Assessment of Genetic Diversity of *Solanum lycopersicum var. cerasiforme* in Veracruz, Mexico

Doi: https://doi.org/10.15446/abc.v29n3.108548

To appear in: Acta Biológica Colombiana

Received Date: 26th July 2023 / 26 de julio de 2023

Revised Date: 04th March 2024 / 04 de marzo de 2024

Accepted Date: 21st June 2024 / 21 de junio de 2024

Please cite this article as: Délices, G., Mota-Vargas, C., Serna-Lagunes, R., Andrés-Meza P., Núñez-Pastrana, R., Galindo-Tovar, M.E., Herrera-Corredor, J.A. Murguía-González, J., and Leyva-Ovalle, O.R. (2024). Assessment of Genetic Diversity of *Solanum lycopersicum var. cerasiforme* in Veracruz, Mexico. *Acta Biol. Colomb*., *29*(3), XX-XX. https://doi.org/10.15446/abc.v29n3.108548

Acta Biol Colomb (2024); **29(3)**: pag–pag

Doi: https://doi.org/10.15446/abc.v29n3.108548
Available online at: http://www.revistas.unal.edu.co/index.php/actabiol

ARTÍCULO DE INVESTIGACIÓN / RESEARCH ARTICLE

ASSESSMENT OF GENETIC DIVERSITY OF *Solanum lycopersicum var. cerasiforme* IN VERACRUZ, MEXICO

Evaluación de la diversidad genética de *Solanum lycopersicum var. cerasiforme* en Veracruz, México

Running title: Microsatellite in tomato

Gino Delices[1a]; Claudio Mota-Vargas[2b]; Ricardo Serna-Lagunes[1c]; Pablo Andrés-Meza[1d]; Rosalía Núñez-Pastrana[1e]; María Elena Galindo Tovar[1f]; José Andrés Herrera Corredor[3g]; Joaquín Murguía-González[1h] and Otto Raúl Leyva-Ovalle[1i*]

[1.] Facultad de Ciencias Biológicas y Agropecuarias, Universidad Veracruzana, Josefa Ortiz de Dominguez s/n, Colonia centro, Amatlán de los Reyes, Veracruz, México, 94945. [2.] Instituto de Ecología A.C., Carretera antigua a Coatepec 351, Col. El Haya, Xalapa, Veracruz, México, 91073. [3.] Colegio de Postgraduados, Campus Córdoba, Carretera federal Córdoba-Veracruz Km. 348.5, Venta Parada, Veracruz, México, 94500.

[a] ORCID: https://orcid.org/0000-0003-3322-9429, gyneco61@yahoo.fr; [b] ORCID: https://orcid.org/0000-0003-2641-8512, claudio.mota@inecol.mx; [c] ORCID: https://orcid.org/0000-0003-1265-9614, rserna@uv.mx; [d] ORCID: https://orcid.org/0000-0002-0575-0084, pandres@uv.mx; [e] ORCID: https://orcid.org/0000-0001-9695-9221, ronunez@uv.mx; [f] ORCID: https://orcid.org/0000-0002-5296-6996, megalindo@uv.mx; [g] ORCID: https://orcid.org/0000-0002-2392-2521, jandreshc@colpos.mx; [h] ORCID: https://orcid.org/0000-0001-5087-4943, jmurguia@uv.mx; [i] ORCID: https://orcid.org/0000-0002-6150-9367, oleyva@uv.mx

* ***For correspondence: oleyva@uv.mx***

Received: 26th July 2023. **Returned for revision**: 04th March 2024. **Accepted**: 21st June 2024

Associate Editor: Xavier Marquínez

Citation/ citar este artículo como: Délices, G., Mota-Vargas, C., Serna-Lagunes, R., Andrés-Meza P., Núñez-Pastrana, R., Galindo-Tovar, M.E., Herrera-Corredor, J.A. Murguía-González, J., and Leyva-Ovalle, O.R. (2024). Assessment of Genetic Diversity of *Solanum lycopersicum var. cerasiforme* in Veracruz, Mexico. *Acta Biol. Colomb.*, *29*(3), XX-XX. https://doi.org/10.15446/abc.v29n3.108548

ABSTRACT

Solanum lycopersicum var. *cerasiforme* is widely distributed in Veracruz, Mexico, and represents genetic material that must be evaluated as a genetic pool of groups of tomato plants that can be used in a management program. The objective of the study was to evaluate the genetic diversity of *S. l.* var. *cerasiforme* from 12 sites in the state of Veracruz, Mexico, using SSR markers. DNA was extracted from the leaf tissue of tomato plants from 12 sites of

occurrence and their genetic diversity was evaluated using 14 SSR markers, amplified with the polymerase chain reaction technique, and visualized in polyacrylamide gels. Genetic diversity and structure were with POPgene, FSTAT, GenAlex, STRUCTURE, and STRUCTURE Harvester software. With 73.74 % Polymorphic Information Content, the observed heterozygosity (H_o=0.2251) was lower than the heterozygosity within, between groups, and the total, an indicator of low genetic diversity, due to the high genetic differentiation coefficient (G_{ST}=0.743) The genetic differentiation index indicated 35 % differentiation (F_{st}= 0.655), with a low differentiation coefficient between (F_{is}= 0.209) and within groups (F_{it}= 0.727). In the dendrogram, four main groups and four subgroups were formed, related based on geographic-genetic distance. The groups of tomato plants showed genetic differentiation and may constitute valuable germplasm for genetic diversity management programs.

Keywords: Genetic resources, Genetic variation, repetitive DNA.

