Selection of genotypes with ornamental potential in an F4 population of ornamental peppers (Capsicum annuum L.) based on multivariate analysis

Maria do Perpetuo Socorro Damasceno Costa¹*¹, Elizanilda Ramalho do Rêgo¹, João Felipe da Silva Guedes¹, Michele Gonçalves de Carvalho¹, Anderson Rodrigo da Silva², Mailson Monteiro do Rêgo¹

¹Federal University of Paraíba, Areia, Brazil  
²Federal Institute of Science and Technology Education in Goiás, Uruaí, Brazil

*Corresponding author. e-mail: agro30costa@gmail.com

Abstract

The commercialization of ornamental pepper plants has been growing in the national and international flower markets. Despite this growth, there are still few available cultivars for ornamental purposes, suggesting a need for breeding programs in this segment. Under the hypothesis of the existence of variability within an F4 family, and of the better discrimination based on qualitative, quantitative, and mixed data, aiming at the selection of potential genotypes for in-pot ornamental cultivation, this study aimed to select ornamental genotypes from an advanced population based on multivariate analysis. One segregating F4 (17.15) population was used to analyze 16 qualitative and quantitative morphological descriptors. Singh’s criterion was adopted to quantify the relative contribution of the traits. Afterward, Tocher’s clustering method based on the generalized distance of Mahalanobis was used for the quantitative variables. Gower’s distance was employed for the qualitative variables, followed by the joint analysis of qualitative and quantitative data. Furthermore, the non-metric multidimensional scaling was applied. Based on the methods used, Tocher’s clustering was the most efficient method in discriminating genotypes based on qualitative data. There is genetic diversity between the individuals of the F4 population of ornamental peppers. According to the analysis performed, genotypes 12, 16, 19, 15, 29, 41, and 44 are indicated to open line in an F5 generation. Tocher’s clustering method is recommended to evaluate the diversity of the F4 population in ornamental peppers for presenting a more significant variability between genotypes, based on qualitative data.

Keywords: diversity, Gower, Mahalanobis, ornamental plants

Introduction

Peppers (Capsicum ssp.) are among the most important horticultural crops of the family Solanaceae (Mahmoud et al., 2015). This crop is native to the American continent and has approximately 38 species, although only five are considered domesticated (Capsicum annuum, C. baccatum, C. chinense, C. frutescens, and C. pubescens) (Dewitt & Bosland, 2009). Among the domesticated species, C. annuum is the most cultivated, presenting greater genetic variability and great versatility of applications and forms of use, among them the use as an ornamental plant (Padilha et al., 2016; Neitzke et al., 2016; Pessoa et al., 2019a).

In the industry of ornamental peppers, the supply of new plant types opens new markets (Rêgo & Rêgo, 2018). In principle, any pepper plant could be used as an ornamental plant, but not every pepper cultivar adapts to in-pot cultivation. In order to find all ornamental attributes in one plant, it is necessary to recombine genes by hybridization (Rêgo & Rêgo, 2016; Neitzke et al., 2016).

In the last decades, ornamental peppers have grown in the national and international flower markets (Ibraflor, 2018). There are few commercial cultivars available for cultivation in small pots since most Capsicum plants are relatively large and are selected for field cultivation in order to produce fruits for in natura consumption of industrial processing (Finger et al., 2012). Therefore, there is a need for breeding programs with ornamental peppers for in-pot cultivation in order to meet the demand of this market.

Through genetic divergence analysis, it is possible to select pepper genotypes that present desirable traits for ornamental purposes, in a segregating population, through successive cycles of self-fertilization obtained from homogeneous lineages (Rêgo et al., 2013). Seven or eight selfing cycles are indicated to obtain homozygotic
ornamental pepper lineages (Seguí-Simarro, 2017; Rêgo & Rêgo, 2018). Therefore, selection is still possible in this F4 population due to the genetic diversity verified in this family, requiring more cycles in order to obtain a homogenous lines.

Biometric models are used to determine the genetic distance between individuals based on the collected data, using multivariate techniques that allow combining data on a set of traits and then selecting the most adapted genotypes to in-pot cultivation (Sudrê et al., 2005; Silva et al., 2015; Mesquita et al., 2016). The joint analysis of variables provides a better understanding of the variability in germplasm banks when different traits are analysed (Rocha et al., 2010; Torres et al., 2015). Gower’s algorithm (Gower, 1971) allows the joint analysis of qualitative and quantitative data, as well as the multidimensional scaling (Pessoa et al. 2019b).

