

ASSESSMENT OF GENETIC DIVERSITY AND IDENTIFICATION OF SUPERIOR DONORS FOR HIGHER YIELD, NUTRITIONAL AND PROCESSING TRAITS IN TOMATO THROUGH HIERARCHICAL CLUSTERING ANALYSIS

Comment [A1]: Add scientific name of the crop

Abstract

The present investigation was carried out in the Department of Horticulture, BAC, BAU, Sabour, Bihar, with 22 diverse genotypes of tomato germplasm collected from diverse locations and sources. The 22 genotypes were assessed for their phenological, morphological, yield, nutritional and processing quality attributes. The hierarchical clustering analysis based on the Euclidean distance and Ward's method were applied on these genotypes separately based on two segments namely the morphological and yield attributes; and the nutritional and processing quality attributes, aimed to find superior genotypes from both segments. Further, through the same analysis, the genetic diversity among the 22 genotypes for these traits were also established. Additionally, the association among the plant traits are also partly reported with a heatmap visualization. Our findings suggest that the genotypes 2019/TODVAR-4, 2019/TODVAR-5, 2019/TODVAR-2, 2019/TODVAR-7, and 2019/TOLOCVRES-4 exhibit superior processing quality attributes. In particular, genotypes with higher yields coupled with favorable nutritional and processing qualities significantly enhance the success of breeding programs. Notably, genotype 2019/TODVAR-4 has been identified as particularly suitable for developing cultivars with excellent processing qualities.

Keywords: Tomato; Genetic diversity; Nutritional & processing quality; Identification of superior genotypes; Hierarchical clustering.

Comment [A2]: Expand and write AND instead of &

1. Introduction

Tomato (*Solanum lycopersicum* L.) is one of the most important warm season fruit vegetable crops grown throughout the world. India holds second position as a leading producer of tomato in the world.

It belongs to the family Solanaceae and has chromosome number $2n=24$. It produces chasmogamous flowers where predominantly self-pollination is seen.

It is mostly known as “Protective food” due to its nutritive value, antioxidant molecules such as carotenoids, particularly

lycopene, ascorbic acid, vitamin E and phenol compounds, mainly flavonoids. It

also contains minerals like iron, phosphorus, and potassium (Brar *et al.*, 1998).

Tomato fruits can be used both as fresh as well as processed form.

It ranks first in the list of processing vegetables in India. Fruits

used for making processed foods include ketchup, puree, powder, paste, sauces, and soups.

Lycopene has been shown in studies to offer significant health benefits, since it protects against the oxidation of free radicals. It may also stop low-

density lipoproteins from oxidising (bad cholesterol). Because lycopene is a powerful antioxidant.

It has been associated to a decreased risk of some forms of cancer in those who consume it because of its nutritional content. Human beings (Chernet,

2013). It protects the human body against disease due to its nutritious value. A typical processing genotype should possess high TSS ($>5^\circ$ Brix), minimum sugar acid ratio (15:1), lycopene ($>10\text{mg}/100\text{g FW}$) with a high colour value of (>2) and low pH (<4.3) that improve quality of the valued-added products, and reduce energy consumption and cost of processing (NAAS, 2020).

By choosing apex genotypes for quick increase in yield and other horticultural traits, genetic variability plays a significant role in crop breeding programmes. The more the genetic diversity in qualitative and quantitative features, the higher the prospects of crop development through selection. The multivariate analysis using hierarchical clustering analysis using Ward's algorithm is a valuable tool to quantify the degree of divergence at genetic level. While formulating the tomato crop improvement program, understanding about the nature and degree of genetic divergence available in the germplasm plays a pivotal role. It is well recognized that the use of diverse parents results in superior hybrids and desirable recombinants. Thus, genetic divergence existing in the population helps in selecting suitable parents for hybridization program.

