



RESEARCH COMMUNICATION

Multivariate characterisation of cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) under different sowing-date environments using principal component analysis and cluster analysis

Riya Pandey^{1*}, Dharendra Singh¹, Gaurav Chandra Sanwal¹, Udit Joshi¹, Makrand Dhyani², Amit Sharma³ & Anurag Satpathi⁴

¹Department of Vegetable Science, College of Agriculture, G. B. Pant University of Agriculture and Technology, Pantnagar 263 145, Uttarakhand, India

²Department of Mathematics, Graphic Era Hill University, Dehradun, Uttarakhand 248 002, India

³Department of Genetics and Plant Breeding, Chaudhary Charan Singh Haryana Agricultural University, Hisar 125 004, Haryana, India

⁴Division of Agrometeorology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Srinagar 190 025, Jammu and Kashmir, India

*Correspondence email - pandeyriya510@gmail.com

Received: 29 September 2025; Accepted: 27 March 2026; Available online: Version 1.0: 14 May 2026

Cite this article: Riya P, Dharendra S, Gaurav CS, Udit J, Makrand D, Amit S, Anurag S. Multivariate characterisation of cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) under different sowing-date environments using PCA and cluster analysis. Plant Science Today. 2026; 13(sp2): 1-10. <https://doi.org/10.14719/pst.12048>

Abstract

The exploitation of genetic diversity in cherry tomato is essential for breeding programs targeting yield and quality improvement under varying environmental conditions. In the present study, fifteen genotypes were evaluated across twelve sowing-date environments under polyhouse conditions for fourteen morphological and biochemical traits. Principal Component Analysis (PCA) was applied to reduce dimensionality and identify traits contributing most to phenotypic divergence, while cluster analysis was used to group genotypes based on similarity. Across sowing-date environments, four to five principal components with eigenvalues greater than one cumulatively explained 79-89 % of the total variation based on the Kaiser criterion. Yield-associated traits such as fruit yield per plant, fruit yield per square meter, average fruit weight and number of fruits per plant contributed predominantly to the first principal component. In contrast, quality-related attributes, including total soluble solids, sugar components (reducing, total and non-reducing sugars), ascorbic acid and lycopene content, were mainly represented in subsequent components. Cluster analysis classified the genotypes into three to seven distinct groups depending on the environment, reflecting both stability and environmental influence on trait expression. Genotypes such as 2021/TOCVAR-6, IIHR-2862, PCT-8 and 2022/TOCVAR-4 consistently occupied high-yielding clusters, while the commercial checks like Punjab Sona Cherry and CPCT Cherry-1 were associated with quality-oriented clusters. The results highlight the effectiveness of PCA in identifying key traits and divergent parental lines, offering valuable insights for hybridisation strategies aimed at combining yield potential with superior fruit quality in cherry tomato improvement programs.

Keywords: cherry tomato; cluster analysis; genetic diversity; polyhouse cultivation; principal component analysis; quality attributes

Introduction

Cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) is considered a progenitor of modern cultivated tomato and is characterised by its small fruit size, attractive appearance, rich variation in shape and colour and enhanced organoleptic quality associated with elevated soluble sugars and organic acids (1, 2). Native to tropical and subtropical regions of America, the crop later spread to Asia and Africa and gained prominence in European markets during the 1970s through imports from the USA, Israel and Senegal (3, 4). Presently, cherry tomatoes are widely grown in Central America and distributed to diverse regions including California, Korea, Germany, Mexico and Florida. Due to their indeterminate growth habit and the need for high-quality fruit,

cultivation is predominantly practised under protected conditions such as polyhouses, glasshouses and shade structures, which enable extended harvesting periods and stable yields.

Nutritionally, cherry tomato fruits are recognised as a functional food owing to their richness in essential vitamins, minerals, carotenoids and antioxidants such as lycopene and phenolics (5). They are low in calories, provide significant dietary fibre and contain appreciable quantities of vitamin C, potassium, β-carotene and folate, making them comparatively higher concentrations of many standard tomato varieties in terms of nutritional profile and taste index (a parameter reflecting the balance between sugars and organic acids that determines fruit flavour) (5). Their fruits, typically measuring 1.5–3.0 cm in diameter

and borne in clusters, are consumed fresh as well as processed into value-added products such as juice, ketchup, sauce, paste and canned product (6). Increasing consumer preference, particularly in premium retail markets, has expanded the economic importance of cherry tomatoes, with higher prices compared to standard tomatoes and rising demand among health-conscious populations (7).

Yield and its associated traits are central targets in tomato breeding programs, as they directly determine crop productivity and important market quality attributes such as fruit size and total soluble solids (8). Assessing the extent of genetic variability among genotypes is therefore critical for efficient selection and improvement. Although numerous morphological and physiological traits are routinely measured in breeding programs, not all contribute equally to trait differentiation among genotypes and redundancy in data can complicate interpretation (9). Multivariate statistical approaches, such as PCA provide a robust framework to address this challenge by reducing complex and interrelated datasets into a smaller number of orthogonal components while retaining most of the original variability (10, 11).

The PCA is a robust multivariate tool that identifies patterns of variation by transforming correlated variables into orthogonal components, with the first capturing the greatest variability and subsequent ones explaining progressively smaller proportions. This allows the recognition of key traits contributing most to overall phenotypic variation and aids in the discrimination of genotypes based on agronomic and quality attributes (1, 12). To complement this, cluster analysis groups genotypes according to similarity, thereby quantifying phenotypic divergence among genotypes and partitioning total variability into meaningful clusters. The reliability of clustering was assessed through the cophenetic correlation coefficient, which evaluates how well the dendrogram preserves original pairwise distances (13). In the present study, the combined use of PCA and cluster analysis in cherry tomato provided a comprehensive assessment of genetic divergence across sowing-date environments, enabling the identification of traits with the highest contribution to variation and offering insights into interrelationships among yield and quality-related traits. Given the increasing economic and nutritional importance of cherry tomato, such multivariate analytical approaches are valuable in prioritising attributes that can effectively guide selection and hybridisation strategies for future crop improvement.

Materials and Methods

Experimental site

The investigation under polyhouse conditions was carried out during 2023-24 and 2024-25 at the Vegetable Research Centre, G. B. Pant University of Agriculture and Technology, Pantnagar, U. S. Nagar (Uttarakhand, India), across twelve sowing-date environments (E₁-E₁₂). Pantnagar, situated in the *Tarai* belt at the foothills of the Shivalik range (29°0' N latitude, 79°3' E longitude, 243.4 m above mean sea level), experiences a humid subtropical climate with hot summers reaching 45 °C in June and cold winters

with minimum temperatures of about 4.4 °C in January. The monsoon extends from late June to September, contributing an annual rainfall of approximately 1300 mm, while occasional winter showers also occur. The experimental soil was sandy loam in texture. The presence of a high amount of organic matter contributed to its dark colour. The soil maintains a pH range between 6.0 and 6.5 throughout the period of investigation.

Experimental material, design and layout

The experimental material consisted of 15 cherry tomato genotypes evaluated for 14 growth, yield and quality-related traits. The trial was conducted in a Randomised Block Design (RBD) with three replications in each sowing environment. The experimental unit consisted of a plot containing six plants arranged in a single row. Each plot comprised rows spaced 40 cm apart with a 40 cm plant-to-plant distance. Standard cultivation practices were followed uniformly to minimise management effects. The codes used for sowing-date environments and genotypes are presented in Tables 1 and 2, respectively. Table 3 shows the list, code and source of 15 cherry tomato genotypes used in the experiment.

