Environment and Ecology 41 (4D) : 3060—3065, October—December 2023
Article DOI: https://doi.org/10.60151/envec/YLSS4838
ISSN 0970-0420

Assessment of Genetic Divergence through Principal Component Analysis and Clustering in Tomato Germplasm Accessions

Greetty Williams, Y. Anbuselvam

Received 16 August 2023, Accepted 3 November 2023, Published on 29 December 2023

ABSTRACT

The base material of this study comprises of 104 tomato accessions including local landraces, varieties and germplasm collections. The collected tomato accessions were evaluated using 13 quantitative traits by Principal Component Analysis (PCA) and Hierarchial clustering. PCA was done to quantify diversity among the germplasm accessions and also the contribution of individual traits towards diversity. In our study, only the first four (PC1, PC2, PC3 and PC4) of the thirteen principal components yielded eigen value more than one indicating the greater influence of identified traits under study. The first six PCs accounts for 84% of variability whereas, PC1 exhibited 41% of total variability. Cluster analysis aids to classify the genotypes based on the grouping pattern of the accessions under evaluation. According to the dendrogram obtained, cluster analysis grouped 104 tomato accessions into two significant clusters. The first cluster consists of 16 genotypes whereas, the second cluster consists of 88 genotypes. Among the genotypes used in this study EC617055, EC617061, EC638302, Periakulam local and EC631390 were found to be best performing in terms of yield and quality. These accessions can be used as a base material in future breeding programs.

Keywords Clustering, Diversity, Germplasm, Principal component analysis, Variability.

INTRODUCTION

Tomato belongs to the diverse Solanaceae family which includes more than three thousand species. In the early sixteenth century, they were considered ornamental plants (Bauchet and Causse 2012), but within 200 years, they became a precious crop with greater social and economic values. Domesticated tomato (*Solanum lycopersicum*) and its 12 wild relatives are the members of *Lycopersicon* clade (Kamenetzky *et al*. 2010). They are natives of the Andean region. The members of this clade were found in wide range of ecological conditions which contributed towards diversity of wild species. This clade also serves as a pre-eminent model in species variation studies and genetic studies for ripening process (Klee and Giovannoni 2011). *Solanum lycopersicum* is cosmopolite in nature and its spread throughout the world.

Yield increment has been the major objective for any breeding program. As a result of rigorous breeding programs, development of high yielding genetically uniform varieties gained attention during early 20th century (Ceccarelli 2012). Artificial selection led to reduction in genetic diversity. There was a huge

Greetty Williams[1],Y. Anbuselvam[2]*

[2]Professor,
[1, 2]Department of Plant Breeding and Genetics, Annamalai University, Chidambaram 608002, Tamil Nadu, India

Email : yanbuselvam@gmail.com
*Corresponding author

relegation of landraces which created a greater void in the genetic diversity of tomato (Farinon *et al.* 2022). But in modern breeding program, the main objective is to get back to the crop wild relatives and ancestors to employ the diversity lost during domestication (Gur and Zamir 2004).

Success of a crop improvement program relies on the source of the parental material. Wild relatives and germplasm accessions serves as a base material for any breeding program (Casañas *et al.* 2017). Crosses between wild and cultivated types generated novel phenotypic diversity. In tomato, they are the source for resilience to varied environmental stress conditions, low input responsiveness, distinctive nutraceutical, nutritional, organoleptic, cultural and historical traits (Ramirez-Villegas *et al.* 2022). Due to this uniqueness, tomato heirlooms and landraces are in breeder's spotlight now and efforts have been taken to breed flavorsome and nutritious tomato fruits. Consequently, studies aiming to characterize tomato germplasm accessions are increasingly gaining attention (Athinodorou *et al.* 2021). The present investigation is aimed to assess genetic diversity in tomato germplasm accessions.

MATERIALS AND METHODS

One hundred four tomato accessions (including germplasm accessions collected from Gene Bank, National Bureau of Plant Genetic Resources, local landraces and a few varieties) served as a base material for this investigation. The acquired seeds were sown in protrays filled with an admixture of organically enriched compost and topsoil. Nursery management practices were carried out, which aided in the production of vibrant seedlings. Seedlings were transplanted on the 30th day after sowing. An augmented design with fifteen blocks and three controls was formed for morphological assessment. Seedlings were planted with a spacing of 60×45 at the plant breeding farm, Department of Plant Breeding and Genetics, Annamalai University, Chidambaram, from January to May 2022.

