SEQUENTIAL PATH ANALYSIS AND RELATIONSHIPS BETWEEN FRUIT YIELD IN WATERMELON

Maryam Abdoli Nasab¹, Mehdi Rahimi¹*, Arzu Karatas² and Sezai Ercisli³

¹Department of Biotechnology, Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, Iran; ²Recep Tayyip Erdogan University, Agricultural and Natural Science Faculty, Department of Horticulture Pazar, Rize-Turkey; ³Ataturk University, Agricultural Faculty, Department of Horticulture, Erzurum, Turkey

*Corresponding author’s e-mail: mehdi83ra@gmail.com

The knowledge of relationship between fruit yield and its components certainly helps to improve the efficiency of a breeding program with appropriate selection criteria. In current study, interrelationships between fruit yield and its components in watermelon were investigated using sequential path coefficient analysis. The experiment containing 38 diverse genotypes of watermelon were evaluated in a randomized complete block design with 3 replications during the year 2016 and 2017. The correlation between flesh weight and fruit yield (0.854 and 0.751) was the highest, followed by fruit length-fruit yield correlation (0.463 and 0.459) in both years. Sequential stepwise multiple regression analysis was performed to organize the predictor variables into first, second and third-order paths based on their respective contributions to the total variation of yield and minimum collinearity. Based on the variance inflation factor and magnitude of direct effects, flesh weight was considered as first-order variables in both years and accounted for 72 and 49 percent of the total variation of fruit yield, respectively. The results of both year 2016 and 2017 also indicated that the plant traits i.e. number of nodes per stem, and fruit length could be considered as selection criterion to increase fruit yield in watermelon.

Keywords: Bootstrap analysis; correlation; regression; sequential path analysis; watermelon.

INTRODUCTION

Watermelon (Citrullus lanatus) is a diploid plant (n = 2x = 22) of the Cucurbitaceae family (Bisognin, 2002). It originates in Africa but has been dispersed in Asia and throughout the world with a long history of cultivation and plantation (Bisognin, 2002; National Research Council, 2008; Grumet et al., 2017). The global production of watermelons in 2018 was about 104 million tons and more than 80% of which was produced in Asia. Iran had the second rank with an annual production of 4.1million tons after China (FAO, 2018). Watermelon is a sweet and popular fruit of summer and people usually consume as fresh or in the form of juice (Ashlesha et al., 2015). In order to achieve the desired goals in plant breeding, recognition of the genetic features of important traits, their specific relationships, and mutual effects is one of the foundations for deciding on the design and implementation of different breeding methods. By identifying these features, the best methods can be selected and somewhat predicted breeding results (Jaskani et al., 2005; Breseghello and Coelho, 2013). Consideration and understanding the relationships between traits for indirect selection in plant breeding programs is very important for the traits that are not easily measured, or traits that have a small/limited inheritance (Çalışkan, 2012).

Although increase in fruit yield is one of the main goals of watermelon breeding, but selection of cultivars based on direct measurement of yield is of little benefit because of the complex genetic control and the effects of environment (Gusmini and Wehner, 2005). Therefore, fruit yield can be indirectly increased through traits that are correlated with it and can be through the simultaneous or non-simultaneous selection of those traits (Heslot et al., 2015). However, there is a positive relationship between fruit yield and a number of its components, but the existence of negative relationships between some of the components of fruit yields cannot be used as an effective factor in increasing fruit yield (Gusmini and Wehner, 2005). Multivariate analyzes are used to describe and evaluate genetic material for optimal use as well as studying the internal relations between traits (Johnson, 1998). These statistical methods are used to determine the effects of independent traits on dependent traits, to determine the contribution of each trait to the total variation, to distinguish and classify traits and genotypes, to reduce the volume of data and the number of major variables in the form of new components and to define the selection indices (Johnson, 1998; Sharma, 2006; Ferreira et al., 2015).