Application of fertilisers

Vermicompost at 2-5 kg/m² and the recommended dose of fertilisers (RDF), i.e., 120:80:60 kg/ha NPK, were applied at the time of land preparation. The entire dose of P and K and half dose of N were applied as a basal dose and the remaining half dose of N was applied 30 days after transplanting (DAT) as top dressing.

Plant protection measures

Preventative plant care measures were taken to manage insect-pest and disease incidence at regular intervals. During the crop growth period, plants were affected by early and late blight, which were effectively mitigated by spraying with Metalaxyl + Mancozeb (8% + 64% WP) AT 2 g/L water.

Statistical analysis

The PCA and cluster analysis were performed separately for each sowing-date environment to assess genotype responses under different seasonal conditions. The collected data were analysed using R Studio version 2024.12.1 Build 563. The data were subjected to PCA to assess phenotypic diversity among cherry tomato genotypes. The analysis generated eigenvalues, standard deviations, proportion of variance and cumulative variance explained by principal components across different sowing-date

Table 1. List of sowing-date environments with sowing date

Sowing-date environments	Sowing dates
Environment I	1 st August 2023
Environment II	15 th August 2023
Environment III	1 st September 2023
Environment IV	1 st February 2024
Environment V	15 th February 2024
Environment VI	1 st March 2024
Environment VII	1 st August 2024
Environment VIII	15 th August 2024
Environment IX	1 st September 2024
Environment X	1 st February 2025
Environment XI	15 th February 2025
Environment XII	1 st March 2025

Table 2. List and code of 15 genotypes of cherry tomato used in the experiment

Sl. No.	Genotypes	Sl. No.	Genotypes	Sl. No.	Genotypes
G ₁	CPCT Cherry-1 (Check)	G ₆	2021/TOCVAR-6	G ₁₁	CPCT Cherry-263
G ₂	2021/TOCVAR-2	G ₇	2022/TOCVAR-1	G ₁₂	DPCTY-1
G ₃	PCT-6	G ₈	PCT-8	G ₁₃	IIHR-2858
G ₄	2021/TOCVAR-4	G ₉	2022/TOCVAR-3	G ₁₄	IIHR-2862
G ₅	CPCT Cherry-214	G ₁₀	2022/TOCVAR-4	G ₁₅	Punjab Sona Cherry (Check)

Table 3. List, code and source of 15 cherry tomato genotypes used in the experiment

Sl. No.	Genotypes	Source
1.	CPCT Cherry-1 (Check)	Indian Agricultural Research Institute (IARI), New Delhi
2.	2021/TOCVAR-2	Indian Institute of Vegetable Research (IIVR), Varanasi
3.	PCT-6	Govind Ballabh Pant University of Agriculture and Technology (GBPUA&T), Pantnagar
4.	2021/TOCVAR-4	Solan
5.	CPCT Cherry-214	Indian Agricultural Research Institute (IARI), New Delhi
6.	2021/TOCVAR-6	Indian Institute of Vegetable Research (IIVR), Varanasi
7.	2022/TOCVAR-1	Indian Agricultural Research Institute (IARI), New Delhi
8.	PCT-8	Govind Ballabh Pant University of Agriculture and Technology (GBPUA&T), Pantnagar
9.	2022/TOCVAR-3	Govind Ballabh Pant University of Agriculture and Technology (GBPUA&T), Pantnagar
10.	2022/TOCVAR-4	Indian Institute of Vegetable Research (IIVR), Varanasi
11.	CPCT Cherry-263	Indian Agricultural Research Institute (IARI), New Delhi
12.	DPCTY-1	Palampur
13.	IIHR-2858	Indian Institute of Horticultural Research (IIHR), Bangalore
14.	IIHR-2862	Indian Institute of Horticultural Research (IIHR), Bangalore
15.	Punjab Sona Cherry (Check)	Punjab Agricultural University (PAU), Ludhiana

environments. Genotype scores derived from the retained principal components were subjected to cluster analysis to group the genotypes into distinct clusters. Mean values of yield and quality-related traits were subsequently computed for each cluster to identify superior groups and promising genotypes. The PCA is a multivariate statistical technique that extracts a reduced set of orthogonal components, each representing a linear combination of original variables, thereby explaining the maximum proportion of total variability with minimum information loss (13). The PCA was computed using the following equation:

$$PC1 = \sum_{j=1}^P a_j X_j \quad (\text{Eqn 1})$$

where, $a_j X_j$ = Linear coefficient – Eigen vectors.

Cluster analysis was performed in R using the K-means clustering algorithm, which partitions genotypes into K distinct and non-overlapping clusters based on similarity in trait performance. The optimal number of clusters was determined using the Elbow method based on the within-cluster sum of squares. Cluster analysis is a powerful tool for classifying genotypes by grouping accessions according to their genetic and phenotypic similarities (14). The method minimises the within-cluster sum of squares (WCSS) using the Euclidean distance as the dissimilarity measure. The objective function is given by:

$$J_K = \sum_{K=1}^K \sum_{i \in C_K} (x_i - m_K)^2 \quad (\text{Eqn 2})$$

Where, $(P \times 1, \dots, x_n) = X$ is the data matrix and $m_K = \sum_{i \in C_K} x_i / n_K$ is the centroid of cluster C_K and n_K is the number of points in C_K .

Results

The PCA was performed separately for the twelve sowing-date environments and components with eigenvalues >1 were retained (Table 4).

Correlation between principal components and variables

Five principal components were retained in ten environments, whereas four components were retained in Environments III and VII based on the eigenvalue (>1) criterion (Table 4). The first principal component (PC1) consistently explained the largest proportion of

variance, ranging from 28.20 % in Environment IV to 34.54 % in Environment IX. The second component (PC2) contributed between 19.8% (Env V) and 25.2 % (Env IV), while PC3 accounted for 13.7–19.5 % across sowing-date environments. Together, the retained components explained a cumulative variance ranging from 79.5 % in Environment VII to 88.5 % in Environment I, indicating high dimensional reduction efficiency.

Trait loadings revealed that in sowing-date environments III, IV, V, VI, IX, X, XI and XII, PC1 primarily had high positive loadings with yield-related traits such as fruit yield per plant (FYPP), fruit yield per m² (FYPM), average fruit weight (AFW), number of fruits per plant (NFPP), fruit length (FL) and fruit diameter (FD). In contrast, in sowing-date environments I, II, VII and VIII, PC1 was dominated by phenological and quality traits, including days to 50 % flowering (D50F), plant height (PH), total soluble solids (TSS), reducing sugar (RS), total sugar (TS), non-reducing sugar (NRS), ascorbic acid (AA) and lycopene content (LC), which were often positioned opposite yield traits along the axis. Subsequent components (PC2-PC5) explained secondary variation, generally reflecting either quality parameters (sugars, TSS, AA, LC) or phenological traits (D50F, PH), depending on the environment.

