Pak. J. Bot., 46(5): 1865-1870, 2014.

ASSESSMENT OF GENETIC DIVERGENCE IN TOMATO THROUGH AGGLOMERATIVE HIERARCHICAL CLUSTERING AND PRINCIPAL COMPONENT ANALYSIS

QUMER IQBAL*, MUHAMMAD YUSSOUF SALEEM, AMJAD HAMEED AND MUHAMMAD ASGHAR

[1]*Nuclear Institute for Agriculture and Biology (NIAB), P.O. Box 128, Jhang Road, Faisalabad, Pakistan*
*****Corresponding author email: qumerhort@gmail.com*

Abstract

For the improvement of qualitative and quantitative traits, existence of variability has prime importance in plant breeding. Data on different morphological and reproductive traits of 47 tomato genotypes were analyzed for correlation, agglomerative hierarchical clustering and principal component analysis (PCA) to select genotypes and traits for future breeding program. Correlation analysis revealed significant positive association between yield and yield components like fruit diameter, single fruit weight and number of fruits plant^{-1}. Principal component (PC) analysis depicted first three PCs with Eigen-value higher than 1 contributing 81.72% of total variability for different traits. The PC-I showed positive factor loadings for all the traits except number of fruits plant^{-1}. The contribution of single fruit weight and fruit diameter was highest in PC-1. Cluster analysis grouped all genotypes into five divergent clusters. The genotypes in cluster-II and cluster-V exhibited uniform maturity and higher yield. The D^2 statistics confirmed highest distance between cluster- III and cluster-V while maximum similarity was observed in cluster-II and cluster-III. It is therefore suggested that crosses between genotypes of cluster-II and cluster-V with those of cluster-I and cluster-III may exhibit heterosis in F_1 for hybrid breeding and for selection of superior genotypes in succeeding generations for cross breeding programme.

Key words: Tomato germplasm/variety, Genetic diversity, Correlation, Fruit Yield.

Introduction

Tomato (*Lycopersicon esculentum* Mill., 2n=2x=24) is one of the most important Solanaceous vegetable crop grown all over the world. It is versatile in nature and used for various cooking purposes. It can be processed in puree, paste, ketchup, sauce, soup etc. The average yield of tomato is very low in the tune of 10.1 tonnes per hectare in Pakistan (*Anon*., 2011a) as compared to 33.6 tonnes per hectare of modern agricultural systems of tomato in the world (*Anon*., 2011b). Besides yield limiting factors, the lack of information on genetic diversity and adaptability misleads to choice of parents suitable for hybridization program. Consequently the hybrids (F_1s) or recombinants (selected at F_2 / later generations) very often do not express full spectrum of genetic trait (s) of interest owing to limited genetic base and inappropriate selection of the parents. This problem can only be overcome if the breeders have substantial information on genetic diversity of source population.

Knowledge about levels and patterns of genetic diversity is very important for diverse applications in plant breeding. Such study focuses on the degree of similarities or dissimilarity in genetic resources (Reif *et al*., 2005; Rashid *et al*., 2008; San-San-Yi *et al*., 2008) leading to set up organization of gene banks and isolation of best parental combinations. Following hybridization, these parental combinations can possibly produce progenies with elevated genetic variability, thereby increasing chances of creating superior genotypes with traits of interest (Mohammadi & Prasanna, 2003; Crossa & Franco, 2004).

In tomato, yield is the cumulative effect of many components contributing individually to yield (Bernousi *et al*., 2011). Different characteristics viz., number of flowers cluster^{-1}, days to first fruit ripening, fruit weight, fruit length, fruit width assume vital importance and must be assessed for genetic divergence aiming to develop high yielding tomato varieties or hybrids. The most commonly used algorithms for this purpose, are canonical variable analysis, principal component analysis and clustering methods (Mohammadi & Prasanna, 2003; Sudre *et al*., 2007). Principal component analysis is frequently used to determine the relative significance of different variables of classification, prior to cluster analysis (Jackson, 1991). Additionally PCA also gives a reduced dimension model that would point out the measured differences among different groups and leads to understanding of variables by telling how much of the total variance is explained by each one. Mahalanobis D^2 statistics is powerful tool for measuring divergence among a set of population on the basis of statistical distance utilizing multivariate measurements. The present study was therefore conducted to categorize the available germplasm into separate clusters or groups on the basis of genetic diversity among their morphological attributes using agglomerative hierarchical clustering and principal component analysis. Having performed analysis, the desirable groups of genotypes could be crossed with confidence to develop either open pollinated or hybrid varieties on commercial scale.

