Assessment of Genetic Divergence in Tomato (*Solanum lycopersicum* L.) through Clustering and Principal Component Analysis

Md Sarowar Alam*[1,3], Shahadat Hossain[2,3], Md Arfan Ali[3], Md Golap Hossain[3] and Md Faridul Islam[1]
[1]Regional Agricultural Research Station, Bangladesh Agricultural Research Institute, Akbarpur, Moulvibazar, Bangladesh
[2]Regional Agricultural Research Station, Bangladesh Agricultural Research Institute, Hathazari, Chattogram, Bangladesh
[3]Department of Arid Land Agriculture, Faculty of Meterology, Environment and Arid Land Agriculture, King Abdulaziz University, P.O. Box: 80208, Jeddah, Saudi Arabia
*Corresponding author's email: asarowar04bau@gmail.com

***Abstract*— The genetic divergence of tomato was studied with twenty three selected genotypes using D^2 statistics and principal component analysis at Regional Agricultural Research Station, BARI, Akbarpur, Moulvibazar, Bangladesh during 2014 to 2015. The genotypes were grouped into 5 clusters and the maximum number of genotypes was included in cluster I and the minimum number in cluster V. The highest intra cluster distance was observed for the cluster II and the lowest for the cluster V. For cluster III, the highest mean values for days for 50% flowering, individual fruit weight, fruit diameter, pericarp thickness, number of locules per fruit, yield per plant, yield ton per hectare were recorded. The first axis largely accounted for the variation among the tomato genotypes (50.88%) followed by second axis (20.33%). The first five axes accounted 91.71 % of the total variation. Considering the magnitude of cluster means for different traits and performance the genotypes of cluster III and V may be considered as parents for future hybridization program for improvement of tomato.**

Keywords*— Tomato; genetic divergence; clustering; D^2 analysis; PCA

I. INTRODUCTION

Tomato (*Solanum lycopersicum* L., 2n=2x=24) is one of the most important and popular vegetable in the world due to its wider adaptability, higher yield potentiality and suitability for diversified uses in fresh as well as processed food industries [1, 2]. It belongs to the family Solanaceae and its center of origin in Peru Equador region [3, 4] and is normally a self-pollinated crop. Tomato is the second most important vegetables after potato [5] and it has great demand as cash crop in the international market [6- 8]. Tomatoes are rich source of minerals such as Ca, P and Fe and vitamin A, C and antioxidants such as lycopene, glutathione etc. [9-13]. Tomatoes are main source of lycopene [14] and the lycopene level in tomato fruit increases 500 times in ripening [15]. High antioxidants of tomato eliminate reactive oxygen species (ROS) and thus lowering the risk of certain chronic diseases such as cancer, strokes etc. in human body [16]. Furthermore, consumption of tomatoes prevents cardiovascular diseases [17, 18] and some other types of cancers, as for example prostate cancer [19, 20].

The success and efficiency of any plant breeding scheme for selecting superior genotypes depends upon the nature and extent of genetic divergence and the extent to heritability of the characters of interest [21]. Better understanding and exploitation of genetic diversity could be helpful to ascertain long term selection gain in plants [22]. Multivariate analysis such as D^2 cluster and factor analysis are useful and effective method for selecting genotypes in any hybridization program. D^2 analysis has been successfully utilized in plant breeding for measuring the diversity in several crops [23]. An understanding of nature and magnitude of variability among the tomato germplasm is a prerequisite for its improvement. Precise knowledge on the type and extent of genetic divergence helps the plant breeder in selecting the diverse parents for purposeful hybridization [24, 25]. Although correlation analysis helps in selection of effective characters with indirect selection of superior genotypes but principal component analysis (PCA) is an efficient multivariate technique to identify and determine the independent principal components that governs plant traits separately. Therefore, PCA also helps the plant breeders for genetic improvement of traits such as yield that have low heritability in any crop improvement program [26, 27]. So, the present study has been undertaken with 23 tomato genotypes to understand the nature and magnitude of genetic divergence and the traits contributing genetic divergence by D^2 analysis for improvement of tomato.

