Indian J. Pl. Genet. Resources 12(3): 290-301, 1999

STUDIES IN WORLD TOMATO (*Lycopersicon esculentum* Mill.) GERMPLASM COLLECTION: METHODOLOGY FOR ESTABLISHING A REPRESENTATIVE CORE SET USING CHARACTERIZATION DATA

UMESH SRIVASTAVA, R L SAPRA AND S V S CHAUHAN[1] AND PREM NARAIN[2], National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi 110 012; [1]Dept. of Botany, School of Life Sciences, Dr. B. R. Ambedkar University (formerly Agra University), Khandelwal Kothi, Civil Lines, Agra 282 002 (Uttar Pradesh); [2]Emeraitus Professor (CSIR), IARI, New Delhi 110 012

The present paper discusses the formation of a core set of tomato (*Lycopersicon esculentum* Mill.) germplasm comprising 2021 accessions and representing variability from 47 countries. Data on 41 morpho-agronomic and economic traits (qualitative as well as quantitative) scored/studied during summer season of 1990, 1991 and 1992 were used in the analysis. Principal Component Analysis (PCA) strategy was used to estimate the inertia and the corresponding Relative Contribution (RC) of each accession. Fourteen principal components with eigen values over 1 and explaining a total of 65 per cent of the variation were selected for computing the inertia and the relative contribution of each accession and improved method based on the RC score requiring low number of accessions to be sampled, was suggested for establishing the core set. The said method resulted into higher values of Shannon Diversity Index (SDI)/Coefficient of variation (CV) for 36 out of 41 characters scored. An optimum sample size of 140 was estimated on the basis of pooled diversity. A sample of 140 accessions (nearly 7% of the base collection) was finally selected for the final core set.

Key words: Tomato, *Lycopersicon esculentum*, germplasm, variability, diversity, principal component analysis, core set, Shannon Diversity Index, relative contribution

Cultivated tomato originated in the New World, while the Andean zone is likely to be the centre of origin of wild tomato. However, the domestication took place in Mexico. The most likely ancestor of cultivated tomato is the cherry tomato (*Lycopersicon esculentum* var. *cerasiforme*). According to Rick (1979), species of *Lycopersicon* were evolved via gene substitution. This fact is based on the systematic studies of *Lycopersicon* and closely related species of *Solanum*. Cultivars introduced from Latin America generally had exposed stigma, facilitating cross pollination. In the later stages, plants with high percentage of fruit set and with short style at the mouth of anther tube were selected. Presently, cultivars with much shortened style exhibiting very high degree of self-pollination are available. In Peru, Equador, Colombia and Chile, cultivated tomatoes exhibit spontaneous crossing with *L. pimpinellifolium* indicating a close genetic relationship. This and other wild species e.g. *L. cheesmanii, L. chilense, L. hirsutum, L. peruvianum* and some *Solanum* species are, now included in tomato breeding programmes (Kalloo, 1991). Tomatoes are used directly as raw vegetables, in sandwiches, salads etc. several processed items like paste, soup, juices, ketchup, drinks, whole peeled tomatoes etc. are prepared on a large scale and enjoy high acceptance as food ingredients. It is an important source of minerals and vitamins as well. The pulp and juice of the fruit is digestible and a mild aperient. It is a promoter of gastric secretion and blood purifier. It is also considered to be intestinal antiseptic and has culinary effect in the entire portion of alimentary canal.

The tomato plant is a classical breeding and genetic material for several reasons. A tremendous amount of genetic diversity exists in *Lycopersicon.* Normally, the cultivated tomato is self-pollinated but natural cross-pollination of 1.9 per cent to 26 per cent was recorded in various localities in India. There is abundant seed set under normal conditions and from full grown plant, 5,000 to 25,000 seeds can be obtained. Tomato has tremendous potential of heterosis of earliness, uniformity, total yield and resistance attributes. Thus, the cultivation of hybrid varieties is economical and remunerative, and therefore there is great demand of hybrid tomatoes. Tomato production is limited under conditions of low and high temperature. The fruit set is reduced when night temperatures exceed 21°C although tomatoes can set fruits at day temperature up to 38°C (Peet and Willit, 1993). The plants can not withstand severe frost. Temperature and light intensity affect the fruit setting, pigmentation and nutritive value of fruits (Chaudhary, 1967).

Greater the diversity in the germplasm material, the better is the chances for evolving promising and stable varieties. Attempts to identify cultivars with high productivity and quality potential were needed to cope with these challenges. Success of any breeding programme is largely dependent on the extent of genetic variability present in the population. The National Bureau of Plant Genetic Resources (NBPGR), over the years, has assembled *ca* 2900 germplasm lines from diverse agro-climatic zones of the world (from 47 countries) and from within India. This includes several wild species viz., *Lycopersicon pimpinellifolium, L. hirsutum, L. chesmanii, L. peruvianum, L. glandulosum, L. chilense, L. chmieliwskii, L. parviflorum, L. esculentum* var. *cerasiforme, L. hirsutum var. glabratum* and *L. parviflorum* var *humifusum.* Thus considerable variability has been generated. Interestingly, introduction of germplasm of tomato has paved way for reasonably successful performance of different varieties in India. Some of the tomato varieties in India have been bred by selection technique. Considering the dearth of variation within *L. esculentum,* it is important to characterize and evaluate world collection of tomato genetic stocks on 41 agro-morphological, economic and yield contributing characters, so that breeders may utilize the same for deliberate breeding programme.