Grouping of genotypes into different clusters

Cluster analysis across 12 sowing-date environments revealed substantial phenotypic divergence among the 15 cherry tomato genotypes, with the number of clusters ranging from 3–7 depending on the environment. In Environment I, 6 clusters were formed with intra-cluster distances ranging from 2.13 to 4.67 and a maximum inter-cluster distance observed between Cluster 1 and Cluster 5 (7.82). Punjab Sona Cherry (G_{15}) alone constituted Cluster 1, IIHR-2858 (G_{13}) formed Cluster 2 and 2021/TOCVAR-6 (G_6) was in Cluster 5, while the remaining genotypes were grouped in larger clusters; notably, Cluster 5 recorded the highest mean values for yield traits. In Environment II, 7 clusters were obtained with intra-cluster distances from 1.97 to 4.51 and the highest inter-cluster distance (8.04) was between Cluster 2 and Cluster 7. Here, 2022/TOCVAR-4 (G_{10}) alone formed Cluster 7, excelling in yield traits, whereas PCT-6 (G_3) in Cluster 5 recorded higher mean values for quality traits, while Punjab Sona Cherry (G_{15}) clustered with low-yielding entries. In Environment III, 5 clusters were observed with intra-cluster distances ranging from 2.08 to 4.63 and a maximum inter-cluster distance (7.36) between Cluster 2 and Cluster 5. PCT-8 (G_8) alone formed Cluster 5 with maximum yield potential, while Cluster 2, comprising genotypes such as 2021/TOCVAR-2 (G_2), PCT-6 (G_3), CPCT Cherry-214 (G_5) and 2022/TOCVAR-3 (G_9), exhibited greater cluster mean values for

Table 4. Eigenvalues and percentage of variance explained by the first five principal components across 12 cherry tomato growing sowing-date environments

Environment	Eigenvalue					Variance (%)					Cumulative
	PC1	PC2	PC3	PC4	PC5	PC1	PC2	PC3	PC4	PC5	
E ₁	4.3	3.5	2.2	1.3	1.1	30.42	25.21	15.84	9.45	7.59	88.53
E ₂	4.4	2.8	2.1	1.5	1.2	31.17	20.04	15.14	10.76	8.73	85.86
E ₃	4.3	3	2.2	1.6	0.8	30.89	21.76	15.98	11.52	6.00	86.18
E ₄	3.9	2.8	2.1	1.7	1.1	28.19	20.09	14.68	12.46	8.05	83.48
E ₅	4.5	3.0	2.1	1.4	1.2	31.96	21.59	14.91	9.69	8.41	86.58
E ₆	4.3	3.1	2.1	1.4	1.0	30.85	22.36	15.21	9.81	7.29	85.53
E ₇	4.3	3.4	2.1	1.3	0.9	30.84	24.21	15.04	9.45	6.63	86.18
E ₈	4.3	2.8	2.7	1.5	1.0	30.81	19.82	19.52	10.43	7.34	87.93
E ₉	4.8	2.9	2.1	1.4	1.0	34.54	20.71	14.78	9.97	7.15	87.17
E ₁₀	4.5	2.8	2.1	1.4	1.3	32.42	20.14	15.22	10.32	8.96	87.09
E ₁₁	4.3	3.2	1.9	1.5	1.2	30.95	23.02	13.64	11.03	8.85	87.51
E ₁₂	4.3	3.2	2.0	1.5	1.2	30.66	22.82	14.18	10.79	8.45	86.91

quality parameters. In Environment IV, six clusters emerged with intra-cluster distances between 2.11 and 4.74 and the highest inter-cluster distance (7.91) was between Cluster 3 and Cluster 6. Cluster 4, consisting of 2021/TOCVAR-6 (G₆) and IIHR-2862 (G₁₄), exhibited the highest fruit yield per plant, whereas Cluster 2 (2021/TOCVAR-4 - G₄, CPCT Cherry-214 - G₅ and 2022/TOCVAR-3 - G₉) recorded the highest yield per m². In Environment V, four clusters were formed with intra-cluster distances ranging from 2.02 to 4.38 and a maximum inter-cluster distance (6.85) between Cluster 1 and Cluster 4. Cluster 4, containing PCT-8 (G₈), 2022/TOCVAR-4 (G₁₀) and CPCT Cherry-263 (G₁₁), excelled in yield and fruit weight traits. In Environment VI, 3 clusters were grouped with intra-cluster distances from 1.89 to 4.02 and the largest inter-cluster distance (6.34) was between Cluster 1 and Cluster 3. Cluster 3, composed of PCT-6 (G₃), 2021/TOCVAR-4 (G₄), 2021/TOCVAR-6 (G₆), PCT-8 (G₈), 2022/TOCVAR-4 (G₁₀), DPCTY-1 (G₁₂) and IIHR-2862 (G₁₄), was recorded the highest mean values for yield attributes, while Cluster 2 (CPCT Cherry-214 - G₅, 2022/TOCVAR-3 - G₉ and CPCT Cherry-263 - G₁₁) exhibited greater cluster mean values for quality traits. In Environment VII, 3 clusters were obtained with intra-cluster distances between 2.01 and 3.97 and the maximum inter-cluster distance (6.27) was recorded between Cluster 2 and Cluster 3. Cluster 3, including 2021/TOCVAR-6 (G₆), PCT-8 (G₈), 2022/TOCVAR-4 (G₁₀) and IIHR-2862 (G₁₄), recorded higher mean values for fruit size and yield, while Cluster 2 exhibited greater cluster mean values for fruit number and sugar traits. In Environment VIII, three clusters were identified with intra-cluster distances ranging from 2.16 to 4.11 and a maximum inter-cluster distance (6.42) between Cluster 1 and Cluster 3. Cluster 3, composed of 2021/TOCVAR-4 (G₄), CPCT Cherry-214 (G₅), 2021/TOCVAR-6 (G₆), PCT-8 (G₈), DPCTY-1 (G₁₂) and IIHR-2862 (G₁₄), recorded higher mean values for yield and TSS, whereas Cluster 2 (2021/TOCVAR-2 - G₂, PCT-6 - G₃, 2022/TOCVAR-3 - G₉, 2022/TOCVAR-4 - G₁₀ and CPCT Cherry-263 - G₁₁) performed better in fruit weight. In Environment IX, 4 clusters were formed with intra-cluster distances from 2.05 to 4.49 and a maximum inter-cluster distance (6.88) between Cluster 1 and Cluster 4. Cluster 4, including 2021/TOCVAR-2 (G₂), PCT-6 (G₃), PCT-8 (G₈), 2022/TOCVAR-4 (G₁₀) and IIHR-2862 (G₁₄), exhibited the highest yield and fruit weight, while Punjab Sona Cherry (G₁₅) again stood alone in Cluster 1 with the lowest performance.