Under the hypothesis of the existence of variability within F4 families, and of the better discrimination based on the analysis of qualitative, quantitative, and mixed data, aiming at selection potential genotypes for in-pot ornamental cultivation, this study aimed to select ornamental genotypes of an advanced breeding population based on multivariate analysis.

Material and Methods

The experiment was performed in a plant nursery belonging to the Laboratory of Plant Biotechnology of the Center of Agricultural Sciences of the Federal University of Paraíba (CCA/UFPB), municipality of Areia-PB, with an elevation of 618 m, and located in the coordinates 06° 57’ 48” S latitude and 35° 41’ 30” W longitude. An F1 family (17.15) from a segregating ornamental pepper population (C. annuum) originated from the cross between parents UFPB 77.2 and UFPB134 was used in the experiment.

Sowing was performed in expanded polystyrene trays with 200 cells filled with the commercial substrate Plantmax HT®. When presenting three pairs of definitive leaves, at 35 days after sowing, the seedlings were transplanted to plastic pots with a volumetric capacity of 900 mL, which contained the commercial substrate.

Both plant and fruit descriptors were evaluated regarding quantitative data. The evaluated traits were plant height (PH), canopy diameter (CD), first bifurcation height (FBH), stem diameter (SD), leaf length (LL), and leaf width (LW). For the fruit, the evaluated traits were pedicel length (PLE), fruit weight (FW), fruit length (FL), largest fruit diameter (LFD), smallest fruit diameter (SFD), pericarp thickness (PT), placenta length (PL) (all measurements were taken in centimeters), number of seeds per fruit (NSF), number of fruits per plant (NFP), and dry matter content (DMC).

Likewise, plant and fruit descriptors were also evaluated for the qualitative data. The plant traits evaluated were stem color (SC), stem shape (SS), growth habit (GH), branching density (BD), leaf color (LC), and leaf shape (LS). For the fruit qualitative descriptors, the evaluated traits were anthocyanin stains on the fruit (ASF), immature fruit color (IFC), fruit color at the intermediate stage (IFC), mature fruit color (MFC), fruit shape (FS), fruit apex shape (FAS), fruit appendix (FA), pedicel persistence with the fruit (PPF), and pedicel persistence with the stem (PPS). The morpho-agronomic characterization was performed according to the recommendations for Capsicum descriptors proposed by the IPGRI (1995). The experimental design was completely randomized. The experimental unit consisted of one plant per pot. One family from an F1 population with 45 plants was evaluated in the experiment.

To quantify the relative contribution of the traits for genetic divergence, Singh’s criterion (1981) was used based on the Mahalanobis generalized distance matrix.

Tocher’s method was used based on quantitative, qualitative, and mixed (qualitative + quantitative) data to analyze the genetic divergence. To validate the consistency of the methods, the phenotypical correlation coefficient was applied by measuring the degree of adjustment between the dissimilarity matrix and the phenotypical matrix, that is, the preservation of the distances resulting from the clustering, in relation to the original distances (Sneath & Sokal, 1973).

Tocher’s clustering of the quantitative variables based on the Mahalanobis generalized distance was obtained through the expression:

$$D_{ij}^{2} = \delta^2 \sum_{kl} \Psi_{kl}$$

(1)

In which $D_{ij}^{2} = \text{Mahalanobis distance between accessions i and i'}$, $\Psi = \text{is the matrix of residual variances and covariance}$, $\delta = \{d_1, d_2, \ldots, d_n\}$, in which $d_n = Y_i - Y_j$ and $d_n$ represents the difference between the mean of two i and i' for each trait evaluated; $Y_i =$ mean of the i-th accession in relation to the j-th parameter.

