

# Selection of BC<sub>1</sub>F<sub>3</sub> populations of Santa Cruz type dwarf tomato plant by computational intelligence techniques

Danilo Araújo Gomes<sup>1</sup> , Gabriel Mascarenhas Maciel<sup>2</sup> , Ana Carolina Silva Siquieroli<sup>3,\*</sup> , Camila Soares de Oliveira<sup>1</sup> , Rafael Resende Finzi<sup>1</sup> , Douglas José Marques<sup>2</sup> 

1. Universidade Federal de Uberlândia – Instituto de Ciências Agrárias – Uberlândia (MG), Brazil.

2. Universidade Federal de Uberlândia – Instituto de Ciências Agrárias – Monte Carmelo (MG), Brazil.

3. Universidade Federal de Uberlândia – Instituto de Biotecnologia – Monte Carmelo (MG), Brazil.

**Received:** Feb. 17, 2021 | **Accepted:** Jul. 1, 2021

**Section Editor:** Freddy Mora

**\*Corresponding author:** carol@ufu.br

**How to cite:** Gomes, D. A., Maciel, G. M., Siquieroli, A. C. S., Oliveira, C. S., Finzi, R. R. and Marques, D. J. (2021) Selection of BC<sub>1</sub>F<sub>3</sub> populations of Santa Cruz type dwarf tomato plant by computational intelligence techniques. *Bragantia*, 80, e4821. <https://doi.org/10.1590/1678-4499.20210046>

**ABSTRACT:** The aim of this study was to estimate genetic divergence and select BC<sub>1</sub>F<sub>3</sub> populations of dwarf tomato plant within the Santa Cruz segment by computational intelligence techniques. The experiment was conducted in a greenhouse in the Vegetable Crop Experimental Station of the Universidade Federal de Uberlândia (UFU), Monte Carmelo, MG, Brazil. A randomized block experimental design was used with 17 treatments and four replications. The genetic material evaluated comprised thirteen dwarf tomato plant populations obtained by a backcross and two self-fertilizations, plus both parents (recurrent and donor), and two commercial check varieties. The traits evaluated were mean fruit weight (MFW), soluble solids content (SSC), fruit diameter (FD), fruit length (FL), fruit shape (FS), pulp thickness (PT), number of locules (NL), distance between internodes, and acylsugar, -carotene, and lycopene content. The data were analyzed by means testing, and genetic divergence was measured using Mahalanobis generalized distance by the unweighted pair group method with arithmetic mean (UPGMA) and through computational intelligence using Kohonen self-organizing maps (SOM). Genetic dissimilarity in relation to the donor parent could be confirmed through both methodologies. However, the SOM was able to detect differences and organize the similarities among the populations in a more consistent manner, resulting in a larger number of groups. In addition, the computational intelligence techniques allow the weight of each variable in formation of the groups to be ascertained. Among the BC<sub>1</sub>F<sub>3</sub> populations, UFU-SC#3 and UFU-SC#5 stood out for agronomic traits, and UFU-SC#10 and UFU-SC#11 stood out for quality parameters.

**Key words:** *Solanum lycopersicum*, dwarfism, neural networks, Kohonen self-organizing maps.

## INTRODUCTION

The tomato plant (*Solanum lycopersicum*) has considerable socioeconomic relevance and is one of the vegetable crops most extensively grown throughout the world (Gerszberg and Hnatuszko-Konka 2017). In 2018, worldwide production was approximately 244 million tons, occupying around 5.8 million hectares (FAO 2019). In Brazil, tomatoes are classified into five groups: Minitomate, Salada, Caqui, Santa Cruz and Saladete (Alvarenga 2013). Out of Brazil, these tomato groups are known as cherry or grape (minitomatoes), round, beefsteak, chonto, and saladette or roma, respectively (Finzi et al. 2020). Among the different segments, hybrids belonging to the Santa Cruz group have fruit with the highest post-harvest durability, together with high yield potential and superior organoleptic traits compared to conventional long shelf-life tomatoes (Shirahige et al. 2010).

The high cost of production and high susceptibility to biotic and abiotic stresses make tomato growing a potentially high-risk operation (Finzi et al. 2020). Some proposals for increasing the yield and quality of the tomato fruit to ensure profitability are changes in plant spacing (Wamser et al. 2012), evaluation of different harvest times (Maciel et al. 2015a),



and the use of stimulants (Seleguine et al. 2016). In addition, there are reports of the benefits of reduction in internode space through plant breeding, resulting in more compact plants and, consequently, a larger number of tomato bunches per linear meter of stem/vine (Finzi et al. 2017a). Obtaining more compact plants that provide higher yield (Finzi et al. 2020) should be one of the aims of tomato plant breeding programs.

The success of breeding programs depends on the existence of genetic variability (Santos et al. 2019). For the purpose of measuring this genetic dissimilarity, various techniques have been used to handle the data during evaluations of dwarf tomato plants (Finzi et al. 2017a; Finzi et al. 2020; Maciel et al. 2015b). However, one of the main obstacles has been the difficulty of discovering which statistical method is most suitable for evaluating multiple traits in the germplasm of dwarf tomato plants. One alternative would be the use of computational intelligence techniques through artificial neural networks (ANN), which have proven to be promising in different areas of breeding (Santos et al. 2019).

Therefore, the aim of this study was to estimate genetic dissimilarity and select  $BC_1F_3$  populations of dwarf tomato within the Santa Cruz segment by computational intelligence techniques.

## MATERIAL AND METHODS

The experiment was conducted from October 2019 to March 2020 at the Vegetable Crop Experimental Station, Monte Carmelo, MG, Brazil (18°42'43.19"S, 47°29'55.8"W, and altitude of 873 masl). The plants were grown in an arch type greenhouse (7 × 21 m), with a ceiling height of 4 m, covered with a 150-micron transparent polyethylene film with an ultraviolet radiation inhibitor additive and lateral curtains of a white anti-aphid screen.

The genetic material evaluated consisted of 13 dwarf tomato plant populations obtained from two self-fertilizations of a previous backcross ( $BC_1F_3$ ) after hybridization of a pre-commercial homozygote line with Santa Cruz type fruit (UFU-TOM-Mother-2) crossed with the dwarf line UFU MC TOM1 (Maciel et al. 2015b), both parents, and the Santa Cruz commercial check varieties Kada and Santa Clara, for a total of 17 treatments. The wild accession *Solanum pennellii* present in this study was used only for comparison of the variable related to resistance to arthropod pests. The  $BC_1F_3$  populations and the parents are from the tomato germplasm bank of UFU. The commercial check varieties and the recurrent parent are characterized by their indeterminate growth habit and red Santa Cruz type fruit. In contrast, UFU MC TOM1 is a homozygote line for dwarf plant size with an indeterminate growth habit and oblong fruit of the mini-tomato type (Finzi et al. 2017b; Maciel et al. 2015b), which was used as a donor parent. Since expression of the dwarf phenotype is of recessive and monogenic origin (Maciel et al. 2015b), backcrosses were performed for transfer of the recessive allele.

