

Characterization of tomato generations according to a three-way data analysis

Ana Paula Del Medico^{1*} , Victoria Guadalupe Cabodevila¹ , Maria Susana Vitelleschi² , Guillermo Raúl Pratta¹ 

- 1.Universidad Nacional de Rosario - Facultad de Ciencias Agrarias - Instituto de Investigaciones en Ciencias Agrarias de Rosario - Zavalla (Santa Fe), Argentina.
- 2.Universidad Nacional de Rosario - Facultad de Ciencias Económicas y Estadística - Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística – Rosario (Santa Fe), Argentina.

ABSTRACT: Availability of a three-way data analysis to characterize two consecutive tomato (*Solanum lycopersicum*) generations is necessary to continue a plant breeding program with less uncertainty. The aim of this work was to analyze tomato fruit quality from F_2 and F_3 populations by two three-way data analysis: multiple factorial analysis (MFA) and generalized procrustes analysis (GPA). These techniques have the same main objective, searching for a common structure, but they achieve it in different ways. This work evaluated 18 tomato genotypes, represented by individual plants in F_2 and selfed families in F_3 . The same quantitative traits related to fruit quality were measured in both generations. The first two axes of the MFA represented 51.0% of the total variability. From the representation of the genotypes on these axes, traits differing from one generation to another were identified. The first two axes of the GPA represented 56.4% of the total variability. This analysis provided a table of analysis of variance (ANOVA), which corroborated the graphic and analytical interpretations of the MFA, a technique that provides the composition of the obtained axes. The comparison between the results obtained from these techniques indicated that both MFA and GPA allowed the detection of genotypes with discrepancies between the two generations. The MFA technique presented the advantage of studying graphically and analytically the nature and degree of phenotypic differences among genotypes in both generations, while the GPA complemented the analysis with an ANOVA, achieving the quantification of statistical significances for the discrepancies or similarities between them.

Key words: three-way data analysis, multiple factorial analysis, generalized procrustes analysis, plant breeding, plant genetic resources.

Received:
May 15, 2019

Accepted:
Aug. 08, 2019

Section Editor:
Freddy Mora

*Correspondence author:
adelmedico@fcecon.unr.edu.ar

INTRODUCTION

The use of wild *Solanum* spp. germplasm to improve not only fruit quality traits but also tomato shelf life was demonstrated by Pereira da Costa et al. (2013) and Mahuad et al. (2013). These authors reported the development of improved fruit shelf life and quality tomato genotypes by crossing *Solanum lycopersicum* cv. Caimanta and the wild species *Solanum pimpinellifolium* accession LA0722. From this interspecific first cycle hybrid, 18 recombinant inbred lines (RILs) were obtained by divergent-antagonist selection (Rodríguez et al. 2006). The second cycle hybrids (SCH, according to de Toledo et al. 1984) from cross between RIL18 and RIL1 was obtained for deriving the F_2 and F_3 populations (Cabodevila et al. 2017b).

When crosses to wild germplasm are accomplished, an exhaustive characterization of the wide phenotypic variability generated as well as its genetic basis becomes necessary (Mahuad et al. 2013). In prebreeding, different sources of variation



may be encountered, such as generations, traits, genotypes as in the situation previously described, but also environment or management of the crops could contribute to the variability of data set under study. In this context, complex database should be analyzed (Halewood et al. 2018).

In this study, a three-way data matrix (generations, genotypes and traits) was constructed, which could be analyzed through the statistical techniques of multiple factorial analysis (MFA) (Escofier and Pagès 1992, 1994) and generalized procrustes analysis (GPA) (Gower 1975).

The first usages of the GPA and MFA were for analyzing data from sensory profiling (Dijksterhuis and Gower 1992; Husson et al. 2001; Russell and Cox 2004; Pagès 2005). Due to these methodologies being powerful for producing a common configuration by consensus, their applications have been applied to others disciplines, including studies related to plant genetic resources and genetic variability (Kroonenberg et al. 2003, Bramardi et al. 2005; Zuliani et al. 2012; Vitelleschi and Chavasa 2015).

The objective of this work was to achieve the characterization of the two consecutive tomato generations evaluated for 10 phenotypic traits in different selected genotypes by GPA and MFA in order to explore the suitability of these three-way analysis techniques for continuing the breeding program with less uncertainty and with statistical support.

MATERIAL AND METHODS

Eighteen selected genotypes from the Tomato Breeding Program of the Cátedra de Genética, Facultad de Ciencias, Agrarias Universidad Nacional de Rosario (33° S, 61° W), represented by 18 individuals in the F_2 generation of the SCH RIL18 × RIL1 and the 18 corresponding F_3 families obtained by selfing (Figs. 1 and 2) were evaluated in a completely randomized design in field condition. In the selection of the F_2 individuals to originate the F_3 families, two criteria were used: On one hand, a “molecular criterion” based on Pereira da Costa et al. (2013) and Green et al. (2016) allowed the selection of those F_2 individuals with the highest number of heterozygous simple sequence repeats in order to evaluate their segregation in the derived F_3 families. On the other hand, a “phenotypic criterion” based on Mahuad et al. (2013) was followed, selecting those individuals with extreme average phenotypic values for fruit shelf life and weight, traits that had been used when obtaining the parental RIL and also maximum average values for the shape index, since segregation of round, flattened and oval tomato was observed in this F_2 generation. Both criteria were independently applied and the 18 F_2 individuals selected to obtain the F_3 families were: a) according to the molecular criterion, the F_2 individuals selected were: I50, I82, I89, II58, II72, IV58, IV61, IV65, IV69, IV75, VI42, VIII6, VIII9, VIII10 and VIII12; b) according to the phenotypic criterion: the individual having the maximum phenotypic value for fruit shelf life was II66, the individual with the maximum average for fruit weight character was I99 and the individual with the highest average for shape index was I42.

