Received: 21 Jun 2022 | Revised: 11 Aug 2022 | Accepted: 16 Aug 2022

Electronic Journal of Plant Breeding

Research Article

Selection of tomato (*Solanum lycopersicum* L.) lines for fresh and processing market through multivariate analysis

J. Purushothaman[1*], T. Saraswathi[1], T. Kalaimagal[2], R. Swarnapriya[1], M. Karthikeyan [3] and A. Anitha [4]

[1]Department of Vegetable Science, HC&RI, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India
[2]Department of Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India
[3]Department of Plant Pathology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India
[4]Department of Nematology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India
***E-Mail**: purushothj14@gmail.com

Abstract
A total of 36 tomato accessions were subjected to principal component analysis (PCA) based on seventeen yield and quality traits. The overall variation was split into seven major principle components, which accounted for 81.62 per cent of the total variation. The bi-plot was constructed using the first two PCs, in which the genotypes CBESL133, CBESL129, CBESL115, CBESL121, CBESL101, CBESL114, CBESL136, CBESL102 and CBESL111 were dispersed across all four quadrates, indicating the greatest genetic divergence. The first two PCs contributed the most divergence due to the yield and yield related traits. In Pearson's correlation analysis, the number of fruits per plant, the number of clusters per plant, plant height, single fruit weight and ascorbic acid content were positive and significantly associated with yield per plant. The lines grouped under PC1 and PC2 were suitable for the yield improvement breeding programme. The lines suitable for processing come under PC4 and PC5.

Keywords: PCA, diversity, tomato, correlation, Multivariate analysis

INTRODUCTION

Tomato (*Solanum lycopersicum* L., 2n=2x=24) is one of the most significant and popular vegetables in the world due to their extensive adaptability, better production potential and appropriateness for a variety of fresh and processed food applications. It is a self-pollinated crop that belongs to the Solanaceae family. Tomatoes are the second most important vegetable after potatoes and they are in high demand as a cash crop around the world (Geraldini *et al*., 2018). Genetic diversity among parents is considered an important factor for obtaining heterotic hybrids (Moll *et al.,* 1962; Khanna and Chaudhary, 1974; Chandra, 1977). Wild species have lost valuable characteristics such as disease resistance and stress tolerance as a result of domestication. Despite the increased yields brought about by domestication, the plant breeding concentration on yield increase (production) has resulted in a loss of genetic variation, as well as a reduction in nutritional value and taste (Zsögön *et al.,* 2018; Schouten *et al*., 2019). Reduced crop diversity poses a risk to farming, especially when most varieties have the same genetic basis for disease and insect resistance. When landraces were domesticated, favorable genotypes were chosen, resulting in the loss of alleles and a reduction in genetic diversity when compared to wild accessions (Blanca, 2015; Lin *et al*., 2014).

The goal of a breeding programme is to improve the yield and quality characteristics, while also conserving

or improving the expression of secondary traits (Nogueira *et al*., 2012). The environment, on the other hand, influences the direct selection of quantitative features, which may result in adverse changes in other traits. The only way to improve this process is to use simultaneous selection (Vasconcelos *et al*., 2010). The economic importance of tomatoes significantly contributes to the processing sector as a result of which the industries have been focusing on research in areas to breed cultivars with genes that aid yield and quality improvement with biotic and abiotic stress adaption (Parmar *et al*., 2017).

The success and effectiveness of any plant breeding strategy for selecting superior genotypes are determined by the nature and degree of genetic divergence, as well as the heritability of the desired traits. Long-term selection gain in plants could benefit from better knowledge and utilization of genetic variation. Plant breeders can pick varied parents for deliberate hybridization using precise information about the kind and level of genetic difference. Although correlation analysis aids in the selection of beneficial features through indirect genotype selection, principal component analysis (PCA) is an effective multivariate technique for identifying and determining the independent principal components that govern plant traits individually. As a result, PCA aids plant breeders in the genetic improvement of low heritability traits such as yield in any crop improvement effort.