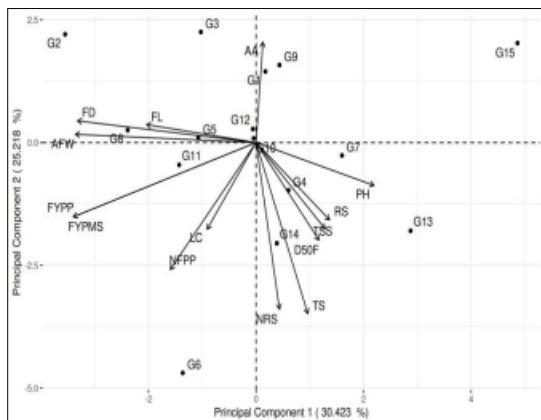
In Environment X, 7 clusters were recorded with intra-cluster distances between 2.12 and 4.76 and a maximum inter-cluster distance (7.85) between Cluster 4 and Cluster 7. Cluster 7, comprising 2021/TOCVAR-6 (G₆) and IIHR-2862 (G₁₄), recorded the highest mean values for yield, while Cluster 6 (2021/TOCVAR-2 - G₂ and CPCT Cherry-263 - G₁₁) recorded the highest fruit weight. In Environment XI, 3 clusters were observed with intra-cluster distances ranging from

1.94 to 3.89 and a maximum inter-cluster distance (6.41) between Cluster 1 and Cluster 3. Cluster 2, comprising most genotypes, recorded higher mean values for yield and fruit weight, whereas Cluster 3 (CPCT Cherry-1 - G₁, 2021/TOCVAR-6 - G₆, DPCTY-1 - G₁₂, IIHR-2858 - G₁₃ and IIHR-2862 - G₁₄) exhibited greater cluster mean values for sugar-related traits. Finally, in Environment XII, 5 clusters were obtained with intra-cluster distances from 2.18 to 4.58 and a maximum inter-cluster distance (7.22) between Cluster 2 and Cluster 5. Cluster 5, consisting of 2021/TOCVAR-6 (G₆) and IIHR-2862 (G₁₄), recorded higher mean values for yield and sugars, while Cluster 4 (2021/TOCVAR-2 - G₂ and CPCT Cherry-263 - G₁₁) recorded maximum fruit weight. (Fig 1 and 2) The results of cluster analysis depicting the grouping pattern of cherry tomato genotypes are presented in Supplementary Table 1, while the corresponding cluster mean values for yield and quality-related traits are detailed in Supplementary Table 2.

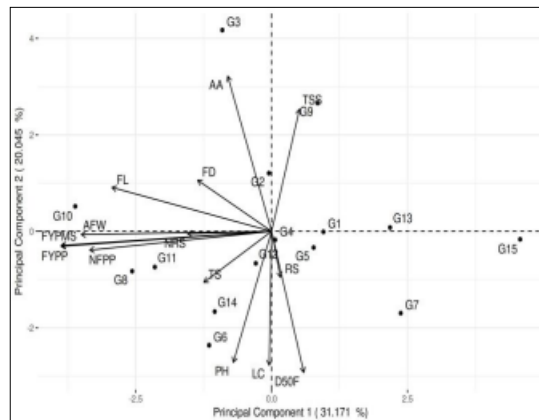
Discussion

Principal component analysis is widely applied in plant breeding as a quantitative tool for dissecting phenotypic diversity and identifying maximally divergent parental lines, thereby facilitating the design of effective hybridisation programs (15). In the present study, PCA revealed that only 4–5 components were sufficient to capture the majority of variability among cherry tomato genotypes across twelve sowing-date environments. The first 2 components cumulatively explained nearly 54 % of the total variation, demonstrating their effectiveness in summarising complex, interrelated trait data. Yield-related traits such as fruit yield per plant, fruit yield per square meter, average fruit weight, fruit length and fruit diameter contributed predominantly to the first principal component, whereas phenological and quality traits, including days to 50 % flowering, plant height, total soluble solids, sugars, ascorbic acid and lycopene content, were more influential in subsequent components. The dominance of yield-related traits in PC1 suggests that productivity attributes exhibited the greatest variability among the evaluated genotypes and therefore played a major role in differentiating them. This pattern is expected in breeding materials where selection pressure is often directed towards yield improvement, resulting in greater dispersion of yield components compared with quality traits.

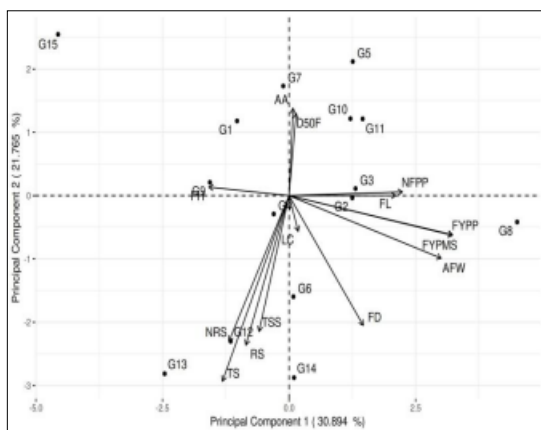
These findings are consistent with earlier reports in tomato and related crops. Research highlighted the predominance of yield-associated traits in PC1 (16, 17). While research indicates that the first 2 components captured a substantial proportion of total variation, reflecting strong trait interrelationships. Similarly, research observed that total fruits per plant and days to first fruiting contributed