Considering the potential and demand of tomato crop, there is an urgent need to identify and develop varieties/genotypes suitable for cultivation under different agro-

climatic conditions of Sabour, Bhagalpur,

Bihar. However, a lot of work has been carried out on crop improvement in tomato elsewhere, but meager work in this line has been carried out for Bhagalpur region. Keeping the above themes in view, this investigation was undertaken to assess the genetic diversity among the evaluated genotypes and to identify superior and distant lines to deploy them in the future breeding programs for processing and nutritional quality.

2. Materials and methods

The present experiment conducted at Department

of Horticulture, BAC, BAU, Sabour, Bhagalpur during *grab* season of the year 2021-

2022. Geographically, Sabour is situated at $25^\circ 15' 40''$ N latitude and $87^\circ 2' 42''$ E longitude with an altitude of 45.57 meter above mean sea level (MSL). This place is characterized by semi-arid and sub-tropical climate with dry summer, average precipitation and cold winter. The experimental

materials included 22 diverse genotypes of tomato collected from diverse locations and sources (Table 1).

Comment [A3]: No check varieties in the study?

Table 1. List of tomato genotypes and their sources

S. No.	Name of Variety	Sources
1	2019/TODVAR-1	AICRP on vegetablecrops
2	2019/TODVAR-2	AICRP on vegetablecrops
3	2019/TODVAR-3	AICRP on vegetablecrops
4	2019/TODVAR-4	AICRP on vegetablecrops
5	2019/TODVAR-5	AICRP on vegetablecrops
6	2019/TODVAR-6	AICRP on vegetablecrops
7	2019/TODVAR-7	AICRP on vegetablecrops
8	2019/TODVAR-9	AICRP on vegetablecrops
9	2019//TOLCVRES-2	AICRP on vegetablecrops
10	2019//TOLCVRES-3	AICRP on vegetablecrops
11	2019//TOLCVRES-4	AICRP on vegetablecrops
12	2019//TOLCVRES-5	AICRP on vegetablecrops
13	2019//TOLCVRES-6	AICRP on vegetablecrops
14	2019//TOLCVRES-7	AICRP on vegetablecrops
15	2019//TOLCVRES-8	AICRP on vegetablecrops
16	PunjabChhuhara	PAU,Ludhiana, Punjab
17	KashiAmrit	IIVR, Varanasi, Uttar Pradesh
18	KashiChayan	IIVR, Varanasi, Uttar Pradesh
19	IIHR-2614	IIHR,Bengaluru, Karnataka
20	Arka Vikas	IIHR,Bengaluru, Karnataka
21	Arka Alok	IIHR,Bengaluru, Karnataka
22	BRDT-1	BAU,Sabour, Bihar

Comment [A4]: Are these varieties or genotypes? Specify

The crop production was carried out by following the standard package of practices recommended for Sabour region of Bihar. The plant morphological traits namely days to 50% flowering (DFF), primary branches plant⁻¹ (PBP), plant height (PH), days to first fruit set (DFFS), days to maturity (DM), average single fruit weight (FW), locules fruit⁻¹ (LF), fruit polar diameter (FPD), fruit equatorial diameter (FED), pericarp thickness (PT), fruits plant⁻¹ (FP), fruit yield plant⁻¹ (FYP) were observed in randomly selected four plants. For the same plants, biochemical attributes namely the total lycopene content (LYC), titrable acidity (TA), total soluble solids (TSS), ascorbic acid (ASC), total carotenoids (CAR) and β -carotene (B-CAR) were determined. The total lycopene content of tomato fruit was determined using Lee's (2001) method with some

modifications. Titrable acidity was determined by using titration method (AOAC, 2000). The TSS was assessed with a digital refractometer. The amount of ascorbic acid in the juice was measured by titrating it against 2,6-dichlorophenolindophenol dye (AOAC, 2000). Total carotenoids and β -carotene contents were estimated as per Sadasivam and Manickam (1996) using composite sample of five fruits from each replication, in a UV-VIS spectrophotometer (Labman) at 452 nm wavelength. The analysis of variance (ANOVA) of the observations recorded on different attributes was performed according to Panse and Sukhatme's standard approach (1985). The significance of the results was determined using the F table values (Fisher and Yates, 1963). Grouping of genotypes and subsequent heatmap for chromic visualization of mean performances of the genotypes were done based on plant morphological, yield, nutritional and processing quality attributes using the cluster analysis through Ward's method and Euclidian distance in RStudio (Version 4.3.3, RStudio Team, 2020).