One of the major issue that Gene Bank managers (curators) face is the need to increase the accessibility of their collection to a large group of users. Developing procedure for reducing the size of a collection to a manageable and accessible level is becoming one of the most important issues in the management and utilization of large germplasm collections. Concept of developing core germplasm set as a minimal, yet diverse set, representing the diversity of the germplasm collection has attracted wide attention in recent years.

MATERIALS AND METHODS

Over *ca* 2900 germplasm lines were assembled, over the years, from diverse agro-climatic zones of the world (from 47 countries) and from within India by the NBPGR. These germplasm were evaluated under field conditions during hot summers (maximum temperature 38-42°C) at NBPGR Experimental Farm, Issapur (annual rainfall 400 mm,; soil texture-sandy load; pH 6.5-8.2) in Delhi. Each accession was grown in 2-row plot with rows 4 m long and 60 cm between rows, during summer seasons of 1990, 1991 and 1992 in augmented block design and donors identified (Umesh Chandra and Thomas, 1993).

Developing a core set for efficient management of germplasm helps in maintaining comparatively fewer accessions rather than entire set of accessions. Principal Component Analysis (PCA) strategy was used to estimate the inertia and the corresponding relative contribution (RC) of each accession. Fourteen principal components with eigen values over 1 and explaining a total of 65 per cent the variation were selected for computing the inertia and relative contribution of each accession. An improved method based on RC score and requiring low number of accessions to be sampled, is suggested for establishing the core set.

Descriptors and Descriptor States used

C1: **Plant growth habit** : (1) Determinate : One or two leaves between flowering clusters and tip ends with an inflorescence; also known as self topping dwarf, (2) Indeterminate: Constancy of 3-4 leaves in between the flowering clusters, and tip does not end with an inflorescence (non-topping) and (3) Spreading: In this lateral axis terminates in a blossom which gives self topping. There appear less three leaves between the flower clusters, also known as self-topping bush/spreading type.

C2: **Leaf Type** : (1)Normal leaf and (2)Potato leaf

C3: **Leaf Size** : (1) Small, (2) Medium and (3) Large

C4: **Leaf colour**: (1) Light Green, (2) Green and (3) Dark Green

C5: **Leaf pubescence**: (1) Smooth and (2) Pubescent

C6: **Leaf cover**: (1) Poor, (2) Moderate, (3) Good and (4) Excellent

C7: **Petiole pubescence**: (1) Smooth and (2) Pubescent

C8: **Stem type**: (1) Round and (2) Angular

C9: **Stem thickness**: (1) Thin, (2) Medium and (3) Thick

C10: **Stem pubescence**: (1) Smooth and (2) Pubescent

C11: **Flower size** : (1) Small, (2) Medium and (3) Large

C12: **Flower colour** : (1) Deep yellow, (2) Light yellow and (3) Reddish yellow

C13: **Style position**: (1) Exerted and (2) Non-exerted

C14: **Pistil type**: (1) Branched and (2) Highly Branched

C15: **Fruit size**: (1) Small (cm), (2) Medium (5-8 cm) and (3) Large (8 cm)

C16: **Fruit shape**: (1) Round, (2) Oblong, (3) Flattened, (4) Pear shaped and (5) Bottle shaped

C17: **Fruit surface**: (1) Smooth and (2) Corrugated

C18: **Blossom-end shape fruit**: (1) Round and (2) Nippled

C19: **Green stem-end of fruit**: (1) Nil, (2) Indented, (3) Flat (Round) and (4) Pointed (Nippled)

C20: **Type of fruit cracking**: (1) None, (2) Slight, (3) Moderate and (4) Severe

C21: **Fruit abnormalities**: (1) Absent and (2) Present

C22: **Type of placenta**: (1) Small, (2) Medium and (3) Large

C23: **Fruit firmness**: (1) Soft, (2) Medium and (3) Firm

C24: **Pulpiness**: (1) Pulpy, (2) Highly pulpy and (3)Juicy

C25: **Skin thickness**: (1) Thin, (2) Medium and (3) Thick

C26: **Seediness**: (1) Low, (2) Medium and (3) High

C27: **Time of maturity**: (1) Synchronous: Attaining fruit maturity at a particular time and all plants in a row/plot attains maturity simultaneously and (2) Non synchronous: Attaining fruit maturity at a different time and all plants in a row/plot attains maturity at a different time

C28: **Plant height (cm)**: Measured just before first harvest, in centimeters from the ground level to the tip of the primary branch on 5 random plants from the middle of the plot and data averaged.

C29: **Number of primary branches (number)**: The number of branches that arise from the main stem were reckoned as primary branches. Total number of primary branches was counted at the time of first fruit harvest on 5 random plants from the middle of the plot and data averaged.

C30: **Days to 50% flowering (days)**: Number of days taken from the date of transplanting to 50 % of plants flowered in the row/ plot. Stigma emergence on the main branch is considered as flowering.

C31: **Days to first fruit set (days)**: As the fruits set for the first time in each plot, data recorded. Number of days taken from the date of transplanting to date of first fruit set on 5 random plants were calculated in a row/ plot and data averaged.