In most studies of path coefficient analysis, researchers had considered the predictor traits as first-order traits and their effect had been calculated with dependent traits such as yield. This method may lead to the collinearity of traits, and thus the
interpretation of the contribution of each trait faced with a problem. Using stepwise regression analysis which is excluded non-significant traits can reduce the collinearity amount between the remaining traits and this procedure can lead to a loss of information. Therefore, a better strategy can be to use a sequential stepwise regression instead of a one-step stepwise regression. In this method, the traits that are eliminated from the first-order path analysis are re-evaluated for the next rank and this is the basis of the sequential path analysis. Also, first-order traits are determined based on having the highest direct effect and minimum collinearity on the dependent trait. Then, the effect of other variables on the first-order traits is studied and the variables affecting these traits are determined (Mohammadi et al., 2003).

Few reports on the correlation analysis and path coefficient analysis existed in watermelon (Maggs-Kölling and Christiansen, 2003; Zheng et al., 2009; Sundaram et al., 2011; Choudhary et al., 2012; Nisha et al., 2018). The study of Bhagyalekshmi et al. (2020) in watermelon revealed that number of fruits per vine, average fruit weight and pulp to seed ratio had high direct positive effect on fruit yield per vine indicating their true positive and significant association with yield. Therefore, the purpose of this study was to analyze the correlation between fruit yield and its components, determine the direct and indirect effects of traits on fruit yield and identify the useful and effective traits in breeding high yielding watermelon genotypes.

MATERIALS AND METHODS

To estimate the correlation between agronomic traits and relationships between yield and other traits, 38 diverse varieties of watermelon (Table 1) were cultivated in a randomized complete block design with three replications in 2016 and 2017 at the research field of Graduate University of Advanced Technology, Kerman, Iran. The row-to-row and plant-to-plant distance were kept 2 m and 0.5 m, respectively.

During the growing season data for various morphological and agronomic traits i.e. fruit length (cm), fruit width (cm), fruit yield (kg), flesh weight (kg), skin thickness (mm), 100-seed weight (gr), seed length (mm), seed width (mm), leaf length (cm), pH, number of leaf, number of male flower and number of nodes per stem were recorded on 5 plants and thus was used in statistical analyzes. Weeds were by hand instead of chemicals.

Normality of the data was investigated through Kolmogorov–Smirnov test by using the SPSS software 24 (IBM-Corp., 2016) and also used to calculate the correlation coefficients of traits and stepwise regression analysis. The phenotypic and genotypic variance of trait estimated based on expected value of combined analysis design through the following formulas.

\[
\sigma^2_{ge}(X) = \frac{M^2_{ge} - M^2_{e}}{r^2}
\]

(1)

\[
\sigma^2_g(X) = \frac{M^2_g - M^2_{ge}}{M^2_{ge}}
\]

(2)

\[
\sigma^2_p(X) = \sigma^2_g(X) + \frac{\sigma^2_{ge}(X)}{M^2_{ge}}
\]

(3)

Where \( \sigma^2_p \) is the phenotypic variance (Vp), \( \sigma^2_g \), genotypic variance (Vg), \( \sigma^2_{ge} \), genotype × environment interaction variance (Vge) and \( \sigma^2_e \), environmental variance (Ve).

Then, the broad-sense heritability \( h^2_b \) was calculated for the traits through the following equation.

\[
h^2_b = \frac{\sigma^2_g}{\sigma^2_p} \times 100
\]

(4)

Initially, the usual path analysis was used, and all traits were considered as the first predictor traits for the fruit yield. Then, the sequential path analysis method was used by stepwise regression analysis to put predictive traits in first, second and third paths. The collinearity amount between traits in each section of the path analysis was measured using the variance inflation factor and tolerance value (Hair et al., 1995). Accordingly, the values of the variance inflation factor greater than 10 and the tolerance value of less than 1 were considered as high collinearity. Amos 24.0 software (Arbuckle, 2016) was used to find out the cause and effect of yield and its components through path analysis. Also, Bootstrap analysis (Efron and Tibshirani, 1993) was used to calculate the standard error of path coefficients.

