, yield, soluble solids, tomato hybrids

Introduction

Tomato (Lycopersicon esculentum Mill.) is one of the most consumed vegetables in the world, and in Brazil its production has expanded markedly in recent years. One of the varieties of great importance is the cherry tomato (Zhao et al., 2010), whose popularity has grown due to increased consumption of minimally processed foods. In addition, it presents nutritional quality higher than traditional tomatoes, as higher concentration of total soluble solids (Srivalli et al., 2016), besides high levels of vitamins A and C (Oliveira et al., 2013) and antioxidant properties (Naveen et al., 2017). Allied to the a fore mentioned characteristics, cherry tomato aggregates high commercial value.

The major importance of cherry tomatoes is mainly related to its inherent characteristics of nutritional and organoleptic quality and excellent taste. However, for selecting superior genotypes, it is interesting that these characteristics are associated with production component traits. Thus, one of the objectives of a breeding program is to understand the association between traits related to nutritional quality and production components. Correlation analysis makes it possible to evaluate the magnitude and direction of the relationship between traits, so if the traits have favorable correlation, it is possible to obtain gains for one of them through indirect selection in a faster way than the direct selection on the desired trait (Cruz et al., 2012).

Despite the importance of the correlation analysis in the selection of genotypes, it does not identify the direct and indirect influences, i.e., they do not provide information on cause and effect. Therefore, to better understanding the associations between traits, it is needed to perform path analysis, which corresponds to the unfolding of the correlation proposed by Wright (1923) and detailed by Li (1975). In this analysis, it is studied the direct and indirect effects of traits on a main variable, whose estimates are obtained through regression equations in which the variables are previously standardized (Cruz et al., 2012).

Many studies have investigated physical-chemical (Akbudak et al., 2009) and yield traits of tomato (Graça et al., 2015; Zhang et al., 2015; Zeist et al., 2017), as well as the correlation and the correlation unfolding between them (Meena and Bahadur, 2015; Ali et al., 2017). However, there are few papers addressing...
the correlations between traits and path analysis on cherry tomatoes.

Therefore, the objective of this study was to evaluate the correlation between primary and secondary production components and their unfolding into direct and indirect effects on the total soluble solids content of cherry tomato genotypes.

**Material and methods**

**Conduction of the experiment**

Five cherry tomato (*S. lycopersicum* var. cerasiforme) lines were crossed in complete diallel scheme of Griffing (1956), Method 1 (parents, F₁ and reciprocal). The parents were chosen by their agronomic and commercial interest characteristics, as shown in Table 1. The parents and the 20 F₁ hybrids obtained by diallel cross were evaluated in hydroponic cultivation at University of Caldas, located in the municipality of Manizales, Department of Caldas, Colombia. The average altitude of the site is 2,160 m, with average temperature of 18°C and relative humidity of 65%.

The experimental design was a randomized block design with three replicates. In each replication was evaluated four plants per treatment. Sowing was carried out in nurseries containing 128 locules, using as substrate peat and rice husk in a ratio 4:1. At 30 days after sowing, the seedlings were transplanted to gutters, planting 10 other plants for each cross in case of establishment loss. The irrigation was performed every 1.5 hours. Fertilization and irrigation were carried out with standard handling and supported by the tomato management manual under protected conditions (Jaramillo et al., 2007). The tutoring was performed at 26 days after transplantation to the final site using fine agricultural fiber. Harvesting began at 90 days after sowing, the seedlings were transplanted to gutters, planting 10 other plants for each cross in case of establishment loss. The irrigation was performed every 1.5 hours. Fertilization and irrigation were carried out with standard handling and supported by the tomato management manual under protected conditions (Jaramillo et al., 2007). The tutoring was performed at 26 days after transplantation to the final site using fine agricultural fiber. Harvesting began at 90 days after transplantation, followed by sanitary pruning of the branches responsible for filling the already collected bunches that were diseased and old.

At maturation stage, the following traits were evaluated: number of bunches per plant (BP), number of flowers per bunch (FLB), number of fruits per bunch (FRB), and number of fruits per plant (FRP). At harvesting, the traits evaluated were: fruit weight (FW, g), evaluated by weighing each of the fruits harvested, obtaining an average weight; fruit yield per plant (YIE, g), obtained by the sum of fruit weights per plant; and total soluble solids content (TSS, °Brix), measured by using a refractometer in samples of fully mature fruits.