C32: Days to first fruit harvest (days): As the fruits attained harvesting stage (turning stage) first time in each plot, the data recorded and averaged. Number of days taken from the date of transplanting to date of first fruit harvest on 5 random plants was calculated in a row/ plot.

C33: Days to maturity (days): Data recorded as the fruits attained maturity stage i.e., plants attained physiological maturity (80% maturity) in each plot. Number of days taken from the date of transplanting to date of plants attaining physiological maturity (80% maturity) in a row/ plot.

C34: Number of clusters per plant (number): Number of clusters found on a plant was counted in 5 random plants in a row / plot and data averaged.

C35: Number of flowers per cluster (number): Number of flowers found in the fourth cluster (on a plant) was counted in 5 randomly selected plants in a row / plot and data averaged.

C36: Number of fruits per cluster (number): Number of fruits found in the fourth cluster (on a plant) was counted in 5 randomly selected plants in a row / plot and data averaged.

C37: Number of fruits per plant (number): Number of fruits counted at each harvest and the cumulative total of all the harvests were taken as number of fruits per plant .The data were recorded on 5 randomly selected plants in a row / plot and averaged.

C38: Fruit yield per plant (g): The fruits picked in each harvest were weighed and cumulative yield was expressed in grams.The data were recorded on 5 randomly selected plants in a row / plot and averaged.

C39: Number of locules per plant: Actual number of locules. The number of locules was counted after cutting (section cut) the fruits transversely and counting the locules separated by septae. The data were recorded on 5 randomly selected plants in a row / plot and averaged.

C40: Pericarp thickness (mm): Measured in millimeters from an equatorial section of the fruit by using Vernier calipers. The data were recorded on 5 randomly selected fruits in a row / plot and averaged.

C41: Total soluble solids (%) : A hand refractometer was used for direct determination of total soluble solids-TSS (00 Brix) from fresh juice of fully ripened fruits. Mean of atleast 3 samples recorded as per solid's read directly from a Brix scale superimposed over the refractive index scale.

RESULTS AND DISCUSSION

Data on 41 morpho-agronomic traits of 2021 well-characterized accessions grown and evaluated at the NBPGR Farm, New Delhi, was used for sampling of the core accessions. Data pertained to 47 countries. There were 410 accessions whose source was not available. Owing to this lack of information, we did not initially classify the accessions on the basis of source country. The standardized data on all the 2021 accessions were subjected to principal component analysis (PCA). The information on (N x k) scores obtained from PCA were used for extracting the accessions accounting for maximum variability as described below. This technique has also been used by Hamon and Noirot, 1990; Hamon *et al.*, 1995; Mahajan *et al.*, 1996; Noirot *et al.*, 1996).

Selection of accessions through Principal Component Analysis

Principal component analysis and Factor analysis are basically dimension reduction techniques that analyze the interrelationships among the variables to study and/or assess the underlying structure of the given data. The principal components are uncorrelated among themselves and having variances equal to the eigen values of the sample correlation matrix. Also for any given $k < p$, principal components have the property that among all possible sets of k linear combinations of the variables, the set of first k principal components accounts for the maximum variation present in the given data. Hence this set is ideal choice to represent the original data on p variables by a smaller number of k variables as this set retains the maximum variation in the original data. Computationally obtaining principal components is essentially obtaining eigen values and eigen vectors of a nonnegative definite matrix.

Let y_{ij} be the score through PCA for i^{th} (i = 1, 2, 3, ... N) accession and jth (j = 1, 2, 3 ...... k) principal components. The inertia of ith accession is defined as:

$$P_i = \sum_{j=1}^{k} y_{ij}^2$$

Then the relative contribution of this ith accession is given by $RC_i = P_i/(N \times k)$. The proposed selection procedure for obtaining the optimum number of accessions required for establishing core set consists of following steps:

i) Arrange the accessions in descending order of Relative Contribution

ii) Plot a curve between the Relative Contribution (RC) of each accession and the corresponding accession number.

iii) Select a point on the curve near the point of inflection where the fall in RC is almost low.

iv) Begin with the number of accessions equal to the number (point) selected in the above step (iii); compute the Shannon Diversity Index (SDI) for each qualitative descriptor and obtain the pooled diversity.

v) Repeat the step (iv) for a number of times with reduced number of accessions.

vi) Plot a curve between number of accessions and the pooled diversity. The peak point on the curve where the diversity is maximum, is thus, is the optimum point.

The base collection comprising 2021 accessions represented variability from 47 countries (Table 1). However, there were 410 accessions whose source was unknown. Frequency distribution of descriptor states (qualitative traits) for each descriptor is given in Table 2. The principal component analysis applied on all the 41 characters simultaneously resulted into 14 Principal components (Table 3, 4, 5) with eigen values greater than 1. All the components were used for computing the inertia of each accession. These components could account for nearly 64.6 pct of the total variation (Table 6). Fig. 1 suggested that the beyond 200 accessions (nearly 10% of the whole collection) the relative contribution of the succeeding accessions is very low and the curve becomes almost parallel to the X-axis.