The matrix with qualitative data was built using Gower’s distance (Gower, 1971) through the following equation:

$$S_i = \frac{\sum_{j=1}^{n} W_{ij} S_{jk}}{\sum_{j=1}^{n} W_{jk}}$$

(2)

In which $S_i$ is the distance between genotypes i and j. $W_{jk} = \text{is a weight attributed to the i j k comparison}$, attributing the value 1 for the valid comparisons and the value 0 for the invalid comparisons (when the value of
the variable is absent in either one or both individuals). \( S_k = 1 \) is the contribution of the \( k \) variable in the similarity between individuals \( i \) and \( j \), with values between 0 and 1. For a nominal variable, if the value of the \( k \) variable is the same for individuals \( i \) and \( j \), then \( S_k = 1 \), otherwise, it will be 0. A later joint analysis was performed with the qualitative and quantitative data.

Non-metric multidimensional scaling (nMDS) was used for graphic representation in the bidimensional space of the distance matrices. The adjustment level of the nMDS mapping was calculated using Kruskal’s Stress1.

\[
\text{Stress}_1 = \left( \frac{\sum_{i=1}^{n-1} \sum_{j=1}^{i-1} (d_{ij} - d_{ij})^2}{\sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^2} \right)^{1/2}
\]  

(3)

In which \( d_{ij} \) is the distance between accessions \( i \) and \( j \); \( d_{ij} \) is the spatial distance between accessions \( i \) and \( j \). The table of stress values proposed by Sturrock & Rocha (2000) was used to validate the result obtained with the nMDS solution by comparing the value obtained with the stress value generated from the matrixes with the same number of objects and the same number of dimensions. All analyses were performed with the R software, version 4.0.1 (R Core Team, 2020).

**Results and Discussion**

Tocher’s optimization method allowed to separate the genotypes studied in family 17.15 into different groups (Table 1), demonstrating variability among them for plant and fruit traits. Variability is essential in the selection of individuals for generation advance (Pessoa et al., 2019b). Mesquita et al. (2016) and Fortunato et al. (2019), working with a segregating population of *C. annuum*, also detected high variability when observing the formation of different groups in their studies. The variability detected in this study reinforces the need to perform more selfing cycles to select elite lines.

Fourteen groups were formed based on Tocher’s clustering for qualitative data, while two groups were formed with the quantitative and mixed data (Table 1).

<table>
<thead>
<tr>
<th>Qualitative Group</th>
<th>Plant</th>
<th>Quantitative Group</th>
<th>Plant</th>
<th>Mixed Group</th>
<th>Plant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1, 18, 40, 25, 38, 7, 9, 19, 28, 22, 35</td>
<td>1, 18, 40, 25, 38, 7, 9, 19, 28, 22, 35</td>
<td>19, 45, 14, 16, 41, 22, 31, 35, 10, 33, 25, 38, 5, 29, 29, 37, 26, 21, 24, 4, 18, 20, 8, 39, 40, 44, 30, 6, 28, 3, 9, 42, 34, 43</td>
<td>30, 42, 16, 19, 33, 3, 45, 8 26, 41, 35, 10, 11, 18, 12, 1, 44, 13, 39, 25, 27, 7, 5, 20, 31, 32, 24, 34, 22, 37, 14, 6, 40, 21, 38, 2, 9, 36, 29, 23, 17, 4, 28, 43</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>14, 17, 4, 12, 13, 23, 39, 6</td>
<td>2</td>
<td>15</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>10, 34, 36, 37, 45, 26, 27, 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>31, 32, 21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2, 5, 24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3, 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>30, 33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>42, 43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results of the clustering analysis based on quantitative, qualitative, and mixed variables were revealed by the cophenetic correlation coefficients (ccc). According to Rohlf (1970), cophenetic correlation (CC) values lower than 0.7 indicate the inadequacy of the clustering method. In this study, the CC values were 0.78 (qualitative data), 0.19 (quantitative data), and 0.73 (qualitative + quantitative). Therefore, divergence evaluation using these data is not safe through this method.
The qualitative data from groups 1 to 8 comprised most genotypes, which were similar for most traits, except for pedicel persistence with the fruit. Although this population is an F₅ generation, some traits did not reach the desired homozygosity, thus detecting variability in this population. Furthermore, genetic variability is the base of genetic improvement and is required for the practice of selection in the advance of segregated generations (Pessoa et al., 2019b). It is worth noting that this population still needs to advance more generations in order to obtain a homogeneous lines that presents interesting traits for in-pot ornamental purposes.