RESULTS AND DISCUSSION

Phenotypic correlations: The correlation coefficients between different traits are presented in Table 2. The results showed that during the year 2016 the traits i.e. number of leaves, number of nodes per stem, fruit length, fruit width and flesh weight; and 2017, the traits e.g. and number of male flowers, number of leaf, fruit length and flesh weight showed positive and significant correlation with fruit yield. Flesh weight (0.854 and 0.751) and fruit length (0.463 and 0.459) both showed the highest positive correlation with fruit yield in the year 2016 and 2017, respectively. The positive and significant correlation coefficient of fruit yield with the
Path analysis in watermelon

Table 2. Correlation coefficients between measured watermelon traits in 2016 (upper diagonal) and 2017 (lower diagonal).

<table>
<thead>
<tr>
<th>Traits</th>
<th>SL</th>
<th>SW</th>
<th>LL</th>
<th>NMF</th>
<th>NL</th>
<th>NNS</th>
<th>pH</th>
<th>HSW</th>
<th>FL</th>
<th>FW</th>
<th>FLW</th>
<th>ST</th>
<th>FY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed length (SL)</td>
<td>1.00</td>
<td>0.847**</td>
<td>0.215</td>
<td>0.434**</td>
<td>0.376*</td>
<td>0.401*</td>
<td>-0.14</td>
<td>0.763**</td>
<td>0.187</td>
<td>0.421**</td>
<td>0.243</td>
<td>0.258</td>
<td>0.174</td>
</tr>
<tr>
<td>Seed width (SW)</td>
<td>0.795**</td>
<td>1.00</td>
<td>0.227</td>
<td>0.204</td>
<td>0.267</td>
<td>0.209</td>
<td>-0.245</td>
<td>0.691**</td>
<td>0.019</td>
<td>0.424**</td>
<td>0.202</td>
<td>0.239</td>
<td>0.165</td>
</tr>
<tr>
<td>Leaf length (LL)</td>
<td>0.162</td>
<td>0.170</td>
<td>1.000</td>
<td>0.338*</td>
<td>0.355*</td>
<td>0.443**</td>
<td>-0.05</td>
<td>0.147</td>
<td>0.122</td>
<td>0.299</td>
<td>0.239</td>
<td>0.017</td>
<td>0.220</td>
</tr>
<tr>
<td>Number of male flower (NMF)</td>
<td>0.406*</td>
<td>0.288</td>
<td>0.338*</td>
<td>1.000</td>
<td>0.839**</td>
<td>0.814**</td>
<td>-0.132</td>
<td>0.384*</td>
<td>0.366*</td>
<td>0.122</td>
<td>0.359*</td>
<td>0.092</td>
<td>0.276</td>
</tr>
<tr>
<td>Number of leaves (NL)</td>
<td>0.389*</td>
<td>0.295</td>
<td>0.368*</td>
<td>0.863**</td>
<td>1.000</td>
<td>0.714**</td>
<td>-0.203</td>
<td>0.332*</td>
<td>0.416**</td>
<td>0.150</td>
<td>0.402*</td>
<td>0.055</td>
<td>0.387*</td>
</tr>
<tr>
<td>Number of nodes per stem (NNS)</td>
<td>0.795**</td>
<td>0.621**</td>
<td>0.969</td>
<td>0.453**</td>
<td>0.412*</td>
<td>0.456**</td>
<td>0.135</td>
<td>1.000</td>
<td>-0.044</td>
<td>0.28</td>
<td>0.277</td>
<td>0.131</td>
<td>0.233</td>
</tr>
<tr>
<td>pH</td>
<td>-0.036</td>
<td>-0.130</td>
<td>0.167</td>
<td>0.178</td>
<td>0.153</td>
<td>0.202</td>
<td>1.000</td>
<td>-0.116</td>
<td>0.047</td>
<td>-0.349*</td>
<td>0.063</td>
<td>-0.083</td>
<td>0.010</td>
</tr>
<tr>
<td>100-seed weight (HSW)</td>
<td>0.789**</td>
<td>0.621**</td>
<td>0.969</td>
<td>0.453**</td>
<td>0.412*</td>
<td>0.456**</td>
<td>0.135</td>
<td>1.000</td>
<td>-0.044</td>
<td>0.28</td>
<td>0.277</td>
<td>0.131</td>
<td>0.233</td>
</tr>
<tr>
<td>Fruit length (FL)</td>
<td>0.283</td>
<td>0.063</td>
<td>0.107</td>
<td>0.359*</td>
<td>0.507**</td>
<td>0.257</td>
<td>0.185</td>
<td>0.140</td>
<td>1.000</td>
<td>0.412*</td>
<td>0.388*</td>
<td>0.328*</td>
<td>0.463**</td>
</tr>
<tr>
<td>Fruit width (FW)</td>
<td>0.475**</td>
<td>0.413**</td>
<td>0.359*</td>
<td>0.243</td>
<td>0.380*</td>
<td>0.408*</td>
<td>-0.068</td>
<td>0.377*</td>
<td>0.422**</td>
<td>1.000</td>
<td>0.390*</td>
<td>0.397*</td>
<td>0.341*</td>
</tr>
<tr>
<td>Flesh weight (FLW)</td>
<td>0.240</td>
<td>0.221</td>
<td>0.189</td>
<td>0.455**</td>
<td>0.489**</td>
<td>0.489**</td>
<td>0.183</td>
<td>0.306</td>
<td>0.450**</td>
<td>0.391*</td>
<td>1.000</td>
<td>0.202</td>
<td>0.854**</td>
</tr>
<tr>
<td>Skin thickness (ST)</td>
<td>0.230</td>
<td>0.191</td>
<td>0.004</td>
<td>0.082</td>
<td>0.156</td>
<td>0.16</td>
<td>0.212</td>
<td>0.153</td>
<td>0.274</td>
<td>0.292</td>
<td>0.210</td>
<td>1.000</td>
<td>0.216</td>
</tr>
<tr>
<td>Fruit yield (FY)</td>
<td>0.100</td>
<td>0.060</td>
<td>0.220</td>
<td>0.329*</td>
<td>0.427**</td>
<td>0.238</td>
<td>0.14</td>
<td>0.194</td>
<td>0.459**</td>
<td>0.263</td>
<td>0.751**</td>
<td>0.202</td>
<td>1.000</td>
</tr>
</tbody>
</table>