In both generations, the same 10 fruit quantitative traits were evaluated, registering the average value of each F_2 individual in at least 18 fruits per plant and the average value of each F_3 family, composed by 11 plants per family and harvesting at least 18 fruits in each plant (Zorzoli et al. 2000, Rodríguez et al. 2006). Fruit quality traits evaluated at the breaker stage (10% of the surface presenting reddish coloration) were weight (in g), diameter (in cm), height (in cm), shape index (ratio of height over diameter), shelf life (days elapsed from harvest until the fruit loses commercial quality, stored at a temperature of 25 ± 3 °C in shelves). For this assessment, 10 fruits per plant were harvested. Also, in fruits harvested at the ripe stage (90% of its surface with red color) the evaluated traits were: Content in soluble solids (in °Brix, percentage of glucose plus fructose of the homogenized juice, measured with a manual refractometer), acidity (measured through the pH of the homogenized juice), firmness (measured on the equatorial plane, in two opposite areas of the fruit with a Shore A durometer, Durofel DFT100, with a 0.10 cm² tip), a/b ratio (or Chroma index, a parameter related to color tone, with “a” being the absorbance at wavelengths of 540 nm and “b” the absorbance at wavelengths of 675 nm) and L value (or percentage of reflectance, a parameter related to color intensity, taking values from +100 for whites to 0 for blacks). The values “a”, “b” and L were determined with a CR 400 chromameter. At least 8 additional fruits per plant were harvested for this assessment, according to fruit size.