Table 1. Source wise frequency distribution of 2011 accessions

Source Country	No. of Accessions	Source Country	No. of Accessions
Argentina	6	New Zealand	29
Australia	92	Nigeria	26
Baghdad	2	Peru	2
Brazil	9	Philippines	2
Bulgaria	29	Poland	18
Canada	48	Portugal	9
China	1	Russia	4
Cuba	2	South Africa	1
Cyprus	4	South America	1
Czechoslovakia	9	Spain	2
Denmark	20	Sri Lanka	2
France	9	Sudan	1
Ghana	2	Sweden	3
Netherlands	27	Switzerland	4
Hong Kong	3	Taiwan	269
Hungary	151	Tangerine	2
India	244	Trinidad	1
Israel	8	Turkey	2
Italy	20	UK	27
Japan	83	USA	382
Mauritius	9	Russia	36
Mexico	1	West Germany	1
Nepal	1	Yugoslavia	5
		Unknown	410

Table 2. **Frequency distribution of descriptor and descriptor states**

Descriptor	Descriptor States (Codes)				
	1	2	3	4	5
C1	421	1373	227		
C2	1969	52			
C3	41	1666	314		
C4	459	1556	6		
C5	1914	107			
C6	748	952	280	41	
C7	42	1979			
C8	34	1987			
C9	12	1195	814		
C10	42	1979			
C11	431	1462	128		
C12	1143	813	65		
C13	1833	188			
C14	1365	656			
C15	472	1400	149		
C16	1417	271	268	44	21
C17	1872	149			
C18	1813	208			
C19	1314	667	38	2	
C20	1079	863	69	10	
C21	1568	453			
C22	490	1412	119		
C23	786	1082	153		
C24	941	425	655		
C25	579	1322	120		
C26	148	1540	333		
C27	627	1394			

Fig. 2 indicates that the captured diversity increases with the reduction in sample size and attains maximum value when the collection size is 140 (nearly 7% of the entire collection). Further reduction in sample size reduces the diversity.

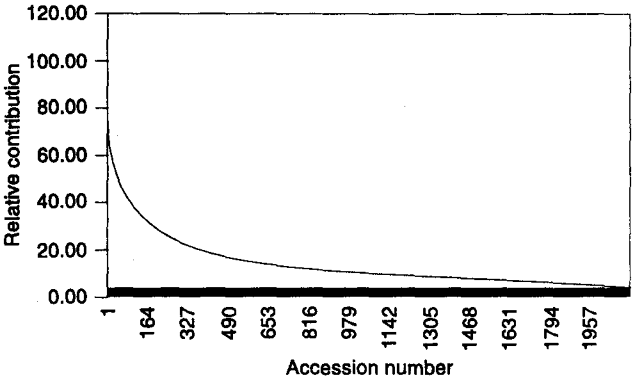

Fig. 1. Showing Relative Contribution (RC) of accessions

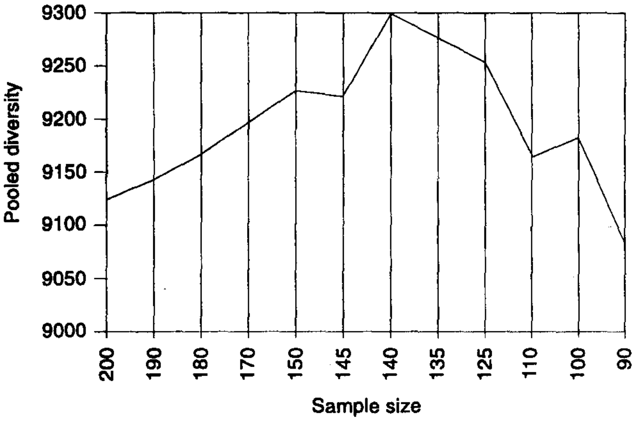

Fig. 2. Showing optimum sample size

The representativeness of sampled accessions was assessed by comparing its diversity with that of the whole collection for each individual qualitative and quantitative descriptor. The Shannon diversity indices and coefficients of variation (CV) for the selected accessions were compared with that of the whole collection (Table 7 and 8).The suggested procedure resulted into higher SDIs for all the qualitative characters for the core set; except for the characters, C5: Leaf Pubescence, C13:Style position and C21:Fruit abnormalities whose values were slightly lower than those of the base collection comprising of 2021 accessions (Table 7). In the same way, higher values of CV pct were observed for all the quantitative characters except for the characters, C29: Number of primary branches and C37: Number of fruits per plant (Table 8). The loss in variability in these two characters was only 0.3 per cent and 4 per cent respectively.