II. MATERIALS AND METHODS

The experiment was conducted at the vegetables research field and laboratory of the Regional Agricultural Research Station, Bangladesh Agricultural Research Institute, Akbarpur, Moulvibazar, Bangladesh during Rabi season from October 2014 to April 2015. Soil texture was sandy clay (43-85%), silt (<50%) and clay (>20%) and high land soil type with pH 4.5. Twenty three tomato genotypes were used in the present study and the experiment was laid out in randomized complete block design (RCBD) with 3 replications. The unit plot size was 4.8 × 1.0 m and plant spacing was 60 × 40 cm. Manure and Fertilizers were applied @10 tons well decomposed cowdung, 550 kg Urea, 450 kg TSP and 250 kg MP, Gypsum 121 kg, Zinc Sulphate 15 kg and Boric acid 12 kg per hectare. Half of

the quantity of cowdung, half amount of TSP and entire amount of gypsum and boric acid were applied during land preparation. The remaining half of cowdung and TSP was applied during pit preparation before a week of planting. The entire Urea and MP were applied in 3 equal installments at 21, 35 and 50 days after transplanting. Irrigation, intercultural operation and pest management were done as and when necessary. Data on plant height (cm), days for 50% flowering, number of flowers per infloresence, number of fruits per cluster, individual fruit weight (g), fruit length (cm) and diameter (cm), pericarp thickness (cm), number of locules per fruit, number of fruits per plant, number of seeds per fruit, % Brix (TSS), yield per plant (kg) and yield (t/ha) were recorded. Genetic diversity was studied following Malanobsis's [23] generalized distance (D^2) extended by Rao [28]. Clustering of genotypes was done according to Tocher's Method [28] and principal component analysis was done according to Rao [29]. Mean data for each character were subjected to multivariate principal component analysis (PCA), principal coordinate analysis (PCO), cluster analysis and canonical variate analysis (CVA) using GENSTAT 5.5 computer software. Average intra cluster distance was calculated by the formula as suggested by Singh and Chaudhury [30].

III. Results and Discussions

On the basis of D^2 analysis, twenty three genotypes of tomato were grouped into five clusters based on D^2 values (Table 1). The distribution pattern indicate that the maximum numbers (8) of genotypes was included in cluster I followed by cluster IV and III. The minimum number of genotype (01) was included cluster V. The grouping pattern of the genotypes was found to be random proving that the geographical and genetic diversity were unrelated. Similar grouping pattern in tomatoes were also confirmed by other researchers [31-33].

Table 1. Distribution of 23 tomato genotypes in five clusters

Cluster	Numbers	Accessions	Percentage
I	8	BARI tomato-3, BARI tomato-14, BARI tomato-15, GPT-011, GPT-015, GPT-017, SL-010, GWT-043	34.7826
II	3	BARI tomato-11, SL-011, SL-012	13.04
III	4	GPT-037, GPT-053, GBT-056, GWT-052	17.3913
IV	7	GPT-009, SL-001, SL-003, SL-008, SL-009, AVTOV-1010, AVTOV-1007	30.4348
V	1	Chimacherry	4.3478

Intra and inter cluster distances are presented in Table 2. The inter cluster distances were higher than the average intra cluster distances which revealed a wide genetic divergence among the tomato genotypes of different groups than those of same cluster. These findings were confirmed by other research in tomato [33-35], in brinjal [36], in lablab bean [37] and in pummelo [38]. The highest inter cluster distance was observed between cluster III and IV (51.52) and followed by cluster I and V (50.24) and the lowest between III and I (5.03) (Table 2).

Table 2. Average intra (bold) and inter cluster distances for 23 tomato genotypes

Clusters	I	II	III	I V	V
I	**0.684**				
II	26.02	**2.597**			
III	5.03	29.54	**0.667**		
IV	7.72	19.40	12.50	**0.929**	
V	50.24	30.11	51.52	46.24	**0**

The highest intra cluster distance was observed for the cluster II (2.597) and the lowest for the cluster V (0). The highest values for inter cluster distance indicated that the accessions belonging to cluster III was far away from those of cluster IV. The minimum inter cluster divergence was observed between III and I indicating that the genotype of these cluster were genetically closer. Hybridization among the genotypes drawn from widely divergent clusters with high yield potential would likely to manifest maximum heterotic combinations as well as new recombination with desired traits. Therefore, the genotypes falling in these clusters (I, III, IV and V) were genetically more divergent. Crossing between the genotypes from these clusters could result in greater number of useful segregates with maximum hybrid vigour [39]. Several authors also reported profound genetic divergence in the tomato genotypes by assessing genetic divergence of quantitative traits following D^2 statistics [31-35]. Similar results were observed in brinjal [36], lablab bean [37] and pummelo [38], lemon [40] and sweetgourd [41].