Seeds were sown in polystyrene trays (200 cell) on October 3, 2019. Seedlings were transplanted 36 days after sowing (DAS) in 5 L capacity plastic pots. Coconut fiber based commercial substrate was used both in the trays and in the pots. Throughout the time of the experiment, crop treatments were performed as recommended for the tomato crop grown in a protected environment (Alvarenga 2013). The recurrent parent and the commercial check varieties were oriented vertically with two stems in a cord training system.

A randomized block experimental design was used, with 17 treatments and four replications. The experimental plots consisted of six plants distributed in double rows at a spacing of 0.3 × 0.3 m. A spacing of 0.8 m was used between the double rows, for a total of 360 plants.

Tomatoes were harvested weekly from January 3 to March 6, 2020, for a total of ten harvests. The fruit from each experimental plot was harvested in the full maturity stage, and the following agronomic traits were evaluated: mean fruit weight (MFW) (g), ratio between the weight of the tomato fruit and the number of tomatoes harvested from the plot; total soluble solids content (SSC) (°Brix), after harvest, the fruit was analyzed regarding total SSC using a portable digital refractometer (Atago PAL-1 3810); fruit diameter (FD) (cm), obtained with the aid of a ruler after cutting the fruit vertically in the middle, measuring its horizontal length/diameter; fruit length (FL) (cm), obtained with the aid of a ruler after cutting the fruit vertically in the middle, measuring its vertical length; fruit shape (FS), obtained by the ratio between the FD and FL (FD/FL). The recurrent parent and the commercial check varieties were used as references of the Santa Cruz segment to allow classification of the fruit; pulp thickness (PT) (cm), obtained with the aid of a ruler after cutting the fruit

vertically in the middle, measuring the length between the fruit peel and the beginning of the locule; number of locules (NL) (locules·fruit<sup>-1</sup>), obtained after cutting the fruit horizontally in the middle, counting the NL; internode length (IL) (cm), obtained by the equation [(plant height/number of nodes)], in two plants at the center of the plot; acylsugar content (AA) (nmol·cm<sup>-2</sup> of leaf area), obtained at 75 DAS, using a sample composed of eight leaf disks (equivalent to 4.2 cm<sup>2</sup>) from each plant of the plot. The discs were collected from leaflets from the upper third of the plants and placed in test tubes. Extraction and quantification followed the method adapted by Maciel and Silva (2014).

Regarding nutritional traits, the  $\beta$ -carotene and lycopene contents were extracted and quantified according to the method adapted from Nagata and Yamashita (1992), Rodriguez-Amaya (2001), Rodriguez-Amaya and Kimura (2004). The fruit pulp was ground and then 1 g of it was placed in a glass vial containing 3 mL of 80% acetone. The samples were kept in the dark at a temperature of 8 °C for 48 h. Absorbances were then obtained by the spectrophotometry method using the wavelength of 450 nm for  $\beta$ -carotene and 470 nm for lycopene.

After presuppositions were validated by analyses of normality (Kolmogorov–Smirnov test), homogeneity (O’Neill and Mathews test), and additivity (Tukey’s non-additivity test), data were transformed by  $\sqrt{x}$  for MFW and  $\log x + 1.0$  for NL, and the true values of these variables were tabulated.

The data were analyzed by genetic dissimilarity obtained by the matrix of Mahalanobis generalized distance and represented by a dendrogram obtained by the unweighted pair group method with arithmetic mean (UPGMA) hierarchical method. Validation of grouping by the UPGMA method was determined by the cophenetic correlation coefficient (CCC).

Computational intelligence was used to determine the similarity among the populations by means of a class of neural networks, using Kohonen self-organizing maps (SOM). Self-organizing maps learning is basically achieved in three stages. Initially, synaptic weights are attributed to the different neurons, and then a competition process occurs. The set of genetic values of each genotype is allocated to the neuron that best represents it (winning neuron). The comparison phase begins with this allocation, in which the winning neuron determines the approximation of the other neurons according to similarity. Finally, the neurons establish which will be the neighbor neurons and pass to the adaptation phase, characterized by weight adjustment for each variable.

Network training used 5000 epochs, with three neurons in each dimension, radius equal to one, allocated in nine organizational neurons (three rows and three columns), topology with a hexagonal neighborhood, Feedforward network architecture with an input layer (means) and an output neuron, and Euclidean distance type activation function. All analyses were performed on the GENES software, integrated with the R and Matlab software (Cruz 2016).

## RESULTS AND DISCUSSION

The dwarf tomato BC<sub>1</sub>F<sub>3</sub> populations differed from the parents and commercial check varieties for the agronomic traits related to tomato fruit (Table 1) by the F test ( $\alpha = 0.05$ ), except for the FS trait, and most of them were similar to the Kada and Santa Clara commercial check varieties of the Santa Cruz segment. The agronomic traits had a coefficient of genotypic determination ( $h^2$ ) greater than 70% and a CVg/CVe ratio greater than 0.90, showing their reliability during the selection process. The present study highlights the progress achieved from carrying out the first backcross and two generations of self-fertilization of superior BC<sub>1</sub>F<sub>3</sub> populations of dwarf tomato plants of the Santa Cruz group.

Mean fruit weight is of considerable importance in the tomato crop and is directly connected with fruit quality and yield (Souza et al. 2012). Separate evaluation of this trait showed an expressive mean increase in the BC<sub>1</sub>F<sub>3</sub> populations in relation to the donor parent, especially in the UFU-SC#3 and UFU-SC#5 populations, with increases of 727.91 and 681.70%, respectively.

Fruit with greater PT is generally firmer and, consequently, more resistant to physical damage during transport and less subject to deformation (Siddiqui et al. 2015). In this study, PT in all the F<sub>3</sub>BC<sub>1</sub> populations was superior to that of the donor parent, with a mean increase of 136%. Another trait associated with fruit firmness is the NL; the smaller the NL, the firmer the fruit (Siddiqui et al. 2015). The populations UFU-SC#1, UFU-SC#3, UFU-SC#4, UFU-SC#7, UFU-SC#9, UFU-SC#11, and UFU-SC#12 had a smaller NL than the commercial check varieties and the recurrent parent.

**Table 1.** Agronomic traits evaluated in dwarf tomato BC<sub>1</sub>F<sub>3</sub> populations, recurrent parent, donor parent, Kada variety, and cultivar Santa Clara.