* and ** Significant at the 0.05 and 0.01 probability levels, respectively.

number of leaves (0.387 and 0.427) and fruit length (0.463 and 0.459) also indicate that increasing total biomass and growth can increase fruit yield (Table 2). The results showed that there are some characters of yield components that correlated with yields. The Mulyani and Waluyo (2020) study also showed that the fruit weight had a positive and significant correlation with the stem length, number of branches, fruit stalk length, fruit length, fruit diameter, thickness of pericarp, number of seeds per plant, and weight of seeds per plant. Due to the fact that leaves play a role in photosynthesis and by increasing photosynthesis and good plant growth, yield can be increased, so a positive and high correlation between leaves and fruit yield can play a role in increasing yield and it can be achieved by selecting plants with larger leaves. Also, increasing the length of the fruit can increase the fruit yield. Flesh weight was also positively correlated with fruit yield. Since flesh weight is ideal for customers, it can be expected that this trait will increase as fruit yield increases. Fruit length and width traits with each other and also with fruit yield and flesh weight showed a positive and significant correlation.

Therefore, increasing fruit yield through related traits will be effective when the heritability of those traits is high and otherwise it will not be effective and time consuming. Hence, the selection value depends on the inheritance of each of the traits of interest to fruit yield, otherwise, direct selection for fruit yield will have higher genetic efficiency. Previous studies have shown a correlation between the traits in watermelon (Maggs-Kölling and Christiansen, 2003; Sundaram et al., 2011; Choudhary et al., 2012).