MATERIALS AND METHODS

Thirty-six tomato accessions from the World Vegetable Centre, Taiwan and NBPGR, India were employed in the experiment which was undertaken in the Orchard of Horticulture College and Research Institute, Tamil Nadu Agricultural University during the late *Rabi,* 2019. The study was conducted in a randomized block design with three replications. Each genotype was planted in rows spaced at 90 cm with a plant to plant distance of 60 cm. All the entries were subjected to the same recommended package of practice to raise a good crop. The replicated values were used for statistical analysis. Five plants were selected from each row to record the observations on plant height (cm), the number of primary branches, days to flowering, days to 50% flowering, the number of flowers per cluster, the number of fruits per cluster, the number of clusters per plant, per cent fruit set, pericarp thickness (cm), the number of locules, T.S.S. (°Brix), ascorbic acid content (mg/100 g), lycopene content (mg/100 g), β-carotene content (mg/100 g), single fruit weight (g), the number of fruits per plant and yield per plant (kg). The principal component analysis (PCA) and correlation analysis among various traits were performed using R Studio software.

RESULTS AND DISCUSSION

Based on 17 quantitative and quality traits, principal component analysis was performed to assess the proportionate value of different components in capturing genetic variation in tomatoes collected from the World Vegetable Centre, Taiwan and NBPGR of India. PCA identifies the variable or characteristic that is responsible for clustering or grouping the population. Generally, the trait that provides maximum variation is considered for selection (Santhy *et al*., 2020). Using R software, PCA was calculated (Hammer *et al*., 2001).

In a scree plot graph, the eigenvalues associated with a factor were plotted in descending order versus the number of principal components to illustrate the per cent variation associated with the principal components. The scree plot graph showed that apart from the first seven principal components, the remaining PCs exhibited a very less amount of variation (**Fig.1**)**.** Seven of the seventeen principal components had an eigenvalue greater than one and accounted for 81.62 per cent of the entire variation in the traits studied. The eigenvalue of PC1, PC2, PC3 PC4, PC5, PC6 and PC7 were 3.609, 2.445, 2.128, 1.782, 1.566, 1.262 and 1.081, respectively (**Table 1**). PC1 accounted for 21.23 per cent of the total variation and almost all the characters under study showed positive loadings except days to flowering, days to 50% flowering, pericarp thickness, number of locules and single fruit weight. Hussain *et al.* (2018) reported similar findings.

PC1 accounted for the most heterogeneity for traits *viz.*, days to first flowering (0.572), the number of flowers per cluster (0.671), the number of fruits per cluster (0.827), per cent fruit set (0.557), ascorbic acid content (0.546), the number of fruits per plant (0.792) and yield per plant (0.571), according to factor loading of principal components. The key components of PC2, which accounted for 14.38 per cent of the total variation, were the number of clusters per plant (0.815) and yield per plant (0.682). The third primary component, pericarp thickness, was shown to be negatively loaded and responsible for 12.52 per cent of the variation (-0.554). PC4 accounted for 10.49 per cent of total variation and was linked with pericarp thickness (0.205), days to flowering (0.654) and

Table 1. Eigenvalues, per cent contribution of quantitative traits towards principal components

Principal components	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7
Eigenvalue	3.609	2.445	2.128	1.783	1.567	1.263	1.081
Per cent variance	21.231	14.383	12.519	10.488	9.215	7.429	6.357
Cumulative percentage of variance	21.231	35.614	48.133	58.621	67.836	75.265	81.622