3. Results and discussion

3.1. Genetic diversity among 22 tomato genotypes based on the morphological and yield attributes

The morphological and yield attributes namely the days to 50% flowering (DFF), days to first fruit set (DFFS), days to fruit maturity (DFM), number of fruits plant⁻¹ (NFP), number of primary braches plant⁻¹ (NPB), plant height at 90 days after sowing (PH90) fruit yield plant⁻¹ (FYP) were recorded from the plants of 22 tomato genotypes. The hierarchical clustering pattern, presented in Fig. 1a, was constructed based on Euclidean distance among the genotypes using the Ward's algorithm. The clustering analysis grouped 22 genotypes into 5 clusters: cluster I with 8 genotypes (2019/TODVAR-2, Punjab chhuhara, 2019/TODVAR-5, 2019/TOLCVRES-7, 2019/TODVAR-6, Kashi Amrit, 2019/TOLCVRES-2 and 2019/TOLCVRES-6); cluster II with 2 genotypes (IIHR-2614 and Arka Vikas); cluster III with 3 genotypes (2019/TODVAR-1, Kashi Chayan (C) and 2019/TODVAR-3); cluster IV with 2 genotypes (Arka Alok and BRDT-1) and cluster V with 7 genotypes (2019/TOLCVRES-5, 2019/TODVAR-7, 2019/TODVAR-4, 2019/TOLCVRES-8, 2019/TOLCVRES-4, 2019/TODVAR-9 and 2019/TOLCVRES-3). From this, it is evident that the evaluated genotypes are diverse and thus suitable for subjecting them for breeding cultivars with high yielding traits. Nankar *et al.* (2020) studied the diversity of 150 accessions of tomato genotypes across the world for their morphological and biochemical traits, through hierarchical clustering method. Their findings revealed better genotypes for future breeding programs in tomato. Similar studies were also performed by Mahesha *et al.* (2006), Meena and Bahadur (2015), Jogi *et al.* (2008), Rana and Singh (2010), and Lekshmi and Celine (2016).

Comment [A6]: Check the spelling of genotypes

Comment [A7]: Check the reference writing format and change accordingly in the whole manuscript

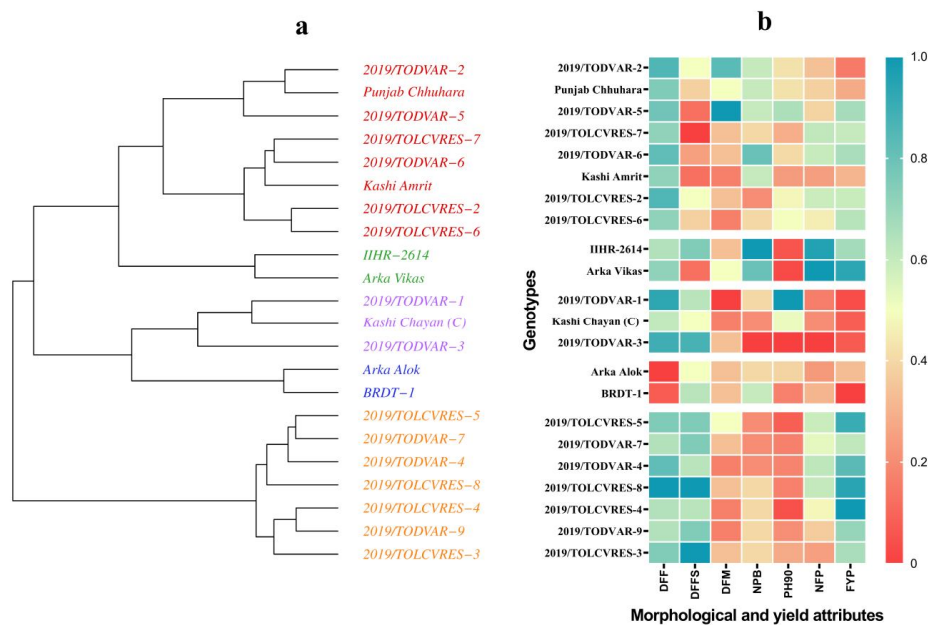


Fig. 1. Genetic diversity based on morphological and yield attributes. a) dendrogram from Ward's algorithm; b) heatmap showing the performance of the 22 genotypes

