DOI : 10.58993/ijh/2025.82.1.4

Indian J. Hortic. 82(1), March 2025: 22-27

Principle component and cluster analysis for heat tolerance in tomato under open condition for yield and seed production

Varun B.H.[*], Rajinder Singh and S.K. Jindal

Department of Vegetable Science, Punjab Agricultural University, Ludhiana 141004, Punjab

ABSTRACT

The nature and magnitude of genetic divergence of 76 genotypes of tomato under heat stress in open condition was estimated using R-Studio. The genotypes under the study showed considerable amount of diversity for all the traits. The first three components of Principle Component Analysis (PCA) explained the maximum variability present among the traits studied (40 percentage of total phenotypic variability). The PC_1, PC_2 and PC_3 components were responsible for 16.63, 11.53 and 11.05% of phenotypic variability, respectively. Based on the diversity analysis the genotypes under investigation were divided into four diverse and genetically distinct clusters. Cluster-1 (50) and Cluster-4 (19) had majority of genotypes under them whereas Cluster-2 (2) and Cluster-3 (5) had smaller number of genotypes under them. The study concluded that the genotypes used had considerable amount of phenotypic-variability for heat tolerance under open condition for yield as well as seed production. The traits such as yield per plant, seed yield per plant, number of fruits per plant, titratable acidity, and plant-height contributed a major portion to the phenotypic-variability among all the traits in this study for heat-tolerance.

Key words: Principle Component Analysis, cluster analysis, genotypes, heat tolerance.

INTRODUCTION

The tomato (*Solanum lycopersicum* L.), commonly known as the “Poor man’s orange” and a member of the Solanaceae family, is a globally significant vegetable crop and the second most consumed vegetable after potatoes (FAOSTAT, 5). In India, Madhya Pradesh leads in tomato cultivation with the largest area (1.17 million ha) and highest production (34.98 million tonnes), while Andhra Pradesh records the highest productivity (NHB, 10). It is abundant in vitamin K, organic acids, potassium, folate and vitamin C. Tomato stands first in processing among vegetables and the products of processing like sauce, ketchup, puree, paste, juice, and chutney are in great demand throughout the year. Tomato is a typical sub-tropical to tropical vegetable crop, though tomato plant can grow under a wide range of temperature, it thrives well under 10-30°C for optimum growth and production it requires about 21-24°C for lycopene development but the temperatures above 35°C is found to have undesirable effect on its growth and production. Above 40°C there is no development of any pigment in fruits and also fruits suffer from disorders (Xu *et al*., 15). High temperatures above 35°C along with high light-intensity and dry winds during hotter months of the year leads to drying of stigma, deformed flowers, and extreme flower drop resulting in poor fruit set, pre-mature fruit drop, sunscald and many physiological disorders (Sánchez-Rodríguez *et al*., 11). Flowering phase is most sensitive to high-temperature; failure of fertilization is due to reduction in pollen-viability and decrease in pollen-germination percentage and failure in pollen tube elongation. Seed yield is drastically reduced under high temperature due to pollen number and viability reduction and loss of stigmatic receptivity, flower abortion and blossom drop (Iwahori, 7). In the context of heat tolerance, correlation can help to access the degree of association between heat-related traits such as yield, seed-yield, lycopene-content and other relevant traits. Principle Component Analysis (PCA) can provide insights into the relative importance of different traits or markers in distinguishing between heat-tolerant and heat-sensitive tomato genotypes. It can also identify potential correlations or trade-offs between different traits related to heat-stress tolerance. Clustering analysis can provide valuable information about the genetic diversity within the tomato genotypes and reveal potential relationships among genotypes with similar heat-stress tolerance profiles. Combining PCA and clustering analysis can provide a comprehensive understanding of the genetic variability in tomato genotypes for heat-stress tolerance. Heat tolerance in tomato is genetically controlled trait of the plant, which is a result of significant genotype – environment interaction (Kumar *et al*., 9). Since tomato cultivars vary in their heat tolerance, breeding for heat-resistant varieties

---

[*] Corresponding author: varunrcbian@gmail.com

remains a key focus in tomato cultivation research. Additionally, management practices such as irrigation, shading, and optimal planting time selection can help mitigate the adverse effects of heat stress on tomato growth and yield.