RESUMEN

Solanum lycopersicum var. *cerasiforme* se encuentra ampliamente distribuido en Veracruz, México y representa material genético que debe ser evaluado como un acervo genético para su conservación y su mejoramiento. Sin embargo, no existe información sobre el *pool* genético de grupos de plantas de tomate silvestre que pueda ser utilizado en un programa de manejo. El objetivo del estudio fue evaluar la diversidad genética de *Solanum lycopersicum* var. *cerasiforme* de 12 sitios del estado de Veracruz, México utilizando marcadores SSR. Se extrajo ADN de tejido foliar de plantas de tomate de 12 sitios de ocurrencia y se evaluó su diversidad genética utilizando 14 marcadores SSR, amplificados con la técnica de reacción en cadena de la polimerasa y visualizados en geles de poliacrilamida. La diversidad y

estructura genética se evaluó con los softwares POPgene, FSTAT, GenAlex, STRUCTURE y STRUCTURE Harvester. Con un 73,74 % de Contenido de Información Polimórfica, la heterocigosidad observada (H_o=0,2251) fue menor a la heterocigosidad dentro, entre grupos y a la total; indicador de baja diversidad génica, debida al alto coeficiente de diferenciación genética (G_{ST}=0,743). El índice de diferenciación genética indicó un 35 % de diferenciación (F_{st}= 0,655), con un bajo coeficiente de diferenciación entre (F_{is}= 0,209) y dentro de grupos (F_{it}= 0,727). En el dendograma se formaron cuatro grupos principales y cuatro subgrupos, relacionados con base en la distancia geográfica-genética. Los grupos de plantas tomate mostraron diferenciación genética y pueden constituir germoplasma valioso para programas de manejo de la diversidad genética.

Palabras clave: ADN repetitivo, recursos genéticos, variación genética.

INTRODUCTION

Tomato is an herbaceous plant belonging to the Solanaceae family (Rai et al., 2012) and is one of the most significant vegetables in the world for fresh consumption and processed products (Seymour and Rose, 2022). It is the fourth most economically important crop globally (Kaushal et al., 2017) in terms of the area under cultivation and the income generated producing 186,821,216 Tons; placing Mexico as the ninth producer worldwide with 3,370,827 Tons and the largest exporter with 1,663,316 Tons (FAOSTAT, 2020).

Tomatoes, specifically *Solanum lycopersicum* var. *cerasiforme*, are widely distributed across various habitats and regions with high edaphoclimatic variability conditions (Délices et al., 2019), boasting a vast genetic diversity. While predominantly self-compatible and autogamous, they exhibit varying degrees of allogamy depending on the geographical region under consideration (Blanca et al., 2012). Globally, it is estimated that 80 % of the germplasm

lacks characterization data, and 95 % lacks agronomic evaluation data (Xu, 2010). In this context, the evaluation of morphological traits is a method used to assess genetic differentiation (Nyadanu et al., 2014). However, morphological characters are significantly influenced by Genotype x Environmental interaction (Shirasawa and Hirakawa, 2013). In this regard, the evaluation of the germplasm of tomato is a valuable resource for crop improvement programs and breeding to meet the current and future needs of humanity (Kaushal et al., 2017).

For an accurate evaluation of genetic diversity, molecular markers play a crucial role (Korir et al., 2014) as they offer abundant information and high efficiency (Al Shaye et al., 2018). Microsatellites or simple sequence repeats (SSR) are effective codominant markers for detecting genetic diversity among tomato germplasm (Mazzucato et al., 2008). This is due to their high levels of allelic diversity at different loci (Benor et al., 2008).

Tomato (*S. l.* var. *cerasiforme*) is a genetic resource that represents a potential pool for genetic improvement. However, conservation and efficient use of germoplasm require accurate genetic potential information (Aguirre et al., 2017) This study aims to evaluate the genetic and structure genetics of plants of *S. l.* var. *cerasiforme* in twelve occurrence sites of Veracruz state, Mexico, using microsatellites markers.

MATERIALS AND METHODS

Collection of biological material: Samples of tomato *S. l.* var. *cerasiforme* were taken from twelve localities with different edaphic and climatic conditions in Veracruz state, Mexico. Samples of mature fruits were taken from September 2016 to November 2017. Each sample was coded using the name of the locality, which included: 1-Tuxpan, 2-Zongolica, 3-Tenejapan, 4-Ixhuatlán, 5-Palenque, 6-Mahuixtlán, 7-Ocotitlán, 8-Coscomatepec, 9-

Córdoba, 10-Xalapa, 11-Pajapán, and 12-Maltrata (Fig. 1). Seeds from fruits of each locality were sown in pots and grown in a greenhouse for morphological and phenotypic characterization (Délices, 2021).

Microsatellite genotyping to describe genetic diversity: For genetic diversity analysis, leaf samples from 25 plants from each locality were sampled and kept in dry ice until stored at -80 °C. The biological material of the plants was collected from leaf and fruit tissue, which was initially stored in dry ice at environmental temperature and subsequently at -80 ^{o}C in the laboratory. For DNA extraction, young leaves tissue was pulverized in liquid nitrogen as Chetty et al. (2013). DNA quality and concentration were determined using a spectrophotometer, in addition, the quality of DNA was confirmed using 3 % agarose gels. The extracted DNA was diluted to a final concentration of 50 ng/µl with 1X TE buffer and stored at 4 °C. For PCR amplification, fourteen microsatellites SSRs were selected from the *Solanum* genomic network (https://solgenomics.net/). PCR amplification was conducted in a 25 µl reaction mix comprising 2.5 µl each of forward and reverse primers, 0.2 µl of Taq polymerase (1 U), 0.5 µl of DNTP mix (dTTPs, dGTPs, dCTPs, dATPs), 5 µl of 10x PCR buffer with $MgCl_2$, 2 µl of DNA template, and 12.3 µl H_2O. The amplifications were carried out with an initial denaturation at 94°C for two minutes, followed by 40 cycles of denaturation at 94 °C for 45 seconds, annealing at 60 °C for 45 seconds, and extension at 72°C for 45 seconds. Subsequently, the amplicons were analyzed by electrophoresis on 5 % agarose gels. Visualization of the DNA banding patterns was performed using the ChemiDoc XRS + Bio-Rad system.