3.2. Genetic diversity among 22 tomato genotypes based on the nutritional and quality attributes of fruits

The harvested fruits of the 22 genotypes were assessed for their nutritional attributes namely lycopene, titrable acidity, total soluble solids (TSS), ascorbic acid (ASC), total carotenoids (CAR) and β -carotene (B-CAR) and processing quality attributes like fruit weight (FW), number of locules fruit⁻¹ (NLF), fruit polar diameter (FPD), fruit equatorial diameter (FED) and pericarp thickness (PT). The Fig. 2a shows the clustering pattern based on these traits, the 22 genotypes were categorized into 5 clusters: cluster I with 3 genotypes (Kashi Amrit, Punjab chhuhara, 2019/TOLCVRES-2); cluster II with 3 genotypes (2019/TODVAR-5, 2019/TOLCVRES-4, 2019/TOLCVRES-3); cluster III with 6 genotypes (2019/TODVAR-9, 2019/TOLCVRES-8, 2019/TOLCVRES-5, IIHR-2614, 2019/TODVAR-4 and Arka Vikas); cluster IV with 7 genotypes (Kashi Chayan (C), Arka Alok, 2019/TODVAR-2, 2019/TODVAR-7, BRDT-1, 2019/TODVAR-3 and 2019/TOLCVRES-7); cluster V with 3 genotypes (2019/TODVAR-1, 2019/TOLCVRES-6 and 2019/TODVAR-6). This shows that the evaluated genotypes are diverse for the nutritional and processing traits too, indicating their prominent potential for utilization as donors in tomato breeding programs aimed at enhanced nutritional and processing quality attributes. Lekshmi and Celine (2016) reported that wide ranges of variation were observed among the characters studied which have a great interest for polyhouse tomato breeding. Genetic divergence analysis was carried out using Mahalanobis D₂ statistics and the 40 tomato genotypes were grouped into eight clusters. Cluster I was the largest cluster with twenty four genotypes followed by cluster II with ten genotypes and all other clusters were solitary. The highest intra- cluster distance

Comment [A8]: Is it titrable or titratable? Check and correct the spelling in the complete manuscript

Comment [A9]: D² not D₂

Comment [A10]: 22 or 40 genotypes?

Comment [A11]: Number of genotypes how varied in the study?

was noticed in cluster II followed by cluster I. The highest inter- cluster distance was observed between clusters VII and VIII, followed by clusters IV and VIII. Lycopene content and **truss** per plant had maximum contribution towards total divergence followed by fruit length, fruit weight and yield per plant.

Comment [A12]: What is the meaning of truss?