Treatments <sup>1</sup>	MFW*	FL	FD	FS	PT	NL	IL	AA
UFU-SC#1	32.64d	3.94e	3.94c	1.00a	0.55c	2.34b	1.38b	40.79a
UFU-SC#2	34.05d	4.13e	3.92c	1.05a	0.61c	2.56c	1.42b	42.78a
UFU-SC#3	43.88c	5.27b	3.96c	1.33c	0.75a	2.32b	1.38b	30.02c
UFU-SC#4	34.00d	4.23e	3.88c	1.09a	0.63b	2.26b	1.28a	39.25b
UFU-SC#5	41.43c	4.55d	4.15c	1.10a	0.71a	2.59c	1.37b	33.79b
UFU-SC#6	26.94e	4.28e	3.43e	1.25b	0.55c	2.43c	1.71c	25.33c
UFU-SC#7	25.05e	3.60f	3.54e	1.02a	0.52c	2.29b	1.44b	26.36c
UFU-SC#8	22.32e	3.58f	3.37e	1.06a	0.54c	2.46c	1.39b	25.64c
UFU-SC#9	23.15e	3.79f	3.50e	1.08a	0.57c	2.23b	1.17a	30.20c
UFU-SC#10	28.99e	4.58d	3.72d	1.24b	0.57c	2.56c	1.24a	37.35b
UFU-SC#11	24.84e	4.04e	3.36e	1.21b	0.55c	2.28b	1.50b	26.90c
UFU-SC#12	25.91e	3.97e	3.75d	1.06a	0.56c	2.15a	1.29a	28.25c
UFU-SC#13	26.80e	4.23e	3.46e	1.22b	0.59c	2.41c	1.47b	29.85c
Recurrent genitor	104.05a	6.71a	5.45a	1.23b	0.74a	5.64d	5.12e	20.75c
Donor genitor	5.30f	3.23g	1.88f	1.72d	0.25d	2.00a	1.18a	32.80b
Santa Clara	55.50b	4.92c	4.59b	1.07a	0.72a	2.76c	4.50d	20.00c
Kada	58.71b	5.21b	4.56b	1.14a	0.75a	2.50c	4.57d	25.10c
<i>Solanum pennellii</i>	-	-	-	-	-	-	-	41.50a
CV(%)	12.55	5.00	4.00	4.43	6.44	7.12	4.09	15.67
h <sup>2</sup>	98.90	98.22	98.93	97.68	97.48	98.72	99.40	86.35
CVg/CVe	4.75	3.72	4.80	3.25	3.11	4.40	16.53	1.26

Note. \*MFW: mean fruit weight (g); FL: fruit length (cm); FD: fruit diameter (cm); FS: fruit shape; PT: pulp thickness (cm); NL: number of locules (locules-fruit<sup>-1</sup>); IL: internode length (cm); AA: acylsugar content (nmols·cm<sup>-2</sup> of leaf area). <sup>1</sup> Mean values followed by different letters in the column differ from each other by the Scott-Knott test at 0.05 significance.

Fruit size in the present study was represented by FD and FL. A mean increase of 96.28% was found for FD and 29.10% for FL. Fruit size directly affects the tomato grower; tomatoes of reduced size are difficult to place on the market in a competitive manner (Nascimento et al. 2013).

Among the dwarf BC<sub>1</sub>F<sub>3</sub> populations evaluated, UFU-SC#3 and UFU-SC#5 showed prominent results in all the traits considered up to this point (Fig. 1). Finzi et al. (2020) also reported success in obtaining BC<sub>1</sub>F<sub>2</sub> populations of dwarf tomato in the Salada segment, corroborating the results of the present study.

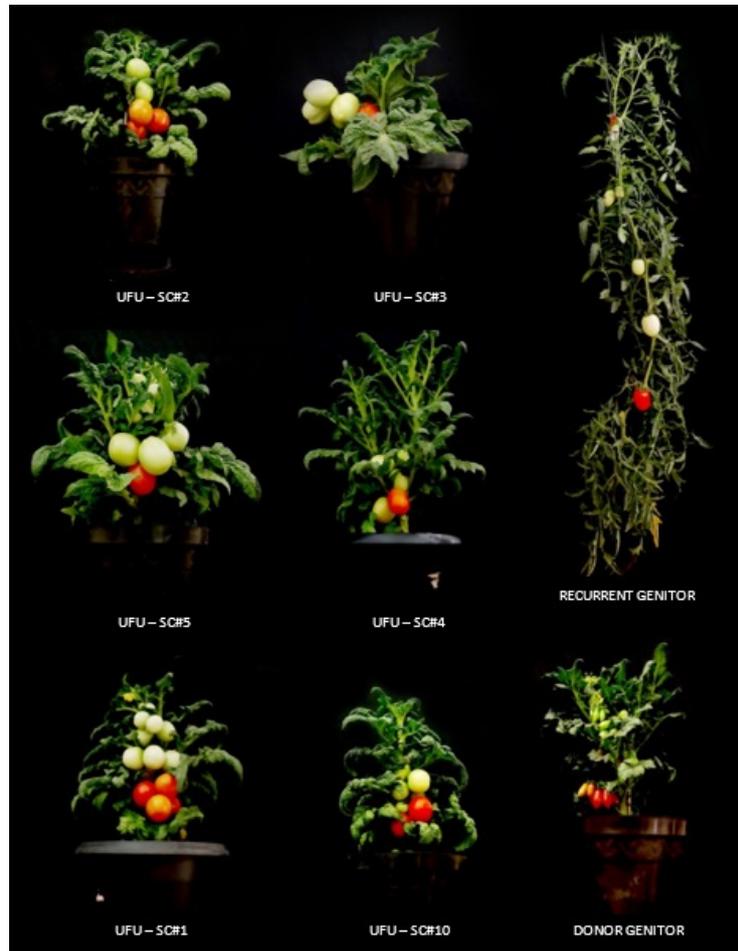
According to Siddiqui et al. (2015), FS is another attribute linked to nutritional quality and uniformity. In this study, the populations UFU-SC#1, UFU-SC#2, UFU-SC#4, UFU-SC#5, UFU-SC#7, UFU-SC#8, UFU-SC#9, and UFU-SC#12 had a rounded FS, characteristic of the Santa Cruz group (Alvarenga 2013).

Tomato breeding programs focusing on *in natura* consumption have adopted the strategy of introgression of alleles not only for fruit traits, but also for insect resistance. Various studies have confirmed that genotypes with high allelochemical contents in the leaves, such as acylsugars and 2-tridecanone, confer resistance to whitefly and other arthropod pests (Andrade et al. 2017; Oliveira et al. 2012; Neiva et al. 2019; Silva et al. 2018).

In the present study, the AA of the UFU-SC#1 (40.79 nmol·cm<sup>-2</sup>) and UFU-SC#2 (42.78 nmol·cm<sup>-2</sup>) populations did not differ from that of the wild species *Solanum pennellii* (41.49 nmol·cm<sup>-2</sup>), commonly used as a donor parent of pest resistance alleles. These materials had higher content than that found in either of the parents, the cultivar Santa Clara and Kada variety, and the other BC<sub>1</sub>F<sub>3</sub> populations. These results show the potential of the dwarf BC<sub>1</sub>F<sub>3</sub> populations with high AA have as sources of resistance to arthropod pests.