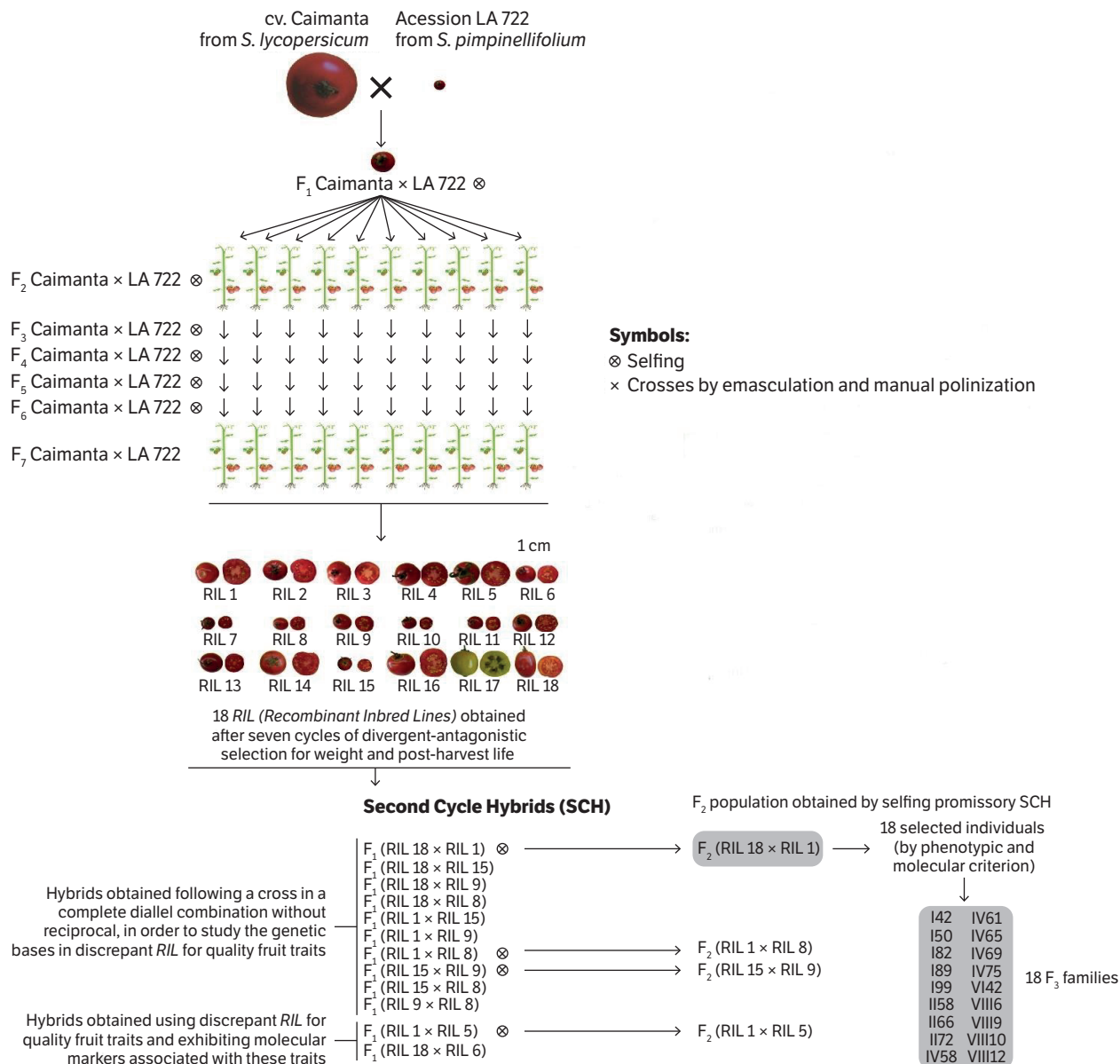


Figure 1. Diagram of crosses to obtain plant material (adapted from Rodríguez et al. 2013).

For data analyses, two three-way data analysis were applied to both generations jointly. The MFA, developed by Escofier and Pagès (1992), is a method that allows the analysis of data tables in which the same group of individuals are described through a group of variables, evaluated in different conditions, moments in time or places. Variables can be quantitative or qualitative, with the only restriction that the nature of these variables must be the same within each group. When the variables are quantitative, it is based on the principal component analysis (PCA) methodology and consists of two stages (Pagès 2004): *a) Preliminary stage (separate analyzes):* each group of variables is analyzed separately through a PCA applied in each K standardized data matrices of order $n \times p$. The first eigenvalue from each of these analyzes is denoted as λ_k^2 ($k = 1, \dots, K$) (Fig. 3) and will be used in the subsequence step; *b) Main stage (global analysis):* consists of making a PCA on the complete data matrix, which results from the juxtaposition of the submatrices weighted by the inverse of the first corresponding eigenvalue, obtained in the PCA of the preliminary stage (Fig. 4). This weighting maintains the structure of each matrix and manages to balance the influence of the different groups of variables. The objective of the MFA technique



Figure 2. Photograph of the F2 generation obtained by selfing the SCH RIL18 × RIL1.

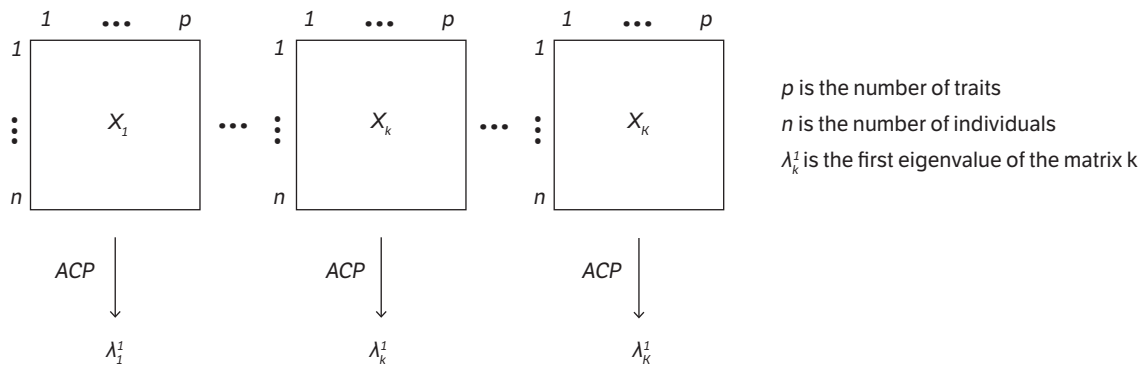


Figure 3. First stage of the multiple factorial analysis (MFA) (adapted from Vitelleschi and Chavasa 2015).

is to highlight the main variability of individuals, the latter being balanced by the various groups of variables. In addition, a global measure of the relationship between the configurations of the groups can be calculated based on the RV coefficient. This coefficient allows to quantify the association between two tables and takes values between zero (the configurations are orthogonal) and one (the distance between the configurations is zero).

The GPA proposed by Gower (1975) harmonizes the individual configurations from each data matrix, through iterative algebraic steps that include translation, rotation and scaling of the coordinates under two premises: maintaining the relative distance between elements of individual configurations and minimize the sums of squares between similar points. After the initial standardization or the translation have been done and all the configurations have been transformed, an

iteration is completed (Fig. 5). The consensus configurations are obtained from the average of all individual transformed configurations. The process is repeated until the change between two consecutive steps in the residual square sums is less than a particular value. Once the iterative process of the GPA ends, the total variability can be partitioned in the form of a table of analysis of variance (ANOVA).

All analyses were accomplished by the FactoMineR Package (Lê et al. 2008) in the R software environment (version 3.5.0, R Core Team 2017) with the script presented in Supplementary Material.