Table 3. Principal Components Analysis with factors 1-5

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
ZC1	–.01727	.11697	.02091	.00399	.00328
ZC2	.06578	.01359	–.04280	.12593	.01789
ZC3	.11941	–.02252	.08863	.32221	–.04880
ZC4	.08385	.11162	.37542	–.09642	–.18432
ZC5	–.12907	–.30494	–.25057	.05897	–.00762
ZC6	.04036	.18401	.25143	.04372	.10891
ZC7	.15280	.03053	.44155	.14721	–.75408
ZC8	.03799	–.03794	.20105	.06319	–.35339
ZC9	.02167	.26414	.22138	.26874	–.18863
ZC10	.15584	.03130	.44459	.14682	–.75057
ZC11	.03761	–.21653	–.05658	.36056	.04980
ZC12	.03171	–.15566	.35131	.13670	.29935
ZC13	.07218	–.07871	.26527	–.20193	.09874
ZC14	.05208	–.04247	.56457	–.33540	.14818
ZC15	.17160	–.45169	.07564	.44467	–.09613
ZC16	.09848	–.22278	.10956	.15090	.12867
ZC17	.03091	–.04334	.30258	–.19684	.05841
ZC18	.01158	–.19777	.03568	–.04293	.30158
ZC19	.10802	–.07819	.56066	–.06523	.32530
ZC20	.03212	–.18951	.35335	–.00924	.42995
ZC21	.02936	.05746	.33992	.11367	.19394
ZC22	.20108	–.49340	.10719	.36221	–.06813
ZC23	.16043	–.51125	–.01847	.23077	–.11218
ZC24	–.07564	.56325	–.00324	–.26475	–.16231
ZC25	.17106	–.60379	.03900	.34739	.13330
ZC26	–.13295	.54090	–.04492	–.24066	–.07653
ZC27	–.00103	–.02601	–.56161	.39541	–.10855
ZC28	–.03952	.54194	.03386	.37358	–.02899
ZC29	–.10846	.40452	.11842	.36968	.24254
ZC30	.94755	.11910	–.12991	–.10897	.05477
ZC31	.94675	.12482	–.13040	–.10650	.05523
ZC32	.96162	.10110	–.10374	–.15047	.04799
ZC33	.88393	.09033	–.09160	–.18993	.00072
ZC34	.12609	.45398	–.22686	.53253	.05132
ZC35	.00795	.37624	.19402	.36301	.19840
ZC36	.03806	.40550	.36884	.25278	.27167
ZC37	.12999	.49157	–.00614	.53649	.18830
ZC38	–.00744	.00889	.21198	.06536	.01423
ZC39	.15244	–.10599	.18686	.29699	–.02243
ZC40	–.02622	–.05437	.07174	–.04834	–.03822
ZC41	–.02598	–.08423	.20610	–.09200	–.09005

Table 4. Principal Components Analysis with factors 6-10

Variable	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10
ZC1	–.21438	–.23809	.30096	–.01353	–.30638
ZC2	.27992	.02987	.05714	.01391	.44183
ZC3	.08973	–.13451	.35419	–.07661	.06106
ZC4	–.13658	–.47919	–.20256	–.19798	.12741
ZC5	–.13387	.69274	.13857	.28332	.00108
ZC6	.31000	.23979	–.11482	.13154	.27704
ZC7	–.16979	.24988	.09639	–.03452	.10036
ZC8	–.11273	.10630	.12937	–.08899	–.37507
ZC9	.26534	–.13345	.04911	–.11567	–.23570
ZC10	–.16840	.24659	.10766	–.03780	.10330
ZC11	–.18983	–.01499	–.04091	–.03349	–.06246
ZC12	–.37052	.28168	–.06103	.06379	–.07478
ZC13	.05867	.04604	.33218	–.24714	–.07321
ZC14	.34042	.13954	–.09308	–.13656	–.16074
ZC15	.14796	–.01388	–.24396	.15043	.04506
ZC16	.37907	.04352	.20052	–.04445	.28390
ZC17	.23776	.04445	.10389	–.01335	.16226
ZC18	.15485	–.10539	.49925	–.17084	.06193
ZC19	–.06397	.03130	–.17865	–.04343	–.09905
ZC20	–.17902	.31823	–.04039	.03682	–.07435
ZC21	–.36344	–.16256	.05544	–.02520	.11346
ZC22	.15307	–.02940	–.16752	.09820	.06046
ZC23	.20435	–.11395	.15354	.01253	.21918
ZC24	.16192	–.02445	.05061	.11493	.13854
ZC25	.05096	–.08242	–.00478	–.10024	–.09034
ZC26	.00912	.09805	–.14803	.20284	.15510
ZC27	–.29947	–.12043	.08141	.04890	.07915
ZC28	.35874	.05246	–.15514	–.02992	–.08964
ZC29	–.04531	.20555	–.10795	.08658	–.02769
ZC30	–.03472	.06049	.00441	.05081	–.01859
ZC31	–.03423	.05689	.00607	.05421	–.01239
ZC32	–.05198	.01192	.00050	.04888	–.01033
ZC33	–.02193	–.02281	–.00465	.02443	–.01982
ZC34	.25775	.15670	.03010	–.16614	–.23165
ZC35	–.35860	–.15242	.16456	.17690	.27658
ZC36	–.28350	–.12844	.16585	.13339	.22799
ZC37	.12931	.12823	.13075	–.12130	–.16553
ZC38	.30782	–.09089	.15702	.41189	–.22939
ZC39	.02909	–.31441	–.46601	.10433	–.06554
ZC40	.11852	–.17802	.12800	.58458	–.17156
ZC41	.07468	–.22342	.20306	.52854	–.17191