Material and Methods

Forty four exotic tomato genotypes collected from Tomato Genetic resource Center (TGRC) along with three local varieties (Pakit, Galia and Naqeeb) were grown in tomato experimental field of Nuclear Institute for Agriculture and Biology (NIAB), Faisalabad, Pakistan in Randomized Complete Block Design with 2 replications. Five to six inch nursery seedlings were transplanted in field keeping Plant to Plant and Bed to Bed distance of 50 cm and 1.5 m, respectively. Seven plants of each genotype per replication were grown by adopting standard agronomic and plant protection practices to maintain healthy crop. The

data for different traits viz. days to maturity (DTM), plant height (PH) in cm, fruit length (FL) in cm, fruit diameter (FD) in cm, single fruit weight (SFW) in g, number of fruit per plant (NFP) and fruit yield per plant (FY) in kg were recorded as per tomato descriptor. Digital vernier caliper was used to measure tomato fruit length and diameter. Finally data were subjected to analysis of variance (Steel *et al*., 1997), cluster analysis by agglomerative hierarchical clustering and principal component analysis using computer software Microsoft Excel along with XLSTAT Version 2012.1.02, Copyright Addinsoft 1995-2012 (http://www.xlstat.com).

Results and Discussion

Patterns of correlations among traits: Analysis of variance indicated significant genotypic mean square values for all traits showing worth of genetic variability (Table 1) to be manipulated for tomato improvement. Simple correlation coefficient values demonstrated significant relationships to design breeding strategy (Table 2). Days to maturity revealed significant positive correlation with fruit length and fruit diameter. However it had significantly negative association with number of fruits per plant. Fruit length displayed highly significant and positive correlation with single fruit weight and fruit diameter but showed highly significant negative correlation with number of fruits plant^{-1}. Single fruit weight and number of fruits per plant had positive correlation with fruit yield. However, single fruit weight showed significant negative association with number of fruits per plant. Fruit diameter had significantly positive correlation with single fruit weight and yield per plant; however it had significantly negative association with number of fruits per plant. Plant height did not show any significant correlation with all traits in the present study. Presence of significant correlations for fruit diameter, single fruit weight and number of fruits per plant visa vie fruit yield reflects that increase in either of yield components will increase in the net fruit yield of tomato. Therefore, it is advocated that these morphological traits can be used for selection of high yielding tomato genotypes. Similar results had also been reported elsewhere (Singh *et al*., 2002; Bernousi *et al*., 2011).

Principal component analysis (PCA): Out of 7 principal components (PCs), three viz. PC-1, PC-II and PC-III had Eigen values >1 and contributed for 81.72% of total cumulative variability among different genotypes (Table 3). The contribution of PC-I towards variability was highest (44.20%) followed by PC-II and PC-III which contributed 22.97% and 14.55% variability respectively. The PC-I showed positive factor loadings for all the traits except number of fruits per plant while PC-II indicated positive factor loading for plant height, fruit diameter, single fruit weight, number of fruits and fruit yield per plant. Traits which contributed positive factor loadings towards PC-III were plant height followed by days to maturity and number of fruits plant^{-1}. It is evident that fruit size related traits (FL, FD and SFW) were those with highest contribution to PC-I whereas number of fruits and fruit yield were the chief contributors to PC-II. Therefore, both PC-I and PC-II could be collectively referred as reproductive axis. Plant height contributed maximum share in PC-III therefore, it could be designated as vegetative axis. These results clearly indicated that PC (s) analysis in parallel to characterization of genetic resources also highlighted certain traits for exercising selection of interest for practical breeding purposes. Similar results were found in earlier article of Krasteva & Dimova (2007). In further support to our findings, Merk *et al*., (2012) reported that first two PCs explained 28% and 16.2% of the variance and were heavily weighted by measures of fruit shape and size in tomato. PC-III explained 12.9% of the phenotypic variance and was determined by fruit color and yield components. The authors concluded that PC analysis using the trait Best Linear Unbiased Predictors (BLUPs) proposed a mean to assess which of the traits explained variation in the germplasm. The same was equally applicable to current findings.