Cluster mean values of 14 different characters are shown in Table 3. Difference in cluster means existed for almost all the characters studied. For cluster III, the highest mean values for days for 50% flowering (61.08), individual fruit weight (99.3 g), fruit diameter (6.16 cm), pericarp thickness (0.75 cm), number of locules per fruit (4.67), yield per plant(2.63 kg), yield (116.99 tons per hectare) were recorded. The highest mean value for plant height (185.50 cm), number of flowers per inflorescence (17.50), number of fruits per cluster (14.0), number of fruits per plant (503.60) and % brix (7.0) were observed in cluster V. It was revealed that parental lines fallen in this cluster III and V having the genetic potentiality to contribute better for yield maximization of improved tomato

Table 3. Cluster mean values for yield and yield contributing characters of 23 tomato accessions

Characters	Cluster means				
	I	II	III	IV	V
Plant height (cm)	120.58	163.65	167.60	120.65	185.50
Days for 50% flowering	60.53	58.77	61.08	59.46	56.50
Number of flowers per inflorescence	8.18	10.55	7.62	6.13	17.50
Number of fruits per cluster	5.70	8.43	4.84	3.64	14.00
Individual fruit weight (g)	64.29	10.84	99.30	58.23	4.30
Fruit length (cm)	5.44	3.56	5.32	5.53	3.21
Fruit diameter (cm)	5.23	2.95	6.16	4.95	2.26
Pericarp thickness (cm)	0.60	0.40	0.75	0.63	0.25
Number of locules per fruit	3.21	2.11	4.67	3.00	2.00
Number of fruits per plant	32.99	161.67	26.46	31.28	503.60
Number of seeds per fruit	115.52	42.89	101.75	64.64	34.34
% Brix (TSS)	5.10	6.27	4.80	4.44	7.00
Yield per plant (kg)	2.06	1.57	2.63	1.78	2.15
Yield (t/ha)	86.03	65.62	116.99	74.51	89.58

varieties in terms of fruit quality and yield contributing traits. For obtaining maximum heterosis in any crop improvement program, crossing among germplasms with outstanding mean characters values drawn from the cluster means were suggested by researchers [42, 43].

Eigen values and percent contribution of each principal component axis among the genotypes accounted through the per cent contribution of these axes (Table 4).

Table 4. Latent root (Eigen value) and percent of variation in respect of fourteen characters in 23 tomato accessions

Plant characters	Eigen value	Percent of Variance	Cumulative Percentage
Plant height (cm)	7.124	50.88	50.88
Days for 50% flowering	2.846	20.33	71.21
Number of flowers per inflorescence	1.392	9.94	81.15
Number of fruits per cluster	0.880	6.28	87.43
Individual fruit weight (g)	0.599	4.28	91.71
Fruit length (cm)	0.380	2.71	94.42
Fruit diameter (cm)	0.262	1.87	96.29
Pericarp thickness (cm)	0.188	1.34	97.63
Number of locules per fruit	0.102	0.73	98.36
Number of fruits per plant	0.093	0.67	99.03
Number of seeds per fruit	0.086	0.62	99.65
% Brix (TSS)	0.033	0.24	99.89
Yield per plant (kg)	0.013	0.09	99.98
Yield (t/ha)	0.002	0.01	99.99

The results revealed that the first axis largely accounted for the variation among the tomato accessions (50.88%) followed by second axis (20.33%). The first five axes accounted 91.71% of the total variation among 14 characters of describing 23 tomato genotypes. The rest nine characters contributed remaining 8.29% of total variation. It was observed in a study that, 76.6% of the total variability present among the 56 genotypes of tomato was explained by the first five component axes [44]. Similar results were also confirmed by other researchers in tomatoes [31-35]. In a study of genetic

diversity of acid lime, fruits per plant, yield per plant, juice volume and juice percentage were major contributing traits towards divergence [45]. The character with maximum contribution to the divergence should be given more priority for selection of parents in breeding and improvement of any crop [46].