Data analysis of diversity and genetic structure: Genetic diversity and population genetic differentiation were analyzed using the GDA software for Windows. The proportion of polymorphism and private alleles for each group were calculated using GDA (Lewis and Zaykin, 2001). The observed number of alleles (n_a), effective number of alleles (n_e), observed heterozygosity (H_o), total heterozygosity (H_t), heterozygosity within populations (H_s), heterozygosity between groups (D_{ST}), and gene flow were calculated using POPgene Ver. 1.31. Indicators of genetic differentiation such as *F*-statistics, $F_{st,}$ and F_{is} were calculated using FSTAT Ver 2.94 (Goudet, 2003) and were considered to have little differentiation when $F_{ST} \leq 0.05$, moderate differentiation when $0.05 < F_{ST} \leq 0.15$, strong differentiation when $0.15 < F_{ST} \leq 0.25$, and very strong differentiation when $F_{ST} > 0.25$ (Mohammadi and Prasanna, 2003).

We used GENALEX version 6.51 to determine the proportion of genetic variability and genetic distance with a principal component analysis (PCA) (Peakall and Smouse, 2012). The unweighted pair group method with arithmetic average (UPGMA) dendrogram was constructed based on genetic distance and identity (Nei, 1973), and Principal Coordinate Analysis (PCoA) was performed using data standardization with GENALEX version 6.51 (Peakall and Smouse, 2012).

The population structure was determined using the Bayesian clustering approach implemented in the STRUCTURE V 2.3.4 software, employing the admixture model with correlated allele frequencies. Twenty independent runs were performed, setting the number of populations (*K*) from one to ten to identify the optimal number of populations present within the 12 collections (samples) (Pritchard et al., 2000). The following parameters were used: a burn-in period of 10,000 and a Markov Chain Monte Carlo replication of 10,000.

The best K-value was estimated using the *ad-hoc* statistic ΔK via the online software Structure Harvester V0.6.94, to classify samples of tomato plants based on their similarity and dissimilarity of genetic diversity (http://taylor0.biology.ucla.edu/structureHarvester) (Evanno et al., 2005).

RESULTS

The analysis of 14 microsatellite loci in 12 occurrence sites of *S. l.* var. *cerasiforme* in Veracruz, México, revealed that all the loci were polymorphic in at least one locality, except for group nine, where no variants were detected, with all 14 markers monomorphic. Because it is considered that when markers turn out to be monomorphic, they do not meet the assumptions of Hardy-Weinberg equilibrium; and according to Peakall and Smouse (2011), for the analysis of genetic diversity, those samples should be excluded. Considering it, group nine was not included in the calculation of diversity indices; however, it was considered in the other analyses.

Allelic diversity ranged from six to 11 alleles, with an average of 7.7 alleles. The effective number of alleles was less than the number of alleles (4.5 on average), an indicator that they are only transmitted to the next generation (Table 1). The observed heterozygosity was lower than total, within and between populations heterozygosity, indicating low genetic diversity (Table 1). This was consistent with heterozygosity between populations (D_{ST}), which is mediated by the genetic differentiation (G_{st}) (Table 1). Polymorphic information content (PIC), evaluated with the panel of 14 microsatellite markers was 73.74 average (Table 1). The mean observed heterozygosity (H_o), total genetic diversity (H_t), heterozygosity between populations (D_{ST}), and heterozygosity within populations were 0.224, 0.763, 0.541, and 0.222, respectively, indicating that the genetic diversity between populations was larger than

within populations. The mean genetic differentiation (G_{st}) was 0.743, suggesting that the majority of genetic diversity was between groups of plants, and it was less within groups of plants.

Genetic Structure: We estimated *F* statistics and found that the mean inbreeding coefficient (F_{is}) was 0.209, indicating inbreeding between individuals within the population and a moderate deficit of heterozygotes. The mean overall inbreeding coefficient (F_{it}) was 0.727, suggesting that the allelic variance in local populations compared to the overall population was due to the inbreeding process. The mean fixation index (F_{st}) was 0.655, indicating strong genetic differentiation between populations, which aligned with the *Gst* value. The gene flow was found to be very low, with a value of N_m= 0.083.

Based on the genetic distance, a cluster analysis was conducted, revealing four main clusters (Fig. 2). The first cluster comprised Tuxpan, Xalapa, Zongolica and Tenejapan; the second cluster included Ixhuatlán and Palenque; the third cluster consisted of Ocotitlán, Coscomatepec, Mahuixtlán, Córdoba and Pajapán; and the fourth cluster solely contained Maltrata (Fig. 2).

Structure admixture analysis on the complete dataset identified a maximum rate of change in the log probability of data was *K*= 4 with Evanno's method. The analysis delineated four plant groups represented by distinct colors (green, yellow, red, and blue), for the 12 collections of *S. l.* var. *cerasiforme* (Fig. 3). The first population (red) included Tuxpan, Coscomatepec, Ocotitlán, Córdoba and Pajapán. The second population (green) comprised individuals from Zongolica, Tenejapán, and Ixhuatlán; the third one (blue) included Mahuixtlán, Xalapa, and individuals from Tuxpan, Maltrata, Ixhuatlán, and Palenque were part of the fourth plant group. The AMOVA detected 64 % of the variation between

populations, 26 % within individuals, and 10 % among individuals; with no variation detected among regions.

Principal component analysis using genetic distance illustrated the dispersion of populations across all four quadrants of the PCOA (Fig. 4), highlighting the genetic diversity observed between populations from various localities in Veracruz. A separation of Maltrata was noted, consistent with observations from the UPGMA dendrogram. Additionally, a lesser genetic distance was observed between populations from the northern and central regions compared to those from the southern region.

DISCUSSION

When genetic diversity within individuals, within plant groups, and between plant groups of tomato species *S. l.* var. *cerasiforme* was examined. Our study revealed a higher allelic diversity (observed number of alleles per locus N_a) compared to Aguirre et al. (2017) who reported 4.92 alleles per locus. The N_a values observed in our study indicate significant genetic variability in the tomato *S. l.* var. *cerasiforme* sampled in Veracruz. Previous studies conducted by He et al. (2003) and García-Martínez et al. (2006) have highlighted the significance of SSR markers in evaluating genetic diversity and variation in tomatoes. In this study, all 14 markers were found to be polymorphic. In other studies, with tomatoes Aguirre et al. (2017), identified thirteen polymorphic markers in 31 introductions of Cherry tomatoes using 36 microsatellite markers. The N_a values observed in our study indicate significant genetic variability in the tomato *S. l.* var. *cerasiforme* sampled in Veracruz.