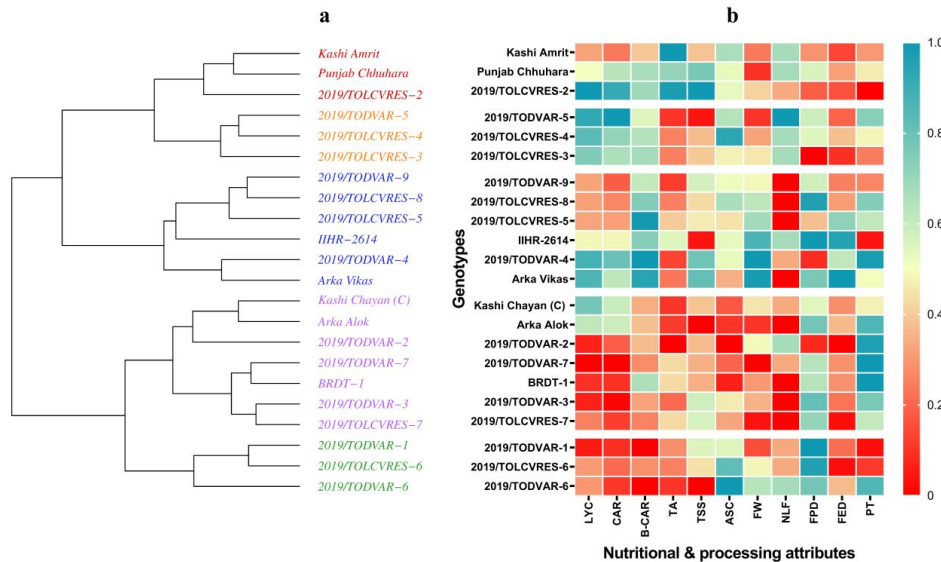


Fig. 2. Genetic diversity based on nutritional and processing quality attributes. a) dendrogram from Ward's algorithm; b) heatmap showing the performance of the 22 genotypes

3.3. Identification of superior donors for breeding towards higher yield, nutritional and processing quality attributes

Fig. 1b shows the mean performance trend of the morphological and yield attributes of 22 genotypes (data not shown). A correlation among the morphological traits were hard found, except for days to 50% flowering and days to first fruit set and yield with fruit related traits. The clusters 2 and 5 had higher yielding genotypes, whereas the genotypes under cluster 3 and 4 had poor performers, while the genotypes under cluster 1 were intermediate to both high and low extremes. The heatmap also showed a negative correlation between fruits yield plant⁻¹ and plant height across all the genotypes. We found it interesting that the poor yielders of cluster 3 and 4 as well as the good yielders of cluster 5 showed lower number of primary branches and number of fruits plant⁻¹. Whereas the genotypes under cluster 2 (IIHR-2614 and Arka Vikas) showed higher yield coupled with higher number of branches plant⁻¹ and number of fruits plant⁻¹. More are the primary branches, more is the number of fruits plant⁻¹, hence is the yield. Thus, the number of primary branches plant⁻¹ is a primary and direct selection trait towards improving yield. The genotypes Arka Vikas, 2019/TOLCVRES-4, 2019/TOLCVRES-8, 2019/TOLCVRES-5 and 2019/TOLCVRES-4 are identified as the higher yielders and therefore can be used in breeding programs for yield maximization, owing to their exceptional performance surpassing the check cv. Kashi Chayan. In addition to the identification of high yielding lines, our results also shed insights on the scopes for improving **othr** genotypes. The line IIFR-2614 showed the highest number of primary

Comment [A13]: Word should be italicise

Comment [A14]: Other

braches plant⁻¹, however it couldn't be a superior yielder and was found shorter in yield than the highest yielders identified and mentioned above. Since, number of primary branches plant⁻¹ is a major target for yield maximization, IIHR-2614 could be used in hybridization efforts with the high yielders of cluster 3 and 5. The increase in number of primary branches plant⁻¹ in these poor branch-bearing genotypes would significantly boost the yield by manifolds. Essentially, the underlying cause for lower yields of IIHR-2614 despite its higher number of primary branches plant⁻¹ needs further investigation to identify the negative drivers of the yield reduction. The genotypes under cluster 4 (Arka Alok and BRDT-1) showed earliest flowering and maturity but are poor yielders. However, early maturity of these genotypes could be incorporated into high yielders of clusters 3 and 5, thereby coupling early maturity and higher yielders, enabling farmers to find places and profits in lean seasons. In addition to the days to 50% flowering, the days to first fruit set and days to fruit maturity are also important traits deciding the early harvest. Though Arka Alok and BRDT-1 showed the earliest flowering, the earliest fruit set and maturity were seen with 2019/TOLCVRES-7 (cluster 1) and 2019/TODVAR-1 (cluster 3). These lines could also be deployed to breed early and high yielding cultivars. Since the fruit yield and its drivers showed no significant correlation (data not shown), breeding superior lines with higher yield and early maturity, with suitable plant architecture (plant height and number branches plant⁻¹) is possible through hybridization.