MATERIALS AND METHODS

The present study was conducted at the Department of Vegetable Science, Punjab Agricultural University, Ludhiana. The study included 76 genotypes, including 4 check varieties: Punjab-Upma, Punjab-Ratta, Punjab-Chhuhara and OS-213 (Table 1). The nursery was sown in January and the crop was transplanted during March for two years 2020-2021 respectively. It has sandy-loamy soil, extremely hot, dry summers, and temperatures that range from 25° to 40°C during the months of March and June. During the first week of March, three replications of the seedlings were transplanted to the experimental field using a RCBD following the cultural operations recommended by PAU. The findings regarding the observations on Days to first anthesis (NDA), total number of fruits per plant, total yield per plant (kg), average fruit weight (g)(AFW), fruit shape index (FSI), plant-height (cm) (30 days and 60 days, respectively) (PH), pollen-viability (%); the test was conducted three times at 15-day intervals (30, 45 and 60-days) (PV), and total soluble solids of fruit (°Brix) (TSS), number of locules (LOC), thickness of pericarp (mm) (PCT), lycopene content (g/100 g)(LYC), titratable acidity (mg/ml)(TA), dry-matter content (DM)(%), number of seeds per fruit (NOS), seed yield per plant (g) (SY), and 1000 seed weight (g) (HSW). Germination (%), ISTA (6).

In Principle Component Analysis (PCA), only the components with more than 1 Eigen-value are used. The variability is broken down into various principle components based on the variation that each principle component explains (Bhattarai *et al*., 3). The hirechacial clustering method is used in this study for grouping of genotypes. The recorded observations were analyzed using the R-studio software and DARwin.

RESULTS AND DISCUSSION

To identify the variability source structure in the data variability, PCA is used. Additionally, it establishes how much the observed trait contributes to overall variability (Table 2). Based on Eigen values, PCA breaks down the total number of phenotypic variables into smaller Principle Components (PC's). The first two PCs are typically known to explain the greatest amount of phenotypic variability in the total variability present. We are using the third PC to interpret the results of the current study because the first two PCs only explain only a small amount of the phenotypic variability.

Table 1. Tomato genotypes used for the study including 4 varieties as check, *viz.*, Punjab Upma, Punjab Chhuhara, Punjab Ratta and Pusa Sadabahar.

No-575	MMZ-743
Rani	MLT-742
Sakata-3201	MTZ-744
Sakata Keshav	MMR-745
Samrudhi	MUC-746
Satyajit	NHP-761
Shivaji	NHP-762
US-1143	NOP-763
PAU-1021	NNF-764
AS-501	NWZ-765
Anabelle	NSP-767
Beltsvilllle-67B-33-1	NSP-768
BFT-525	NSP-766
BFT-523	NSP-769
BFT-524	PSZ-801
BGV-3010	PLRT-5
BS-521	PLRT-6
BT-522	PAU-120
Campbell-22	Pusa-Sadabahar
CH-221	RNI-842
CB-541	RMZ-843
CLN-2679E	Roma
CLN-37-8-1	SRD-864
CR-542	SMZ-865
CR-543	SHV-866
PAU-114	TPA-882
DVRT-2	TPA-883
EC-971784	TPA-881
PAU-2381	TPA-884
IPNIL-145	TPA-885
ISS-663	TLH-886
IPT-664	USP-902
KS-701	VPA-921
KS-702	VUZ-922
LTZ-1	Punjab-Upma
LST-17	Punjab-Chhuhara
LST-36-1	Punjab-Ratta
LST-65-2	OS-213

Table 2. PCA-variability components and contribution to variability of each principal components to the traits based on PCA loadings.