All standard horticultural practices for tomato production were taken up to raise the crop. Thirteen traits, viz., plant height, thickness of pericarp, size of core, pedicel length, pedicel scar, fruit length, fruit width, plant yield, fruit weight, days to fifty percent flowering, number of locules, number of days to first picking and wilt susceptibility were observed from five randomly selected plants in each accession based on the tomato descriptors IPGRI (1996). In order to categorize variation and the contribution of traits towards total variation, the collected phenotypic data is subjected to Principal Component Analysis and Hierarchical cluster analysis following Ward's method was done using R studio software version (v1.4.1717) to find the association among accessions. The PCA–biplot was obtained using "ggplot2" (Wickham *et al.* 2016), "Factoextra" (Kassambara and Mundt 2017) and "FactomineR" (Lê *et al.* 2008) packages of R.

RESULTS AND DISCUSSION

The Principal Component Analysis is a powerful tool to identify minimum components which explains maximum variability (Shoba *et al.* 2019). It also quantifies the significance of each dimensions and displays the variability in a data set visually appealing (Lakshmi *et al.* 2022). Practically, PCA is a vital tool used to choose parental lines for hybridization (Ahmadizadeh and Felenji 2011). In our study, thirteen traits were subjected to PCA and thirteen principal components have been obtained. (Table 1) presents the Eigen value and percentage of variance explained by each component. Principal Components having Eigen values more than one and percentage of variance more than four can be considered as main

Table 1. Eigen value and percentage of variance.

Principal component	Eigen value	Percentage of variance	Cumulative percentage of variance
PC1	5.332	41.019	41.019
PC2	1.665	12.809	53.828
PC3	1.253	9.636	63.464
PC4	1.112	8.552	72.016
PC5	0.887	6.822	78.838
PC6	0.692	5.321	84.159
PC7	0.622	4.781	88.94
PC8	0.438	3.366	92.306
PC9	0.399	3.068	95.374
PC10	0.296	2.278	97.652
PC11	0.179	1.378	99.03
PC12	0.126	0.97	100
PC13	0	0	100

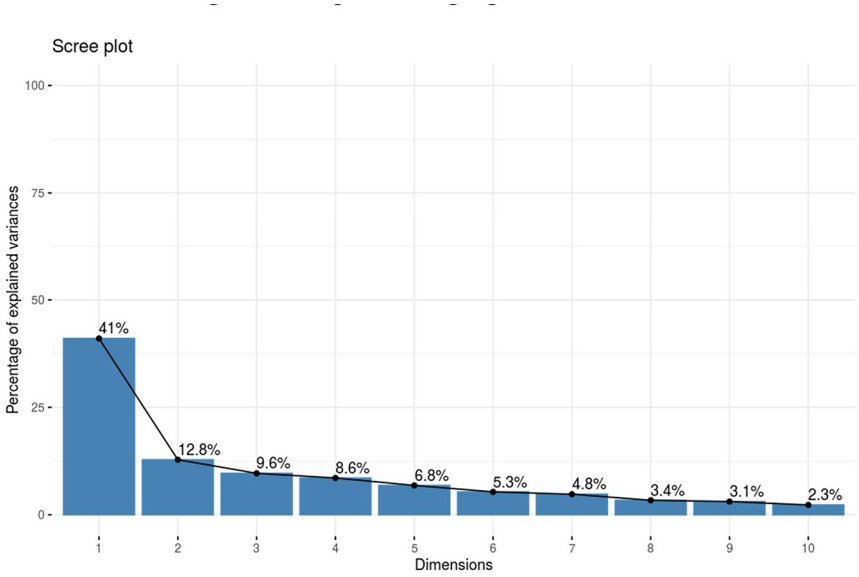

Fig. 1. Scree plot showing Eigen value variation.