Furthermore, we observed higher total heterozygosity (H_T) across the loci SSR19, SSR28, and SSR86 in our study compared to those reported by Aguirre et al. (2017) in various tomato introductions. All markers in our study were highly informative, displaying a high level of

polymorphism with *PIC* values exceeding 0.5. Bredemeijer et al. (2002) reported *PIC* values of 0.40 when assessing 500 tomato varieties with microsatellites, while García-Martínez (2006) reported *PIC* values ranging from 0.035 and 0.775 using amplified fragment length polymorphism (AFLP) to evaluate tomato germplasm. Our PIC values were higher than those reported by some authors but like others, such as Meng et al. (2010) and Kwon et al. (2009). Notably, the markers SSR19, SSR28, SSR86, SSR92, and SSR248 used in our study were not polymorphic in a previous study by Rai et al. (2016) assessing genetic variation in tomato genotypes (*Solanum lycopersicum* L.) using SSR molecular markers. This discrepancy may be attributed to the greater genetic diversity present in tomatoes *S. l.* var. *cerasiforme* compared to cultivated tomatoes, with cultivated tomatoes representing only 5 % of the genetic diversity of wild tomatoes (Miller and Tanksley, 1990). Kwon et al. (2009) also noted high polymorphism in 28 tomato varieties using these markers.

We observed high genetic diversity, with an H_t value of 0.763, which was higher than the H_t values reported by Domingos (2011) in evaluating genetic diversity introductions from eight different locations in Mexico (0.32), in nine introductions from Ecuador (0.20), and in the evaluation of genetic diversity among different tomato varieties by Korir et al. (2014) (0.49). The H_o was lower than H_t, indicating the presence of inbreeding between individuals. The inbreeding index varied from -0.473 to 0.703 with markers Lega007, SSR248, SSR19, and SSR448 detecting a high number of heterozygotes.

A positive F_{is} value of 0.209 was recorded, indicating an inbreeding index and a deficit of heterozygotes. This deficit was attributed to non-random mating within the groups or localities, where individuals within the same group were more closely related than expected. The genetic variation analysis revealed that most of the genetic variation between

individuals stemmed from heterozygosity, as the variation within the individuals (26 %) exceeded that between individuals (10 %).

A high proportion of variation between populations was observed with an F_{st} value of 0.665, indicating significant genetic differentiation between populations according to Wright, (1984). This value was notably higher than the lowest F_{st} value of 0.3474 reported by Aguirre et al. (2017) in a previous study. The 64 % variation between populations aligned with the F_{st} value of 0.655, signifying a genetic differentiation of 65.5 % between populations. The observed differences between groups may be linked to the geographical distances of the collection sites and variations among individuals potentially influenced by an adaptative process as has been explained by Vilas et al. (2015), who demonstrated that a maximum allelic leads to a higher adaptative potential.

Tomato, *S. l.* var. *cerasiforme*, being self-compatible and autogamous, has some level of allogamy based on its geographical region, contributing to the observed variation between groups and within individuals in those groups.

The dendrogram displayed a clear separation between the groups, indicating genetic differentiation among them. Maltrata separation can be explained by the geographic separation of this valley from the other localities in the center of the state. In addition, genetic differentiation may be due to the adaptation of this group to a variable biogeographic area in climatic and orographic terms, since it includes characteristic climates of the Nearctic region and influence of the Neotropical region, it presents a soil of arid origin with little precipitation, but a favorable temperature for its development. Also, represents an important area for the tomatoes, since they were transported from the coasts of the Gulf of Mexico in Veracruz to Puebla (Jenkins, 1948).

These results were supported by the principal coordinate analysis (PCoa), which corroborated the findings from the cluster analysis. Aguirre et al. (2017) also found similar results, with clustering analysis revealing distinct groups based on genetic distance and identity in the *S. l.* var. *cerasiforme*.

The structure analysis detected a gene flow process, and all four plant groups identified by the STRUCTURE analysis showed admixture. These tomato populations exhibited a high capacity for adaptation due to the high allelic diversity and the low migration rate (*Nm*= 0.083). According to Caballero and García-Dorado (2013), allelic diversity is a predictor of long-term and total adaptation. It is explained by the long history of tomatoes in Mesoamerica. According to Long (2022), this species arrived in Mesoamerica as a spontaneous plant through natural dispersal by animals as birds, without human intervention. In the new environment, it found a favorable ecological niche for its development, and when man arrived in Mesoamerica a domestication process started

CONCLUSIONS

The present study revealed significant genetic diversity in the tomato groups analyzed. The highest genetic diversity was found in the central region of Veracruz. The populations exhibited a high capacity for adaptation, suggesting their ability to survive environmental changes.

Furthermore, some of the tomato populations in Veracruz showed admixture, indicating their potential value as germplasm for crop improvement programs. The SSR markers used in this study proved to be highly informative in assessing genetic diversity, as evidenced by the results obtained.