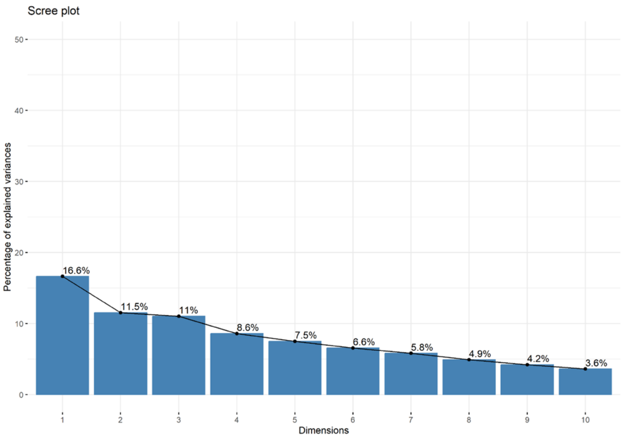
PCA	Trait	PC1	PC2	PC3	S.D	E.v
Dim.1	No. of days to anthesis	0.07443	0.33542	0.06201	1.824	3.33
Dim.2	Pollen-viability (%) (30-days)	0.14916	0.23383	0.02673	1.519	2.31
Dim.3	Pollen-viability (%) (45-days)	0.24718	0.10273	0.03353	1.485	2.21
Dim.4	Pollen-viability (%) (60-days)	0.13968	0.21011	0.00744	1.310	1.72
Dim.5	Plant height (cm) (30-days)	0.24253	0.19609	0.00788	1.225	1.50
Dim.6	Plant height (cm) (75-days)	0.25185	0.33365	0.16022	1.145	1.31
Dim.7	Fruit shape index	0.14353	0.32516	0.08881	1.078	1.16
Dim.8	No. of locules	0.02357	0.24877	0.00211	0.990	0.98
Dim.9	Pericarp thickness (cm)	0.02249	0.04095	0.52944	0.917	0.84
Dim.10	TSS (ºBrix)	0.33567	0.21034	0.27796	0.849	0.72
Dim.11	Titratable acidity (mg/100 ml)	0.35105	0.02606	0.03363	0.815	0.66
Dim.12	Lycopene (mg/100 g)	0.28598	0.14797	0.25180	0.777	0.60
Dim.13	Dry matter (%)	0.18362	0.33838	0.04097	0.711	0.51
Dim.14	No. of fruits	0.34606	0.12384	0.36018	0.700	0.49
Dim.15	Yield (kg/plants)	0.24652	0.36370	0.09572	0.627	0.39
Dim.16	Av. fruit weight (g)	0.07530	0.10945	0.53772	0.574	0.33
Dim.17	No. of seeds per fruit	0.24052	0.04837	0.15715	0.551	0.30
Dim.18	Seed yield per plant (g/plant)	0.34233	0.22942	0.23304	0.501	0.25
Dim.19	100-seed wt.	0.18379	0.23982	0.16605	0.471	0.22
Dim.20	Germination (%)	0.03547	0.07517	0.04747	0.410	0.17

E.v. - Eigen-values; S.D. - Standard-deviation

In this study nearly 40% of phenotypic variability was shown by the first three PC's. PC1, PC2 and PC3 accounted for 16.63, 11.53, and 11.05% of the total phenotypic variability, respectively. As a result, we used the first three PCs to interpret the distribution of phenotypic variability in our study. To construct the Scree plot and PCA-Biplot, PCs with more than one Eigen value were chosen based on Kaiser's rule. Due to the very limited amount of information they contain about a single variable, values with Eigen values less than 1 are excluded from the data.