**Statistical analyses**

**Analysis of variance**

Data were submitted to analysis of variance (ANOVA) considering information between and within plot, according to the model described in Equation 1:

\[ Y_{ijk} = \mu + G_i + B_j + e_{ij} + \delta_{ijk} \]  

Wherein: \( Y_{ijk} \): value observed in the k-th plant, in the i-th genotype, of the j-th block; \( \mu \): overall mean observed; \( G_i \): effect of the i-th genotype (random effect); \( B_j \): effect of the j-th block (fixed effect); \( e_{ij} \): error of the i-j plot; \( \delta_{ijk} \): error of the k plant, of the i-th genotype, in the j-th block.

The experimental precision of each trait was evaluated by the estimation of selective accuracy (SA) as recommended by Resende and Duarte (2007) and described in Equation 2:

\[ A = \sqrt{1 - \frac{1}{F_c}} \]  

Wherein: \( F_c \) is the calculated F value obtained for the genotypes effect by ANOVA.

**Correlations**

The phenotypic correlations (\( r_F \)) between the pairs of traits were estimated according to Equation 3:

\[ r_F = \frac{COV_{F(xy)}}{\sqrt{\delta_{Fx}^2} \times \delta_{Fy}^2} \]  

Where: \( COV_{F(xy)} \) is the phenotypic covariance between traits X and Y, \( \delta_{Fx}^2 \) is the phenotypic variance of trait X, and \( \delta_{Fy}^2 \) is the phenotypic variance of trait Y.

The genotypic correlations (\( r_G \)) between the pairs of traits were estimated according to Equation 4:

\[ r_G = \frac{COV_{G(xy)}}{\sqrt{\delta_{Gx}^2} \times \delta_{Gy}^2} \]  

where: \( COV_{G(xy)} \) is the genotypic covariance between traits X and Y, \( \delta_{Gx}^2 \) is the genotypic variance of trait X, and \( \delta_{Gy}^2 \) is the genotypic variance of trait Y.

Correlation network was used in order to graphically express the functional relationship between the estimates of phenotypic and genotypic correlations coefficients between the 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 applying a cut-off value of 0.50, which means that only \( |r_{xy}| \geq 0.50 \) have their edges highlighted. At last, positive correlations were highlighted in green, while negative correlations were represented in red.

Subsequently, a multicollinearity diagnosis of the phenotypic correlation matrix was performed, which revealed moderate to strong multicollinearity according to the classification of Montgomery and Peck (2001). Therefore, to perform the path analysis, a constant \( k = 0.05 \) was added to the diagonal of the \( X'X \) matrix, which provided weak multicollinearity (condition number < 100).
Path analysis

Path analysis, considering total soluble solids (TSS, °Brix) as principal dependent variable, was performed using the model described in Equation 5:

\[ TSS = \beta_1 BP + \beta_2 FLP + \cdots + \beta_6 YIE + p_e \]  

Eq. 05

In which: \( \beta_1, \beta_2, \ldots, \beta_6 \) are the direct effects estimators of the traits number of bunches per plant (BP), number of flowers per bunch (FLP), number of fruits per bunch (FRB), number of fruits per plant (FRP), fruit weight (FW) and fruit yield per plant (YIE) on the total soluble solids (TSS), \( p_e \) is the residual effect of the analysis. Thus, the normal equation system was used to estimate the direct and indirect effects of each explanatory variable on TSS, according to Equation 6:

\[
\begin{bmatrix}
1.05 & \vdots & \vdots & \vdots \\
\vdots & \ddots & \vdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \vdots & 1.05 \\
1.05 & \vdots & \vdots & \vdots \\
\vdots & \ddots & \vdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \vdots & 1.05 \\
\end{bmatrix}
\begin{bmatrix}
\hat{\beta}_1 \\
\hat{\beta}_2 \\
\vdots \\
\hat{\beta}_6 \\
\end{bmatrix}
= 
\begin{bmatrix}
\hat{TSS} \\
\hat{YIE} \\
\vdots \\
\hat{FRP} \\
\end{bmatrix}
\]  

Eq. 06

The coefficient of determination (R²) of the path analysis was obtained by the Equation 7:

\[ R^2 = \hat{\beta}_1 TSS_{BF} + \cdots + \hat{\beta}_6 TSS_{YIE} \]  

Eq. 07

The residual effect (\( \hat{p}_e \)) of the path analysis was obtained by Equation 8:

\[ \hat{p}_e = \sqrt{1 - R^2} \]  

Eq. 08

Softwares

The statistical analyses were performed with the Genes (Cruz, 2013) and Rbio (Bhering, 2017) softwares.