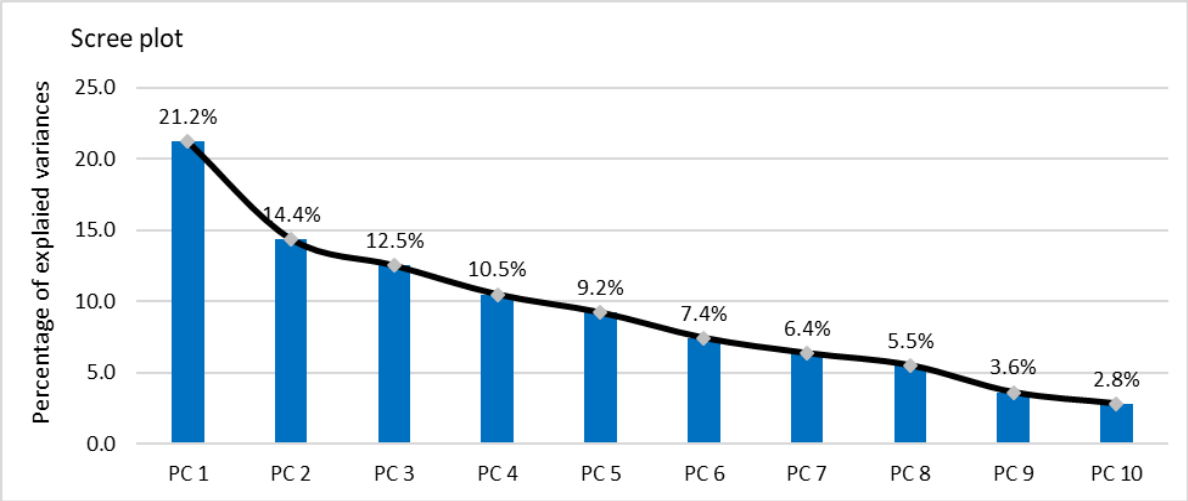

Fig. 1 . Scree plot showing contribution of various principal components towards divergence

days to 50% flowering (-0.632). The number of locules (0.581), TSS (0.721) and lycopene content (**-0.527)** were linked to PC 5 (9.22%). PC 6 was shown to be adversely correlated with plant height (0.532) and β-carotene content (-0.627). Similarly, the seventh principal component was associated with a per cent fruit set (**0.592)** (**Fig. 2**). Similar results were reported by Iqbal *et al.* (2014).

To make visualization easier, genotypes and variables were integrated into a single bi-plot graph, which was based on two primary principal components (PC1 and PC2). The PCA bi-plot graph revealed that the most discriminatory variables were the number of fruits per cluster, yieldplant, the number of fruits plant, the number of flowers cluster, per cent fruit set, ascorbic acid content, days to 50% flowering and plant height, accounting for 35.61 per cent of total variability. The genotypes *viz.,* CBESL133, CBESL129, CBESL115, CBESL121, CBESL101, CBESL136, CBESL102, CBESL103 and CBESL111 were placed farthest from bi-plot origin and accounted for more variability for traits of respective principal components than other genotypes (**Fig. 3.)** (Rai *et al*., 2017).

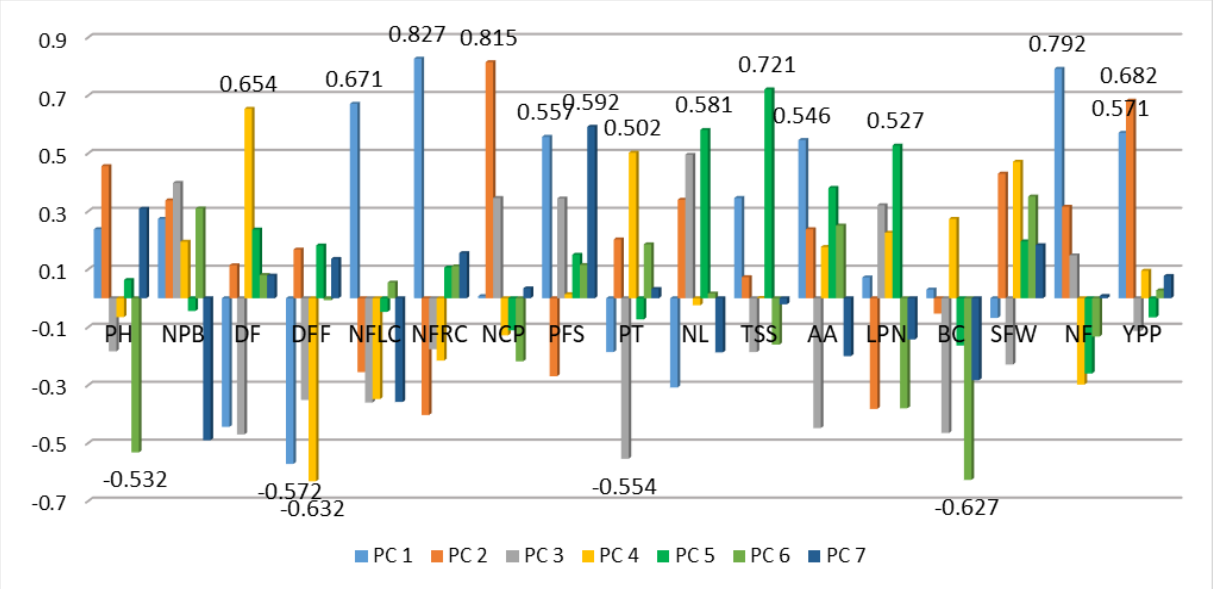

Fig. 2. Per cent contributions of seventeen quantitative characters towards principal components