Fig. 2b gives a comprehensive picture on the phenotypic variability, genotypic diversity, correlation and lines that scored maximum for nutritional and processing qualities. There was hardly a correlation among the nutritional traits and processing quality traits, except for positive correlation among the total carotenoids, β -carotene and lycopene; fruit weight and β -carotene; fruit equatorial diameter and β -carotene (data not shown). The desirable fruit weight for processing is >80g. In our investigation, only the genotypes 2019/TODVAR-4 had an average fruit weight of 80.75 g (data not shown). The desirable number of locules fruit⁻¹ is 2-4, which was observed in all the genotypes, except in 2019/TODVAR-5. The pericarp thickness greater than 0.4cm is desirable for processing cultivars. Most of the genotypes under the study showed a pericarp thickness > 0.4cm. Particularly, 2019/TODVAR-7 and BRDT-1 showed the highest pericarp thickness of 7.42 mm, followed by 2019/TODVAR-4 (7.33 mm) and 2019/TODVAR-2 (7.26 mm). The desirable lycopene in fruit for processing purpose is >8.5 mg 100g⁻¹. The highest lycopene contents were seen with 2019/TOCVRES-2 (10.98), followed by 2019/TODVAR-5 (10.45) and 2019/TODVAR-4 (10.15). The titrable acidity needs to be as low as 0.40%. The genotypes 2019/TODVAR-2 showed the least acidity (0.18%), followed by 2019/TODVAR-5, 2019/TODVAR-6 and Kashi Chayan (C) (0.25%). The desirable TSS content is > 5.5° Brix. The ascorbic acid content for processing types is recommended to be >25 mg 100mg⁻¹. In our investigation, none of the genotypes had an ascorbic acid content > 25 mg 100mg⁻¹. However, 2019/TODVAR-6 (24.28) and 2019/TOCVRES-4 (23.27) and 2019/TOCVRES-6 (21.44) showed greater ascorbic acid contents.

Our findings reveal that the genotypes namely 2019/TODVAR-4, 2019/TODVAR-5, 2019/TODVAR-2, 2019/TODVAR-7 AND 2019/TOLOCVRES-4 are superior in the processing quality attributes. However, the genotypes with higher yield and good nutritional and processing qualities make the breeding efforts more fruitful. In this context, the

Comment [A15]: Check the spelling of acidity also

genotype 2019/TODVAR-4 is also found to be suitable for developing cultivars with good processing qualities. It is important to note that these genotypes are better performers, both by means of yield; and nutritional and processing attributes, thereby are potential donors in future breeding programs. These genotypes would form a population upon which new cultivars could be developed, that would possibly surpass the existing cultivars for processing namely Punjab Chhuhara and Kashi Chayan (C). Further efforts in this direction with these genotypes would fulfil the never-ending and ever-increasing demand for food and nutritional security of the human kind.