Table 5. Latent vectors for fourteen characters of 23 tomato genotypes

Characters	Vector-I (Z_1)	Vector-II (Z_2)
Plant height (cm)	0.071	0.005
Days for 50% flowering	0.018	-0.150
Number of flowers per inflorescence	0.058	-0.046
Number of fruits per cluster	-1.618	0.155
Individual fruit weight (g)	-0.285	-0.200
Fruit length (cm)	-1.443	0.774
Fruit diameter (cm)	1.321	0.770
Pericarp thickness (cm)	-4.188	-10.475
Number of locules per fruit	-0.638	1.125
Number of fruits per plant	0.078	-0.056
Number of seeds per fruit	-0.039	-0.056
% Brix (TSS)	-0.927	-0.961
Yield per plant (kg)	-0.979	-18.687
Yield (t/ha)	0.114	0.355

The canonical variate analysis (CVA) (Table 5) revealed that in vector I (Z_1), the important characters responsible for genetic divergence in the major axis of differentiation fruit diameter, yield ton per hectare, number of flowers per inflorescence, number of fruits per plant, plant height (cm) and days for 50% flowering. In vector II (Z_2), number of locules per fruit, fruit length and diameter (cm), number of fruits per cluster, yield (t/ha) and plant height had positive impact towards divergence. In a study with tomato, the Vector-I (Z_1) explained characters like number of fruits per cluster, number of fruits per plant, plant height, intermodal distance and harvest duration which are positively related with yield and characters like average fruit weight, yield per plant, pericarp thickness and locular wall thickness were observed in Vector-II (Z_2) [44].

The characters plant height, fruit diameter and yield per hectare showing positive value in both the vectors contributed maximum towards divergence. So, the divergence in the present materials due to these three traits will offer a good scope for improvement of tomato varieties through selection of parents.

CONCLUSION

Crosses involving parents belonging to most diverse genotypes are expected to exhibit maximum heterosis and create wide variability in genetic architecture. Considering magnitude of genetic distance, contribution of different traits toward the total divergence, magnitude of cluster means for different traits and performance the genotypes of cluster III and V may be considered as parents for future hybridization program in tomato improvement.

ACKNOWLEDGMENT

The authors are thankful to Regional Agricultural Research Station (RARS), Bangladesh Agricultural Research Institute (BARI), Akbarpur, Moulvibazar, Bangladesh for support and cooperation to conduct this experiment.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] HeC., Poysa, V. and Yu, K.. Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersiconesculentum* cultivars. Theoretical and Applied Genetics, 2003. 106, 363-373.

[2] NwosuD. J, Onakoya, O. A., Okere, A. U., Babatunde, A. O. and Popoola, A. F.. Genetic variability and correlations in rainfed tomato (*Solanum spp*.) accessions in Ibadan, Nigeria. Greener Journal of Agricultural Sciences, 4(5), 211-219, 2014. http://dx.doi.org/10.15580/GJAS.2014.5.031014141

[3] JenkinsJ. A. . The origin of cultivated tomato, Economic Botany, Vol.2, 379, 1948.

[4] Rick,C. M. Origin of cultivated tomato, current status of the problem (p. 180). Abstract XI International Botanical Congress,1969.

[5] FAO. Food and Agriculture Organization of the United Nations, Production 46 Yearbook 12, Rome, Italy, 2013.

[6] Sekhar L, Salimath PM, Sridevi O. and Patil AA. Genetic diversity among some productive hybrids of tomato. Karnataka J. Agric. Sci. 21(2): 264-265. 2008.

[7] Hannan, M. M., Biswas, M. K., Ahmed, M. B., Hossain, M., &Islam, R.. Combining ability analysis of yield and yield components in tomato (*Lycopersicon esculentum Mill*.). Turkish Journal of Botany, 31(6),559-563, 2007a.