According to PCA for phenotypic variables for the first two PCs, the phenotypic traits that will contribute the most to phenotypic variability are yield (0.24652), seed-yield (0.34233), number of fruits (0.34606), titratable acidity (0.35105), TSS (0.33567), and plant-height (60-days) (0.24253). Average fruit weight, pericarp-thickness, and germination % were discovered to be three traits that contribute very little to the phenotypic variability among the first three PCs. In comparison to genotypes with lower values, those with higher values for traits like yield per plant, seed yield, number of fruits per plant, titratable acidity, TSS, and plant-height (60-days) will exhibit greater phenotypic variability. The expression of phenotypic variability was also moderately influenced by traits like fruit shape index, dry-matter content, 1000-seed weight, pollen-viability (30-days), and pollen-viability (30th-May), while traits like germination percentage, number of locules per fruit, and pericarp thickness contributed very little to phenotypic variability. Therefore, breeders can directly choose a genotype with higher variability during the selection process based on these traits. Among all the genotypes, SHV-866, SRD-864, NHP-761, NSP-768, US-1143, NOP-763, Shivaji, ISS-663, LTZ-1, Sakata Keshav and Satyajjit were found to have maximum variability for fruit yield and quality traits. Also, genotypes CLN-2679E, SRD-864, Campbell-22, MMR-745, EC 971784, LST-36-1 and LST-17 were found to have higher seed yield per plant along with higher number of quality seeds.

The PCA Scree Plot is constructed based on phenotypic-variability to display the percentage of phenotypic-variability explained by each PC. It is obvious that the variability explained by the Scree plot significantly decreases after the first three PCs. In this study nearly 40% of phenotypic variability was shown by the first three PC's. PC1, PC2 and PC3 accounted



Scree Plot – for PCA

Fig. 1. Scree plot for % variation explained by principle components.

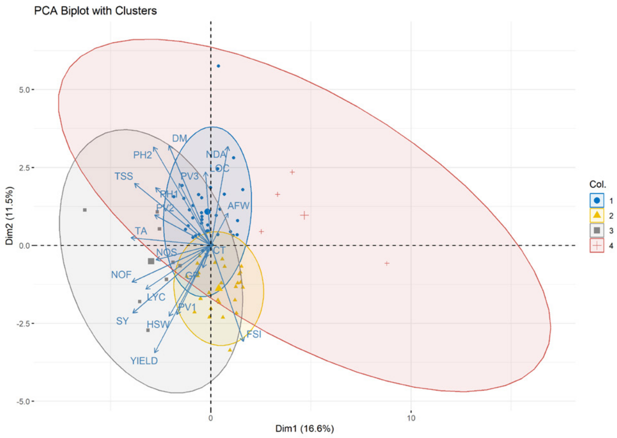


Fig. 2. PCA Biplot for fruit and seed yield.

for 16.63, 11.53, and 11.05% of the total phenotypic variability, respectively.

The traits that have a significant impact on increasing the yield are indicated by the variability displayed by the first two PCs, which contain both genotypic and phenotypic-variability. It is known that traits (vectors) that are perpendicular or have a 90°angle with the yield vector have a positive and significant impact on both seed yield and yield. Additionally, it provides information on the genotypes that significantly correlate with the traits, aiding in the selection process.

The Biplot analysis clearly showed that traits like fruit yield and fruit number per plant had the longest vectors, indicating a greater phenotypic contribution to the yield vector. Moderate expression towards yield was also shown by traits like 1000 seed weight, pollen-viability (30 days), lycopene-content, titratable acidity, and number of seeds. The traits that showed negative expression towards the yield trait included dry-matter content, fruit shape index, average fruit weight, and days to first anthesis as indicated above in the biplot (Fig. 2). The traits 1000-seed weight, lycopene content, number of fruits, number of seeds, and pollen-viability (45 days) have significant vector lengths and lie below the perpendicular range of seed-yield, indicating that they have a significant and direct influence on seed yield per plant.