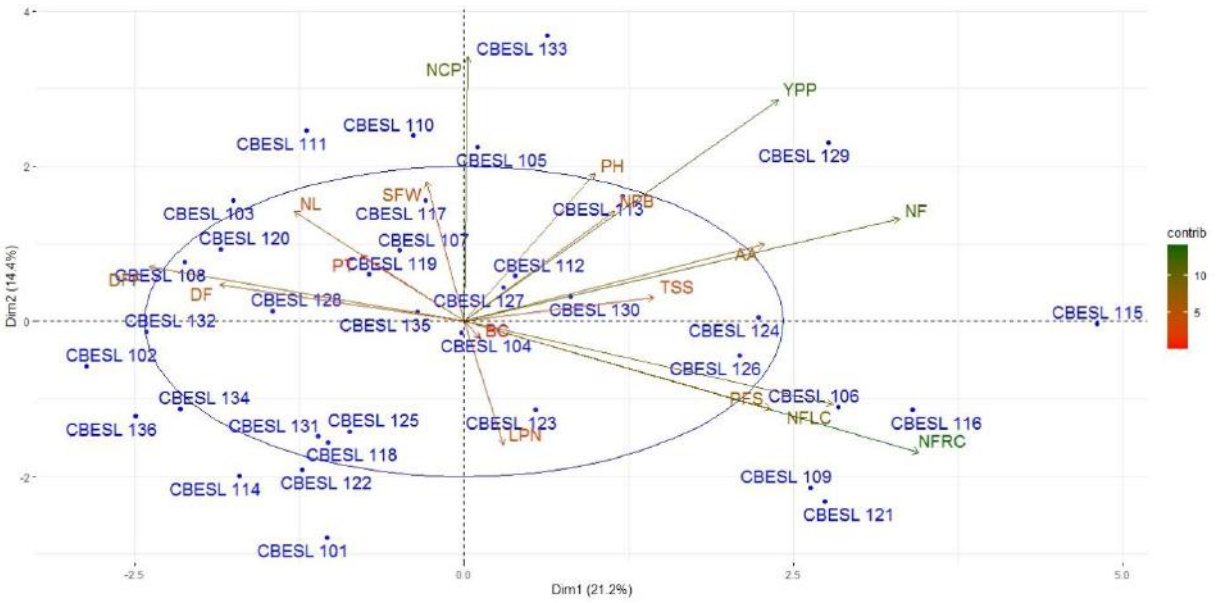

Fig. 3. Genotype scatter plot illustrating the relationship between PC1 and PC2

In the present study, the genotypes were selected on the basis of more than one PC score among the seven principal components (**Table 2**). In PC1, the positive scores ranged from 1.20 (CBESL113) to 4.81(CBESL115). In PC2, the positive values of the components ranged from 1.56 (CBESL103 & CBESL117) to 3.68 (CBESL113). Based on PCA, which highlights the characters with maximum variability, it can be concluded that PC1 and PC2 showed maximum variability for days to 50% flowering, the number of flowers per cluster, the number of fruits per cluster, per cent fruit set, ascorbic acid content, the number of fruits per plant, the number of clusters per plant and yield per plant and the top-ranked genotypes were CBESL103, CBESL105, CBESL106, CBESL109, CBESL110, CBESL111, CBESL113, CBESL115, CBESL116, CBESL117, CBESL121, CBESL124, CBESL126, CBESL129 and CBESL133. Hence, the genotypes which come under PC1 and PC2 would be useful for yield and yield related traits enhancement in the future breeding programme. The genotypes that come under PC3 and PC4 can be used to develop breeding lines for long distance transportation due to the highest pericarp thickness.