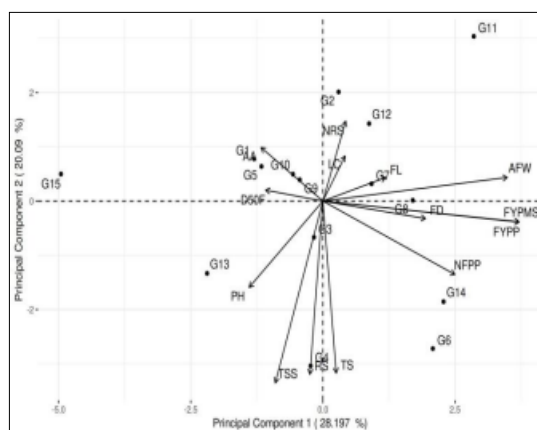
Environment 1



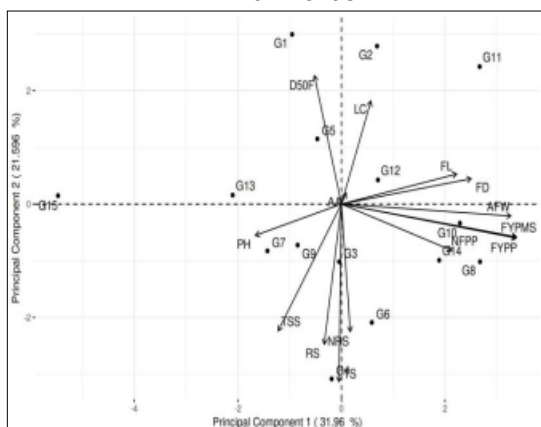
Environment 2



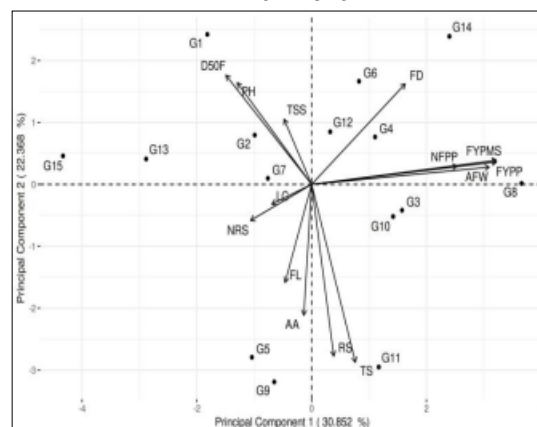
Environment 3



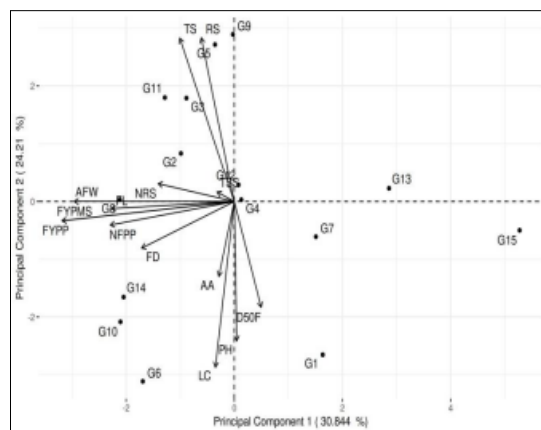
Environment 4



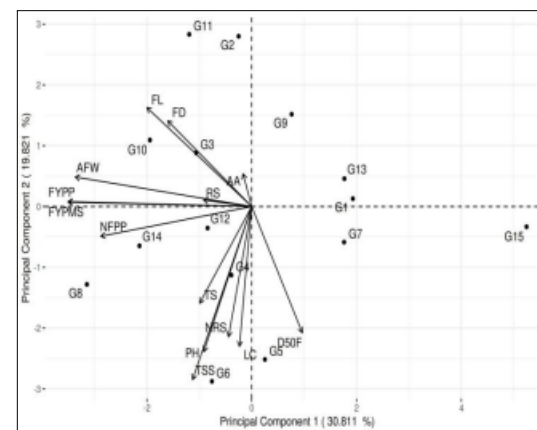
Environment 5



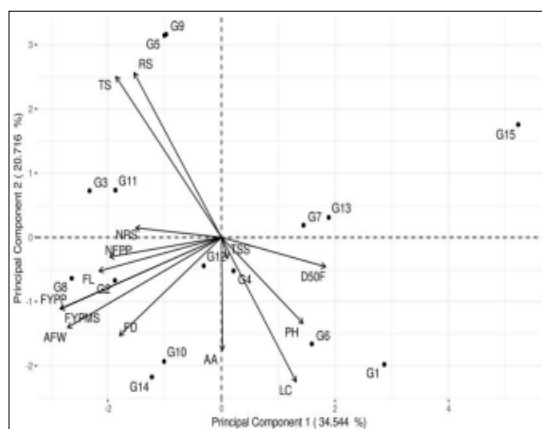
Environment 6



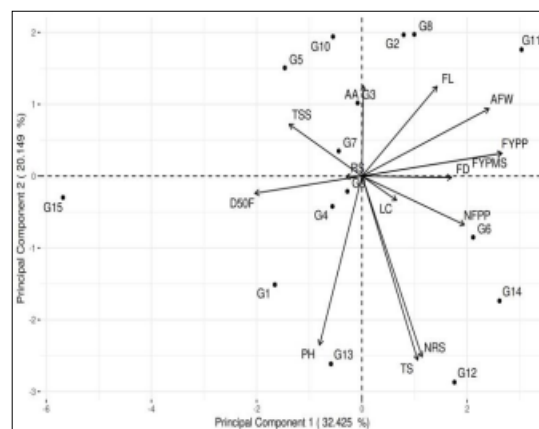
Environment 7



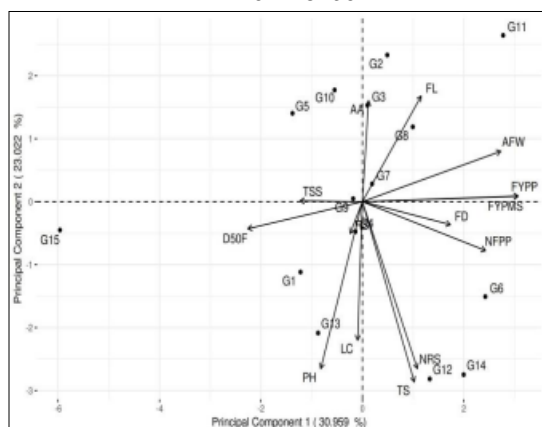
Environment 8



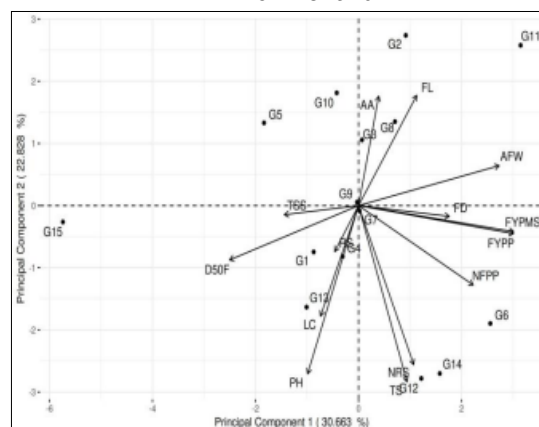
Environment 9



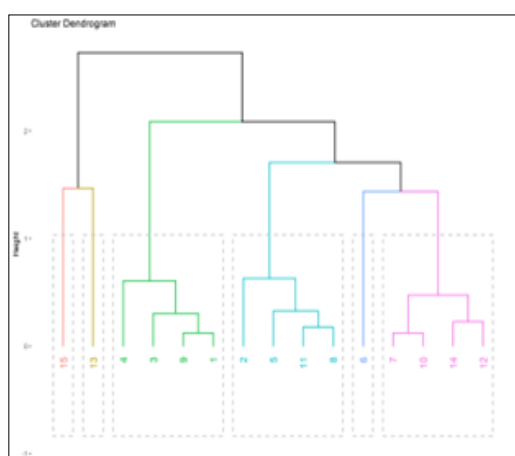
Environment 10



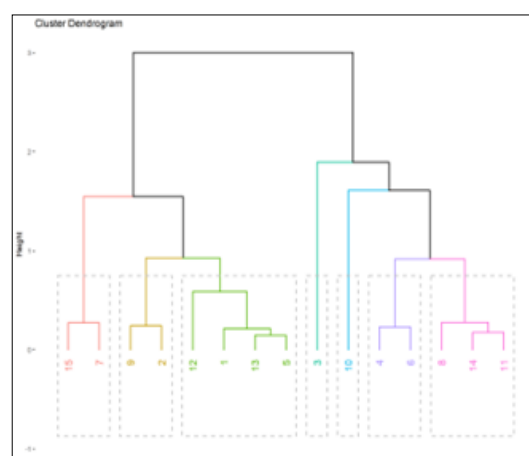
Environment 11



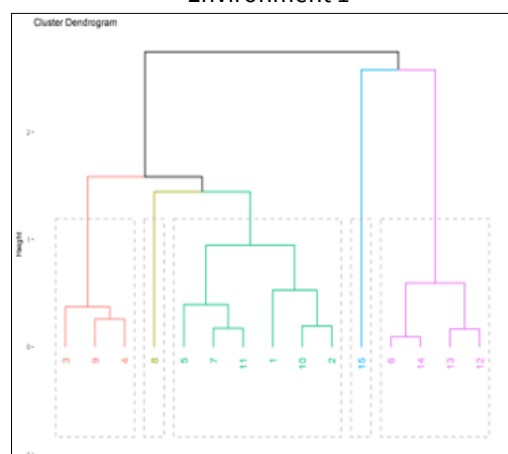
Environment 12

Fig. 1. PCA biplot showing the association between genotypes and traits of cherry tomato under different environments .