**Usual and sequential path analysis:** The results of usual path analysis showed collinearity between independent traits indicating the inadequacy of this model (first model) to show the actual contribution of each independent trait in both years (2016 and 2017) (Table 3). Direct effects of predictor traits on watermelon fruit yield in conventional path analysis and indicators of collinearity in two years had presented in Table 3. There was strong collinearity between some traits, especially those with high direct effects on fruit yield. For

Table 3. Direct effects of predictor traits on watermelon fruit yield in conventional path analysis and indicators of collinearity in two years.

<table>
<thead>
<tr>
<th>Traits</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct effect</td>
<td>Tolerance</td>
</tr>
<tr>
<td>SL</td>
<td>-0.151**</td>
<td>0.143</td>
</tr>
<tr>
<td>SW</td>
<td>0.020**</td>
<td>0.174</td>
</tr>
<tr>
<td>LL</td>
<td>0.074**</td>
<td>0.696</td>
</tr>
<tr>
<td>NMF</td>
<td>-0.233**</td>
<td>0.145</td>
</tr>
<tr>
<td>NL</td>
<td>0.125**</td>
<td>0.207</td>
</tr>
<tr>
<td>NNS</td>
<td>-0.023**</td>
<td>0.190</td>
</tr>
<tr>
<td>pH</td>
<td>-0.097**</td>
<td>0.629</td>
</tr>
<tr>
<td>HSW</td>
<td>0.187**</td>
<td>0.329</td>
</tr>
<tr>
<td>FL</td>
<td>0.263**</td>
<td>0.365</td>
</tr>
<tr>
<td>FW</td>
<td>-0.135**</td>
<td>0.320</td>
</tr>
<tr>
<td>FLW</td>
<td>0.809**</td>
<td>0.600</td>
</tr>
<tr>
<td>ST</td>
<td>0.040**</td>
<td>0.693</td>
</tr>
</tbody>
</table>

*ns and **: Non-significant and significant at the 0.01 probability levels, respectively.
example, the variance inflation factor for flesh weight was 11.67 and 12.13 during the year 2016 and 2017 respectively (Table 3). The reason for this high collinearity was the high correlation between flesh weight and fruit yield in both years (Table 2). The variance-influence factor of number of male flower, number of leaf, number of nodes per stem, fruit length and fruit width was equal to 16.89, 14.83, 15.27, 12.74 and 13.12 in 2016, and 14.86, 17.25, 16.24, 13.71 and 12.43 in 2017 and showed high collinearity (Table 3). The reason for this high collinearity was the high correlation of these traits with other traits justifying fruit yield.

By using sequential path analysis, the collinearity of traits significantly decreased in this study (Table 4 and 5). Compared to usual path analysis results, sequential path analysis has simplified the relationship between traits and their contribution to justifying fruit yield. The results showed a significant reduction in values of the variance-influence-factor of the first model compared to the second model in both years (Table 4 and 5). In this study, stepwise regression analysis minimizes the collinearity of variables and, by reducing the mixing of effects, correctly identifies the actual participation rate of each variable in different paths. The advantage of sequencing path analysis over usual path analysis has also been proven in other studies (Mohammadi et al., 2003; Asghari-Zakaria et al., 2006; Dalkani et al., 2011).

In sequential path analysis, traits were divided into first, second and third rankings in each year (Fig. 1 and 2). Based on the tolerance, variance inflation factor and the magnitude of direct effects, the flesh weight trait was selected as the first variable in both years for justifying the dependent trait of fruit yield (Table 6 and 7). In the next step, the flesh weight was considered as a dependent variable and the others traits were considered as an independent variable and by stepwise regression was determined the second-order variables for flesh weight in both years (Table 6 and 7). To determine the third rank, the second rank variables were considered as a dependent traits in stepwise regression analysis. The direct effects of the 1000 bootstrap samples showed a close agreement with the directly observed effects in both years (Table 6 and 7). Low standard error and the direct effects biased showed the power of sequential path analysis to explain the real contribution of each trait in the fruit yield. The direct effects of variables are presented in Tables 6 and 7, and the results of t test indicate that all of them are significant.