Table 1. Analysis of variance for different characteristics in tomato genotypes.

SOV	DF	DTM (days)	PH (cm)	FL (cm)	FD (cm)	SFW (g)	NFP	FY (kg)
Replication	1	1.53	112.86	0.001	0.007	7.76	1149.40	3.038
Genotype	46	103.32*	4374.37*	1.691*	1.734*	2625.26*	1653.59*	4.253*
Error	46	27.05	106.95	0.016	0.043	98.41	271.41	1.100

* = Significant at 5% level of probability

Table 2. Correlation matrix among different characteristics in tomato genotypes.

Variables	DTM	PH	FL	FD	SFW	NOF	FY
DTM	1						
PH	0.2050	1					
FL	0.4290**	-0.0570	1				
FD	0.3110*	0.2520	0.4150**	1			
SFW	0.2810	0.1750	0.5510**	0.9390**	1		
NOF	-0.4770**	0.0970	-0.4400**	-0.5020**	-0.4880**	1	
FY	-0.0810	0.2510	0.1950	0.4950**	0.4540**	0.3040*	1

*, ** = Significant at 5 and 1% level of probability

Table 3. Principal component analysis for different characteristics in tomato genotypes.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigen value	3.09	1.61	1.02	0.69	0.39	0.16	0.03
Variability (%)	44.20	22.97	14.55	9.89	5.60	2.31	0.48
Cumulative %	44.20	67.17	81.72	91.61	97.21	99.52	100.00
Eigenvector:							
Variables	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**
DTM	0.316	-0.330	0.502	0.415	0.593	0.114	-0.054
PH	0.136	0.367	0.791	-0.111	-0.448	-0.081	-0.030
FL	0.398	-0.180	-0.210	0.640	-0.569	0.011	0.176
FD	0.517	0.180	-0.081	-0.344	0.174	0.133	0.726
SFW	0.523	0.140	-0.182	-0.223	-0.041	0.495	-0.614
NOF	-0.363	0.508	0.005	0.391	0.091	0.646	0.178
FY	0.224	0.643	-0.195	0.291	0.289	-0.548	-0.172

Table 4. Distribution of tomato genotypes in different clusters.

Cluster	Genotypes
I	Bryan Self Topper, C5, Fireball, Galia, M-82, NCEBR-6, NC HS 1,New Hampshire Victor, New Yorker, Ontario 7710, Peto-9543, Primabel, Naqeeb, Roza, Saladette, UC-134, UC-204B, UC-204C, UC-82, UC-N28, UC-TR51, Walter, BL-35
II	Cal Ace, Floradade, NC84173, NC 265-1, T-5, T-9, UC-T338, UC-TR44, Verna Orange, VF-36
III	Dwarf Stone, Edkawi, Homestead 24, NCEBR-5, Red River, Stokesdale, Vendor, White Beauty
IV	Gold Nugget
V	Lukullus, Marglobe, Rutgers, Stirling Castle, Pakit