Chemical components such as acidity, ascorbic acid, lycopene, β-carotene, TSS and total sugar influence the quality and flavour of processed foods and have been shown to change substantially with variety (Balasubramanian, 1984). High total soluble solids (4-8° Brix), acidity less than 0.4 per cent, pH less than 4.5, consistent red colour, smooth surface, free of wrinkles, small core, firm flesh and uniform ripening are all desirable traits for a tomato cultivar to be used for processing (Adsule *et al*., 1980). Thus, high-quality tomatoes should be processed to produce the greatest results. Likewise, the genotypes from PC5 and PC6 showed good variation for the number of locules, TSS, lycopene content, plant height and β-carotene content in that the top-ranked genotypes are CBESL101, CBESL102, CBESL103, CBESL108, CBESL112, CBESL117, CBESL121, CBESL124, CBESL129 and CBESL120. These genotypes are useful for the development of varieties which are suitable for processing or quality improvement. These results have been emphasized by many researchers *viz*., Sattar *et al.* (2011), Lohani *et al*. (2012) and Rai *et al*. (2017). To find the best genotypes, bi-plot has been employed in potato (Ahmadizadeh and Felenji, 2011) and sweet potato (Afuape *et al.,* 2011; Sethuraman *et al*., 2007).

The seventh principal component represents a per cent fruit set. The top ranked genotypes from PC7 would be useful for the yield improvement breeding programme. Per cent fruit set flaunted a positive significant association with the number of fruits per cluster and the number of fruits per plant, while negative significant relationship with days to 50% flowering. Thus, this trait could be considered to influence the earliness of the crop simultaneously.

Significant association among various traits gives an insight into the scope of simultaneous improvement of traits and their direct and indirect effects will lead to simultaneous improvement of yield and quality. Conclusively association studies among various traits are depicted in **Fig.4**. Days to flowering, days to 50%

Table 2. Selection of genotypes on the basis of PC score in each component having positive values & more than >1.0 in each PCs

PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7
CBESL106 (2.84)	CBESL103 (1.56)	CBESL102 (1.58)	CBESL104 (1.78)	CBESL101 (1.59)	CBESL101 (2.58)	CBESL103 (1.65)
CBESL109 (2.63)	CBESL105 (2.24)	CBESL103 (1.37)	CBESL105 (1.92)	CBESL108 (2.04)	CBESL102 (1.47)	CBESL111 (1.04)
CBESL113 (1.20)	CBESL110 (2.39)	CBESL122 (1.78)	CBESL107 (1.44)	CBESL112 (1.48)	CBESL103 (1.33)	CBESL118 (1.03)
CBESL115 (4.81)	CBESL111 (2.46)	CBESL123 (1.45)	CBESL111 (1.64)	CBESL116 (4.26)	CBESL112 (1.68)	CBESL121 (1.38)
CBESL116 (3.41)	CBESL113 (1.62)	CBESL124 (1.56)	CBESL116 (1.61)	CBESL117 (1.16)	CBESL121 (1.66)	CBESL123 (1.58)
CBESL121 (2.74)	CBESL117 (1.56)	CBESL130 (3.22)	CBESL119 (1.84)	CBESL129 (1.27)	CBESL124 (2.10)	CBESL130 (1.12)
CBESL124 (2.24)	CBESL129 (2.30)	CBESL132 (1.31)	CBESL122 (1.16)	CBESL130 (2.03)	CBESL129 (1.03)	
CBESL126 (2.10)	CBESL133 (3.68)	CBESL133 (2.19)	CBESL128 (2.18)			
CBESL129 (2.77)		CBESL136 (2.03)	CBESL135 (1.01)			