Results and discussion

Analysis of variance revealed highly significant mean squares (P≤0.01) for all traits (Table 2), which show the existence of genetic variability among the genotypes. Similar results were observed by Sobreira et al. (2009), Sobreira et al. (2010), Saleem et al. (2013) and Sari et al. (2017). The selective accuracy estimates obtained show high experimental precision, according to the criteria of Resende and Duarte (2007).

The high heritability estimates for BP, FLP, FRB, FRP and FW demonstrate that the direct selection based on these variables is efficient, since this parameter corresponds to the proportion of the total variability that is of genetic nature (Wright, 1976; Falconer and Mackay, 1996), indicating that these traits are little influenced by the environment. Moderate heritability for YIE and TSS indicated favorable influence of environment rather than genotypes, and hence selection of superior genotypes for these traits would not be rewarding in early generations.

Phenotypic and genotypic correlations were graphically expressed in Figure 1, a procedure that allows to easily seeing the interrelation between the evaluated traits. Knowing the correlations among traits for selection provides information to the breeder about the degree of association among the traits of economic importance, since the selection on one trait changes the behavior of the other (Cruz et al., 2012; Teodoro et al., 2014; Ribeiro et al., 2016). It is important to emphasize that the genotypic correlations presented similar magnitude and meaning to their respective phenotypic correlations. This indicates that the phenotypic expression of the evaluated genotypes was predominantly of genetic nature.

The BP showed to be strongly correlated with YIE and TSS, indicating that the yield and total soluble solids content may be higher the larger the number of bunches. This trait also presented high positive correlation with the FRP, an expected behavior because the larger the number of fruits, the larger the number of bunches produced by the plant, since the development of fruits occurs in bunches.

The highest positive correlation observed was YIE x TSS, suggesting that higher yield may result in higher TSS fruits. This result is of great interest to breeders, since when selecting genotypes with higher yield, genotypes with higher flavor quality would be simultaneously selected. However, this result disagrees with what has been reported in the literature. Rodrigues et al. (2010) and Singh et al. (2017) found no significant genotypic correlation for YIE x TSS, while Naveen et al. (2017) reported the existence of negative genotypic correlation among these traits.

There was a moderate negative correlation for FRB x FRP, suggesting that the larger the number of fruits per bunch, the lower the number of fruits per plant, which has no biological sense. Thus, the correlation estimates can produce misconceptions about the relation between two traits and are not a true cause-and-effect measure. In this sense, a biometric technique that stands out to direct the multiple trait selection is path analysis. Study of path coefficients enable the breeders to focus on the variables that show high direct effect on the principal variable (Cruz et al., 2012). Teodoro et al. (2014) mention that this analysis provides a detailed knowledge of the traits influence and justify the existence of positive and negative correlations of high and low magnitude among the studied traits. Therefore, path analysis was performed in order to determine the cause-and-effect relationship between the production components (explanatory variables) and the total soluble solids (dependent variable), since the study of this relationship has not yet been reported in the literature.
Condition number obtained through the multicollinearity diagnosis from the correlation matrix of the explanatory variables indicates moderate multicollinearity according to Montgomery and Peck (2001). In the presence of multicollinearity, the variances associated with the path coefficient estimators can reach values that are excessively high, becoming unreliable. In addition, the parameter estimates may assume absurd values or no coherence with the studied biological phenomenon (Carvalho et al., 1999; Cruz et al., 2012). Aiming at reducing the problems brought by multicollinearity, it was performed the ridge path analysis, whose estimates are shown in Table 3.

The use of path analysis in tomato is recent and is restricted to consider the fruit yield as the principal dependent variable (Rodrigues et al., 2010; Saleem et al., 2013; Gopinath and Vethamoni, 2017; Naveen et al., 2017; Singh et al., 2017). The coefficient of determination was high magnitude (R² = 0.92), which indicates that practically all variation in SST was explained by the auxiliary traits considered in the path analysis. The traits YIE and BP presented the greatest direct effects and in the same direction of their correlations on the TSS trait. The indirect selection via BP and YIE should also be considered, since these traits also influenced the correlations obtained by the others through the indirect effects. In plant breeding, it is important to identify the traits with the greatest direct effect in direction favorable to selection, so that the correlated response through indirect selection is efficient (Cruz et al., 2012).