The remaining groups were constituted by isolated genotypes. These genotypes were the most divergent and presented differences in the variables (fruit color at the intermediate stage, fruit shape, pedicel persistence with the fruit, pedicel persistence with the stem, growth habit, leaf color, and leaf shape). Qualitative traits such as color, shape, and pedicel persistence with the fruit or stem are of great importance for the production of in-pot ornamental peppers since pepper plants present fruits and leaves of varied colors, which favor their choice by the customer. Rêgo et al. (2011) affirm that the color of leaves and fruits at the intermediate stage and the growth habit are decisive traits for the consumer to choose and purchase an ornamental pepper plant. Furthermore, Neitzke et al. (2010) report that these same traits are the most important regarding the aesthetical aspect of the plant with ornamental potential.

Based on the analysis performed, the genotypes present ornamental potential and, some of them should be selected to open line in an F₅ generation. The efficiency of Tocher’s clustering in the analysis of segregating Capsicum populations was confirmed by several authors and is widely employed in the study of segregating populations (Rêgo et al., 2012; Barroso et al., 2012; Silva Neto et al., 2014; Rêgo et al., 2015; Sharma et al., 2017; Htwe et al., 2017; Fortunato et al., 2019; Pessoa et al., 2019b). Although this family is an F₅ generation, it is still possible to select individuals that stand out within this family using Tocher’s clustering method.

The genotypes formed two groups regarding the mixed data. Group 1 comprised most genotypes, and genotype 37 presented traits of ornamental interest, such as small size, long pedicel length, small leaves, and fruits with different coloration stages. Genotypes that exhibit these traits are recommended for breeding programs of ornamental peppers. These traits make the plants more attractive to the eyes of the customer when purchasing an ornamental pepper plant (Rêgo et al., 2010). Group 2 was composed of genotype 15 only. This genotype was distanced from the remainder for presenting divergence in the quantitative variables (fruit weight (0.34cm), leaf length (3.83cm), seeds per fruit (18), fruits per plant (33), and pericarp thickness (0.10cm)) and in the qualitative variables (fruit color at the intermediate stage and fruit shape). Except for leaf length, all remaining traits are interesting for ornamental peppers. Smaller leaves are more preferred by the customer when purchasing an ornamental pepper plant for in-pot cultivation (Barroso et al., 2012; Fortunato et al., 2019).

Through Singh’s method (1981), it was determined that 3 of the 16 quantitative traits contributed with 59.60% of the genetic divergence, while 13 traits contributed with 40.40% (Figure 1). The variables that most contributed to the divergence were fruit weight (28.40%), pericarp thickness (22.00%), and plant height (9.20%) (Figure 1). Similar results were found by Mesquita et al. (2016) in their study with segregating populations of Capsicum spp., in which fruit weight was one of the traits that contributed the most to genetic divergence between genotypes. The traits that contributed the least to genetic divergence were stem diameter (2.10%), seeds per fruit (1.80%), and smallest fruit diameter (1.20%). In the present study, some fruit-related traits presented smaller contributions for genetic divergence. Such results can be explained by the fact that this F₅ population presented some homozygous loci compared to early generations, presenting greater homogeneity between plants (individuals).

**Figure 1.** Estimations of the relative contribution of each variable (Sj) for the genetic divergence between accessions of Capsicum annuum L. for the family 17.15, with 16 quantitative plant and fruit morpho-agronomic variables of ornamental pepper. PH= plant height; CD= canopy diameter; FBH= first bifurcation height; SD= stem diameter; LL= leaf length; LW= leaf width; PLE= pedicel length; FW= fruit weight; FL= fruit length; LFD= largest fruit diameter; SFD= smallest fruit diameter; PT= pericarp thickness; PL= placenta length; NSF= number of seeds per fruit; NFP= number of fruits per plant; DMC= dry matter content.
The non-metric multidimensional scaling analysis allowed separating genotypes in the graph for the quantitative and mixed data, except for the qualitative data, which presented a high stress level. The stress values were 18.00% and 12.88% for the quantitative and mixed data, respectively (Figure 2). The lower the stress value, the more reliable the position of the points in the generated image, representing the calculated distances, with little distortion in the data with the reduction of dimensions (Clarke & Warwick, 2001). According to Haouari & Ferchichi (2008), these values are considered acceptable, indicating a good ordering without any real perspective of a mistaken interpretation. According to the table proposed by Sturrock & Rocha (2000), with n = 45 genotypes and k = 2 dimensions, there is a probability (p < 0.01) that the genotypes are randomly arranged in the bidimensional space.