Similar study was reported by Barbu *et al.* (2) also depicted similar results as the present study. Bhattarai *et al.* (3) reported a study in which the first five principle components explained 92% of variability present in the total variability. Xu *et al.* (15) reported that first 4 principle components explain 93.75% of the variability of the total variability. This study is relatable to present study; here also 1st 4 PC's show maximum variability in the total variability. Evgenidis *et al.* (4) PCA was carried out in 40 diverse tomato genotypes. PC2 included phenotypic traits and tolerance to inbreeding, while PC1 included yield components, heterosis, general and specific combining ability. The results of the correlation analysis revealed that PC1 and the hybrids had the strongest correlation, with PC1 accounting for 62.93% of the total variance and PC2 for the remainder. 37 tomato accessions with 15 morphological and biochemical characteristics of the accessions and their effects on fruit weight and dry-matter content, was established using the PCA Krasteva *et al.* (8). Four general factors were found to have a 73.4% impact on mass change and a 4.7% impact on dry-matter content in the analysis. The first factor for fruit mass, which included variables like vit-C, sugars, total-acidity, and dry-matter content, explained 22.8% of the variance. The second factor, which accounted for 19.7% of the variance, took into account aspects like plant height, fruit length, and shape.

Dendrogram is used in the current study to illustrate the genetic diversity that exists in the genotypes. Using the DARwin software for the analysis of the observations, the Hirechacial-Clustering method is used in this particular study to group or cluster genotypes (Fig. 3). The software had predicted a measurement of 280 for the Euclidean distance. This distance also calculates the separation between clusters. The chart is divided into four distinct groups by this line. Additionally, the K-means Clustering method was used to determine the ideal number of clusters for the genotypes (Table 3). The genotypes were divided into four groups using this method, each of which differ genetically from the others. Cluster-1 contained the most genotypes, while Cluster-2 contained the fewest genotypes. The number of genotypes present in Clusters 3 and 4 was moderate. Alam *et al.* (1) D^2 statistics and PCA were used to investigate the genetic divergence of the tomato plant. The genotypes were divided into 5 clusters, with cluster

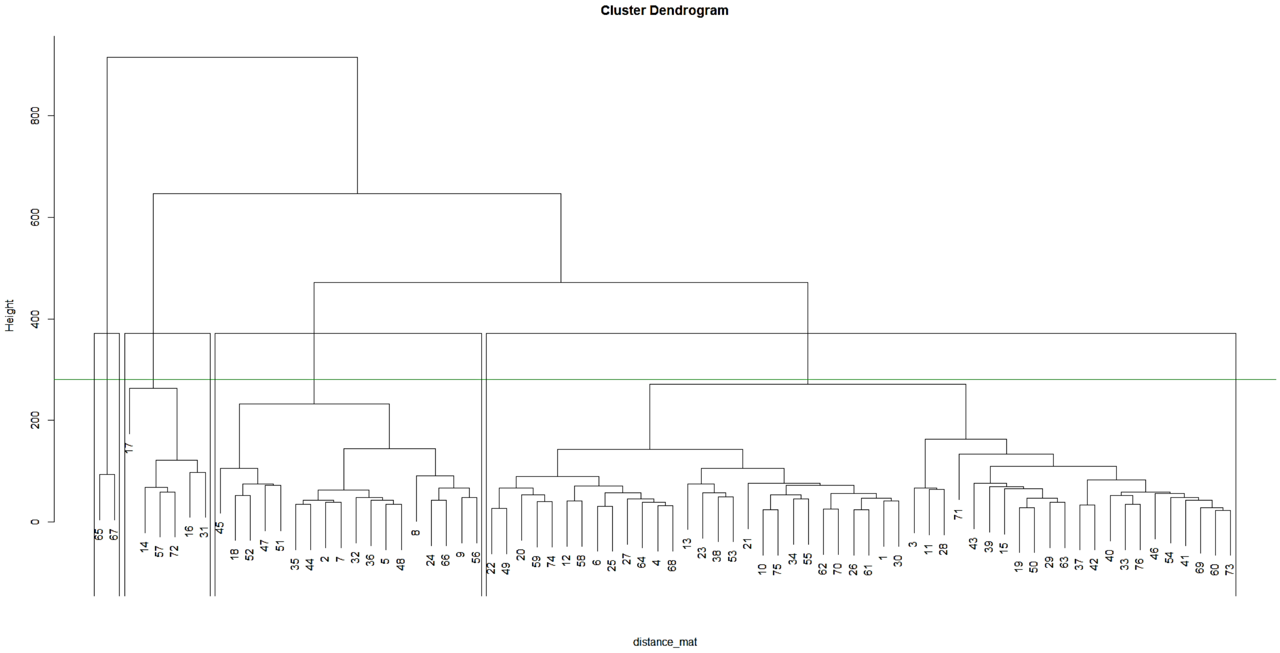


Fig. 3. Diversity study (Dendrogram) for the genotypes showing the phenotypic-variability for the traits under heat-stress.