In relation to IL, all the dwarf BC<sub>1</sub>F<sub>3</sub> populations had values lower than those observed in the recurrent parent, cultivar Santa Clara, and Kada variety, which exhibited IL greater than or equal to 4.50 cm. Among the BC<sub>1</sub>F<sub>3</sub> populations, UFU-SC#4,

UFU-SC#9, UFU-SC#10, and UFU-SC#12 stood out with shorter ILs. Tomato breeding aiming to obtain tomato plant varieties with reduced IL and, consequently, better plant architecture is a future market trend (Sun et al. 2019). Panthee and Gardner (2013) and Finzi et al. (2017a) reported obtaining tomato plants with compact architecture that resulted in higher yield.



**Figure 1.** Comparison of the phenotypes of the donor parent, recurrent parent, and the BC<sub>1</sub>F<sub>3</sub> populations UFU-SC#1, UFU-SC#2, UFU-SC#3, UFU-SC#4, UFU-SC#5, and UFU-SC#10 regarding agronomic parameters.

Regarding SSC,  $\beta$ -carotene content (CC), and lycopene content (LC), the dwarf tomato BC<sub>1</sub>F<sub>3</sub> populations differed from the donor parent, recurrent parent, Kada variety, and cultivar Santa Clara only for SSC and LC (Table 2).

The SSC is the main post-harvest quality constituent in tomato fruit, and is directly linked to fruit flavor (Beckles 2012). In this study, among the dwarf F<sub>3</sub>BC<sub>1</sub> populations, UFU-SC#2 (5.22 °Brix), UFU-SC#7 (5.22 °Brix), UFU-SC#8 (5.47 °Brix), UFU-SC#9 (5.10 °Brix), UFU-SC#10 (5.23 °Brix), UFU-SC#11 (5.34 °Brix), and UFU-SC#13 (5.21 °Brix) were most prominent. For the fresh market segment, SSC of 3.0 °Brix is considered ideal (Schwarz et al. 2013). Thus, all the dwarf F<sub>3</sub>BC<sub>1</sub> populations are promising for the development of dwarf lines directed to the Santa Cruz segment with high SSC.

Expressive differences among the genotypes were not found regarding content of  $\beta$ -carotene, an important nutrient that acts in prevention of diverse diseases (Baldet et al. 2014). However, for content of lycopene, a carotenoid whose main source is the tomato fruit and that acts in protection against cardiovascular diseases and diverse types of cancer (Salvia-Trujillo et al. 2016), the dwarf populations UFU-SC#4 (2.58 mg·100 mg<sup>-1</sup>), UFU-SC#8 (2.94 mg·100 mg<sup>-1</sup>), UFU-SC#10 (2.87 mg·100 mg<sup>-1</sup>), and UFU-SC#11 (2.58 mg·100 mg<sup>-1</sup>) had higher values. Thus, the dwarf BC<sub>1</sub>F<sub>3</sub> populations UFU-SC#8, UFU-SC#10, and UFU-SC#11 showed promise regarding fruit quality parameters because they exhibited SSC and LC superior to those found in the commercial check varieties, and did not differ from them regarding  $\beta$ -carotene content.

**Table 2.** Fruit quality traits evaluated in dwarf tomato BC<sub>1</sub>F<sub>3</sub> populations, recurrent parent, donor parent, Kada variety, and cultivar Santa Clara.

Treatments	SSC	CC	LC
UFU-SC#1	4.25c	2.00a	2.44b
UFU-SC#2	5.22b	1.48a	2.19b
UFU-SC#3	4.46c	1.78a	2.32b
UFU-SC#4	4.38c	1.77a	2.58a
UFU-SC#5	4.11c	1.36a	2.13b
UFU-SC#6	4.03c	1.60a	2.25b
UFU-SC#7	5.22b	1.47a	2.10b
UFU-SC#8	5.47b	1.92a	2.94a
UFU-SC#9	5.10b	1.12a	1.90b
UFU-SC#10	5.23b	2.10a	2.87a
UFU-SC#11	5.34b	1.50a	2.58a
UFU-SC#12	4.48c	1.25a	2.00b
UFU-SC#13	5.21b	1.41a	2.11b
Recurrent genitor	5.67b	2.24a	2.91a
Donor genitor	6.65a	1.68a	2.75a
Santa Clara	6.15a	1.98a	2.72a
Kada	5.74b	2.05a	2.21b
CV(%)	7.26	28.91	14.03
h <sup>2</sup>	93.63	43.82	74.84
CVg/CVe	1.92	0.45	0.87

Note. SSC: soluble solids content (°Brix); CC: β-carotene content (mg·100 mg<sup>-1</sup>); LC: lycopene content (mg·100 mg<sup>-1</sup>). Mean values followed by different letters differ from each other by the Scott-Knott test at 0.05.

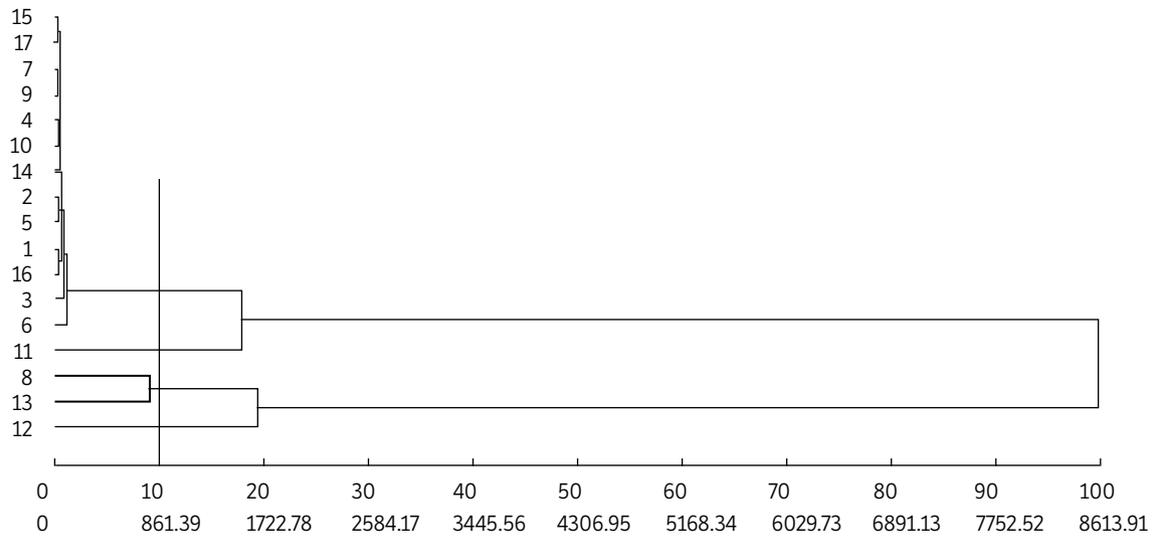
Two more backcrosses are suggested for the next steps, just as performed by Gonçalves Neto et al. (2010), resulting in dwarf lines with commercial standard fruit, and later, hybrids belonging to the Santa Cruz group coming from such lines, such as obtained by Finzi et al. (2017a). Thus, it is important to use selection strategies that allow superior dwarf populations to be obtained, as well as the use of different measures of dissimilarity (Araújo et al. 2016).