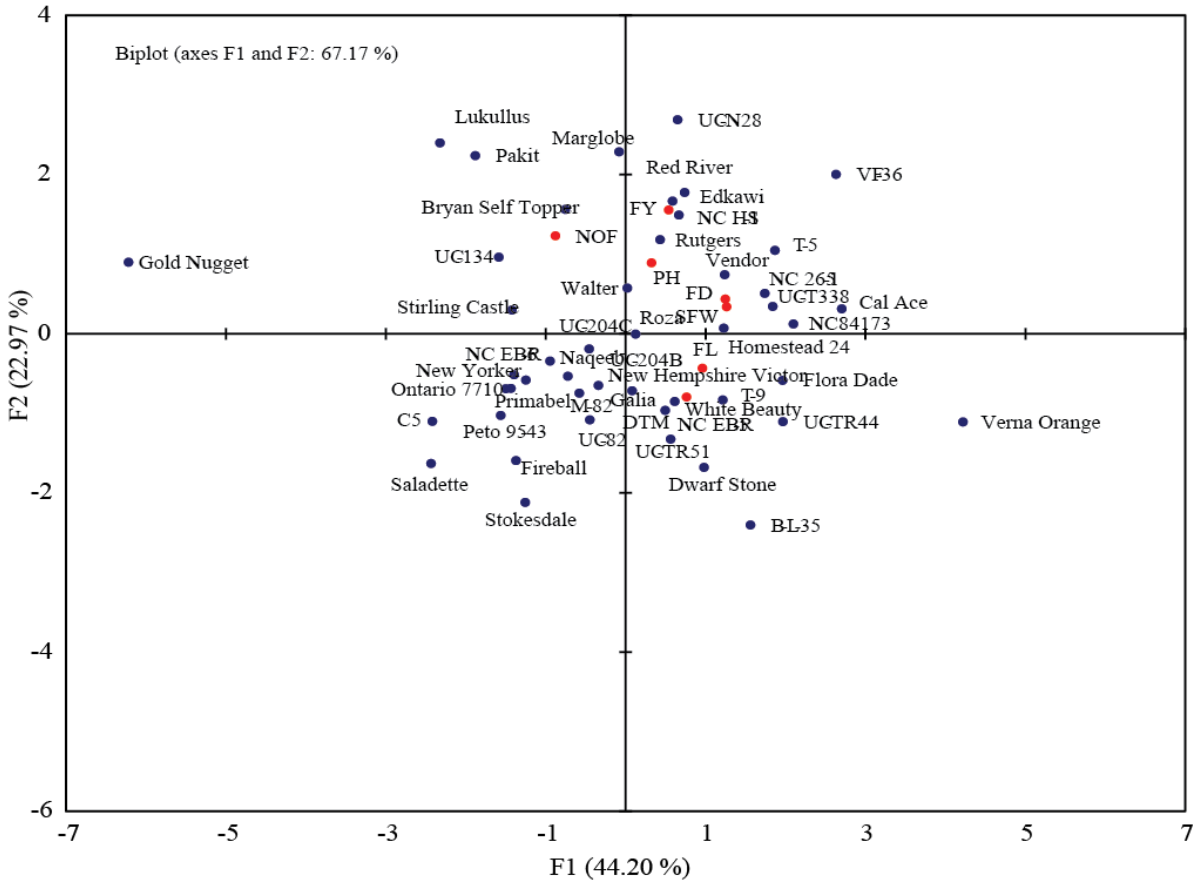

Fig. 1. Bi-plot of tomato genotypes for first two principal components.