4. Conclusion

Based on the morphological and yield attributes, the clustering analysis grouped 22 genotypes into 5 clusters: cluster I with 8 genotypes (2019/TODVAR-2, Punjab chhuhara, 2019/TODVAR-5, 2019/TOLCVRES-7, 2019/TODVAR-6, Kashi Amrit, 2019/TOLCVRES-2 and 2019/TOLCVRES-6); cluster II with 2 genotypes (IHR-2614 and Arka Vikas); cluster III with 3 genotypes (2019/TODVAR-1, Kashi Chayan (C) and 2019/TODVAR-3); cluster IV with 2 genotypes (Arka Alok and BRDT-1) and cluster V with 7 genotypes (2019/TOLCVRES-5, 2019/TODVAR-7, 2019/TODVAR-4, 2019/TOLCVRES-8, 2019/TOLCVRES-4, 2019/TODVAR-9 and 2019/TOLCVRES-3). Similarly, the clustering pattern based on the nutritional and processing traits, the 22 genotypes were categorized into 5 clusters: cluster I with 3 genotypes (Kashi Amrit, Punjab chhuhara, 2019/TOLCVRES-2); cluster II with 3 genotypes (2019/TODVAR-5, 2019/TOLCVRES-4, 2019/TOLCVRES-3); cluster III with 6 genotypes (2019/TODVAR-9, 2019/TOLCVRES-8, 2019/TOLCVRES-5, IHR-2614, 2019/TODVAR-4 and Arka Vikas); cluster IV with 7 genotypes (Kashi Chayan (C), Arka Alok, 2019/TODVAR-2, 2019/TODVAR-7, BRDT-1, 2019/TODVAR-3 and 2019/TOLCVRES-7); cluster V with 3 genotypes (2019/TODVAR-1, 2019/TOLCVRES-6 and 2019/TODVAR-6). This shows that the evaluated genotypes are diverse for their morphological, nutritional and processing traits along with yield levels, indicating their prominent potential for utilization as donors in tomato breeding programs aimed at enhanced nutritional and processing quality attributes coupled with high yields. Our results recommend that genotypes namely 2019/TODVAR-4, 2019/TODVAR-5, 2019/TODVAR-2, 2019/TODVAR-7 AND 2019/TOLCVRES-4 are superior in the processing quality attributes. Particularly, the genotypes with higher yield and good nutritional and processing qualities make the breeding efforts more fruitful. In this context, the genotype 2019/TODVAR-4 is also found to be suitable for developing cultivars with good processing qualities.

References

1. AOAC (2000) Official methods of analysis. Washington, DC: Association of Official Analytical Chemists.
2. Brar PS, Singh H, Singh H (1998) Variability and correlation studies in different varieties of tomato (*Lycopersicon esculentum* Mill.). Punjab Vegetable Grower, 33:23-6.
3. Chernet S, Belew D, Abay F (2013) Genetic variability and association of characters in tomato (*Solanum lycopersicon* L.) genotypes in Northern Ethiopia. Int. J. Agric. Res. 8(2):67-76.
4. Fisher RA, Yates F (1953) Statistical tables for biological, agricultural and medical research. Hafner Publishing Company.

Comment [A16]: There bygive space

Comment [A17]: Scientific name should be italicised in all the references

5. Jogi P, Shukla N, Sahu M, Mehta N (2008) Correlation and path analysis for fruit traits in tomato (*Lycopersicon esculentum* Mill.). *Haryana Journal of Horticultural Sciences*, 37(3/4):301-2.
6. Lee HS (2001) Characterization of carotenoids in juice of red navel orange (Cara Cara).
7. Lekshmi SL, Celine VA (2016) Genetic diversity studies in tomato (*Solanum lycopersicum* L.) under protected conditions. *International Journal of Current Microbiology and Applied Sciences*, 5(4):212-7.
8. Mahesha DK, Apte UB, Jadhav BB (2006) Studies on genetic divergence in tomato (*Lycopersicon esculentum* Mill.). *CROP RESEARCH-HISAR*, 32(3):401.
9. Meena OP, Bahadur V (2015) Genetic associations analysis for fruit yield and its contributing traits of indeterminate tomato (*Solanum lycopersicum* L.) germplasm under open field condition. *Journal of Agricultural Science*, Mar 1;7(3):148.
10. NAAS (2022) Need for Breeding Tomatoes Suitable for Processing. Strategy No. 16, National Academy of Agricultural Sciences, New Delhi, India.
11. Nankar AN, Tringovska I, Grozeva S, Ganeva D and Kostova D (2020) Tomato Phenotypic Diversity Determined by Combined Approaches of Conventional and High-Throughput Tomato Analyzer Phenotyping. *Plants*, 9: 197. <https://doi.org/10.3390/plants9020197>.
12. Rana DK, Singh RV, Mishra AC (2008) Genetic variability in tomato (*Lycopersicon esculentum* Mill.) genotypes. *Progressive Horticulture*. 40(1):53-7.
13. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
14. S. Sadasivam and A. Manickam (1996) "Biochemical Methods," New Age International (P) Limited, New Delhi, Vol. 2. pp. 124-126.