PC (Sao *et al.* 2019). PCs with Eigen value greater than one can be selected (Shoba *et al.* 2019). Only the first four (PC1, PC2, PC3 and PC4) of the thirteen principal components yielded Eigen value more than one indicating the greater influence of identified traits in the phenotype of the genotypes under study (Nachimuthu *et al.* 2014). The scree plot (Fig.1) aids in categorizing variances for the first ten principal component axes. The first six PCs accounts for 84% of variability whereas, PC1 exhibited 41% of total variability. The factor loadings explained by first five principal components are indicated in (Table 2).

Table 2. Factor loadings explained by first five principal components. PH – Plant height, TP - Thickness of pericarp, SC - Size of core, PL - Pedicel length, PS - Pedicel scar, FL - Fruit length, FW - Fruit width, PY - Plant yield, FW - Fruit weight, DFF - Days to fifty percent flowering, NOL - Number of locules, NODFP - Number of days to first picking, WS - Wilt susceptibility.

Variables	PC1	PC2	PC3	PC4	PC5
PH	-0.003	0.242	-0.692	0.173	-0.068
TP	0.244	-0.091	0.22	0.501	0.270
SC	0.297	-0.211	-0.011	0.052	-0.305
PL	0.208	-0.341	-0.233	0.200	0.454
PS	0.230	-0.216	-0.405	0.144	0.270
FL	0.353	0.182	0.223	0.201	0.057
FW	0.377	-0.068	0.160	-0.035	-0.170
PY	0.375	-0.054	0.007	-0.058	-0.195
FWT	0.370	-0.004	0.188	-0.038	-0.154
DFF	-0.303	-0.457	0.131	0.247	-0.125
NOL	0.160	-0.445	-0.310	-0.292	-0.400
NODFP	-0.303	-0.457	0.131	0.247	-0.125
WS	0.060	-0.263	0.113	-0.635	0.515

In the present study, fruit length, fruit width, plant yield and fruit weight were the contributing traits for PC1. Higher the coefficient, either positive or negative the discrimination of accessions will be more effective. In PC1, yield and yield attributing traits like fruit length, fruit width, plant yield and fruit weight contributed more towards the total variation. Similar pattern of contribution by yield attributing traits in PC1 was also reported by Sanni *et al.* (2012), Ojha *et al.* (2017). Many authors (Mahesha *et al.* 2006, Prashanth *et al.* 2008, Ene *et al.* 2022) reported the importance of traits like fruit weight, fruit yield per plant in contributing towards genetic diversity in tomato. They also suggested that these traits have wider scope in tomato yield enhancement by direct selection. Desirable traits coming together in single principal component has the tendency to cling together which offers chance for their utilization in crop breeding.

Plant height, days to fifty percent flowering and number of days to first picking exhibited negative

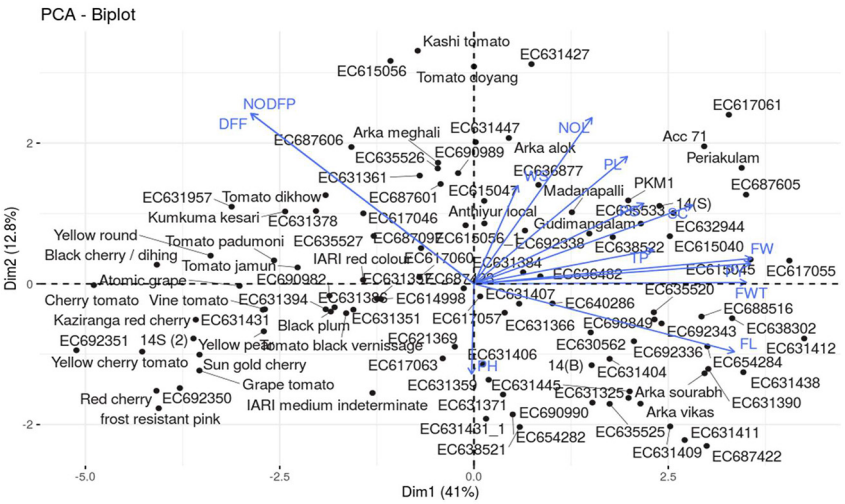

Fig. 2. Distribution of genotypes across two components.

contribution in PC1. Nearly 54% of variation was explained by PC1 and PC2 which indicates a strong relationship between the traits under study (Lakshmi *et al.* 2022). In PC2 most of the traits displayed negative contribution except plant height and fruit length. Plant height, size of the core, pedicel length, pedicel scar and number of locules exhibited positive contribution whereas, the other traits showed negative contribution towards PC3. The biplot of PC1 and PC2 clearly depicts the interaction among traits and also with each genotype (Fig. 2). Vector length depicts the contribution of various traits towards total divergence. Lengthier the vector greater will be its contribution towards diversity. In this study, traits like days to fifty percent flowering and number of days to first picking showed long vector length indicating its higher contribution towards diversity followed by the yield and yield attributing traits such as fruit length, fruit width, fruit weight.