The dissimilarity estimated by the Mahalanobis generalized distance among the dwarf plants ranged from 5.32 (UFU-SC#11 and UFU-SC#13) to 655.07 (donor parent and UFU-SC#3), showing the genetic diversity among the dwarf populations (data not shown). Selection of lines with the greatest divergence possible is recommended for breeding programs (Cruz et al. 2012). The dendrogram obtained by the UPGMA method (Fig. 2) and the Kohonen SOM network model was used for visualization of this dissimilarity (Fig. 3).

For the UPGMA grouping method, the cutoff in the dendrogram was performed considering 10% of genetic variability, which allowed division of the genotypes into four different groups, a criterion defined considering the abrupt change of level (Cruz et al. 2012).

Group I was composed of all the dwarf BC<sub>1</sub>F<sub>3</sub> populations, group II consisted of the donor parent, group III was composed of the recurrent parent and the Kada variety, and group IV consisted of only the cultivar Santa Clara. The success of the first backcross can be confirmed from the separation of the donor parent and the dwarf BC<sub>1</sub>F<sub>3</sub> populations, due to improvement in agronomic performance of the BC<sub>1</sub>F<sub>3</sub> populations. Maciel et al. (2018) compared different methods of multivariate analysis for evaluation of genetic dissimilarity in tomato and found the effectiveness of this method in separation of groups.

In addition to the statistical methods traditionally applied in separation of groups, such as the Tocher optimization method and the UPGMA, ANN techniques have recently been used for classification of genotypes. The Kohonen SOM figure prominently in evaluation of genetic dissimilarity. The nonlinear structure of this technique allows detection of more complex traits in the dataset (Oliveira et al. 2020).

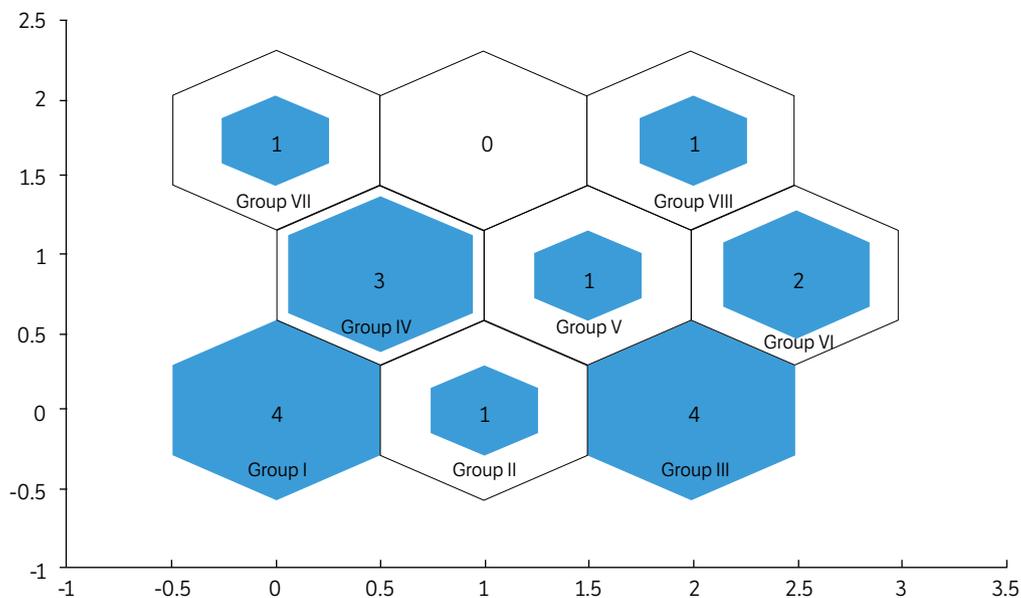


**Figure 2.** Genetic dissimilarity represented by a dendrogram obtained by the unweighted pair group method with arithmetic mean (UPGMA) hierarchical method.

Note. 1 = UFU-SC#1; 2 = UFU-SC#2; 3 = UFU-SC#3; 4 = UFU-SC#4; 5 = UFU-SC#5; 6 = UFU-SC6#; 7 = UFU-SC#7; 8 = UFU-TOM-Mother-2; 9 = UFU-SC#8; 10 = UFU-SC#9; 11 = donor genitor; 12 = cultivar Santa Clara; 13 = variety Kada; 14 = UFU-SC#10; 15 = UFU-SC#11; 16 = UFU-SC#12; 17 = UFU-SC#13.

The map is organized in a topological structure that reflects the similarity among the genotypes under study (Oliveira et al. 2020). Thus, neurons near each other have similar populations, more divergent populations are represented by neurons in the more extreme regions, and the intermediate populations constitute the center of the map.

Using the SOM method, from the nine neurons established with three rows and three grids for the command, the 17 populations were classified and eight classes were filled, forming eight groups (Fig. 3).

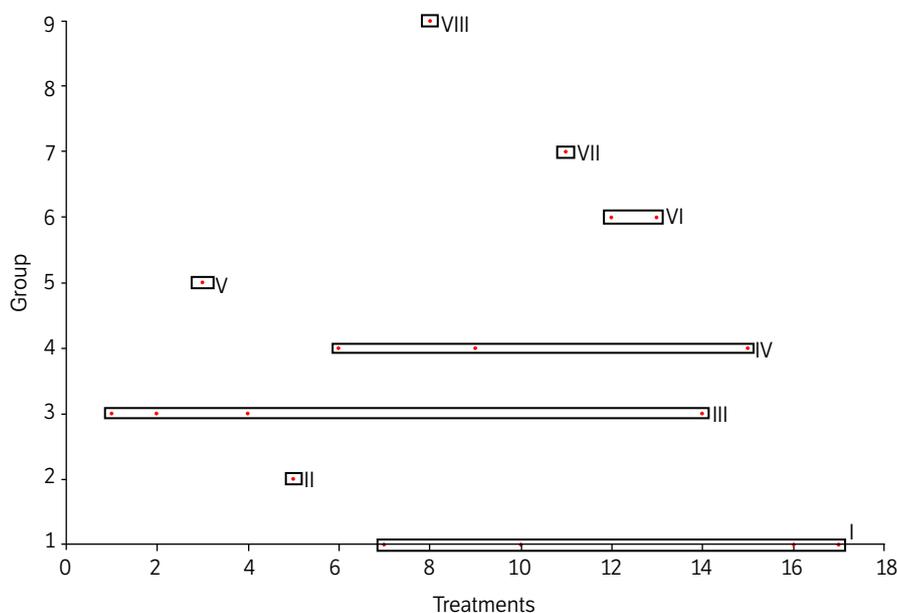


**Figure 3.** Kohonen self-organizing map for nine classes of classification, formed by the artificial neural network.

Neuron classes one and three (row I column I and row I column III, respectively) were constituted by four populations; one population was allocated to each of classes two, five, seven, and nine (row I column II, row II column II, row III column I, and row III column III, respectively), standing out for their dissimilarity in relation to the others. Three populations were

allocated to class four (row II column I); class six (row II column III) was composed of two populations; and classification of populations did not occur for class eight (row III column II). This indicates high divergence among the populations belonging to classes seven and nine, since neighbor neurons indicate proximity between materials.