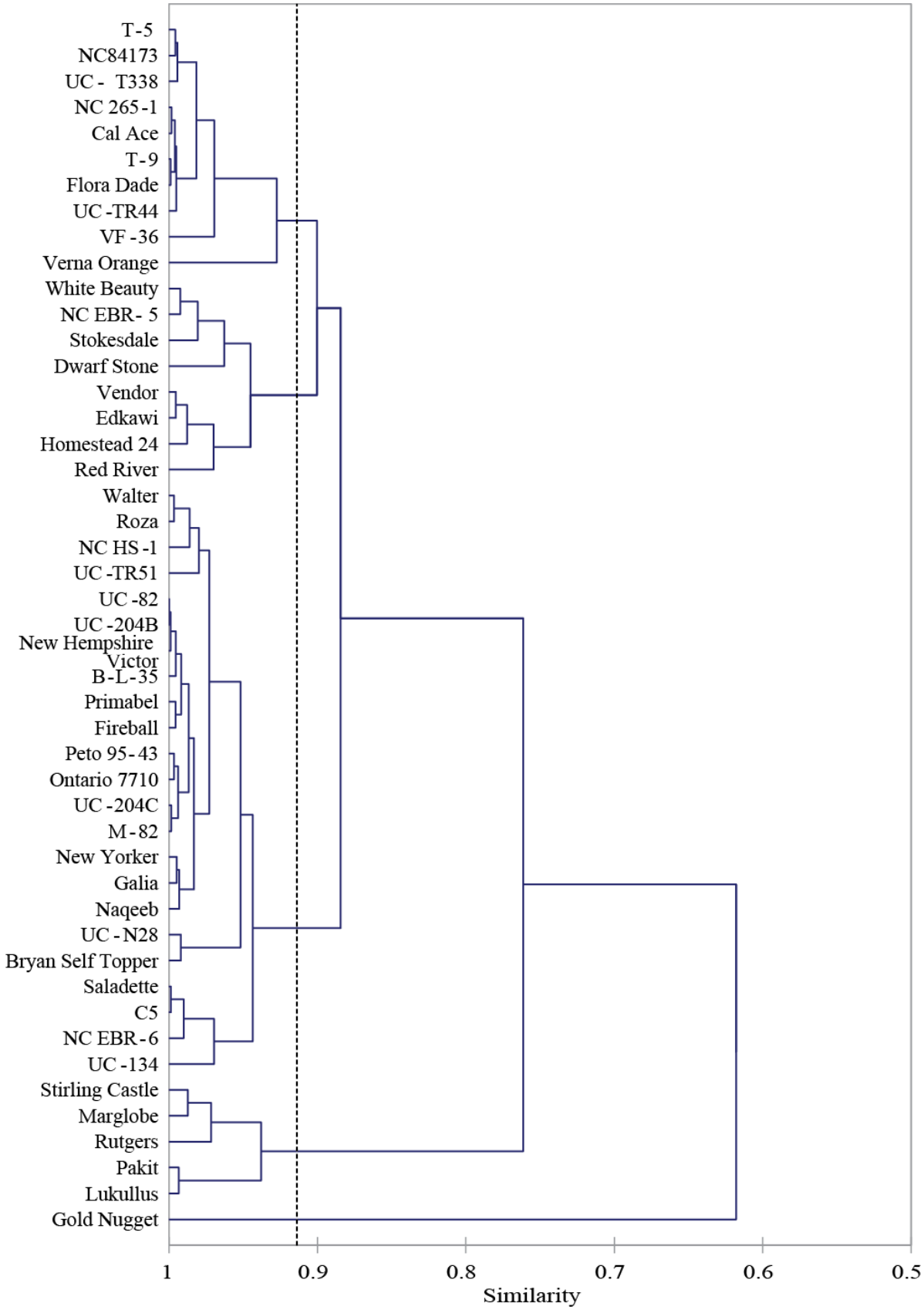

Fig. 2. Tree diagram based on seven traits for different tomato genotypes.

Table 5. Mean values of different traits of tomato genotypes in cluster analysis.

Cluster	DTM	PH	FL	FD	SFW	NOF	FY
I	163.978	59.739	5.070	4.811	63.593	60.737	3.224
II	168.550	85.300	5.810	6.335	133.565	39.035	4.155
III	169.313	125.125	5.006	5.406	87.219	46.306	3.150
IV	143.500	29.500	2.700	2.650	10.500	154.000	1.100
V	163.800	188.400	4.420	4.760	54.130	93.780	4.010

Table 6. D^2 statistics among different clusters.

	Cluster-I	Cluster-II	Cluster-III	Cluster-IV	Cluster-V
Cluster-I	0				
Cluster- II	77.750	0			
Cluster- III	71.208	61.563	0		
Cluster- IV	113.426	179.264	165.261	0	
Cluster- V	133.177	141.292	85.932	176.656	0

The first two principal components who contributed 67.17% towards total variance were plotted on PC-I x-axis and PC-II on y-axis to detect the association between different clusters (Fig. 1). It can be seen that days to maturity was significantly positive correlated with fruit length while fruit yield was positively correlated with plant height, fruit diameter and single fruit weight. However, number of fruits per plant was negatively correlated with all other traits.

Cluster analysis: Clustering of genotypes based on studied traits is presented in Fig. 2. Cluster analysis grouped 47 tomato genotypes into 5 clusters as shown in Table 4. Cluster-I comprised of 23 genotypes followed by 10, 8 and 5 genotypes respectively in cluster-II, III and cluster-V. However, Gold nugget was skipped as it falls in cluster-IV (Table 5). The genotypes in cluster-I were short statured as compared to all other genotypes of cluster-II, III and cluster-V. Similarly cluster-II comprised of genotypes with higher fruit length, fruit diameter, single fruit weight and fruit yield. The genotypes in cluster-III were tall with large size fruits (FL and FD) while the genotypes in cluster-V possessed larger plants with more number of fruits and higher yield. Cluster-IV had only one genotype (gold nugget) that differs significantly from other tomato genotypes for almost all the traits. Pairwise Mahalanobis distances (D^2 statistics) are presented in Table 6. Genotypes of cluster-V elucidated maximum diversity against genotypes of cluster-III followed by cluster-II. However, minimum differences were observed between cluster-II and cluster-III due to least value of genetic divergence. It is evident from current study that cluster analysis can be regarded as efficient tools to categorize germplasm and renders reliable basis in choice of base material to plan future breeding strategies as reported earlier (Susic *et al*., 1998; Feng-Mei *et al*., 2006) in tomato. However, the authors believe that while making choice of base material, one must take care of genetic barriers and breeding methods to get expected genetic improvements for desired traits. Results of present study revealed that multivariate analysis helps to place the genotypes in different clusters on the basis of PC(s) values.