AUTHOR'S PARTICIPATION

GD sampling, molecular testing, and document writing; CMV sampling, molecular testing; RSL molecular testing, phylogenetic analysis, and document writing; PAM experimental design, statistical analysis; RNP experimental design, molecular testing, and document writing; MEGT phylogenetic analysis and document writing; JAHC experimental design, statistical analysis, document writing; JMG experimental design and document writing; ORLO experimental design, phylogenetic analysis, and document writing.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Emrique Laclette for providing us with access to the Laboratory of Genomic and Transcriptomic. We also thank Josue Lira for their assistance during our investigation stay at INECOL. Furthermore, we would like to extend our appreciation to CONAHCYT for their financial support through the student grant.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

Aguirre, N. C., López, W., Orozco-Cárdenas, M., Coronado, Y. M., Vallejo-Cabrera, F., Aguirre, N. C., López, W., Orozco-Cárdenas, M., Coronado, Y. M. and Vallejo-Cabrera, F. (2017). Use of microsatellites for evaluation of genetic diversity in cherry tomato. *Bragantia*, *76*(2), 220-228. https://doi.org/10.1590/1678-4499.116

Al Shaye, N., Migdadi, H., Charbaji, A., Alsayegh, S., Daoud, S., AL-Anazi, W. and Alghamdi, S. (2018). Genetic variation among Saudi tomato (*Solanum lycopersicum*

L.) landraces studied using SDS-PAGE and SRAP markers. *Saudi Journal of Biological Sciences*, *25*(6), 1007-1015. https://doi.org/10.1016/j.sjbs.2018.04.014

Benor, S., Zhang, M., Wang, Z. and Zhang, H. (2008). Assessment of genetic variation in tomato (*Solanum lycopersicum* L.) inbred lines using SSR molecular markers. *Journal of Genetics and Genomics*, *35*(6), 373-379. https://doi.org/10.1016/S1673-8527(08)60054-5

Blanca J, Cañizares J, Cordero L, Pascual L, Diez MJ, et al. (2012) Variation Revealed by SNP Genotyping and Morphology Provides Insight into the Origin of the Tomato. *PLOS ONE, 7*(10): e48198. https://doi.org/10.1371/journal.pone.0048198

Bredemeijer, G., Cooke, J., Ganal, M., Peeters, R., Isaac, P., Noordijk, Y., Rendell, S., Jackson, J., Röder, M., Wendehake, K., Dijcks, M., Amelaine, M., Wickaert, V., Bertrand, L. and Vosman, B. (2002). Construction and testing of a microsatellite database containing more than 500 tomato varieties. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *105*(6-7), 1019-1026. https://doi.org/10.1007/s00122-002-1038-6

Caballero, A. and García-Dorado, A. (2013). Allelic Diversity and Its Implications for the Rate of Adaptation. *Genetics*, *195*(4), 1373-1384. https://doi.org/10.1534/genetics.113.158410

Chetty, V. J., Ceballos, N., Garcia, D., Narváez-Vásquez, J., Lopez, W. and Orozco-Cárdenas, M. L. (2013). Evaluation of four *Agrobacterium tumefaciens* strains for the genetic transformation of tomato (*Solanum lycopersicum* L.) cultivar Micro-Tom. *Plant Cell Reports*, *32*(2), 239-247. https://doi.org/10.1007/s00299-012-1358-1

Délices, G., Otto, R., Nuñez Pastrana, R., Andrés Meza, P., Serna-Lagunes, R. and Gamez Pastrana, R. (2019). Biogeografía del tomate *Solanum lycopersicum* var. Cerasiforme (Solanaceae) en su centro de origen (sur de América) y de domesticación (México). *Revista de Biología Tropical*, *67*(4), 1023-1036. https://doi.org/10.15517/rbt.v67i4.33754

Délices, G., Leyva-Ovalle, O. R., Mota-Vargas, C., Nunez-Pastrana, R., Andres-Meza, P., & Herrera-Corredor, J. A. (2021). Morphological characterization of wild populations of Solanum lycopersicum var. cerasiforme in the tomato domestication area. *Emirates Journal of Food and Agriculture*, *33*(4), 303-313. https://doi.org/10.9755/ejfa.2021.v33.i4.2684

Domingos, J. P. (2011). Análisis de la variabilidad en las especies del subgénero *Eulycopersicon* más relacionadas con el tomate cultivado. [Tesis doctoral]. Editorial Universitat Politècnica de València. https://doi.org/10.4995/Thesis/10251/10914

Evanno, G., Regnaut, S. and Goudet, J. (2005). Detecting the number of clusters of individuals using the software Structure: A simulation study. *Molecular Ecology*, *14*(8), 2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

FAOSTAT. (2020). Statistics division. Food and Agriculture Organization of the United Nations. https://www.fao.org/faostat/en/#home

García-Martínez, S., Andreani, L., Garcia-Gusano, M., Geuna, F. and Ruiz, J. J. (2006). Evaluation of amplified fragment length polymorphism and simple sequence repeats for tomato germplasm fingerprinting: Utility for grouping closely related traditional cultivars. *Genome*, *49*(6), 648-656. https://doi.org/10.1139/g06-016

Goudet, J. F. S. T. A. T. (1995). FSTAT (version 1.2): a computer program to calculate F-statistics. Journal of heredity, 86(6), 485-486. https://doc.rero.ch/record/294716/files/86-6-485.pdf

He, C., Poysa, V. and Yu, K. (2003). Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon sculentum* cultivars. *Theoretical and Applied Genetics*, *106*(2), 363-373. https://doi.org/10.1007/s00122-002-1076-0

Jenkins, J.A. (1948). The origin of the cultivated tomato. Econ. Bot., 2, 379–392. https://doi.org/10.1007/BF02859492

Kaushal, A., Singh, A., and Jeena, A.S. (2017). Genetic diversity in tomato (Solanum lycopersicum L.) genotypes revealed by simple sequence repeats (SSR) markers. Journal of Applied and Natural Science, 9(2), 966-973. https://doi.org/10.31018/jans.v9i2.1305

Korir, N. K., Diao, W., Tao, R., Li, X., Kayesh, E., Li, A., Zhen, W. and Wang, S. (2014). Genetic diversity and relationships among different tomato varieties revealed by EST-SSR markers. *Genetics and Molecular Research*, *13*(1), 43-53. https://doi.org/10.4238/2014.January.8.3

Kwon, Y.-S., Park, S.-G. and Yi, S.-I. (2009). Assessment of genetic variation among commercial tomato (*Solanum lycopersicum* L) varieties using SSR markers and morphological characteristics. *Genes & Genomics*, *31*(1), 1-10. https://doi.org/10.1007/BF03191132