Angle between the trait vectors decides the direction of correlation between the traits (Bhargava *et al.* 2021). The genotypes that are present in the opposite direction of the yield and yield attributing vectors are considered as poor performers (Sao *et al.* 2019). In this present investigation, out of thirteen traits under study days to fifty percent flowering and number of days to first picking showed negative correlation towards plant yield. The genotypes that are present along in the same quadrant of the yield attributing trait vectors are considered as good yielders and the genotypes that are located in opposite direction to these vectors can be considered as inferior genotypes for these traits (Lakshmi *et al*. 2022). In our study, most of the genotypes present in the left side of the biplot is overlapping and this indicates the less variability between the genotypes (Ojha *et al.* 2017).

Hierarchial clustering

Cluster analysis aids to classify the genotypes based on the grouping pattern of the accessions under evaluation (Nankar *et al.* 2020). Hierarchial cluster analysis using thirteen quantitative traits is presented in (Fig. 3). According to the dendrogram obtained, cluster analysis grouped 104 tomato accessions into two significant clusters. The first cluster consists of 16 genotypes whereas, the second cluster consists of 88 genotypes. The second cluster is the largest. The members of cluster I are high yielding with high mean values for average fruit weight and individual plant yield. In similar studies with *Solanum surattense* by Dheebisha *et al. (*2023) and in tomato by Evgenidis *et al.* (2011), genotypes with high yield and yield components aggregated in a single cluster, this result is in consonance with the clustering pattern of genotypes in the present study. Cluster I have two subclusters, the first subgroup IA has 15 genotypes and the next

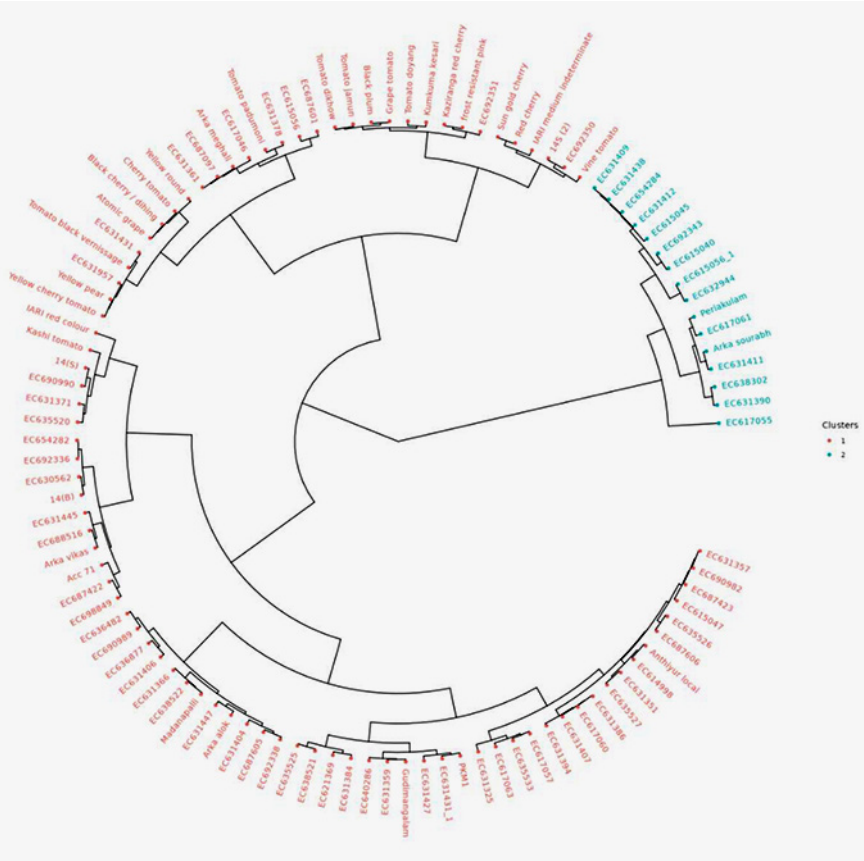

Fig. 3. Hierarchial clustering of 104 genotypes.