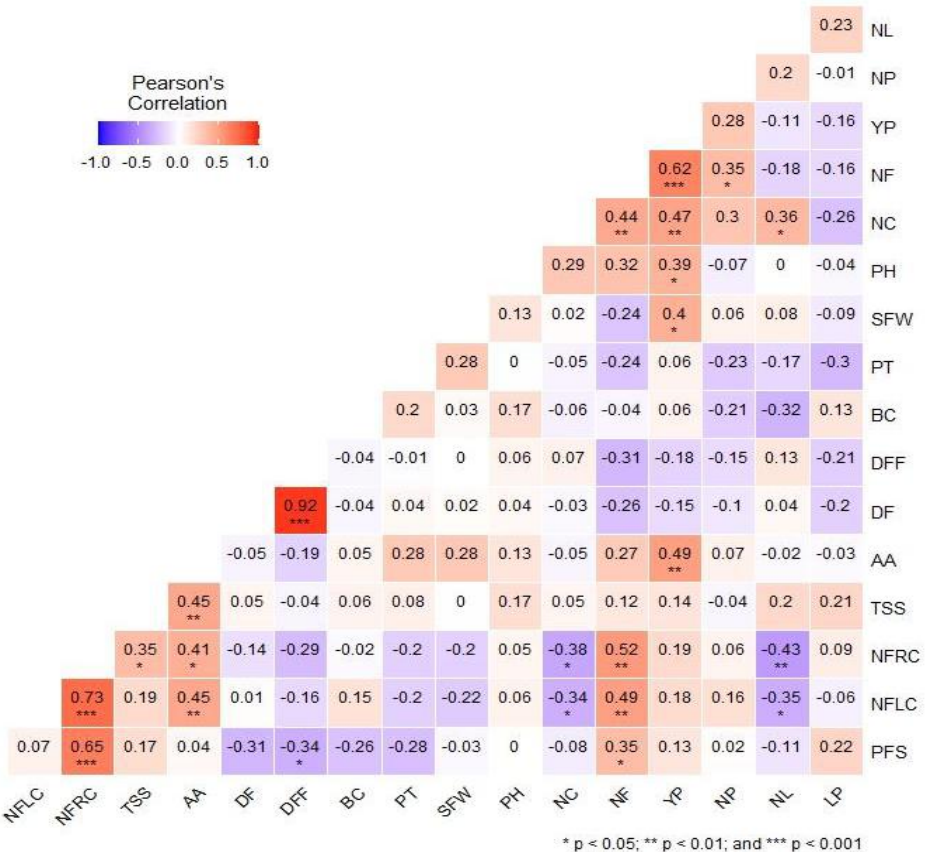

Fig. 4. Pearson's correlation coefficients among yield and yield attributes of tomato genotypes

PH - Plant height; NPB - Number of primary branches; DF - Days to flowering; DFF - Days to fifty per cent flowering; NFLC - Number of flowers per cluster; NFRC - Number of fruits per cluster; NCP - Number of clusters per plant; PFS - Per cent fruit set; PT - Pericarp thickness; NL - Number of locules; TSS – Total soluble solids; AA - ascorbic acid content; LPN -Lycopene content; BC - β-carotene content; SFW – Single fruit weight; NF - Number of fruits per plant; YPP - Yield per plant

flowering, the number of locules and lycopene content are negatively and non-significantly associated with yield per plant. The quality characters like, lycopene and β-carotene content were not significantly associated with any of the traits. The only way to incorporate lycopene and β-carotene in any breeding programme is to select parents which are superior for these quality attributes *i.e.* genotypes from PC5 and PC6.

Ascorbic acid content is positively and significantly associated with TSS, the number of fruits and flowers per cluster which reveals that breeding to improve ascorbic acid content can simultaneously improve the number of fruits per cluster, the number of flowers per cluster and TSS (**Fig. 4**). Similar finding was observed by Singh *et al.* (2018). The number of fruits per cluster, the number of flowers per cluster and the number of clusters per plant were significantly and positively associated with the number of fruits per plant which is highly and significantly correlated with yield per plant. Improving the above-mentioned traits will improve yield per plant indirectly. The results were on par with the findings of Hussain *et al.* (2018) and De-Souza *et al.* (2012).