The classification of the tomato populations in the classes and the constitution of eight filled groups through the Kohonen map can be visualized in Fig. 4. Group I was composed of the dwarf  $BC_1F_3$  populations UFU-SC#7, UFU-SC#9, UFU-SC#12, and UFU-SC#13; group II of the UFU-SC#5 population; group III of the populations UFU-SC#1, UFU-SC#2, UFU-SC#4, and UFU-SC#10; group IV of the populations UFU-SC#6, UFU-SC#8, and UFU-SC#11; group V of the UFU-SC#3 population; group VI of the cultivar Santa Clara and Kada variety; group VII of the donor parent; and group VIII of the recurrent parent.



**Figure 4.** Kohonen self-organizing map, classification of the observations of the groupings ( $3 \times 3$  of radius 1) of the genotypes in the classes formed by the artificial neural network.

Note. Treatments (1 = UFU-SC#1; 2 = UFU-SC#2; 3 = UFU-SC#3; 4 = UFU-SC#4; 5 = UFU-SC#5; 6 = UFU-SC#6; 7 = UFU-SC#7; 8 = UFU-TOM-Mother-2; 9 = UFU-SC#8; 10 = UFU-SC#9; 11 = donor parent; 12 = cultivar Santa Clara; 13 = Kada variety; 14 = UFU-SC#10; 15 = UFU-SC#11; 16 = UFU-SC#12; and 17 = UFU-SC#13).

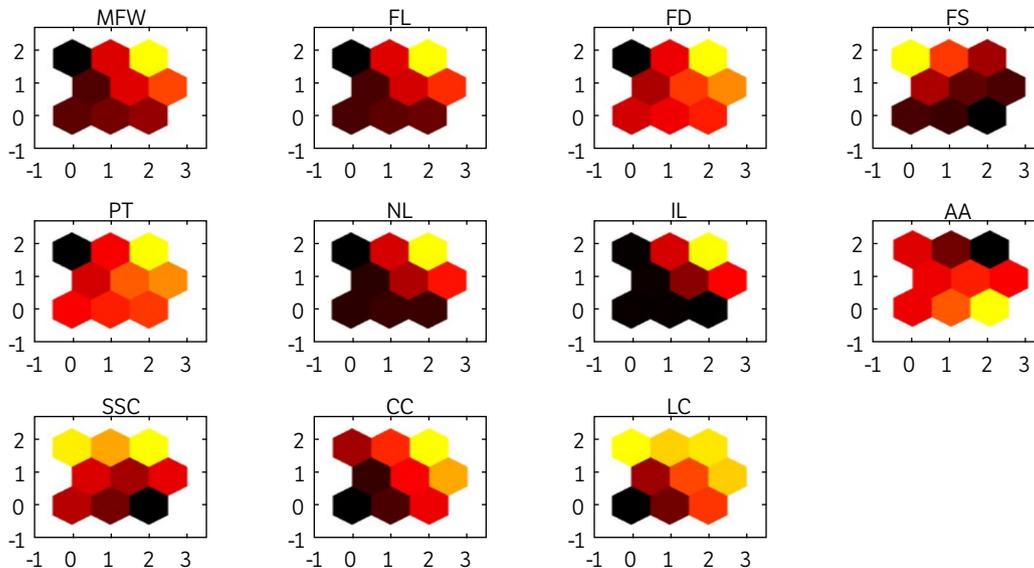
Thus, as observed by the UPGMA method, all the dwarf  $BC_1F_3$  populations differed from the donor parent; they were allocated in different groups. This reaffirms the fact that the first backcross led to increases in the  $BC_1F_3$  populations, recovering part of the genetic constitution of the recurrent parent. Nevertheless, using UPGMA, all the dwarf  $BC_1F_3$  populations were allocated in a single group, whereas for SOM, the  $BC_1F_3$  populations were distributed among groups I, II, III, IV, and V.

Among the  $BC_1F_3$  populations, the populations UFU-SC#3 and UFU-SC#5 can be highlighted, which, by the SOM method, were allocated to groups V and II, respectively. The separation of these populations from the others is explainable by their better performance for MFW and PT, with the UFU-SC#3 population also exhibiting greater FL (Table 1). Thus, the SOM method was more efficient for evaluation of genetic dissimilarity and selection of individuals with superior traits for obtaining dwarf tomato lines. In addition, the efficacy of computational intelligence is validated by the fact of showing the extreme genetic dissimilarity between the donor parent (group VII) and the recurrent parent (group VIII). Not only are they found in different groups, but they also have an empty class between them, which indicates the large distinction between these populations (Fig. 3).

Another difference between the methods was in regard to the allocation of the recurrent parent, Kada variety, and the cultivar Santa Clara. By the UPGMA method, the Kada variety and the recurrent parent were allocated in a single group and the cultivar Santa Clara in an isolated group. By the Kohonen method, the Kada variety and the cultivar Santa Clara were allocated in a single group and the recurrent parent in an isolated group. The latter grouping is more consistent with what is observed by the Scott-Knott test (Table 1), in which the Kada variety and the cultivar Santa Clara were similar regarding MFW, FD, FS, PT, NL, IL, and AA.

These differences can be explained by the high simulation ability of the neural networks, which broaden the input data and estimate new values, with different synaptic weights for each neuron, organizing the groups by order of proximity (Oliveira et al. 2020). Santos et al. (2019), using SOM for evaluation of genetic divergence, concluded that the method was efficient for evaluation of genetic diversity in rice genotypes.

From the maps of the weights and association of each input variable (traits evaluated) with the neurons of the output layer (Fig. 5), the influence that each variable exercises on the network neurons can be observed. The lighter colors indicate greater weights and, therefore, large importance. In contrast, dark colors represent lower importance of a determined trait on the neuron. In addition, neurons with similar color patterns indicate similar responses; that is, by means of analysis of the color pattern, inferences can be made regarding correlation among traits.



**Figure 5.** Weight of the input variables on the network neurons, in which lighter colors represent a greater effect of a variable on a determined neuron.

Note. MFW: mean fruit weight (g); FL: fruit length (cm); FD: fruit diameter (cm); FS: fruit shape; PT: pulp thickness (cm); NL: number of locules (locules-fruit<sup>-1</sup>); IL: internode length (cm); AA: acylsugar content (nmols·cm<sup>-2</sup> of leaf area), SSC: soluble solids content (°Brix); CC: -carotene content (mg·100 mg<sup>-1</sup>); LC: lycopene content (mg·100 mg<sup>-1</sup>).

Analysis of color intensity showed that MFW, FL, and FD, just as NL and IL, exhibited correlation with each other, because they have the same pattern of similarity among the populations, represented by colors. According to Ribeiro et al. (2016), knowledge of the correlation among traits provides information regarding the degree of association among them, in which selection of one trait can change the response of another.