![Figure 1. Sequential path model indicating the interrelationships among fruit yield with related traits in watermelon genotypes in 2016. Traits abbreviations are according to Table 2.](image-url)
Path analysis in watermelon

Figure 2. Sequential path model indicating the interrelationships among fruit yield with related traits in watermelon genotypes in 2017. Traits abbreviations are according to Table 2.

Flesh weight as a first-rank variable justifies 72% of fruit yield variations in year 2016 and 49% in year 2017 (Table 6 and 7). Path analysis for second-order traits showed that in the year 2016, number of nodes per stem had a positive effect on flesh weight, and in total 18% of the variation of this trait was justified. The direct effect of number of nodes per stem on flesh weight (0.445) indicates the efficiency of this trait as a selection index to increase these trait. While in 2017, number of nodes per stem and fruit length showed positive effect on flesh weight and 43% of its variation was justified.

When the third rank variables were considered as predictors and second-ranking traits were considered as a dependent, the results showed that 72% of the variation of number of nodes per stem in 2016 was explained by two number of male flower and fruit width traits. The direct effect of number of male flower was high and positive, while the direct effect of fruit width was low and positive (Table 6). Also, in 2017, the results showed that 61% of the variation in number of nodes per stem was explained by the number of leaf trait, and its direct effect was found high and positive (Table 7). While number of leaf and skin thickness were recognized as effective traits for fruit length and 33% of variation was justified and showed moderate and positive direct effects (Table 7).

Also, the heritability of seed length, seed width, leaf length, number of male flower, number of leaf, number of nodes per stem, pH, 100-seed weight, fruit length, flesh weight, skin thickness and fruit yield were equal to 56.72%, 64.33%, 68.08%, 73.57%, 72.13%, 76.89%, 72.32%, 74.22%, 74.92%, 76.62%, 57.96%, 56.41% and 50.82%, respectively. In another study (Anburani et al., 2019), it was observed high heritability (broad sense) for 100 seed weight (97.51%), fruit weight (96.78%), fruit length (94.21%), flesh thickness (93.86%), number of male flowers(84.59%) and yield per plant (65.58%) in one year. Since the traits that were effective in causal analysis on fruit yield have good heritability, so these traits can be used as a selection criterion to increase fruit yield in watermelon.

Conclusion: In current research, sequential path analysis model was used to determine the effective traits on fruit yield by avoiding the mixed effects of traits due to the collinearity between the traits. Flesh weight was considered as the most important trait in both years. It showed the most direct effect on fruit yield, and then number of nodes per stem, which had the most impact on flesh weight in both years, and was considered and introduced as an important selective criterion for increasing fruit yield. Considering the large watermelons as a marketable feature, these traits can be used as a suitable measure for increasing yield and production of large watermelons. But in addition to fruit yield, there are other important indicators that reduce production costs and increase the income per unit area. These factors include maturity, early maturity, and shortening the length of the harvest period. Besides, qualitative traits such as glucose and beta-carotene are important for watermelon breeding, which should be considered in subsequent studies.
Conflict of Interest: The authors declare that they have no conflict of interest.

Acknowledgments: We gratefully acknowledge the research funding provided for this project (No. 9203924) by Iran National Science Foundation (INSF) and Graduate University of Advanced Technology, Kerman, Iran.

REFERENCES


Arbuckle, J.L. 2016. IBM SPSS® Amos™ 24 user’s guide. Amos Development Corporation, Crawfordville, Florida, USA.


IBM-Corp. 2016. IBM SPSS Statistics for Windows, Version 24.0. IBM Corporation, Armonk, New York, USA.


[Received 11 April 2020; Accepted 12 Aug. 2020 Published 25 Oct. 2020]