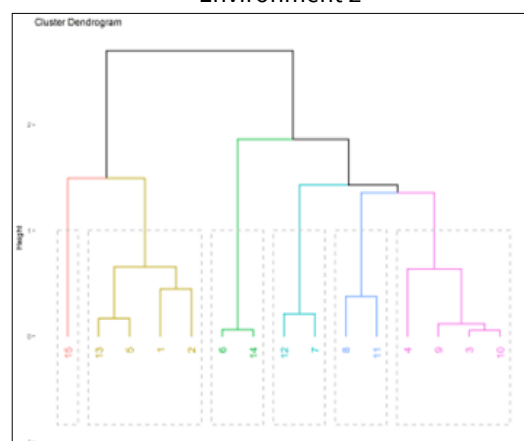
Environment 1



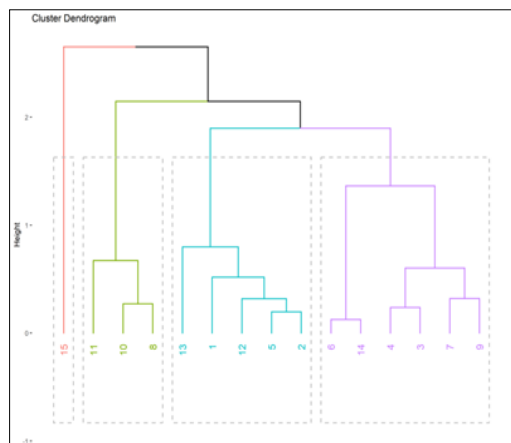
Environment 2



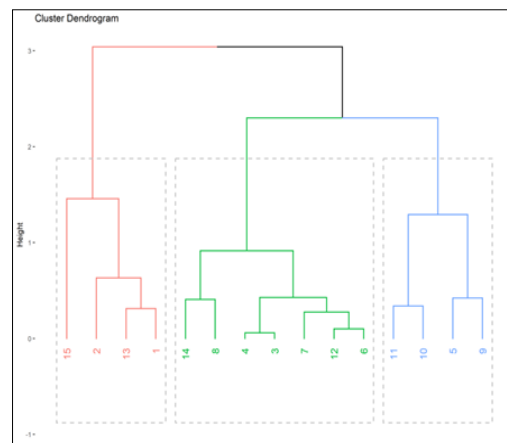
Environment 3



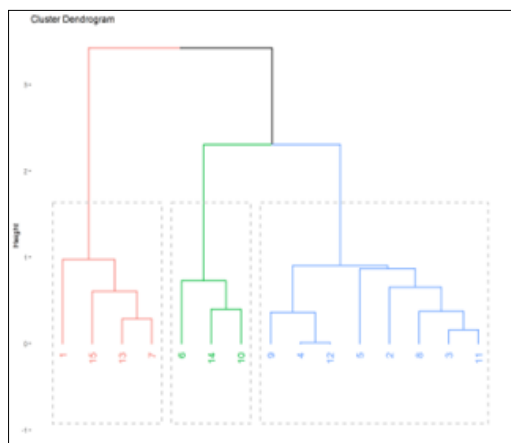
Environment 4



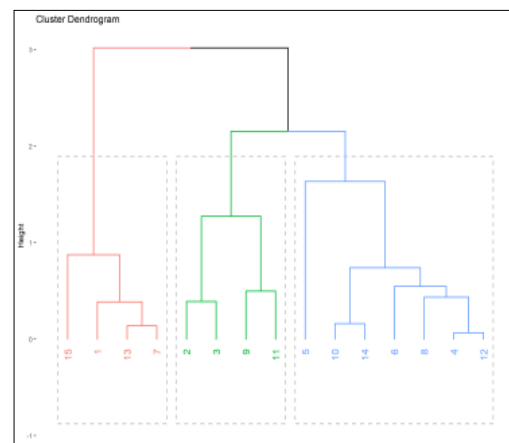
Environment 5



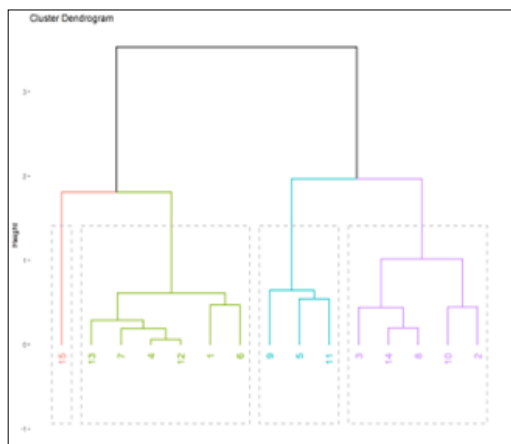
Environment 6



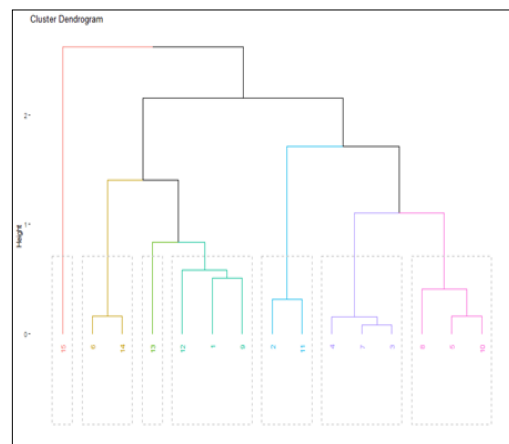
Environment 7



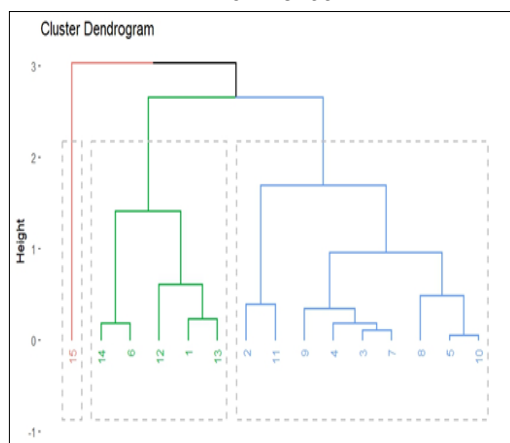
Environment 8



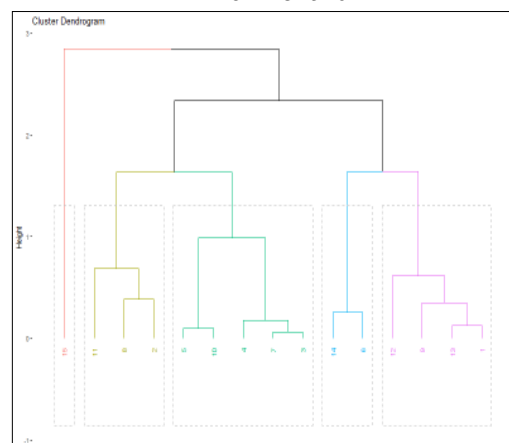
Environment 9



Environment 10



Environment 11



Environment 12

Fig. 2. Dendrogram representing genetic divergence among cherry tomato genotypes under different Environment.

positively to principal components in tomato (18, 19). Whereas research has demonstrated that reported plant height, fruit length and fruit width are major contributors (20, 21). Research has identified fruit yield per plant as a key contributor. Research confirmed that PC1 accounted for maximum variation in tomato and chilli, respectively (22, 23). These studies, along with the current results, emphasise that a limited number of principal components can effectively capture the majority of phenotypic variability, enabling breeders to identify key traits for selection. The clustering of yield traits within the same component also indicates a strong association among fruit number, fruit size and total yield, suggesting that these traits are physiologically interrelated and jointly influence productivity in cherry tomato.