The indirect effects via YIE indicate that selection based on this trait can indirectly influence the TSS trait through BP, FLB and FW. Therefore, the estimates obtained by path analysis reveal that the selection based on yield is an efficient strategy for these genotypes, since it simultaneously implies the selection of individuals with higher soluble solids content. This result is of great relevance from the point of view of the cherry tomato breeding, since the yield and fruit flavor quality are the main traits of interest of the breeding programs. Moreover, the YIE is easier to measure than the TSS, which contributes to the reduction of resources and time spent in the evaluation stages.

<table>
<thead>
<tr>
<th>Table 1. Description of cherry tomato genotypes used as parents.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction</strong></td>
</tr>
<tr>
<td>IAC426*</td>
</tr>
<tr>
<td>LA1480**</td>
</tr>
<tr>
<td>IAC1688</td>
</tr>
<tr>
<td>IAC420</td>
</tr>
<tr>
<td>LA3158</td>
</tr>
</tbody>
</table>

*IAC: introductions proceeding from the Agronomic Institute of Campinas, Brazil; **LA: introductions proceeding from the Tomato Genetics Resources Center (TGRC) of the University of California, USA.

<table>
<thead>
<tr>
<th>Table 2. Analysis of variance for the traits number of bunches per plant (BP), number of flowers per bunch (FLB), number of fruits per bunch (FRB), number of fruits per plant (FRP), fruit weight (FW), fruit yield per plant (YIE) and total soluble solids (TSS) evaluated in 25 cherry tomato genotypes.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SV</strong></td>
</tr>
<tr>
<td>Blocks</td>
</tr>
<tr>
<td>Genotypes</td>
</tr>
<tr>
<td>Error (between)</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SA</td>
</tr>
<tr>
<td>h²</td>
</tr>
</tbody>
</table>

*Significant at 0.01 probability level by F test. SV: sources of variation; DF: degrees of freedom; SA: selective accuracy; h²: heritability.
Table 3. Direct and indirect effects of the number of bunches per plant (BP), number of flowers per bunch (FLB), number of fruits per bunch (FRB), number of fruits per plant (FRP), fruit weight (FW) and fruit yield per plant (YIE) on the total soluble solids (TSS) evaluated in 25 cherry tomato genotypes.

<table>
<thead>
<tr>
<th>Effect</th>
<th>BP</th>
<th>FLB</th>
<th>FRB</th>
<th>FRP</th>
<th>FW</th>
<th>YIE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct on SST</td>
<td>0.330</td>
<td>0.001</td>
<td>0.113</td>
<td>0.072</td>
<td>0.153</td>
<td>0.463</td>
</tr>
<tr>
<td>Indirect via BP</td>
<td>---</td>
<td>0.202</td>
<td>0.008</td>
<td>0.237</td>
<td>0.176</td>
<td>0.301</td>
</tr>
<tr>
<td>Indirect via FLB</td>
<td>0.001</td>
<td>---</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Indirect via FRB</td>
<td>0.003</td>
<td>0.061</td>
<td>---</td>
<td>-0.055</td>
<td>0.022</td>
<td>0.031</td>
</tr>
<tr>
<td>Indirect via FRP</td>
<td>0.052</td>
<td>0.020</td>
<td>-0.035</td>
<td>---</td>
<td>0.021</td>
<td>0.037</td>
</tr>
<tr>
<td>Indirect via FW</td>
<td>0.082</td>
<td>0.096</td>
<td>0.030</td>
<td>0.044</td>
<td>---</td>
<td>0.108</td>
</tr>
<tr>
<td>Indirect via YIE</td>
<td>0.423</td>
<td>0.375</td>
<td>0.128</td>
<td>0.240</td>
<td>0.327</td>
<td>---</td>
</tr>
<tr>
<td>Total (genetic correlation)</td>
<td>0.906</td>
<td>0.755</td>
<td>0.251</td>
<td>0.542</td>
<td>0.707</td>
<td>0.965</td>
</tr>
</tbody>
</table>

R² 0.922
Residual effect 0.280

Figure 1. Phenotypic (A) and genotypic (B) correlations network among the traits number of bunches per plant (BP), number of flowers per bunch (FLB), number of fruits per bunch (FRB), number of fruits per plant (FRP), fruit weight (FW), fruit yield per plant (YIE) and total soluble solids (TSS) evaluated in 25 cherry tomato genotypes.

Conclusion

The selection of yielding genotypes is the main strategy to increase the soluble solids content in cherry tomatoes.
References