Table 5. Principal Components Analysis with factors 11-14

Variable	Factor 11	Factor 12	Factor 13	Factor 14
ZC1	.27519	–.09256	–.39206	.17958
ZC2	–.47688	.37227	–.01913	.12349
ZC3	–.13686	.34131	–.32956	.36720
ZC4	.04269	–.01832	.17575	–.12734
ZC5	.03661	–.05468	.01144	.10502
ZC6	.13517	–.22795	.22482	.33286
ZC7	–.08404	–.02527	.02316	–.11069
ZC8	.21921	.03542	.03532	–.20690
ZC9	.13530	.09807	.11569	.31605
ZC10	–.07895	–.02293	.03298	–.11309
ZC11	.21342	.46416	.04631	–.31194
ZC12	.09866	.24550	.04017	–.07174
ZC13	–.16034	.02199	.20600	.01016
ZC14	–.14104	–.05099	–.10628	–.08065
ZC15	.07445	.10083	–.24624	–.04024
ZC16	.38800	.12893	–.05476	–.11294
ZC17	.44865	.11621	.06134	–.13767
ZC18	.12646	.05754	.20853	–.16969
ZC19	–.18152	.06006	–.04061	–.02254
ZC20	–.09337	.03704	–.10866	.03830
ZC21	–.02758	.14232	.14150	.30465
ZC22	.10606	–.09097	–.06648	–.10028
ZC23	.05121	–.32441	–.05650	.12220
ZC24	.03441	.20669	–.18627	–.08146
ZC25	–.08572	–.20974	.10414	.03264
ZC26	.20984	.12682	–.20991	–.01007
ZC27	.00438	.03377	.20887	–.02408
ZC28	.09962	–.11524	.21303	.00208
ZC29	.21742	.08092	.21618	.19327
ZC30	.02826	.00259	.01226	.03078
ZC31	.02637	.00146	.01362	.02941
ZC32	.02748	.00262	.01618	.01439
ZC33	.03209	.01324	.02125	–.00511
ZC34	–.14168	.04533	.04344	–.18203
ZC35	.02106	–.23178	–.06440	–.06782
ZC36	–.03634	–.26580	–.11514	–.12689
ZC37	–.19042	–.08220	–.07876	–.23800
ZC38	–.07209	–.12717	–.28694	–.15309
ZC39	.00745	.11293	–.07246	.09231
ZC40	–.16342	.10590	.26711	.09285
ZC41	–.05369	.11710	.30166	–16876

Table 6. Showing communality, eigen value, pct variation and cumulative per cent variation

Variable	Communality	Factor	Eigen value	Var. (%)	Cum. (%)
ZC1	.57197	1	3.84925	9.4	9.4
ZC2	.68206	2	3.40634	8.3	17.7
ZC3	.66868	3	2.69180	6.6	24.3
ZC4	.59770	4	2.65213	6.5	30.7
ZC5	.78874	5	2.22184	5.4	36.1
ZC6	.60489	6	1.83974	4.5	40.6
ZC7	.94186	7	1.60741	3.9	44.6
ZC8	.45489	8	1.34412	3.3	47.8
ZC9	.52781	9	1.29622	3.2	51.0
ZC10	.94133	10	1.22907	3.0	54.0
ZC11	.58740	11	1.15952	2.8	56.8
ZC12	.56373	12	1.10312	2.7	59.5
ZC13	.38337	13	1.07180	2.6	62.1
ZC14	.68651	14	1.00640	2.5	64.6

Table 7. Shannon diversity index for the core and whole collection

Descriptor	SDI (Selected accessions)	SDI for all the accessions
C1	0.441733	0.362636
C2	0.272712	0.051929
C3	0.393863	0.229134
C4	0.339448	0.241139
C5	0.066914	0.089940
C6	0.544108	0.467036
C7	0.265295	0.043893
C8	0.217322	0.037091
C9	0.371877	0.307222
C10	0.262614	0.043893
C11	0.407211	0.320739
C12	0.341018	0.347085
C13	0.198969	0.134404
C14	0.286509	0.273728
C15	0.438325	0.341447
C16	0.584837	0.398271
C17	0.18358	0.114294
C18	0.279213	0.143947
C19	0.359853	0.312905
C20	0.358847	0.314724
C21	0.229616	0.231091
C22	0.432818	0.330429
C23	0.456173	0.389637
C24	0.474716	0.455566
C25	0.425291	0.348931
C26	0.380346	0.302125
C27	0.286509	0.268956

Shannon diversity index (SDI)

Table 8. Variability in the core-set accessions as well as full-set for quantitative descriptors

	140 Accessions			2021 Accessions		
Descriptor	C.V	Minimum	Maximum	C.V	Minimum	Maximum
C28	38.69	21.0	180.0	34.39	21.0	180.0
C29	25.69	3.0	8.0	25.98	2.0	10.0
C30	12.90	49.0	98.0	9.61	49.0	98.0
C31	11.56	56.0	106.0	8.59	55.0	106.0
C32	8.59	79.0	127.0	6.68	78.0	127.0
C33	8.18	96.0	142.0	6.83	95.0	142.0
C34	36.69	10.0	57.0	44.46	3.2	59.6
C35	16.46	3.0	9.0	15.33	3.0	9.0
C36	29.56	2.0	8.0	25.52	1.0	8.0
C37	52.79	3.0	235.3	56.78	3.0	250.0
C38	33.29	124.0	5600.0	28.06	124.0	6445.0
C39	36.70	2.0	6.0	30.36	2.0	6.0
C40	26.91	1.0	4.8	22.13	0.1	4.8
C41	16.68	1.200	6.8	14.32	1.2	6.800