I having the most genotypes and cluster V having the fewest. For a number of traits, including days for 50% flowering, individual fruit weight, fruit diameter, pericarp-thickness, number of locules per fruit, yield per plant, and yield tonne per hectare, Cluster III displayed the highest mean values. The diversity study (dendrogram) for the genotypes showing the phenotypic-variability for the traits under heat-stress (Fig. 3). Four diverse groups enclosed within blue boxes were predicted based on the Euclidean distance (Green line).

The cluster plot was constructed based on K-means clustering predicting 4 diverse groups which displays the different genotype groups for all the traits being studied. In this cluster-1, there is only one genotype, 17 (BGV-3010), which is placed in a separate group, indicating that it is either underperforming compared to other genotypes or that it is highly diverse. As a result, it can be eliminated from the selection process since it might not perform well. Additionally, genotypes 11 (AS-501), 57 (PSZ-801) and 71 (TPA-884) form a cluster in cluster-2, which is also very diverse and exhibits little variation across the board for all the traits being studied (Fig. 4). Cluster-3 contains 17 genotypes that exhibit moderate trait variability, while cluster-4 contains 55 genotypes that exhibit a high degree of diversity and trait variability. This variability of genotypes can be used to select an individual who exhibits the desired traits for heat-tolerance, while also being genetically and phenotypically superior.

Table 3. Clusters formed based on K-means clustering method.

Cluster	No. of genotypes present	Cluster means SS
Cluster-1	50	3,65,558.2
Cluster-2	2	77,954.87
Cluster-3	5	1,84,936
Cluster-4	19	3,12,010.3

The study concluded that genotypes used for the current investigation was found to be highly variable from the conclusions drawn from various components, both phenotypically and genotypically. The utilization of selection can be made to identify genotypes that perform better by producing an economical yield

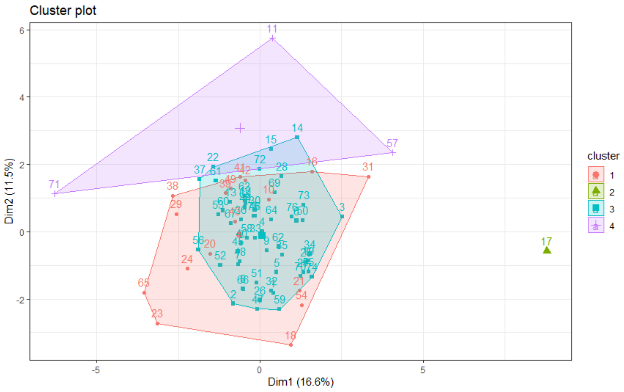


Fig. 4. Cluster plot for dendrogram for all the genotypes under study (K-means clustering method).

of fruit and seed along with desirable nutritional or quality traits and being able to withstand heat stress. According to the aforementioned methods, it is best to look for genotypes that have early flowering behavior, high yield, more fruits per plant, higher average fruit weight, higher seed yield, more seeds per fruit, higher pollen-viability, medium plant growth, and high lycopene content when choosing heat-tolerant lines. Among all 76 genotypes, SHV-866, SRD-864, NHP-761, NSP-768, US-1143, Shivaji, ISS-663, LTZ-1, Sakata Keshav and Satyajjit were found to have maximum variability for fruit yield and quality traits . Also, genotypes CLN-2679E, SRD-864, Campbell-22, MMR-745, EC 971784, LST-36-1 and LST-17 were found to have higher seed-yield per plant along with higher number of quality seeds.