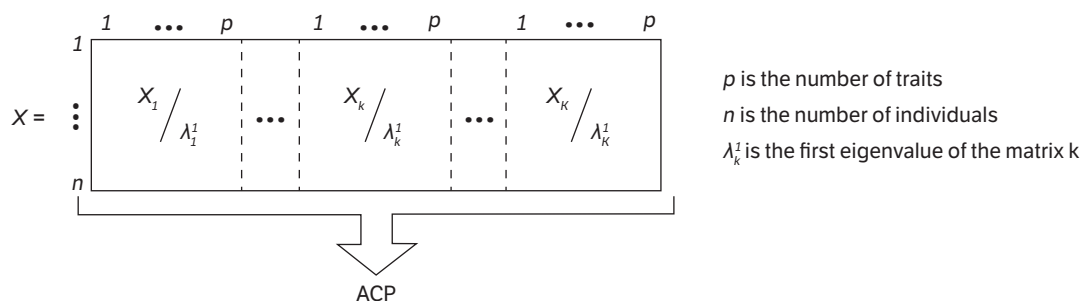


Figure 4. Second stage of multiple factorial analysis (MFA) (adapted from Vitelleschi and Chavasa 2015).

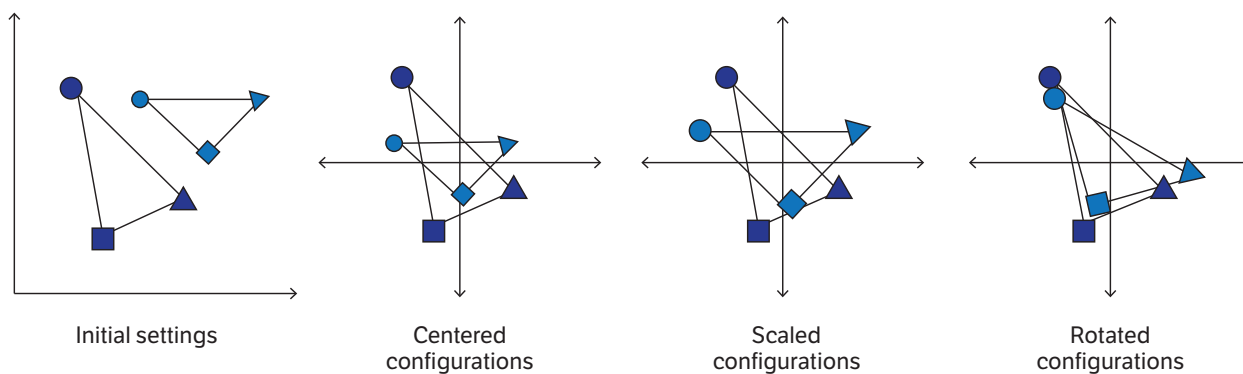


Figure 5. Outline of the transformations made in the generalized procrustes analysis (GPA) (adapted from Dijkstra and Gower 1992).

RESULTS AND DISCUSSION

Initially, to summarize and visualize the data structure in order to facilitate the identification of values trends, summary measures were presented according to generation (Table 1). Central position measurements of the variables (average and median) were slightly higher in the F_2 generation, except for the trait a/b. In addition, in each generation, the averages were similar to the medians. On the other hand, it was observed that all traits, except shelf life, showed greater variability in the F_2 generation, as shown by standard deviation, being remarkable that the greatest differences between generations were for the traits: firmness, weight and soluble solids content. Summarizing, average values of all variables were higher in the F_2 generation than in F_3 , except for a/b. However and in contrast with the expected under selfing, i.e., a broadening in phenotypic variability due to an increment in homozygosity, the variance for all traits was greater in F_2 than in F_3 , except for shelf life. In previous studies (Cabodevila et al. 2017a; Pereira da Costa et al. 2016), this relationship between SCH and their segregating generations was verified. In general, average and variability are greater in parents than in progenies.

Considering the three-way data analysis for both MFA and GPA, the factorial planes conformed by the first two global axes were considered in this work, because similarities and differences among generations for main traits can be observed in these two axes for the evaluated genotypes.

The first eigenvalue of MFA ($\lambda = 1.78$) collected a percentage of inertia of 33.67%, while the second ($\lambda = 0.92$) was 17.31%. The first eigenvalue was similar in the number of generations, which indicated that the first major axis of global direction of greater inertia was common to the two generations.

In the space of the variables (Fig. 6), the first global axis showed a characterization of the tomato fruit size. Height, weight and diameter traits mostly contributed to its formation in both generations. The second axis performed a global characterization of the external quality of the fruit. The variables mostly contributing to its formation were a/b, L and shelf life in generation F_3 and firmness in generation F_2 . The first trait contributed negatively and the other three contributed positively. It can also be observed in Fig. 6 that some vectors that represent the same variable in both generations had small angles and similar lengths, as in the case of the characters weight, height, diameter and shape index, indicating that these variables had more stable behavior from one generation to the next. In contrast, L and shelf life variables exhibited greater size angles between the two generations, which would indicate that they had a less similar behavior than the others.