[8] Solieman, T. H. I., El-Gabry, M. A. H., & Abido, A. I.. Heterosis, potence ratio and correlation of some important characters in tomato (*Solanum lycopersicum L*.). Scientia Horticulturae, 150, 25-30.http://dx.doi.org/10.1016/j.scienta.2012.10.024, 2013.

[9] Rai GK, Kumar R, Singh AK, Rai PK, Rai M, Chaturvedi AK and Rai AB. Changes in antioxidant and phytochemical properties of tomato (*Lycopersicon esculentum Mill*.) under ambient condition. Pakistan Journal of Botany. 44(2): 667-670. 2012.

[10] Dhaliwal, M.S. Singh, D. Cheema. Line x tester analysis for yield and processing attributes in tomato, Journal of Research, Vol.40, 49-53, 2003.

[11] Sainju, M. U., Dris, R., & Singh, B.. Mineral nutrition of tomato. Food Agriculture and Environment, 1(2), 176-183, 2003.

[12] Naika, S., de Jeude, J. L., de Goffau, M., Hilmi, M., & van Dam, B.. Cultivation of Tomato: Production, processing and marketing (p. 92). Agromisa Foundation and CTA, Wageningen, 2005.

[13] Akinfasoye, J., Dotun, A., Ogunniyan, J., & Ajayi, E. O.. Phenotypic relationship among agronomic characters of commercial tomato (*Lycopersicum esculentum*) hybrids. American-Eurasian Journal of Agronomy, 4(1), 17-22, 2011.

[14] Fraser, P. D., Romer S, Shipton C. A., Mills P. B Kiano., J. W., Misawa N., Drake R. G., Schuch W., Bramley P. M..Evaluation of transgenic tomato plants expressing an additional phytoene synthase in a fruit-specific manner, Proceedings of the National Academy of sciences, Vol.99,1092-1097, 2002.

[15] Bai Y., Lindhout P.. Domestication and breeding of tomatoes:what have we gained and what can we gain in the future?,Annals of Botany, Vol.100, 1085-1094, 2007.

[16] Capanoglu E., J. Beekwilder, D. Boyacioglu, R. C. De Vos, R.D. Hall. The effect of industrial food processing on potentially health-beneficial tomato antioxidants, Critical reviews in food science and nutrition, Vol.50, 919-930, 2010.

[17] Arab L., Steck S. Lycopene and cardiovascular disease, The American Journal of Clinical Nutrition, Vol.71, 1691s-1695s, 2000.

[18] Jarquín-Enríquez L., E. Mercado-Silva, J. Maldonado, J. Lopez-Baltazar. Lycopene content and color index of tomatoes are affected by the greenhouse cover, ScientiaHorticulturae, Vol.155, 43-48, 2013.

[19] Barber N. J. and Barber J. . Lycopene and prostate cancer, Prostate Cancer Prostatic Dis., Vol.5, 12, 2002.

[20] J.Shi, Le Maguer M., Bryan M.., Lycopene from tomatoes. In: Shi, J., Ghazza, Le Maguer, M. (Eds.), Functional Foods. Biochemical and Processing Aspects, Vol.2. CRC Press, Ottawa, Canada, 135-166, 2002.

[21] Dudley JW and Moll RH. Introduction and use of estimates of heritability in soybean. Agronomy Journal. 47: 314-318. 1969.

[22] Chowdhury MA, Vandenberg V and Warkentin T. Cultivar identification and genetic relationship among selected breeding lines and cultivars in chick pea (*Cicer arietinum L*.). Euphytica.. 127(3): 317-325. 2002.

[23] P.C.Mahalanobis,. Historical note on the D^2 statistics. Sankhya, 19: 237-239,1949.

[24] V.A. Arunachalam. Genetic distances in plant breeding. Indian Journalof Genetics. 4:226-236,1981.

[25] SamsuddinA.K.M.. Genetic diversity in relation to heterosis and combining analysis in spring wheat. Theoretical and Applied Genetics. 70: 306-308, 1985.

[26] Leilah A A and Al-Khateeb S A.. Statistical analysis of wheat yield under drought conditions. Journal of Arid Environments 61: 483-496, 2005.

[27] Golparvar A R, Ghasemi-Pirbalouti A and Madani H.. Genetic control of some physiological attributes in wheat under drought stress conditions. Pakistan Journal of Biological Sciences 9(8): 1442-1446, 2006.