subgroup IB has only one genotype (EC617055) and it is a solitary subcluster. This genotype is high yielder and ranks first in average fruit weight (210 g) among the 104 accessions studied. This cluster contains genotypes with better agronomic characteristics and yield performances hence, selection will be effective when yield is the target.

Cluster II has two subgroups, the first subgroup IIA has 32 genotypes, whereas the next subgroup is the largest and has 56 genotypes in it. Cluster IIA comprises of genotypes producing small sized fruits, most of the members of this cluster are cherry type. In this study, one genotype belonging to *Solanum pimpinellifolium* is included and this genotype is grouped along with other cherry type tomatoes in this cluster. They are accommodated in the same cluster because cherry tomato types where the genetic admixture of cultivated accessions and *S. pimpinellifolium* (Peralta and Spooner 2006). Cluster IIB comprises of genotypes with medium sized fruits and low to medium yielding ability. In this study, the grouping pattern of tomato accessions is based on their agronomic and yield performances. Grouping pattern is not in the basis of the source, origin or the geographical distribution as the accessions were distributed randomly in the clusters. Similar results were also reported by Ene *et al.* (2022) in tomato. This pattern of distribution is the sign for broad genetic base of the tomato accessions (Vargas *et al.* 2020).

CONCLUSION

Multivariate analysis aids in quantifying diversity among the germplasm accessions and also the contribution of individual traits towards diversity. PCA

helps in ranking the genotypes based on the PC scores. EC617055, EC617061, EC638302, Periakulam local and EC631390 are best performing genotypes in terms of yield and quality. These accessions can be used as a base material in future breeding programs. From the present study, it is clearly evident that cluster analysis is an effective and efficient tool to assort genotypes based on their yield performances. It also provides an authentic foundation in selecting base materials for breeding programs.

REFERENCES

Ahmadizadeh M, Felenji H (2011) Evaluating diversity among potato cultivars using agro-morphological and yield components in fall cultivation of Jiroft area. *Am J Environ Sci* 11 : 655—662.

Athinodorou F, Foukas P, Tsaniklidis G, Kotsiras A, Chrysargyris A, Delis C, Kyratzis AC, Tzortzakis N, Nikoloudakis N (2021) Morphological diversity, genetic characterization, and phytochemical assessment of the cypriot tomato germplasm. *Plants* 10 : 1698.

Bauchet G, Causse M (2012) Genetic diversity in tomato (*Solanum lycopersicum*) and its wild relatives. *Genetic Diversity* Pl 8 : 134—162.

Bhargava K, Shivani D, Pushpavalli S, Sundaram R, Beulah P, Senguttuvel P (2021) Genetic variability, correlation and path coefficient analysis in segregating population of rice. *Electron J Pl Breed* 12 : 549—555.

Casañas F, Simó J, Casals J, Prohens J (2017) Toward an evolved concept of landrace. *Front Pl Sci* 8 : 145.

Ceccarelli S (2012) Landraces: Importance and use in breeding and environmentally friendly agronomic systems. Agrobiodiversity conservation: Securing the diversity of crop wild relatives and landraces. CABI Wallingford UK, pp 103—117.

Dheebisha C, Nalina L, Rajamani K, Geethanjali S, Boopathi NM, Chandrakumar K, Reddy RN (2023) Genetic variability, association and path analysis for yield and fruit quality components in yellow-berried nightshade (*Solanum surattense*). *Electron J Pl Breed* 14 : 419—428.

Ene CO, Abtew WG, Oselebe HO, Ozi FU, Ikeogu UN (2022) Genetic characterization and quantitative trait relationship using multivariate techniques reveal diversity among tomato germplasms. *Food Sci Nutri* 10 : 2426—2442.