Table 1. Summary measures for the variables under study according to generation and the corresponding statistical significance test.

| Variables | Average F_3 | Average F_2 | Standard deviation F_3 | Standard deviation F_2 | Median F_3 | Median F_2 | p-value test t | p-value test U |
|----------------|---------------|---------------|--------------------------|--------------------------|--------------|--------------|----------------|----------------|
| a/b | 1.246 | 1.144 | 0.079 | 0.114 | 1.270 | 1.170 | - | 0.0039 |
| Height | 2.406 | 2.712 | 0.383 | 0.487 | 2.445 | 2.790 | 0.0432 | - |
| Diameter | 2.473 | 2.564 | 0.36 | 0.401 | 2.455 | 2.535 | 0.4779 | - |
| Firmness | 49.164 | 57.519 | 4.333 | 6.195 | 49.175 | 58.020 | <0.0001 | - |
| Shape index | 0.98 | 1.065 | 0.063 | 0.079 | 0.96 | 1.060 | 0.0011 | - |
| L | 37.258 | 39.336 | 0.664 | 1.060 | 37.215 | 39.410 | <0.0001 | - |
| Weight | 9.319 | 10.832 | 3.433 | 5.193 | 9.540 | 10.540 | 0.3098 | - |
| pH | 4.697 | 4.867 | 0.105 | 0.353 | 4.685 | 4.790 | - | 0.2289 |
| Soluble solids | 6.463 | 7.639 | 0.752 | 1.460 | 6.510 | 7.255 | - | - |
| Shelf life | 18.179 | 19.707 | 3.953 | 3.878 | 17.140 | 18.900 | 0.2501 | 0.0065 |

Note: t-test of average comparison was performed for the variables with normal distribution; otherwise, nonparametric U Mann-Whitney test was carried out.

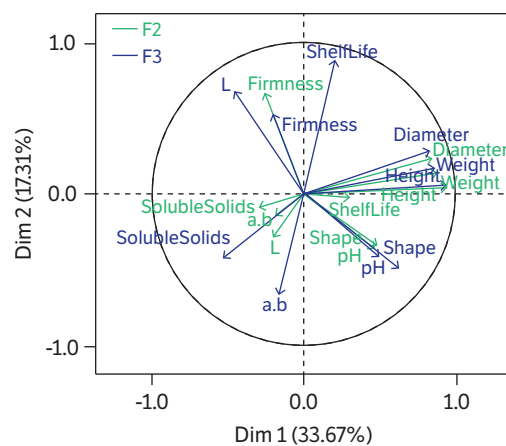


Figure 6. Representation of the correlations of the traits of the two generations on the first two global axes of the multiple factorial analysis (MFA).

In the space of the genotypes, the trajectories of each genotype in the two generations can be observed in Fig. 7. The trajectories are represented by the midpoints (consensus configuration) and the extreme points (relative positions that the genotypes occupy in each generation). The genotypes with the longest trajectories are II58, II66, IV61 and IV65, indicating that the characteristics of these genotypes differed from generation to generation. The genotype II58 differed, from generation to generation, regarding the external quality of the fruit; while genotypes II66 and IV61 differed in the size of the fruit, from one generation to another. In the case of the genotype IV65, it had differences between generations both in size and quality. On the other hand, genotypes I50, I89, VI42, VIII9 and VIII12 were those that had similar characteristics in the two generations.

The F_2 and F_3 generations were similarly positioned according to the first PCA but they differed according to the second one (data not shown). In agreement, the RV coefficient was 0.503, indicating that there is a similar structure of phenotypic variances and covariances between generations. As previously discussed, this joint correspondence is more notable for some variables than for others.

The plane constituted by the first two axes of the GPA represents 56.4% of the total variability of the data. The first global axis opposed genotypes I99 and IV69; while the second one differentiated genotypes II58 from I42 and VIII10. Also, it was observed that the midpoints of the genotypes II66, II72 and IV65 and, on the other hand, I89, IV61 and VI42 were very close to each other, implying that they had similar characteristics (Fig. 8). Figure 8 also reflects the trajectories of each genotype in the two generations represented by the midpoints (consensus configuration) and the endpoints (relative positions that occupy the genotypes in each generation). It was observed,

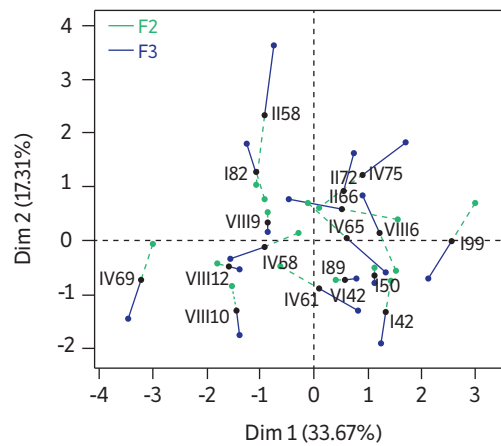


Figure 7. Representation of the medium and partial genotypes on the first two global axes of the multiple factorial analysis (MFA).