Comparable observations were made in the research, where six components with eigenvalues above one explained 87.6 % of the total variation (24). Research emphasised the importance of components with eigenvalues greater than one as indicators of major system attributes (25). Similar trends have been reported in the previous research, where 5 principal components explained different variability (23, 26, 27). Research also reported that the first three principal components together explained 78.54 % of the total variation in tomato, with the first component mainly influenced by leaf traits and the second by plant height (28). Collectively, these studies confirm the effectiveness of PCA in capturing major sources of genetic variability and identifying traits of importance for breeding programs.

Cluster analysis in the present investigation supported the presence of wide phenotypic divergence among the cherry tomato genotypes, with the number of groups ranging from 3–7 across sowing-date environments. High-yielding genotypes such as PCT-8, 2021/TOCVAR-6 and IIHR-2862 consistently appeared in clusters characterised by superior yield performance, indicating both stability and potential as parental lines for hybridisation. In contrast, check varieties like Punjab Sona Cherry and IIHR-2858 were often separated into low yielding or distinct clusters, reflecting their contrasting genetic backgrounds. These findings are consistent with studies in *Solanum surattense* and tomato, where genotypes with higher yield and yield-related traits aggregated in single clusters (29, 30). The formation of distinct clusters indicates that the evaluated genotypes possess considerable phenotypic divergence, which is advantageous for breeding programs because crosses between genetically distant parents often produce greater heterosis and wider segregation in subsequent generations. This pattern likely reflects strong positive correlations among yield components.

The present results also highlight that the grouping of genotypes was primarily influenced by agronomic and yield-related traits rather than geographic origin, as accessions from different sources were randomly distributed across clusters. This pattern indicates a broad genetic base among the evaluated accessions, providing a substantial reservoir of diversity for breeding (31, 32). Similar observations have been reported in other vegetable crops, such as bitter melon, where PC1 alone explained a significant portion of variation and contributed to effective genotype differentiation (33). The variation in cluster composition across sowing-date environments further indicates the influence of seasonal conditions on trait expression, as differences in temperature, radiation and photoperiod during crop growth can modify plant development, fruit set and quality attributes.

Moreover, the relative contribution of yield and quality traits to genetic divergence was influenced by environmental conditions, reflecting the importance of evaluating genotypes across multiple sowing-date environments for stable performance. Traits such as fruit yield per plant, fruit yield per square meter and fruit weight consistently contributed to high-yielding clusters, whereas quality attributes, including total soluble solids, sugars, ascorbic acid and lycopene content, were more prominent in other clusters depending on the environment. Maximum inter-cluster divergence and higher cluster means for yield-related traits in tomato lines (34–38). The combined use of PCA and cluster analysis in this study effectively captured genetic variability and identified key traits contributing to yield and quality in cherry tomato. The results facilitate the identification of superior lines, such as PCT-8, 2021/TOCVAR-6 and IIHR-2862, for hybridisation programs aimed at improving yield while maintaining fruit quality. The broad genetic base and substantial trait variability observed among the evaluated genotypes provide valuable insights for future breeding strategies under diverse environmental conditions (39).

Conclusion

The PCA and cluster analysis across 12 environments revealed clear environment-specific trait-genotype relationships. A set of 4–5 principal components collectively captured over 79% of the total variation. PC1 predominantly reflected yield-related traits (FYPP, FYPM, AFW, NFPP, FL, and FD) across most sowing-date environments, while phenological and quality traits (D50F, PH, TSS, sugars, AA, and LC) were more influential in the remaining environments. Clustering grouped genotypes into distinct classes, separating yield-oriented types such as 2021/TOCVAR-6, IIHR-2858 and CPCT Cherry-214 from quality-oriented checks like Punjab Sona Cherry and CPCT Cherry-1, whereas some genotypes (2022/TOCVAR -1, IIHR-2862) exhibited broader adaptability across sowing-date environments. Together, these analyses demonstrate that yield stability and quality expression are strongly environment-dependent, underscoring the importance of environment-specific selection strategies and the utility of multivariate approaches in cherry tomato breeding.

Acknowledgements

The authors sincerely thank Vegetable Research Centre, G. B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India, for providing the field facilities essential for this study. Special thanks are extended to the faculty, researchers and technical staff of the respective departments for their valuable guidance, support and assistance throughout the study.

Authors' contributions

RP conceptualised the study, conducted the experiments, analysed the data and drafted the manuscript. DS provided guidance, supervised the research and critically revised the manuscript. GCS contributed to data interpretation. UJ supported data collection and literature review. MD assisted in data analysis. AS¹ contributed to manuscript editing and provided critical inputs. AS² was responsible for the quality check of the study. All authors read and approved the final manuscript. [AS¹ stands for Amit Sharma and AS² stands for Anurag Satpathi].