Conclusion

It was concluded that genotypes of cluster-V and cluster-III are complementary for maximum traits and could be selected for hybridization to develop promising F_1 hybrids or transgressive segregants in succeeding generations.

Acknowledgements

Authors are thankful to TGRC for providing the seeds of forty seven tomato genotypes used in present study.

References

Anonymous. 2011a. Agricultural Statistics of Pakistan. Government of Pakistan. Ministry of Food, Agriculture and Livestock. Islamabad.

Anonymous. 2011b. FAO Statistics. www.fao.org/corp/statistics/en

Bernousi, I., A. Emami, M. Tajbakhsh, R. Darvishzadeh and M. Henareh. 2011. Studies on genetic variability and correlation among the different traits in *Solanum lycopersicum* L. *Not. Bot. Hort. Agrobot. Cluj*., 39(1): 152-158.

Crossa, J. and D.J. Franco. 2004. Statistical methods for classifying genotypes. *Euphytica*, 137: 19-37.

Feng-Mei, J., J. Xue, J. Yan-Hong and D.L. Zhong-Qi. 2006. The cluster analysis on tomato germplasms. *Acta Agric. Bor. Sin*., 21(6): 49-54.

Jackson, J. 1991. *A User's Guide to Principal Components*. John Wiley & Sons.

Krasteva, L. and D. Dimova. 2007. Evaluation of a canning determinate tomato collection using cluster analysis and principal component analysis (PCA). *Acta Hort*., 729: 89-93.

Mahalanobis, P.C. 1936. On the Generalized distance in statistics. In: *Proc. Nat. Institute Sci*., India, 2:49-55.

Merk, H.L., S.C Yarnes, A.V. Deynez, N. Tong, N. Menda, L.A. Mueller, M.A. Mutschler, S.A. Loewen, J.R. Myers and D.M. Francis. 2012. Trait diversity and potential for selection indices based on variation among regionally adapted processing tomato germplasm. *J. Am. Soc. Hort. Sci*., 137(6): 427-437.

Mohammadi, S.A. and B.M. Prasanna. 2003. Analysis of genetic diversity in crop plants - salient statistical tools and considerations. *Crop Sci*., 43: 1235-1248.

Rashid, M., A.A. Cheema and M. Ashraf. 2008. Numerical analysis of variation among basmati rice mutants. *Pak. J. Bot*., 40(6): 2413-2417.

Reif, J.C., A.E. Melchinger and M. Frisch. 2005. Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. *Crop Sci*., 45:1-7.

San-San-Yi, S.A., Jatoi, T. Fujimura, S. Yamanaka, J. Watanabe and K.N. Watanabe. 2008. Potential loss of unique genetic diversity in tomato landraces by genetic colonization of modern cultivars at a non-center of origin. *Plant Breeding*, 127:189-196.

Singh, J.K., J.P. Singh, S.K. Jain, J. Aradhana and A. Joshi. 2002. Studies on genetic variability and its importance in tomato (*Lycopersicon esculentum* Mill.). *Prog. Hort*., 34: 77-79.

Steel, R.G.D., J.H. Torrie and D.A. Dick. 1997. Principles and Procedures of Statistics-a biometrical approach. McGraw Hill Book Co., New York.

Sudre, C.P., E. Leonardecz, R. Rodrigues, A.T.D.A. Junior, M.D.C.L. Moura and L.S.A. Gonçalves. 2007. Genetic resources of vegetable crops: a survey in the Brazilian germplasm collections pictured through papers published in the journals of the Brazilian Society for Horticultural Science. *Hortic. Bras*., 25: 496-503.

Susic, Z., J. Zdravkovic, N. Pavlovic and S. Prodanovic. 1999. Selecting features for estimating genetic divergence of tomato genotypes (*Lycopersicum esculentum* Mill.). *Genetika*, 31(3): 235-244.

(Received for publication 22 April 2013)