AUTHORS' CONTRIBUTION

Conceptualization and designing of the research work (VBH, RS, SKJ); Execution of field/lab experiments (VBH); Data collection (VBH); Analysis of data and interpretation (VBH, RS, SKJ); Preparation of manuscript (VBH, RS, SKJ).

DECLARATION

The authors affirm that they do not have any potential conflicts of interest.

REFERENCES

1. Alam, M.S., Hossain, S., Ali, M.A., Hossain, M.G. and Islam, M.F. 2020. Assessment of genetic divergence in tomato (*Solanum lycopersicum* L.) through clustering and principal component analysis. *J. Agri. Sci. Engg. Inno.* (JASEI) **1**: 10-14.
2. Barbu, V., Neagu, C. and Dragan, M. 2015. Principal component analysis of some parameters used for lycopene extraction from tomatoes. *Acta Aliment.* **44**: 473-81.
3. Bhattarai, K., Louws, F.J., Williamson, J.D. and Panthee, D.R. 2016. Diversity analysis of tomato genotypes based morphological traits with commercial breeding significance for fresh market production in eastern USA. *Aust. J. Crop Sci.* **10**: 1098.
4. Evgenidis, G., Traka-Mavrona, E. and Koutsika-Sotiriou, M. 2011. Principal component and cluster analysis a tool in the assessment of tomato hybrids and cultivars. *Int. J. Agron.* **20**: 123-30.
5. Food and Agriculture Organization of the United Nations 2023. Agricultural production statistics 2000-2022. FAOSTAT https://www.fao.org/statistics/highlights-archive/highlights-detail/agricultural-production-statistics-%282000-2022%29/en
6. ISTA 2016. International Rules for Seed Testing. Bassersdorf, Switzerland.
7. Iwahori, S. 1965. High-temperature injuries in tomato development of normal flower buds and morphological abnormalities of flower buds treated with high-temperature*. J. Japanese Soc. Hort. Sci.* **34**: 33-41.
8. Krasteva, L. and Dimova, D. 2007. Evaluation of a canning determinate tomato collection using cluster analysis and principal component analysis (PCA). *Acta Hort.* **729**: 89-93.
9. Kumar, M., Yadav, R., Behera, T., Arora, A. and Talukdar, A. 2022. Physiological and biochemical response of thermo-sensitive and tolerant tomato genotypes to high temperature stress. *Indian J. Hort*. **74**: 388-92.
10. National Horticulture Board (NHB) 2023-24. Ministry of Agriculture and Farmer Welfare, https://www.nhb.gov.in/.
11. Sánchez-Rodríguez, E., Rubio-Wilhelmi, M.M., Cervilla, L.M., Blasco, B., Rios, J.J., Rosales, M.A. and Romero, L. 2016. Genotypic differences in some physiological parameters symptomatic for oxidative stress under moderate drought in tomato plants. *Plant Sci.* **171**: 498-04.
12. Sharma, D., and Sharma, H.R. 2013. Production and evaluation of tomato hybrids using diallel genetic design. *Indian J. Hort*. **70**: 531-37.
13. Singh, U., Patel, P.K., Singh, A.K., Tiwari, V., Kumar, R., Rai, N. and Singh, B. 2015. Screening of tomato genotypes under high-temperature stress for reproductive traits. *Veg. Sci.* **42**: 52-55.
14. Sinha, A., Singh, P., Kumar, R. and Bhardwaj, A. 2022. Assessment of genetic diversity for yield and its attributing traits in tomato (*Solanum lycopersicum* L.) genotypes for protected conditions. *Indian J. Agri. Res*. **56**: 480-84.
15. Xu, J., Henry, A., Sreenivasulu, N. and Struik, P.C. 2017. Effect of temperature on fruit set and reproductive development of tomato (*Solanum lycopersicum*) adapted to high temperatures. *Euphytica* **213**: 1-14.

(Received : May, 2024; Revised : March, 2025;
Accepted : March, 2025)