Table 9. Showing accession No. source and the Relative Contribution (RC) of the top 140 accessions

S. No.	Accession No.	Source	RCx $(2021\times14)^{-1}$
1	EC251663	TAIWAN	109.82
2	EC104243	PHILIPPINES	87.03
3	LIA-722	—	82.77
4	EC99931	RUSSIA	80.75
5	EC251682	TAIWAN	79.71
6	EC1129	AUSTRALIA	75.96
7	EC2687	AUSTRALIA	75.82
8	EC251621	TAIWAN	74.29
9	EC278677	USA	72.80
10	PSR-468	INDIA	71.66
11	EC251683	TAIWAN	71.54
12	PSR-450	INDIA	70.34
13	EC212687	TAIWAN	69.93
14	EC3176	ARGENTINA	67.10
15	NC 57804	INDIA	67.07
16	PSR-444	INDIA	64.35
17	PSR-451	INDIA	64.10
18	EC13904	USA	61.33
19	EC154460	—	60.2
20	IC23585	INDIA	60.20
21	IC1703	INDIA	58.60
22	EC23885	—	58.15
23	EC125754	—	58.07
24	EC1703	—	57.73
25	EC162515	—	57.40
26	EC159036	BULGARIA	56.70
27	EC251711	TAIWAN	55.78
28	EC162508	—	55.66
29	IC159142	INDIA	55.31
30	EC12524	RUSSIA	54.38
31	EC162598	—	54.12
32	EC163670	—	53.61
33	2466-27	INDIA	53.19
34	EC164660	—	53.17
35	IC3709	INDIA	52.70
36	EC260308	—	52.14
37	PSR-446	INDIA	52.12
38	EC3709	—	52.10
39	EC163676	—	51.40
40	EC162518	—	51.09
41	EC164172	—	50.87
42	EC8739-1	SWITZERLAND	50.79
43	M28/81	INDIA	50.56
44	EC130053B	NEW ZEALAND	50.11
45	EC155996	—	49.81
46	EC159048	BULGARIA	49.37
47	EC56186	CANADA	48.36
48	EC164666	—	48.04
49	EC32271	—	48.03
50	U16-99	INDIA	47.97
51	IC127936	INDIA	47.84
52	EC32271	—	47.54

53	EC164686	—	47.33
54	EC164657	—	46.69
55	EC52062	USA	46.23
56	No 25	INDIA	45.74
57	EC126766	RUSSIA	45.69
58	EC241141	TAIWAN	45.64
59	EC125557	—	45.39
60	EC170372	USA	44.69
61	11309/P2	—	44.23
62	PSR-439	INDIA	42.91
63	A1	INDIA	41.54
64	EC61742	CANADA	41.46
65	EC69876	USA	41.35
66	EC130053	NEW ZEALAND	41.34
67	EC737/1	—	41.24
68	EC9208	—	40.89
69	DM-3	INDIA	40.59
70	D-0449	INDIA	40.08
71	EC260637	NETHERLANDS	39.75
72	EC144295	NETHERLANDS	39.01
73	IC 39786	INDIA	38.68
74	EC168010	TAIWAN	38.66
75	EC57442	CANADA	38.59
76	EC164846	—	38.15
77	IC1340	INDIA	37.33
78	EC241156	TAIWAN	37.30
79	EC161838	—	37.04
80	IC4197	INDIA	36.50
81	EC16342	ITALY	36.27
82	EC8822	DENMARK	36.07
83	EC251679	TAIWAN	36.02
84	EC761	—	35.68
85	EC232428	USA	35.42
86	EC27975	JAPAN	35.42
87	EC23885-B	—	35.33
88	EC118282	HONG KONG	35.17
89	EC251680	TAIWAN	35.13
90	EC164677	—	35.05
91	EC232427	USA	34.95
92	IC5539	INDIA	34.73
93	EC159993	USA	34.69
94	DPP-3/204	INDIA	34.31
95	EC141817	—	33.43
96	EC35520	HUNGARY	33.19
97	EC160193	BRAZIL	32.48
98	EC6592	USA	32.28
99	EC164859	—	32.05
100	EC163616	—	31.99
101	EC177294	FRANCE	31.98
102	D.M.M.	—	31.67
103	IC6486	INDIA	31.62
104	EC7286	AUSTRALIA	31.59
105	EC145549	TAIWAN	31.56
106	EC491	UK	31.05
107	EC241147	TAIWAN	30.64
108	EC145623	RUSSIA	30.19
109	EC168004	TAIWAN	30.18
110	EC164853	—	30.13
111	EC1087	—	29.37
112	EC276	—	29.33
113	EC27951	JAPAN	29.31
114	EC35219	HUNGARY	29.09
115	EC241146	TAIWAN	28.91
116	EC145624	RUSSIA	28.71
117	EC6592	USA	28.53
118	LA BONITA	—	28.48
119	PSR-438	INDIA	28.14
120	EC163589	—	28.09
121	EC37302	USA	28.04
122	EC386504	—	28.00
123	EC159966	USA	27.96
124	EC32612	DENMARK	27.94
125	EC106281	USA	27.66
126	EC113838	TAIWAN	27.52
127	EC117381	USA	27.26
128	EC50348	NIGERIA	27.20