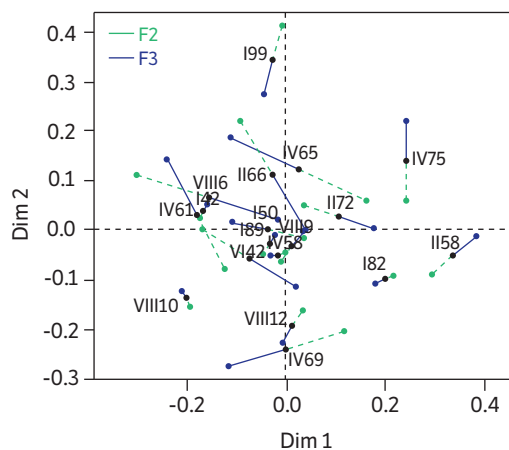


Figure 8. Trajectories of each tomato genotype in the two generations on the first two global axes of the generalized procrustes analysis (GPA).

as in MFA, that some trajectories had greater magnitude, indicating that the characteristics of these genotypes differed from generation to generation. For example, genotypes II66, IV61 and IV65 presented greater discrepancies between the two generations. On the contrary, some trajectories exhibited small magnitudes, revealing that the characteristics of these genotypes were similar in both generations, for example, I42, I50, I82, VIII9 and VIII12.

The ANOVA associated with the GPA allowed to corroborate the previous interpretations (Table 2). The genotypes I99, II58 and IV69 had the largest sum of squares consensus, exhibiting its differential behavior with the rest of the genotypes, from which they moved away in one or another dimension. In the case of the residual sum of squares, it was observed that genotypes II66, IV61 and IV65 exhibited greater discrepancies between the two generations. On the contrary, I42, I50, I82, VIII9 and VIII12 had similar characteristics between the two generations since they had low values of residual sum for squares.

To compare the two consensus configurations obtained after applying MFA and GPA, the correlation between the distance matrices presented by the tomato genotypes (average individuals in GPA and MFA) was calculated in these configurations corresponding to the first two factorial axes. The result was 0.995, indicating that the configurations found for each of the two techniques showed high concordance in the characterization of the tomato genotypes. However, the MFA allowed determining that the variables related to fruit size contributed to the formation of the first axis; while the second axis was determined by attributes related to the fruit quality. Also, it was possible to estimate a parameter in multivariate form ($RV = 0.503$), indicating that the two generations had a structure with similarities for characters that contribute to the first axis, but also with some differences in characters associated with the second axis. On the other hand, while MFA is a purely descriptive method, the GPA allowed to make inferences respecting the significance of the differences among genotype by means of the respective ANOVA table for the consensus.

Considering that the additive variance for a given character in a population implies the existence of different reproductive values among individuals of that population, it appears that GPA evidenced such an additive variance for at least some traits since consensus between F_2 and F_3 genotypes varied. GPA was not able to identify the

Table 2. Analysis of the variance associated with the GPA.

| Genotypes | Consensus | Residual | Total |
|-----------|-----------|----------|---------|
| I42 | 5.813 | 0.223 | 6.036 |
| I50 | 3.065 | 0.521 | 3.586 |
| I82 | 4.782 | 0.292 | 5.074 |
| I89 | 4.828 | 1.877 | 6.705 |
| I99 | 8.739 | 0.66 | 9.399 |
| II58 | 7.555 | 0.988 | 8.544 |
| II66 | 2.466 | 1.825 | 4.290 |
| II72 | 1.429 | 1.182 | 2.611 |
| IV58 | 3.205 | 1.268 | 4.474 |
| IV61 | 2.823 | 1.319 | 4.142 |
| IV65 | 3.040 | 1.321 | 4.360 |
| IV69 | 12.148 | 0.654 | 12.802 |
| IV75 | 3.709 | 1.238 | 4.947 |
| VI42 | 2.103 | 0.862 | 2.965 |
| VIII6 | 5.435 | 1.290 | 6.726 |
| VIII9 | 2.462 | 0.576 | 3.037 |
| VIII10 | 5.324 | 1.219 | 6.542 |
| VIII12 | 3.428 | 0.331 | 3.759 |
| Total | 82.355 | 17.645 | 100.000 |

specific traits having a significant additive variance, which was possible by means of MFA. On the other hand, in the MFA, the RV parameter was associated to an estimate of the global correspondence (or covariance) between both generations for the characters studied. Previous studies used this coefficient for the analysis of the genotype \times environment interaction (Vitelleschi and Chavasa 2015; Zuliani et al. 2012).

The coefficient RV can be associated with the general heritability of the data set (Del Medico et al. 2019). This estimation would make possible to carry out, objectively, a multi-character selection to continue with new cycles of self-fertilization and, thus, obtaining a joint response in the components of fruit quality. This response would depend on the contribution, of each of these components, to the total phenotypic variation and the multivariate heritability, whose magnitude is given by the RV coefficient. However, since each selection cycle, it is expected that the component of additive genetic variation will decrease, if a satisfactory response is obtained, this estimated value is valid for a limited number of generations (six or seven, depending on the selection differential applied). In fact, this occurs with the estimations of heritability in a conventional way (Kearsey and Pooni 1996).

Finally, the great variability generated in this cross could be more efficiently managed if analyzed with an integrated approach, by applying multivariate methods such as GPA and MFA as it was accomplished in this work. Through the application of the GPA, an associated ANOVA was obtained (Table 2), which allowed to corroborate the graphic interpretations (Cabodevila et al. 2017a).

CONCLUSION

The two multivariate techniques MFA and GPA showed results with a high degree of equivalence despite being methods of different origins and with totally different theoretical bases. In consequence, their applications in a plant breeding program, as in this case of the tomato crop, allow continuing the characterization and selection with a high degree of certainty and strong statistical support.