In general, AFW, FL, FD, PT, NL, and IC did not contribute to the formation of the neuron that clustered the donor parent (group VII), which suggests that they were the traits that least contributed to classification of this population. In addition, the classes of neurons that formed groups I, II, and III were characterized mainly by FD, PT, and AA, respectively. The AFW and FL characteristics were important for characterizing group V (UFU-SC#3). Representation of groups VI (Kada variety and cultivar Santa Clara) and VII (donor parent) in lighter colors for most of the traits confirms their importance for evaluation of dissimilarity among these populations that are characterized by highly contrasting phenotypes.

Both methods (UPGMA and SOM) confirmed the genetic dissimilarity in relation to the donor parent. However, SOM was able to detect differences and organize the similarities among the populations in a more consistent manner, resulting in a larger number of groups compared to the UPGMA method. Regarding agronomic traits among the BC<sub>1</sub>F<sub>3</sub> populations, UFU-SC#3 and UFU-SC#5 stood out, which were actually allocated to separate groups when using SOM. In relation to quality parameters, the UFU-SC#10 and UFU-SC#11 populations stood out. In addition, analysis of color intensity indicated that all the traits were important for quantifying dissimilarity among the populations.

## CONCLUSION

The dwarf BC<sub>1</sub>F<sub>3</sub> populations UFU-SC#3, UFU-SC#5, UFU-SC#8, UFU-SC#10, and UFU-SC#11 showed most promise for obtaining lines and, subsequently, hybrids of the Santa Cruz type tomato.

The Kohonen self-organizing map, which uses computational intelligence, proved to be more suitable for classifying and clustering the dwarf tomato plant populations.

Computational intelligence is a promising approach that is easy to use and highly flexible, contributing efficiently to the study of genetic dissimilarity and selection strategies in plant breeding programs.

## AUTHORS' CONTRIBUTION

**Conceptualization:** Gomes, D. A., Maciel, G. M. and Oliveira, C. S.; **Methodology:** Gomes, D. A., Maciel, G. M., Siquieroli, A. C. S. and Oliveira, C. S.; **Investigation:** Gomes, D. A., Maciel, G. M., Siquieroli, A. C. S., Oliveira, C. S. and Finzi, R. R.; **Writing – Original Draft:** Gomes, D. A.; **Writing – Review and Editing:** Siquieroli, A. C. S., Maciel, G. M. and Finzi, R. R.; **Funding Acquisition:** Maciel, G. M., Siquieroli, A. C. S. and Marques, D. J.; **Resources:** Maciel, G. M., Siquieroli, A. C. S. and Marques, D. J.; **Supervision:** Maciel, G. M. and Gomes, D. A.

## DATA AVAILABILITY STATEMENT

All datasets were generated or analyzed in the current study.

## FUNDING

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

[<https://doi.org/10.13039/501100002322>]

Finance Code 001

Conselho Nacional de Desenvolvimento Científico e Tecnológico

[<https://doi.org/10.13039/501100003593>]

Grants No. 304082/2018-0

## ACKNOWLEDGMENTS

Not applicable.

## REFERENCES

Alvarenga, M. A. R (2013). Tomate: produção em campo, em casa-de-vegetação e em hidroponia. Lavras: UFLA.

Andrade, M. C., Silva, A. A., Neiva, I. P., Oliveira, I. R. C., Castro, E. M., Francis, D. M. and Maluf, W. R. (2017). Inheritance of type IV glandular trichome density and its association with whitefly resistance from *Solanum galapagense* accession LA1401. *Euphytica*, 213, 52. <https://doi.org/10.1007/s10681-016-1792-1>

- Araújo, J. C., Telhado, S. F. P., Sakai, R. H., Ledo, C. A. S. and Melo, P. C. T. (2016). Univariate and multivariate procedures for agronomic evaluation of organically grown tomato cultivars. *Horticultura Brasileira*, 34, 374-380. <https://doi.org/10.1590/S0102-05362016003011>
- Baldet, P., Ferrand, C. and Rothan, C. (2014). Vitamins in Fleshy Fruit. In P. Nath, M. Bouzayen, A. K. Mattoo, and J. C. Pech (Eds.), *Fruit Ripening: Physiology, Signalling and Genomics* (p. 127-150). Wallingford: CABI.
- Beckles, D. M. (2012). Factors affecting the postharvest soluble solids and sugar content of tomato (*Solanum lycopersicum* L.) fruit. *Postharvest Biology and Technology*, 63, 129-140. <https://doi.org/10.1016/j.postharvbio.2011.05.016>
- Cruz, C. D. (2016) Genes Software – extended and integrated with the R, Matlab and Selegen. *Acta Scientiarum Agronomy*. 38, 547-552. <https://doi.org/10.4025/actasciagron.v38i3.32629>
- Cruz, C. D., Carneiro, P. C. S. and Regazzi, A. J. (2012) Modelos biométricos aplicados ao melhoramento genético. Viçosa: UFV.
- [FAO] Food and Agriculture Organization of the United Nations. (2019). [Accessed Sept. 15, 2020]. Available at: <http://www.fao.org/faostat/en/#data/QC>
- Finzi, R. R., Maciel, G. M., Silva, E. C., Luz, J. M. Q. and Borba, M. E. A. (2017a). Agronomic performance of mini-tomato hybrids from dwarf lines. *Ciência e Agrotecnologia*, 41, 15-21. <https://doi.org/10.1590/1413-70542017411021416>
- Finzi, R. R., Maciel, G. M., Luz, J. M. Q., Clemente, A. A. and Siquieroli, A. C. S. (2017b). Growth habit in mini tomato hybrids from a dwarf line. *Bioscience Journal*, 33, 52-56. <https://doi.org/10.14393/BJ-v33n1a2017-35763>
- Finzi, R. R., Maciel, G. M., Peres, H. G., Silva, M. F., Peixoto, J. V. M. and Gomes, D. A. (2020). Agronomic potential of BC<sub>1</sub>F<sub>2</sub> dwarf round tomato populations. *Ciência e Agrotecnologia*, 44, e028819. <https://doi.org/10.1590/1413-7054202044028819>
- Gerszberg, A. and Hnatuszko-Konka, K. (2017). Tomato tolerance to abiotic stress: A review of most often engineered target sequences. *Plant Growth Regulation*, 83, 175-198. <https://doi.org/10.1007/s10725-017-0251-x>
- Gonçalves Neto, A. C., Silva, V. F., Maluf, W. R., Maciel, G. M., Nízio, D. A. C., Gomes, L. A. A. and Azevedo, S. M. (2010). Resistência à traça-do-tomateiro em plantas com altos teores de açúcares nas folhas. *Horticultura Brasileira*, 28, 203-208. <https://doi.org/10.1590/S0102-05362010000200011>
- Maciel, G. M. and Silva E. C. (2014). Proposta metodológica para quantificação de açúcares em folíolos de tomateiro. *Horticultura Brasileira*, 32, 174-177. <https://doi.org/10.1590/S0102-05362014000200009>
- Maciel, G. M., Fernandes, M. A. R., Hillebrand, V. and Azevedo, B. N. R. (2015a). Influência da época de colheita no teor de sólidos solúveis em frutos de minitomate. *Scientia Plena*, 11, 120203. <https://doi.org/10.14808/sci.plena.2015.120203>
- Maciel, G. M., Silva, E. C. and Fernandes, M. A. R. (2015b). Ocorrência de nanismo em planta de tomateiro do tipo grape. *Revista Caatinga*, 28, 259-264. <https://doi.org/10.1590/1983-21252015v28n429rc>
- Maciel, G. M., Peixoto, J. V. M., Rocha, J. P. R., Almeida, R. S., Oliveira, C. S. and Mendonça, T. F. N. (2018). Multivariate techniques in the determination of genetic diversity in pest-resistant mini tomato genotypes. *Horticultura Brasileira*, 36, 504-509. <https://doi.org/10.1590/s0102-053620180413>
- Nagata, M. and Yamashita, I. (1992). [Simple method for simultaneous determination of chlorophyll and carotenoids in tomato fruit]. *Nippon Shokuhin Kogyo Gakkaishi*, 39, 925-928. Japanese. <https://doi.org/10.3136/nskkk1962.39.925>
- Nascimento, A. R., Soares Júnior, M. S., Caliar, M., Fernandes, P. M., Rodrigues, J. P. M. and Carvalho, W. T. (2013). Qualidade de tomates de mesa cultivados em sistema orgânico e convencional no estado de Goiás. *Horticultura Brasileira*, 31, 628-635. <https://doi.org/10.1590/S0102-05362013000400020>