Lewis, P. O. and Zaykin, D. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from https://plewis.github.io/software/

Long, J. (2022). De tomates y jitomates en el siglo XVI. *Estudios de Cultura Náhuatl*, *25*, 239-252. https://nahuatl.historicas.unam.mx/index.php/ecn/article/view/78174

Mazzucato, A., Papa, R., Bitocchi, E., Mosconi, P., Nanni, L., Negri, V., Picarella, M. E., Siligato, F., Soressi, G. P., Tiranti, B. and Veronesi, F. (2008). Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. *Theoretical and Applied Genetics*, *116*(5), 657-669. https://doi.org/10.1007/s00122-007-0699-6

Meng, F-J., Xu, X-Y., Huang, F-L. and Li, J-F. (2010). Analysis of genetic diversity in cultivated and wild tomato varieties in Chinese market by RAPD and SSR. *Agricultural Sciences in China*, *9*(10), 1430-1437. https://doi.org/10.1016/S1671-2927(09)60234-0

Miller, J. C. and Tanksley, S. D. (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *TAG. Theoretical and Applied Genetics*, *80*(4), 437-448. https://doi.org/10.1007/BF00226743

Mohammadi, S. A. and Prasanna, B. M. (2003). Analysis of Genetic Diversity in Crop Plants—Salient Statistical Tools and Considerations. *Crop Science*, *43*(4), 1235-1248. https://doi.org/10.2135/cropsci2003.1235

Nei, M. (1973). Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences*, *70*(12), 3321-3323. https://doi.org/10.1073/pnas.70.12.3321

Nyadanu, D., Aboagye, L. M., Akromah, R., Osei, M. K. and Dordoe, M. B. (2014). Agromorphological characterisation of *Gboma* eggplant, an indigenous fruit and leafy vegetable in Ghana. *African Crop Science Journal*, *22*(4), 281-289. https://www.ajol.info/index.php/acsj/article/view/110423

Peakall, R. and Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, *28*(19), 2537-2539. https://doi.org/10.1093/bioinformatics/bts460

Pritchard, J., Stephens, M. and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945-959. https://doi.org/10.1093/genetics/155.2.945

Rai, G. K., Jamwal, D., Singh, S., Parveen, A., Kumar, R. R., Singh, M., Rai, P. K. and Salgotra, R. K. (2016). Assessment of genetic variation in tomato (*Solanum lycopersicum* L.) based on quality traits and molecular markers. *SABRAO Journal of Breeding & Genetics*, *48*(1): 80-89.

Rai, G. K., Kumar, R., Singh, A. K., Rai, P. K., Rai, M., Chaturvedi, A. K. and Rai, A. B. (2012). *Changes in antioxidant and phytochemical properties of tomato (Lycopersicon esculentum* Mill*.) Under ambient condition*. *44*, 667-667.

Seymour, G.B. and Rose, J.K.C. (2022). Tomato molecular biology – special collection of papers for molecular horticulture. *Mol Horticulture 2*, 21. https://doi.org/10.1186/s43897-022-00042-z

Shirasawa, K. and Hirakawa, H. (2013). DNA marker applications to molecular genetics and genomics in tomato. *Breeding Science*, *63*(1), 21-30. https://doi.org/10.1270/jsbbs.63.21

Vilas, A., Pérez-Figueroa, A., Quesada, H. and Caballero, A. (2015). Allelic diversity for neutral markers retains a higher adaptive potential for quantitative traits than expected heterozygosity. *Molecular Ecology*, *24*(17), 4419-4432. https://doi.org/10.1111/mec.13334

Wright, S. (1984). *Evolution and the Genetics of Populations, Volume 4: Variability Within and Among Natural Populations*. University of Chicago Press.

Xu, Y. (2010). *Molecular Plant Breeding*. CABI. https://doi.org/10.1079/9781845933920.0000

TABLES AND FIGURES.

Table 1. The observed number of alleles (***na***), effective number of alleles (***na***), observed heterozygosity (H_o), heterozygosity within populations (H_e), total heterozygosity (H_t), heterozygosity between populations (D_{st}), Polymorphic information content (P_{IC}) and Genetic differentiation coefficient (G_{st}), overall populations of *S. l.* var. *cerasiforme.*

Locus	***na***	n_e	H_o	H_e	H_t	D_{st}	P_{IC}	G_{st}
Lega007	9	4.65	0.644	0.7882	0.801	0.332	0.764	0.482
SSR28	8	3.77	0.1695	0.7375	0.678	0.497	0.695	0.772
Lega003	9	6.20	0.060	0.843	0.834	0.747	0.823	0.909
SSR92	7	3.90	0.109	0.747	0.771	0.609	0.713	0.817
SSR65	7	4.33	0.102	0.772	0.812	0.655	0.739	0.834
SSR248	9	6.03	0.663	0.838	0.825	0.355	0.815	0.445
SSR306	6	4.29	0.068	0.770	0.764	0.616	0.735	0.836
SSR19	11	7.88	0.432	0.877	0.863	0.497	0.861	0.631
SSR448	9	4.13	0.514	0.761	0.746	0.364	0.734	0.502
SSR218	7	3.55	0.068	0.721	0.736	0.572	0.685	0.810
LEta003	7	3.01	0.085	0.671	0.659	0.507	0.645	0.799
SSR86	6	3.73	0.102	0.735	0.749	0.651	0.690	0.889
LEaat002	6	5.03	0.068	0.805	0.817	0.652	0.771	0.836
LEaat001	7	3.14	0.068	0.685	0.632	0.519	0.653	0.846
Mean	7.714	4.546	0.2251	0.768	0.763	0.541	0.7374	0.743