In the quantitative data, the most distanced genotypes from the remainder were genotypes 17, 12, 36, 23, and 19 (Figure 2). Genotype 17 presented a large size (29 cm) and the largest canopy diameter (34 cm). Genotype 12 presented the largest placenta length (0.78 cm) and the highest number of fruits per plant (62). Genotype 36 presented the longest leaf length (3.72 cm) and the longest leaf width (1.41 cm). Genotype 19 presented a small size (22 cm). Among the most distant individuals, those that presented interesting traits for in-pot ornamental peppers were genotypes 12 and 19, exhibiting the highest values for the number of fruits per plant and small size, which, according to several authors (Büttow et al., 2010; Barroso et al., 2012; Rêgo & Rêgo, 2016; Nascimento et al., 2019), are attractive traits to customers.

Furthermore, it favors the obtainment of genetic gains when practicing selection in a segregating generation (Silva Neto et al., 2014). In the mixed analysis, based on the graph, most genotypes were very close, which indicates that these genotypes present similar traits. In turn, genotypes 15,
42, and 43 remained more distant. These individuals exhibited differences in the quantitative (fruit weight, leaf length, seeds per fruit, fruits per plant, dry matter content, and pericarp thickness) and qualitative variables (fruit color at the intermediate stage and fruit shape) (Figure 2). However, the data from this study demonstrate that individuals 42 and 43 are more similar between each other than individual 15, considering all variables analyzed. Nascimento et al. (2015), Lima et al. (2019), and Fortunato et al. (2019), using other multivariate techniques in segregating Capsicum populations, also selected genotypes based on qualitative traits.

The non-metric multidimensional scaling analysis for mixed data provided partially similar results to those found by Tocher’s clustering, in which genotype 15 was isolated from the remainder in both methods, allowing the identification of genetic diversity and the selection of this genotype.

It is worth noting that, when using the mixed analysis with an acceptable stress level, it is possible to better discriminate the desirable genotypes for selection, allowing greater efficiency in the knowledge of divergence between genotypes and a better comprehension of the traits considered, providing more accurate conclusions, from the statistical point of view, on the genetic relationship between the genotypes studied (Moura et al., 2010). Therefore, this technique is a viable option for diversity studies in segregating Capsicum populations, especially when combined with the joint analysis of qualitative and quantitative data to select desirable genotypes in breeding programs. There are some reports in the literature of the use of this analysis in genetic divergence studies for breeding programs of garlic (Silva et al., 2014), fava bean (Barroso et al., 2019), and also with a segregating population of Capsicum (Pessoa et al., 2019b). Multivariate analysis allows separating individuals more precisely. Its use in qualitative traits is important since the colors exhibited by an ornamental plant are attractive to customers.

Therefore, the use of different clustering methods serves to identify the most sensitive method to determine the number of groups and the method that can more efficiently detect the existence of divergent genotypes (Faria et al., 2012). In this study, the method that obtained greater discrimination between genotypes was Tocher’s clustering for qualitative data. This variation occurs due to the contrasting traits of the parents used to obtain the segregating population, which, even in an advanced generation, still presents high segregation for some traits, especially the qualitative ones, which are categorical, allowing to efficiently discriminate them.

**Conclusion**

There is genetic diversity in the individuals of the F4 segregating population of ornamental peppers.

Based on the genetic variability detected by the multivariate analysis, genotypes 12, 16, 19, 15, 29, 41, and 44 should be selected to open line in an F5 generation.

Tocher’s clustering method is recommended to evaluate the diversity of an F4 population of ornamental peppers since it presents greater variability between genotypes for qualitative data.

**Acknowledgments**

The authors thank the Post-Graduation Program in Agronomy of the Federal University of Paraíba (UFPB). We also thank the National Council for Scientific and Technological Development (CNPq) and the Coordination for the Improvement of Higher Education Personnel (CAPES), for their financial assistance.

**References**


Gower, J.C.A. 1971. General coefficient of similarity and


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

All the contents of this journal, except where otherwise noted, is licensed under a Creative Commons Attribution License attribution-type BY.