A highly significant and positive correlation was observed for the number of fruits per plant, the number of clusters per plant and ascorbic acid content with yield per plant. Plant height, single fruit weight, days to flowering and days to 50 %t flowering were significantly and positively related to yield per plant among which days to flowering and days to 50 % flowering showed a highly significant correlation which is corroborated by Khapte and Jansirani (2014) and Nagariya *et al.* (2015). Islam *et al.* (2010) estimated a positive significant relationship between single fruit weight and yield, similarly Haydar *et al.* (2007) and Reddy *et al.* (2013) also reported the same. The number of flowers and per cent fruit set showed a positive and highly significant correlation with the number of fruits per cluster. The trait, the number of fruits per plant was positively and highly significantly associated with the number of clusters per plant, the number of flowers per cluster and the number of fruits per cluster. The number of fruits per cluster was highly significantly and negatively associated with the number of locules. Ascorbic acid content was positively significantly correlated with TSS, number of flowers per cluster and number of fruits per cluster. Selection has to be carried out for the number of fruits per plant, the number of clusters per plant, plant height, single fruit weight and ascorbic acid content in order to achieve good quality and high yielding genotypes. These factors could be considered when selecting parents for a hybridization programme aimed at widening the genetic base of the population as well as developing elite lines or heterotic F_1 hybrids.

ACKNOWLEDGEMENT

The authors are grateful to the Department of Vegetable Science for the financial and technical support of our research.

REFERENCES

Adsule, P.G., Dan, A. and Tikko, S.K. 1980. Studies on the evaluation of tomato cultivars for making juice. *Indian Food Packer*, **34**: 18-20.

Afuape, S.O., Okocha, P.I. and Njoku, D. 2011. Multivariate assessment of the agro-morphological variability and yield components among sweet potato (*Ipomoea batatas*L.) Landraces. *Afr. J. Plant Sci.,* **5**(2): 123-132.

Ahmadizadeh, M. and Felenji, H. 2011. Evaluation diversity among potato cultivars using agro-morphological and yield components in fall cultivation of jiroft area. *Am.-Eurasian j. agric. environ. Sci.,* **11**(5): 655-662.

Balasubramanian, T. 1984. Studies on quality and nutritional aspects of tomato. *J Food Sci Technol*., **21**:419-421.

Blanca, J., Montero-pau, J., Sauvage, C., Bauchet, G., Illa, E. and Díez, M. J. 2015. Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *Bmc genomics,* **16**: 257. [Cross Ref]

Chandra, S. 1977. Comparison of Mahalanobis's method and metroglyph technique in the study of genetic divergence in *Linum usitatissimum* L. Germ plasm collection. *Euphytica*, **26**: 141-148. [Cross Ref]

De-souza, L.M., Melo, P.C.T., Luders, R. R. and Melo, A.M. 2012. Correlations between yield and fruit quality characteristics of fresh market tomatoes. *Hortic. Bras*., **30**(4): 627-631. [Cross Ref]

Geraldini, F., Deleo, J.P., Julião, L., Maragon, M. and Boteon, M. 2018. Anuário 2017-2018. *Hort. Brasil.*, **16**(174): 1417.

Hammer, O., Harper, D.A.T. and Ryan, P.D. 2001. Past: paleontological statistics software package for education and data analysis. *Palaeontolelectr*., **4**: 1-9.

Haydar, A., Mandal, M.A., Ahmed, M.B., Hannan, M.M., Karim, R., Razvy, M.A., Roy, U.K. and Salahin, M. 2007. Studies on genetic variability and interrelationship among the different traits in tomato (*lycopersicon esculentum* mill.). *Middle-east j. Sci. Res*., **2**(3-4): 139-142.

Hussain, I., Khan, A.S., Ali, S., Farid, A., Ali, N., Ali,S., Masaud, S., Hussain, I., Azeem, K., and Raza, H. 2018. Genetic diversity among tomato accessions based on agro-morphological traits. *Sains Malaysiana,* **47**(11): 2637–2645. [Cross Ref]

Iqbal, Q., Saleem, M.Y., Hameed, A. and Asghar, M. 2014. Assessment of genetic divergence in tomato through agglomerative hierarchical clustering and

principal component analysis. *Pak. J. Bot.*, **46**(5): 1865-1870.

Islam, B.M.R., Ivy, N.A., Rasul, M.G. and Zakaria, M. 2010. Character association and path analysis of exotic tomato (*Solanum lycopersicum* L.) genotypes. *Bangl. J. Plant breed. Genet.*, **23**(1): 13-18. [Cross Ref]

Khanna, K. R. and Chaudhary, R.C. 1974. The nature of gene action and combining ability for some vegetative characters in tomato. *Euphytica,* **23**:159-165. [Cross Ref]

Khapte, P. S. and Jansirani, P. 2014. Correlation and path coefficient analysis in tomato (*Solanum lycopersicum* L.). *Electronic Journal of Plant Breeding*, **5**(2): 300-304.