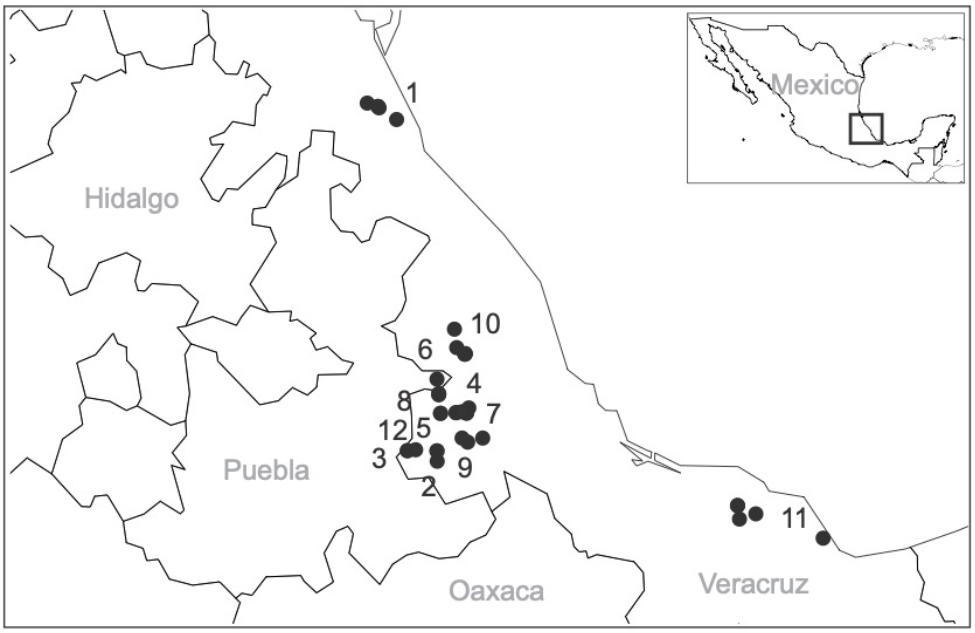


Figure 1. Localities where the species have been collected in Veracruz, México (1-Tuxpan, 2-Zongolica, 3-Tenejapan, 4-Ixhuatlán, 5-Palenque, 6-Mahuixtlán, 7-Ocotitlán, 8-Coscomatepec, 9-Córdoba, 10-Xalapa, 11-Pajapan, and 12-Maltrata).

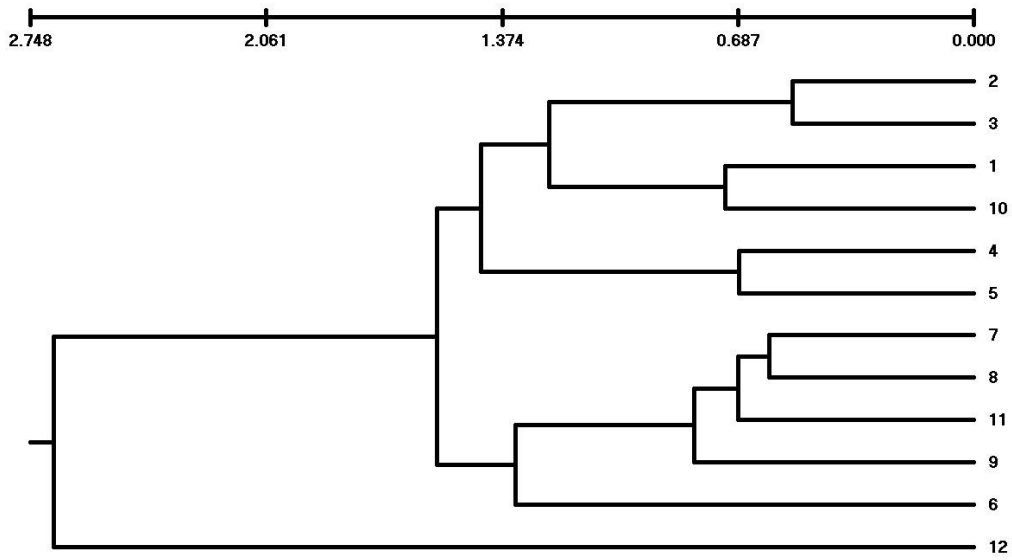
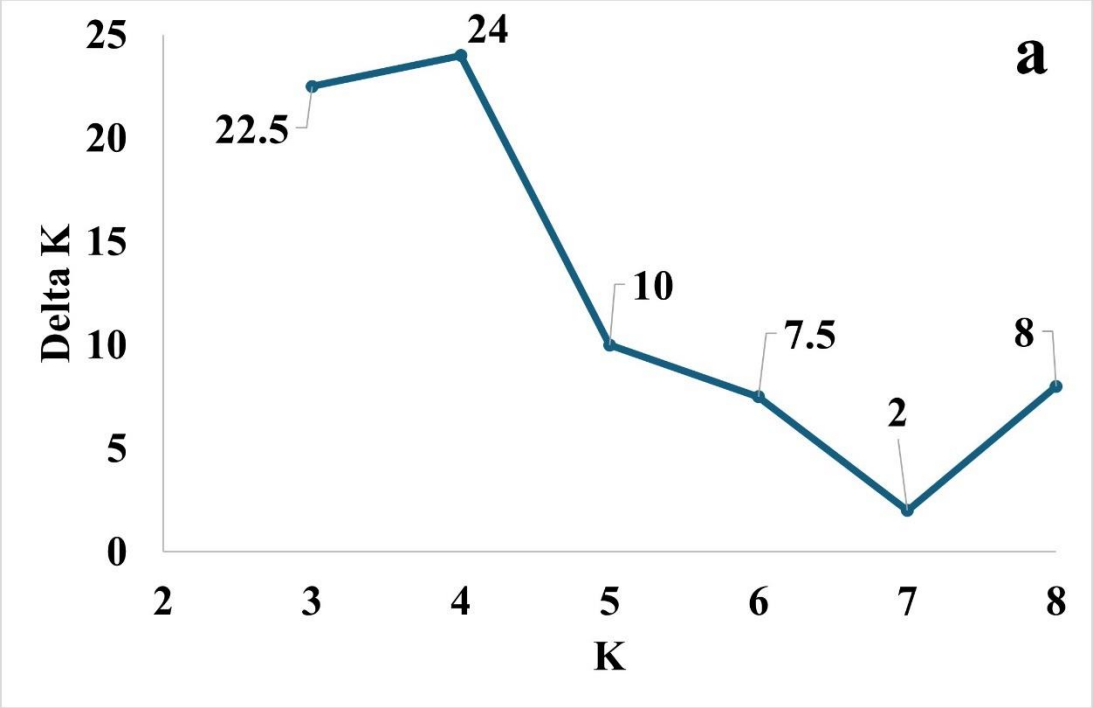


Figure 2. The UPGMA clustering method of twelve groups of tomatoes, evaluating genetic diversity using 14 SSRs markers.