FUNDING

Consejo Nacional de Investigaciones Científicas y Técnicas
[<https://doi.org/10.13039/501100002923>]

Universidad Nacional de Rosario
[<https://doi.org/10.13039/100009573>]

Agencia Nacional de Promoción Científica y Tecnológica
[<https://doi.org/10.13039/501100003074>]

AUTHORS' CONTRIBUTION

A. P. Del Medico wrote the manuscript with support from V. G. Cabodevila, M. S. Vitelleschi and G. R. Pratta. V. G. Cabodevila carried out the experiment and performed the measurements. M. S. Vitelleschi verified the analytical methods. G. R. Pratta encouraged A. P. Del Medico to investigate availability of a three-way data analysis to characterize two consecutive tomato generations and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

REFERENCES

- Bramardi, S. J., Bernet, G. P., Asíns, M. J. and Carbonell, E. A. (2005). Simultaneous agronomic and molecular characterization of genotypes via the generalised procrustes analysis: an application to cucumber. *Crop Science*, 45, 1603-1609. <https://doi.org/10.2135/cropsci2004.0633>
- Cabodevila, V. G., Cacchiarelli, P. and Pratta, G. R. (2017a). Identification of agronomic interesting QTL in the segregating generations of a tomato second cycle hybrid. *Revista de la Facultad de Ciencias Agrarias de la Universidad Nacional de Cuyo*, 49, 1-18.
- Cabodevila, V. G., Picardi, L. A. and Pratta, G. R. (2017b). A multivariate approach to explore the genetic variability in the F₂ segregating population of a tomato second cycle hybrid. *Journal of Basic and Applied Genetics*, 28, 7-18.
- Del Medico, A. P., Cabodevila, V. G., Vitelleschi, M. S. and Pratta, G. R. (2019). Multivariate estimate of heritability for quality traits in tomatoes by three-way data analysis. *Pesquisa Agropecuária Brasileira*, 54, e00064. <https://doi.org/10.1590/s1678-3921.pab2019.v54.00064>
- Dijksterhuis, G. B. and Gower, J. C. (1992). The Interpretation of generalized procrustes analysis and allied methods. *Food Quality and Preference*, 3, 67-87. [https://doi.org/10.1016/0950-3293\(91\)90027-C](https://doi.org/10.1016/0950-3293(91)90027-C)
- Escofier, B. and Pagès, J. (1992). Análisis factoriales simples y múltiples. *Objetivos, métodos e interpretación*. Bilbao: Universidad del País Vasco.
- Escofier, B. and Pagès, J. (1994). Multiple factor analysis (AFMULT package). *Computational Statistics and Data Analysis*, 18, 121-140. [https://doi.org/10.1016/0167-9473\(94\)90135-X](https://doi.org/10.1016/0167-9473(94)90135-X)
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40, 33-51. <https://doi.org/10.1007/BF02291478>
- Green, G. Y., Pereira da Costa, J. H., Cambiaso, V., Pratta, G. R., Zorzoli, R. and Rodríguez, G. R. (2016). Single and joint effect of the basal region of chromosome 2 and centromeric region of chromosome 8 on morphological and fruit quality traits in tomato. *Euphytica*, 210, 327-339. <https://doi.org/10.1007/s10681-016-1689-z>
- Halewood, M., Chiurugwi, T., Sackville Hamilton, R., Kurtz, B., Marden, E., Welch, E., Michiels, F., Mozafari, J., Sabran, M., Patron, N., Kersey, P., Bastow, R., Dorius, S., Dias, S., McCouch, S. and Powell, W. (2018). Plant genetic resources for food and agriculture: opportunities and challenges emerging from the science and information technology revolution. *New Phytologist*, 217, 1407-1419. <https://doi.org/10.1111/nph.14993>
- Husson, F., Le Dien, S. and Pagès, J. (2001). Which value can be granted to sensory profiles given by consumers? *Methodology and results*. *Food Quality and Preference*, 12, 291-296. [https://doi.org/10.1016/S0950-3293\(01\)00014-3](https://doi.org/10.1016/S0950-3293(01)00014-3)
- Kearsey, M. J. and Pooni, H. S. (1996). *The genetical analysis of quantitative traits*. London: Chapman and Hall.
- Kroonenberg, P. M., Dunn, W. J. and Commaneur, J. J. F. (2003). Consensus molecular alignment based on generalized procrustes analysis. *Journal of Chemical Information and Computer Sciences*, 43, 2025-2032. <https://doi.org/10.1021/ci0302916>
- Lê, S., Josse, J. and Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, 25, 1-18. <https://doi.org/10.18637/jss.v025.i01>
- Mahuad, S. L., Pratta, G. R., Rodríguez, G. R., Zorzoli, R. and Picardi, L. A. (2013). Preservation of *Solanum pimpinellifolium* genomic fragments in recombinant genotypes improved the fruit quality of tomato. *Journal of Genetics*, 92, 195-203. <https://doi.org/10.1007/s12041-013-0245-z>
- Pagès, J. (2004). Multiple factor analysis: Main features and application to sensory data. *Revista Colombiana de Estadística*, 27, 1-26.
- Pagès, J. (2005). Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley. *Food Quality and Preference*, 16, 642-649. <https://doi.org/10.1016/j.foodqual.2005.01.006>
- Pereira da Costa, J. H., Rodríguez, G. R., Liberatti, D. R., Mahuad, S. L., Marchionni-Basté, E., Picardi, L. A., Zorzoli, R. and Pratta, G. R. (2016). Tomato second cycle hybrids as a source of genetic variability for fruit quality traits. *Crop Breeding and Applied Biotechnology*, 16, 289-297. <https://doi.org/10.1590/1984-70332016v16n4a44>