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q. and Zheng, Z. 2014. Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.,* **46**: 1220–1226. [Cross Ref]

Lohani, M., Singh, D. and Singh, J. P. 2012. Genetic diversity assessment through principal component analysis in potato (S*olanum tuberosum* L.). *Veg. sci.,* **39**(2): 207-209.

Moll, R.H., Salhuan, W.S. and Robinson, H.F. 1962. Heterosis and genetic diversity in variety crosses of maize. *Crop Sci.*, **2**:197-209. [Cross Ref]

Nagariya, N.K., Mukherjee, S. and Sharma, N. 2015. Genetic variability, heritability and genetic advance in tomato (*Solanum lycopersicon* L.). *Ann. Biol.*, **31**: 91-93.

Nogueira, A.P.O., Sediyama, T., Sousa L.B., Hamawaki, O.T., Cruz, C.D., Pereira, D.G. and Matsuo, E. 2012. Path analysis and correlations among traits in soybean grown in two dates sowing. *Biosci. J.,* **28**(6): 877-888.

Parmar, N., Singh, K.H., Sharma, D., Singh, L., Kumar, P., Nanjundan, J., Khan, Y.J., Chauhan, D.K. and Thakur, A.K. 2017. Genetic engineering strategies for biotic and abiotic stress tolerance and quality enhancement in horticultural crops: a comprehensive review. *Biotech.*,**7**(4): 239-342**.** [Cross Ref]

Rai, A.K., Vikram, A. and Pal, S. 2017. Genetic characterization of tomato (*Solanum lycopersicum* L.) Germplasm for yield and quality traits through principal component analysis. *Res. J. Agric. Sci.,***8**(5): 1171-1174. [Cross Ref]

Reddy, B.R., Reddy, M.P., Begum, H. and Sunil, N. 2013. Genetic diversity studies in tomato (*Solanum lycopersicum* L.). *Iosr j. Agric. Veterinary sci.*, **4**(4): 53-55. [Cross Ref]

Santhy, V., Rathinavel, K., Saravanan, M., Mithila, M. and Priyadharshini, C. 2020. Genetic diversity assessment of extant cotton varieties based on principal component analysis (pca) and cluster analysis of enlisted DUS traits. *Electronic Journal of Plant Breed*ing, **11**(2): 430-438. [Cross Ref]

Sattar, M.A., Uddin, M.Z., Islam, M.R., Bhuiyan, M.K.R. and Rahman,M. S. 2011. Genetic divergence in potato (*Solanum tuberosum* L.). *Bangladesh j. agric. Res.,* **36**(1): 165-172. [Cross Ref]

Schouten, H.J., Tikunov, Y., Verkerke, W., Finkers, R., Bovy, A., Bai, Y. and Visser, R.G.F. 2019. Breeding has increased the diversity of cultivated tomato in the Netherlands. *Front. Plant sci.,***10**:1606. [Cross Ref]

Sethuraman, S. P., Ray, R.C., Panda, S.H. and Naskar, S.K. 2007. Application of principal component analysis for sensory characterization of sweet potato curd. *J. root crops.,* **33**(1): 62-66.

Singh, A.K., Solankey, S.S., Akhtar, S., Kumari, P. and Chaurasiya, J. 2018. Correlation and path coefficient analysis in tomato (*Solanum lycopersicum* L.). *Int.j.curr.microbiol.app.sci., Special issue-***7**: 4278-4285.

Vasconcelos, E.S., Ferreira, R.P., Cruz, C.D., Moreira, A., Rassini, J.B. and Freitas, A.R. 2010. Estimates of genetic gain by different selection criteria in alfalfa genotypes. *Rev. Ceres.*, **57**(2): 205-210. [Cross Ref]

Zsögön, A., Čermák, T., Naves, E. R., Notini, M. M., Edel, K. H. and Weinl, S. 2018. *De novo* domestication of wild tomato using genome editing. *Nat. Biotechnol.,* **36**: 1211–1216. [Cross Ref]