129	EC141823	—	27.17
130	PSR-427	INDIA	27.00
131	EC26106	CANADA	26.90
132	EC164854	—	26.89
133	EC41243	USA	26.87
134	33/81-30	INDIA	26.79
135	EC129605/P2-1	NEW ZEALAND	26.73
136	EC162520	—	26.70
137	EC1148	AUSTRALIA	26.65
138	EC61691	USA	26.61
139	EC129594	NEW ZEALAND	26.59
140	EC31515	CZECHOSLOVAKIA	26.43

EC and IC numbers represent exotic collections obtained from other countries and indegenous collections from within India respectively and maintained at NBPGR, New Delhi, India

The purpose for a core collection is to select a small representative sample of genetic variation from a large germplasm collection to assist the germplasm curators and users. In the present study the, accessions in the representative core set constituted approximately 7 per cent of the whole collection. A sample fraction of 10 per cent has been proposed by Brown (1989a) as the optimum sample size. The optimum sample fraction, however, depends largely on the degree of redundancy among accessions and certain other factors (Yonezawa *et al.*, 1995). The reason for lower fraction in our case could be accounted for repeat introductions of the material from exotic sources in the past several years.

Higher values of indices (SDI) for majority of the qualitative characters in the selected entries as compared to the whole collection is indicative of better representation of the existent diversity by the selected accessions. The quantitative variability was also fairly represented by the selected accessions for almost all the characters. Our method of selecting accessions, slightly differs from the one suggested by Hamon and Noirot (1990). In the later method, we first search two accessions with the highest relative contribution (RC) and simultaneously we go on adding accessions one by one with the next highest RC until the SDI on them approximately equates the corresponding values for the entire cluster. The accessions sampled from various-clusters within a cluster constitute the sample core subset. The accessions from various sample core sub-sets are pooled to form the final sample core set. Our method first locates a point on the RC curve (Fig. 1) where the fall in RC stabilizes. After locating the point on the RC curve, the peak point is located on the pooled diversity curve (Fig. 2) where the pooled diversity (sum or average of SDI of all the characters) is maximum. Our method resulted into higher values of SDI/CV for 36 characters out of 41 characters (Table 7 and 8) with lower number of accessions. For tomato core, an optimum sample size of 140 was estimated on the basis of pooled diversity. A sample of 140 accessions (nearly 7 % of the base collection of 2021 germplasm lines) was finally selected for the final core. Thus, the suggested method of selecting accessions can be safely used, even when the passport information is partially available or not available.

REFERENCES

Brown, A.H.D. 1989a. The case for core collections. *In*: The use of plant genetic resources. Brown, A.H.D., Frankel, O.H., Marshall, D.R. and William, J.T. (eds.). Cambridge University Press, Cambridge. p 136-156.

Chaudhary, B. 1967. Vegetables. National Book Trust, New Delhi-16.

Hamon, S. and M. Noirot. 1990. Some proposed procedures for obtaining a core collection using a quantitative plant characterization data. Paper presented at the International workshop on okra held in New Delhi, India, October, p. 8-12.

Hamon, S., M. Noirot and F. Anthony. 1995. Developing a coffee core collection using the principal component score strategy with quantitative data. *In*: Core collections of plant genetic resources. Hodgkin, T., Brown, A.H.D., Th., J.L. van Hintum and Morales, A.A.V. (eds.). John Wiley and Sons, p. 117-126.

Kalloo, G. (Ed.). 1991. Genetic improvement of tomato. Springer-Verlag, Berlin Heidelberg. Germany. 358 p.

Mahajan, R.K., I.S. Bisht, R. C. Agrawal and R. S. Rana. 1996. Studies of South Asian okra collection: Methodology for establishing a representative core set using characterization data. *Genet. Resources Crop Evol.* 43: 249-255.

Noirot, M., S. Hamon and F. Anthony. 1996. The principal component scoring A new method of constituting a core collection using quantitative data. *Genet. Resources Crop Evol.* 43 : 1-6.

Peet, Mary M. and Danial H. Willit. 1993. Evaluating high night temperature effects on tomato. *In* : C George Kuo (Ed.). Adaptation of food crops to temperature and water stress. AVRDC, Taiwan. 13-18 Aug. 1992. Publication No. 93-410. p 175-187.

Rick, C. M. 1979. Biosystematic studies in Lycopersicon and closely related species of *Solanum.* J.G.Hawkes, R. N. Lester and A.D. Skelding (Eds.). The Biology and Taxonomy of Solanaceae. Academic Press Inc. London.

Umesh Chandra and T. A. Thomas. 1993. Sources of tomato germplasm adaptable to the hot summer season in India. *In*: C. Geerge Kuo (Ed.). Adaptation of food crops to the temperature and water stress. AVRDC, Taiwan, 13-18 Aug. 1992. Publ. No. 93-410, p 501-502.

Yonezawa, K., T. Nomura and H. Morishima. 1995. Sampling strategies for use in stratified germplasm collections. *In*: T. Hodgkin, A.H.D. Brown, Th. J.L. van Hintum and Morales, E.A.V. (eds.) Core collections of plant genetic resources. International Plant Genetic Resources Institute (IPGRI). A Wiley-Sayce Publication.