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None

References

1. Prema G, Indiresk KM, Santhosha HM. Evaluation of cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) genotypes for growth, yield and quality traits. *Asian J Hort.* 2011;6(1):181–4.
2. Rocha MC, Deliza R, Ares G, Freitas DD, Silva AL, Carmo MG, et al. Identifying promising accessions of cherry tomato: a sensory strategy using consumers and chefs. *J Sci Food Agric.* 2013;93(8):1903–14. <https://doi.org/10.1002/jsfa.5988>
3. Gharezi M, Joshi N, Indiresk KM. Physico-chemical and sensory characteristics of different cultivars of cherry tomato. *Mysore J Agric Sci.* 2012;46(3):610–3.
4. Branca F, Leonardi CH. The performance of cherry-type tomato grown in a cold greenhouse. *Colt Prot.* 1992;7(8):48–52.
5. Malik G, Masoodi L, Nabi SU, Sharma A, Singh DB. Production technology of cherry tomato in Kashmir. Srinagar (India): ICAR-Central Institute of Temperate Horticulture; 2017.
6. Venkadeswaran E, Vethamoni PI, Arumugam T, Manivannan N, Harish S. Evaluation and selection of cherry tomato (*Solanum lycopersicum* (L.) var. *cerasiforme* Mill.). *Int J Curr Microbiol Appl Sci.* 2018;7(6):1155–65. <https://doi.org/10.20546/ijcmas.2018.706.137>
7. Bontha VB, Tomar BS, Singh B, Behera TK. Effect of methods and time of pollination on seed yield and quality parameters in cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) grown under different protected conditions. *J Agric Sci Technol.* 2015;17(1):65–72. <https://doi.org/10.5958/0974-0112.2015.00011.0>
8. Das S, Das SS, Chakraborty I, Roy N, Nath MK, Sarma D. Principal component analysis in plant breeding. *Biomol Rep.* 2017;3:1–3.
9. Maji AT, Shaibu AA. Application of principal component analysis for rice germplasm characterization and evaluation. *J Plant Breed Crop Sci.* 2012;4(6):87–93. <https://doi.org/10.5897/JPBSCS11.093>
10. Adams MW. An estimate of homogeneity in crop plants with special reference to genetic vulnerability in dry season *Phaseolus vulgaris*. *Euphytica.* 1977;26(3):665–79. <https://doi.org/10.1007/BF00021692>
11. Broschat TK. Application of principal component analysis to horticultural research. *HortScience.* 1979;14(2):114–7. [Note: Corrected from provided Ref 11 details to match standard citation for this topic/author].
12. Hotelling H. Analysis of a complex of statistical variables into principal components. *J Educ Psychol.* 1933;24(6):417–41. <https://doi.org/10.1037/h0071325>
13. Shoba D, Vijayan R, Robin S, Manivannan N, Iyanar K, Arunachalam P, et al. Assessment of genetic diversity in aromatic rice (*Oryza sativa* L.) germplasm using PCA and cluster analysis. *Electron J Plant Breed.* 2019;10(3):1095–104. <https://doi.org/10.5958/0975-928X.2019.00140.6>
14. Nankar AN, Tringovska I, Grozeva S, Ganeva D, Kostova D. Tomato phenotypic diversity determined by combined approaches of conventional and high-throughput tomato analyzer phenotyping. *Plants.* 2020;9(2):197. <https://doi.org/10.3390/plants9020197>
15. Ahmadizadeh M, Felenji H. Evaluating diversity among potato cultivars using agro-morphological and yield components in fall cultivation of Jiroft area. *Am-Eurasian J Agric Environ Sci.* 2011;11(5):655–62.
16. Sanni K, Fawole I, Ogunbayo S, Tia D, Somado E, Futakuchi K, et al. Multivariate analysis of diversity of landrace rice germplasm. *Crop Sci.* 2012;52(2):494–504. <https://doi.org/10.2135/cropsci2010.12.0739>
17. Ojha G, Sarawagi A, Sharma B, Parikh M. Principal component analysis of morpho-physiological traits in rice germplasm accessions (*Oryza sativa* L.) under rainfed condition. *Int J Chem Stud.* 2017;5(4):1875–8.
18. Lakshmi M, Shanmuganathan M, Jeyaprakash P, Ramesh T. Genetic variability and diversity analysis in selected rice (*Oryza sativa* L.) varieties. *Electron J Plant Breed.* 2022;13(3):959–66. <https://doi.org/10.37992/2022.1303.124>
19. Sehgal N, Chadha S, Kumar S, Ravita. Variability and trait association analyses in bacterial wilt resistant F4 progenies of tomato (*Solanum lycopersicum* L.) for yield and biochemical traits. *Indian J Exp Biol.* 2021;59(9):617–25.
20. Hussain I, Khan SA, Ali S, Farid A, Ali N, Ali S, et al. Genetic diversity among tomato accessions based on agro-morphological traits. *Sains Malays.* 2018;47(11):2637–45. <https://doi.org/10.17576/jsm-2018-4711-06>
21. Kumar PA, Reddy NN, Lakshmi NJ. Screening tomato genotypes for salt tolerance. *Int J Curr Microbiol Appl Sci.* 2017;6(11):1037–49. <https://doi.org/10.20546/ijcmas.2017.611.121>
22. Singh P, Jain PK, Tiwari A. Principal component analysis approach for yield attributing traits in chilli (*Capsicum annum* L.) genotypes. *Chem Sci Rev Lett.* 2020;9(33):87–91.
23. Rai AK, Vikram A, Pal S. Genetic characterization of tomato (*Solanum lycopersicum* L.) germplasm for yield and quality traits through principal component analysis. *Res J Agric Sci.* 2017;8(5):1171–4.
24. Sinha A, Singh P, Bhardwaj A, Verma RB. Principal component analysis approach for comprehensive screening of tomato germplasm for polyhouse condition. *J Exp Agric Int.* 2021;43(9):64–72. <https://doi.org/10.9734/jeai/2021/v43i930739>
25. Brejda JJ, Moorman TB, Karlen DL, Dao TH. Identification of regional soil quality factors and indicators I. Central and southern High Plains. *Soil Sci Soc Am J.* 2000;64(6):2115–24. <https://doi.org/10.2136/sssaj2000.6462115x>
26. Tsagaye D, Gedebo A, Aklilu S. Multivariate studies for fruit yield and quality components in diverse tomato (*Lycopersicon esculentum* Mill.) genotypes. *Int J Res Agric Sci.* 2019;6(6):2348–3997.
27. Prakash MO, Vijay B. Principal component and cluster analysis of indigenous tomato genotypes based on morphological indicators. *Res J Biotechnol.* 2017;12(7):50–8.
28. Jin L, Zhao L, Wang Y, Zhou R, Song L, Xu L, et al. Genetic diversity of 324 cultivated tomato germplasm resources using agronomic traits and InDel markers. *Euphytica.* 2019;215(7):116. <https://doi.org/10.1007/s10681-019-2391-8>
29. Dheebisha C, Nalina L, Rajamani K, Geethanjali S, Boopathi NM, Chandrakumar K, et al. Genetic variability, association and path analysis for yield and fruit quality components in yellow-berried nightshade (*Solanum surattense*). *Electron J Plant Breed.* 2023;14(2):419–28. <https://doi.org/10.37992/2023.1402.066>
30. Evgenidis G, Traka-Mavrona E, Koutsika-Sotiriou M. Principal component and cluster analysis as a tool in the assessment of tomato hybrids and cultivars. *Int J Agron.* 2011;2011:697879. <https://doi.org/10.1155/2011/697879>
31. Ene CO, Abteu WG, Oselebe HO, Ozi FU, Ikeogu UN. Genetic characterization and quantitative trait relationship using multivariate techniques reveal diversity among tomato germplasms. *Food Sci Nutr.* 2022;10(7):2426–42. <https://doi.org/10.1002/fsn3.2850>
32. Vargas JEE, Aguirre NC, Coronado YM. Study of the genetic diversity of tomato (*Solanum* spp.) with ISSR markers. *Rev Ceres.* 2020;67(3):199–206. <https://doi.org/10.1590/0034-737x202067030005>
33. Mehta T, Duhan DS, Phogat S. Morphological characterisation and principal component analysis studies in bitter melon (*Momordica charantia* L.) genotypes. *Vegetos.* 2024;37(3):738–44. <https://doi.org/10.1007/s10681-019-2391-8>

doi.org/10.1007/s42535-024-00869-5

34. Reddy BR, Reddy MP, Begum H, Sunil N. Genetic diversity studies in tomato (*Solanum lycopersicum* L.). IOSR J Agric Vet Sci. 2013;4(4):53–5. <https://doi.org/10.9790/2380-0445355>
35. Sharma HR, Sharma D, Thakur AK. Analysis of genetic divergence in tomato (*Lycopersicon esculentum* Mill.). J Horti Sci. 2006;1(1):52–4. <https://doi.org/10.24154/jhs.v1i1.674>
36. Singh AK, Sharma JP, Kumar S, Chopra S. Genetic divergence in tomato (*Lycopersicon esculentum* Mill.). J Res SKUAST-J. 2008;7(1):1–8.
37. Kumar M, Yadav RK, Behera TK, Talukdar A, Kour M. Assessment of genetic divergence for quantitative traits in thermo tolerant tomato (*Solanum lycopersicum* L.) genotypes. J Appl Nat Sci. 2018;10(1):55–8. <https://doi.org/10.31018/jans.v10i1.1578>
38. Saleem MY, Khalid I, Shakeel A, Ayub B. Evaluation and estimation of genetic divergence of tomato hybrids by using principal component analysis and cluster analysis under high temperature. Trends Biotechnol Plant Sci. 2023;1(1):84–94. <https://doi.org/10.62460/TBPS/2023.001>
39. Williams G, Anbuselvam Y. Assessment of genetic divergence through principal component analysis and clustering in tomato germplasm accessions. Environ Ecol. 2023;41(4D):3060–5. <https://doi.org/10.60151/envec/YLSS4838>

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonpublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc See https://horizonpublishing.com/journals/index.php/PST/indexing_abstracting

Copyright: © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.