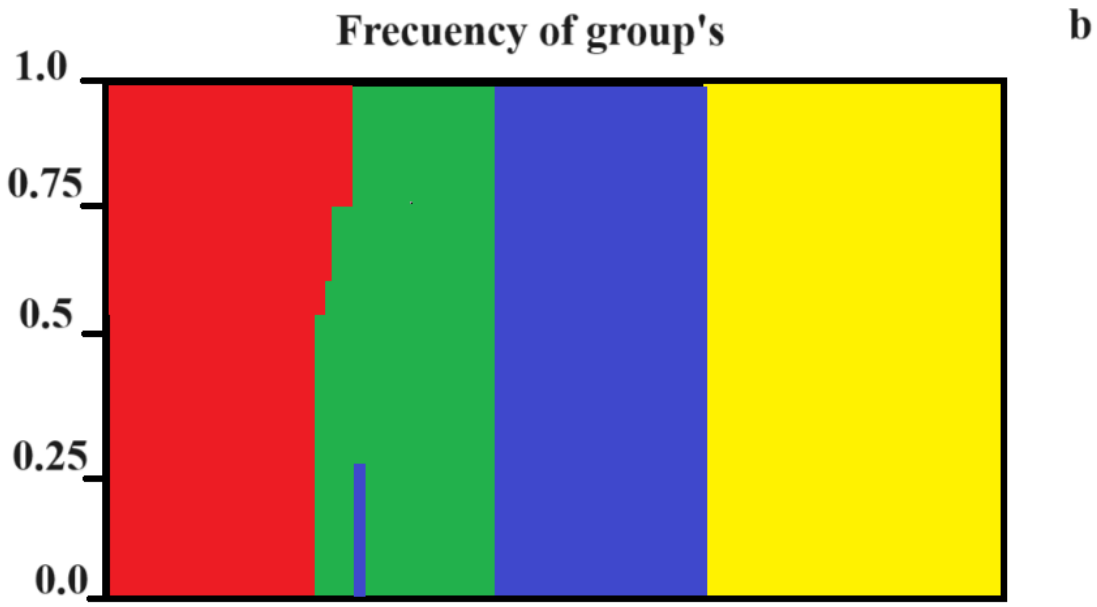


Figure 3. (a) Delta *K* values for different numbers of populations assumed (*K*) in the STRUCTURE analysis. (b) Classification of twelve group of tomatoes from different localities of Veracruz into four populations K= 4 using STRUCTURE v2.3.4

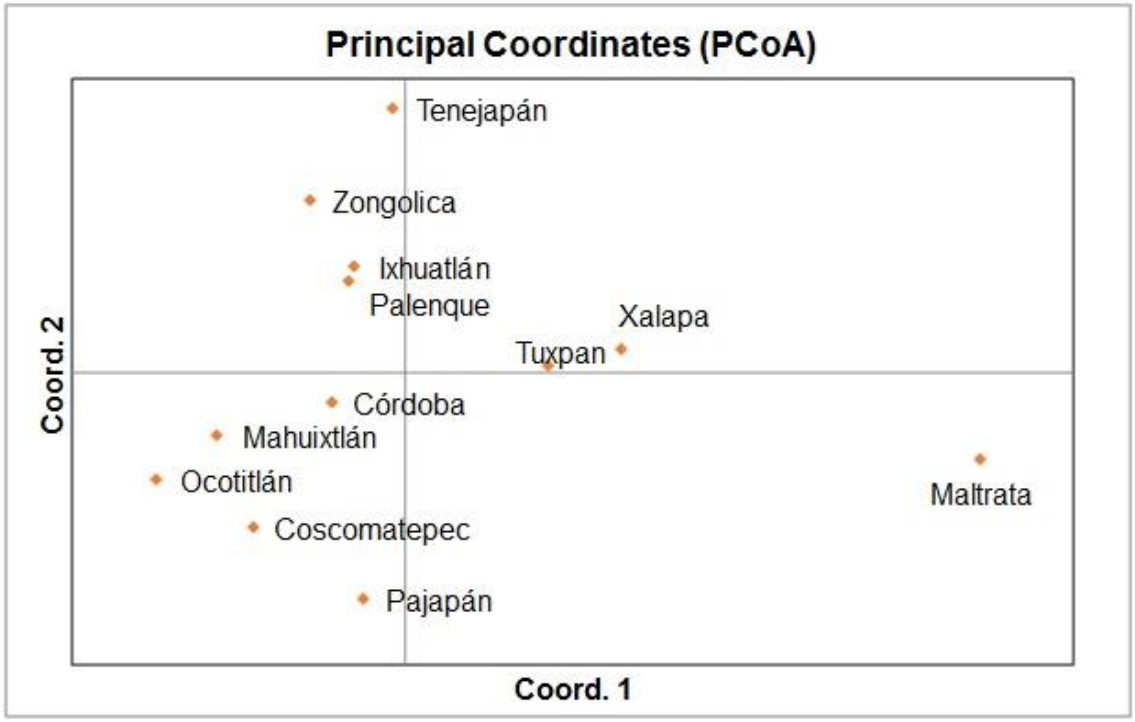


Figure 4. Principal coordinate analysis (PCoA) showing the detection of four groups, the first coordinate explaining 28.43 % and the second coordinate 23.36 % of the variation.