- Neiva, I. P., Silva, A. A., Resende, J. F., Carvalho, R. C., Oliveira, A. M. S. and Maluf, W. R. (2019). Tomato genotype resistance to whitefly mediated by allelochemicals and *Mi* gene. *Chilean Journal of Agricultural Research*, 79, 124-130. <https://doi.org/10.4067/S0718-58392019000100124>
- Oliveira, M. S., Santos, I. G. and Cruz C. D. (2020). Self-organizing maps: a powerful tool for capturing genetic diversity patterns of populations. *Euphytica*, 216, 49. <https://doi.org/10.1007/s10681-020-2569-0>
- Oliveira, C. M., Andrade Júnior, V. C., Maluf, W. R., Neiva, I. P. and Maciel, G. M. (2012). Resistance of tomato strains to the moth *Tuta absoluta* imparted by allelochemicals and trichome density. *Ciência e Agrotecnologia* 36, 45-52. <https://doi.org/10.1590/S1413-70542012000100006>
- Panthee, D. R. and Gardner, R. G. (2013). 'Mountain Honey' hybrid grape tomato and its parent NC 6 grape breeding line. *Hortscience*, 48, 1192-1194. <https://doi.org/10.21273/HORTSCI.48.9.1192>
- Ribeiro, L. P., Teodoro, P. E., Correa, C. C. G., Oliveira, E. F., Silva, F. A. and Torres, F. E. (2016). Correlations and genetic parameters in maize hybrids. *Bioscience Journal*, 32, 48-54. <https://doi.org/10.14393/BJ-v32n1a2016-29388>
- Rodriguez-Amaya, D. B. (2001). *A Guide to Carotenoids Analysis in Food*. Washington: ILSI.
- Rodriguez-Amaya, D. B. and Kimura, M. (2004). *HarvestPlus handbook for carotenoid analysis*. Washington: HarvestPlus.
- Salvia-Trujillo, L. and McClements, D. J. (2016). Enhancement of lycopene bioaccessibility from tomato juice using excipient emulsions: Influence of lipid droplet size. *Food chemistry*, 210, 295-304. <https://doi.org/10.1016/j.foodchem.2016.04.125>
- Santos, I. G., Carneiro, V. Q., Silva Junior, A. C., Cruz, C. D. and Soares, P. C. (2019). Self-organizing maps in the study of genetic diversity among irrigated rice genotypes. *Acta Scientiarum Agronomy*, 41, e39803. <https://doi.org/10.4025/actasciagron.v41i1.39803>
- Schwarz, K., Resende, J. T. V., Preczenhak, A. P., Paula, J. T., Faria, M. V. and Dias, D. M. (2013). Desempenho agrônômico e qualidade físico-química de híbridos de tomateiro em cultivo rasteiro. *Horticultura Brasileira*, 31, 410-418. <https://doi.org/10.1590/S0102-05362013000300011>
- Seleguine, A., Vendruscolo, E. P., Campos, L. F. C. and Faria Júnior, M. J. A. (2016). Efeito do paclobutrazol sobre o crescimento de plantas e produção de tomate (*Solanum lycopersicum* L.) em ambiente protegido. *Scientia Agropecuaria*, 7, 391-399. <https://doi.org/10.17268/sci.agropecu.2016.04.04>
- Shirahige, F. H., Melo, A. M. T., Purqueiro, L. F. V., Carvalho, C. R. L. and Melo, P. C. T. (2010). Produtividade e qualidade de tomates Santa Cruz e Italiano em função do raleio de frutos. *Horticultura Brasileira*, 28, 292-298. <https://doi.org/10.1590/S0102-05362010000300009>
- Siddiqui, M. W., Ayala-Zavala, J. F. and Dhua, R. S. (2015). Genotypic variation in tomatoes affecting processing and antioxidant attributes. *Critical Reviews in Food Science and Nutrition*, 55, 1819-1835. <https://doi.org/10.1080/10408398.2012.710278>
- Silva, A. A., Andrade, M. C., Maluf, W. R., Moraes, J. C. and Rezende, J. F. (2018). Resistance of tomato plant genotypes with high foliar allelochemical contents to the leafminer *Liriomyza trifolii*. *Arquivos do Instituto Biológico*, 84, e0892015. <https://doi.org/10.1590/1808-1657000892015>
- Souza, L. M., Paterniani, M. E. A. G. Z., Melo, P. C. T. and Melo, A. M. T. (2012). Diallel cross among fresh market tomato inbreeding lines. *Horticultura Brasileira*, 30, 246-251. <https://doi.org/10.1590/S0102-05362012000200011>
- Sun, X.-R., Liu, L., Zhi, X.-N., Bai, J.-R., Cui, Y.-N., Shu, J.-S. and Li, J.-M. (2019). Genetic analysis of tomato internode length via mixed major gene plus polygene inheritance model. *Scientia Horticulturae*, 246, 759-764. <https://doi.org/10.1016/j.scienta.2018.11.044>
- Wamser, A. F., Mueller, S., Suzuki, A., Becker, W. F. and Santos, J. P. (2012). Produtividade de híbridos de tomate submetidos ao cultivo superadensado. *Horticultura Brasileira*, 30, 168-174. <https://doi.org/10.1590/S0102-05362012000100028>