Pereira da Costa, J. H., Rodríguez, G. R., Pratta, G. R., Picardi, L. A. and Zorzoli, R. (2013). QTL detection for fruit shelf life and quality traits across tomato segregating populations in tomato. *Scientia Horticulturae*, 156, 47-53. <https://doi.org/10.1016/j.scienta.2013.03.015>

R Core Team. (2017). R: A language and environment for statistical computing. Software available at: <http://cran.r-project.org>

Rodríguez, G. R., Pereira da Costa, J. H., Pratta, G. R., Zorzoli, R. and Picardi, L. A. (2013). Recursos genéticos y genómicos para mejorar la calidad del fruto en tomate. *Agromensajes*, 35, 30-36.

Rodríguez, G. R., Pratta, G. R., Zorzoli, R. and Picardi, L. A. (2006). Recombinant lines obtained from an interspecific cross between *Lycopersicon* species selected by fruit weight and fruit shelf life. *Journal of the American Society for Horticultural Science*, 131, 651-656. <https://doi.org/10.21273/JASHS.131.5.651>

Russell, C. G. and Cox, D. N. (2004). Understanding middle-age consumers' perceptions of meat using repertory grid methodology. *Food Quality and Preference*, 15, 317-329. [https://doi.org/10.1016/S0950-3293\(03\)00073-9](https://doi.org/10.1016/S0950-3293(03)00073-9)

Toledo, J. F. F., Pooni, H. S. and Jinks, J. L. (1984). Predicting the properties of second cycle hybrids produced by intercrossing random samples of recombinant inbred lines. *Heredity*, 53, 283-292. <https://doi.org/10.1038/hdy.1984.87>

Vitelleschi, M. S. and Chavasa, V. (2015). Análisis factorial múltiple para la caracterización de variedades de trigo pan en diferentes ambientes. *FABICIB*, 19, 113-120. <https://doi.org/10.14409/fabicib.v19i0>

Zorzoli, R., Pratta, G. R. and Picardi, L. A. (2000). Variabilidad genética para la vida postcosecha y el peso de los frutos en tomate para familias F₃ de un híbrido interespecífico. *Pesquisa Agropecuária Brasileira*, 35, 2423-2427. <https://doi.org/10.1590/S0100-204X200001200013>

Zuliani, P., Bramardi, S. J., Lavalle, A. and Defacio, R. (2012). Maize landraces characterization using generalized procrustes analysis and multiple factor analysis. *Rev. FCA UNCUYO*, 44, 49-64.

SUPPLEMENTARY MATERIAL

Script applied for analyzing the databases

```
tomato <- read.table("FILE", header=TRUE, sep=";", na.strings=";", dec=".", row.names=1, strip.white=TRUE)
#Multiple Factorial Analysis
tomato.MFA<-tomato[, c("a.bF3", "FirmnessF3", "LF3", "pHF3", "SolubleSolidsF3", "WeightF3", "DiameterF3", "HeightF3",
"ShapeF3", "ShelfLifeF3", "a.bF2", "FirmnessF2", "LF2", "pHF2", "SolubleSolidsF2", "WeightF2", "DiameterF2", "HeightF2",
"ShapeF2", "ShelfLifeF2")]
res<-MFA(tomato.MFA, group=c(10, 10), type=c("s", "s"), ncp=5, name.group=c("F3", "F2"), num.group.sup=c(),
graph=FALSE)
plot.MFA(res, axes=c(1, 2), choix="var", new.plot=TRUE, lab.var=TRUE, habillage="group", lim.cos2.var=0)
plotMFApartial(res, axes=c(1, 2), lab.ind=TRUE, lab.par=FALSE, lab.var=TRUE, habillage="group", partial="all")

#Generalized Procrustes Analysis
tomato.GPA<-tomato[, c("a.bF3", "FirmnessF3", "LF3", "pHF3", "SolubleSolidsF3", "WeightF3", "DiameterF3", "HeightF3",
"ShapeF3", "ShelfLifeF3", "a.bF2", "FirmnessF2", "LF2", "pHF2", "SolubleSolidsF2", "WeightF2", "DiameterF2", "HeightF2",
"ShapeF2", "ShelfLifeF2")]
res<-GPA(tomato.GPA, group=c(10, 10), name.group=c("F3", "F2"), scale=TRUE, graph=FALSE)
plotGPApartial(res, axes=c(1, 2), lab.ind.moy=TRUE, habillage="group", title="")
res$RV
res$ANOVA
```