Evgenidis G, Traka-Mavrona E, Koutsika-Sotiriou M (2011) Principal component and cluster analysis as a tool in the assessment of tomato hybrids and cultivars. *Int J Agron.*

Farinon B, Picarella ME, Siligato F, Rea R, Taviani P, Mazzucato A (2022) Phenotypic and genotypic diversity of the tomato germplasm from the Lazio region in central Italy, with a focus on landrace distinctiveness. *Front Pl Sci* 13: In press

Gur A, Zamir D (2004) Unused natural variation can lift yield barriers in plant breeding. *PLoS Biol* 2 : 245.

IPGRI (1996) Descriptors for Tomato (*Lycopersicon* spp.). Bioversity International.

Kamenetzky L, Asís R, Bassi S, de Godoy F, Bermudez L, Fernie AR, Van Sluys MA, Vrebalov J, Giovannoni JJ, Rossi M (2010) Genomic analysis of wild tomato introgressions determining metabolism-and yield-associated traits. *Pl Physiol* 152 : 1772—1786.

Kassambara A, Mundt F (2017) Factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.

Klee HJ, Giovannoni JJ (2011) Genetics and control of tomato fruit ripening and quality attributes. *Annu Rev Genet* 45 : 41—59.

Lakshmi M, Shanmuganathan M, Jeyaprakash P, Ramesh T (2022) Genetic variability and diversity analysis in selected rice (*Oryza sativa* L.) varieties. *Electron J Pl Breed* 13 : 959—966.

Lê S, Josse J, Husson F (2008) FactoMine R: An R package for multivariate analysis. *J Stat Softw* 25 : 1—18.

Mahesha D, Apte U, Jadhav B (2006) Genetic variability in tomato (*Lycopersicon esculentum* Mill.). *Res* 7 : 771.

Nachimuthu VV, Robin S, Sudhakar D, Raveendran M, Rajeswari S, Manonmani S (2014) Evaluation of rice genetic diversity and variability in a population panel by principal component analysis. *Ind J Sci Technol* 7 : 1555—1562.

Nankar AN, Tringovska I, Grozeva S, Ganeva D, Kostova D (2020) Tomato phenotypic diversity determined by combined approaches of conventional and high-throughput tomato analyzer phenotyping. *Plants* 9 : 197.

Ojha G, Sarawgi A, Sharma B, Parikh M (2017) Principal component analysis of morpho-physiological traits in ricegermplasm accessions (*Oryza sativa* L.) under rainfed condition. *Int J Chem Studies* 5 : 1875—1878.

Peralta IE, Spooner DM (2006) History, origin and early cultivation of tomato (Solanaceae). *Genet Improvement Solanaceous Crops* 2 : 1—27.

Prashanth S, Jaiprakashnarayan R, Ravindra M, Madalageri M (2008) Genetic divergence in tomato (*Lycopersicon esculentum* Mill.). *Asian J Hort* 3 : 290—292.

Ramirez-Villegas J, Khoury CK, Achicanoy HA, Diaz MV, Mendez AC, Sosa CC, Kehel Z, Guarino L, Abberton M, Aunario J (2022) State of *ex situ* conservation of landrace groups of 25 major crops. *Nat Pl* 8 : 491—499.

Sanni K, Fawole I, Ogunbayo S, Tia D, Somado E, Futakuchi K, Sié M, Nwilene F, Guei R (2012) Multivariate analysis of diversity of landrace rice germplasm. *Crop Sci* 52 : 494—504.

Sao R, Saxena RR, Sahu PK (2019) Assessment of genetic variation and diversity in rice germplasm based on principal component analysis. *J Pl Dev Sci* 11 : 725-—730.

Shoba D, Vijayan R, Robin S, Manivannan N, Iyanar K, Arunachalam P, Nadarajan N, Pillai MA, Geetha S (2019) Assessment of genetic diversity in aromatic rice (*Oryza sativa* L.) germplasm using PCA and cluster analysis. *Electron J Plant Breed* 10: 1095—1104.

Vargas JEE, Aguirre NC, Coronado YM (2020) Study of the genetic diversity of tomato (*Solanum* spp.) with ISSR markers. *Rev Ceres* 67 : 199—206.

Wickham H, Chang W, Wickham MH (2016) Package 'gg plot 2'. Create elegant data visualizations using the grammar of graphics. *Version* 2 : 100—189.