[28] Rao, C. R.. Advanced Statistical Method in Biometric Research. Ednl. John Wiley and Sons, New York, 1952.

[29] C.R. Rao. The use and interception of principal analysis in applied research. Sankhya 22: 317-318,1964.

[30] R. K. Singh and B. D. Chaudhary. Biometrical methods in quantitative genetic analysis. Kalyani Publ., New Delhi,1985.

[31] Kumar N., Bhardwaj M.L., Sharma A. and Kumar N.. Assessment of Genetic Divergence in Tomato (*Solanum lycopersicum L*.) through Clustering and Principal Component Analysis under Mid Hills Conditions of Himachal Pradesh, India. Int. J. Curr. Microbiol. App. Sci . 6(5): 1811-1819, 2017.

[32] Ullah M. Z, Hassan L., Singha T. and A. K. Patwary.Genetic divergence in tomato lines (*Solanum lycopersicum L*.). J. Bangladesh Agril. Univ. 13(1): 61–64, 2015.

[33] Islam S., Hossain M. K., Hasan R., Bashar A., Huque AKM M. and Alam N.. Quantification of Genetic Diversity Analysis for the Improvement of Cultivated Tomato Genotypes. Bangladesh J. Bot. 45(3): 485-491, 2016.

[34] Basavaraj S, Patil BC, Salimath PM, Hosamani RM and Krishnaraj PU. Genetic divergence in tomato (*Solanum lycopersiconL*.). Karnataka J. Agric. Sci. 23(3): 538-539, 2010.

[35] Evgenidis G, Traka-Mavrona E and Koutsika-Sotiriou M. Principal component and clusters analysis as a tool in the assessment of tomato hybrids and cultivars. Internat. J. Agron. 2011: 1-7, 2011.

[36] Silambarasan V., Eswaran R., Senthilkumar N., Thangavel P. and Thirugnanakumar S.. Studies on genetic diversity in brinjal (*Solanum melongena L*.). Plant Archives. 20 (1),pp. 9-15, 2020.

[37] Salim M., Hossain S., Alam S., Rashid J. A. and Islam S.. Estimation of genetic divergence in lablab bean (*Lablab purpureus L.)* Genotypes. Bangladesh Journal of Agricultural Research. 38(1): 105-114,2013.

[38] AlamM S, MiaMS, SalimM, Rashid JA and RahmanMS. Genetic divergence of indigenous pummelo genotypes. Res. Agric. Livest. Fish. 3 (1): 37-43, 2016.

[39] Mehta N and Asati BS. Genetic divergence for fruit characters in tomato (*Lycopersicon esculentum Mill.*) Agric. Sci. Digest. 28(2): 141-142, 2008.

[40] IsmailK. M.. Genetic diversity and molecular characterization of lemon. Ph D Dissertation. Dept. of Horticulture, Bangladesh Agricultural University,Mymensingh. 233 P, 2008.

[41] RahamanE. H. M. S., Rabbani M. G. and GarveyE. J.. Genetic diversity of sweet gourd through multivariate analysis. Bangladesh Journal of Agricultural Science,33(2): 197-204, 2006.

[42] Roy A and Sharma RN. Multivariate analysis in chilli (*Capsicum annum L*.). Ann. Agric. Res. 17(1): 30-32, 1996.

[43] Kumar M, Buckseth T, Thakur MS and Thakur KS. Genetic divergence and cluster analysis in tomato (*Solanum lycopersicum*). Progr. Agric. 13(1): 114-117, 2013.

[44] Rai A. K., Vikram A. and Pal S.. Genetic Characterization of Tomato (*Solanum lycopersicum L*.) Germplasm for Yield and Quality Traits through Principal Component Analysis. Research Journal of Agricultural Sciences. 8(5): 1171-1174, 2017.

[45] Ranpise S. A. and DesaiU. T. Genotypic and phenotypic variability in acid lime (*C. aurantifolia Swingle*). Journal of Maharashtra Agriculture University 28 (1): 21-23, 2003.

[46] JagadevP. N., Samal K. M. and LenkaL. Genetic divergence in rape mustard. Indian Journal of Genetics and Plant Breeding 51:465-466,1991.