Developing Core Set in Brinjal (*Solanum melongena* L.) with Limited Traits

Gunjeet Kumar, RK Mahajan*, RL Sapra[1], KK Gangopadhyay, BL Meena, SKTiwari, SK Mishra and SK Sharma

National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi-110 012, India

[1] *Indian Agricultural Research Institute, New Delhi-110 012, India*

Large collections of genetic resources held in Gene Banks poses a challenge for the maintenance of both the accessions as well as the information documented for the germplasm characterized. The accessibility and knowledge of the landrace collections are the prerequisite for an efficient utilization of the genetic resources. Different sample sizes and sampling strategies, either random or non random, were proposed to obtain core sets in brinjal from entire collection of 622 landraces. The proposed strategy emphasizes on relatively smaller number (5) of highly variable descriptors for computing inertia score through Principal Component Score Strategy (PCSS) and thereby capturing maximum genetic diversity in the core set. This strategy is compared with three other strategies i.e. Random, Stratified and Principal Component Score Strategy based on the entire set (13) of quantitative descriptors. This strategy will help in developing core set with limited resources or when information on some descriptors is not available.

Key Words: Brinjal, Core set, Diversity, Principal component score strategy, Random sampling, Stratified random sampling

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.224.50 on dated 9-Feb-2023

Introduction

Brinjal (*Solanum melongena* L.) is an important vegetable crop of Central, Southern and South-East Asia, and in number of African countries (Kalloo, 1988). It is a good source of minerals and vitamins and has several medicinal properties (Khan, 1979). India and Indo-China is considered to be the centre of diversity (Vavilov, 1951; Daunay *et al.,* 2001). The important brinjal growing countries are China, India, Egypt, Turkey, Japan, Italy, Sudan, Indonesia, Philippines and Spain. In India, it is grown in 0.5 million hectare area with a total production of over 8 million tons (FAOSTAT, 2007).

Ex-situ conservation of plant genetic resources can result in large collections that are difficult to characterize, evaluate, utilize and maintain. An important objective for curators is to find a way to preserve the widest range of genetic diversity within crop species as well as to improve the knowledge and utilization of the genetic resources. To overcome management difficulties, the identification and use of core collection has been suggested. The National Bureau of Plant Genetic Resources (NBPGR), New Delhi has the responsibility of collection, characterization and conservation of the brinjal diversity in the country and presently maintains over 2,500 accessions. Much of this diversity was augmented in the recent past during 1999-2005. The diversity was also enriched from exotic sources. Recently, Kumar *et al*. (2008) studied the morphological diversity of 622 accessions of brinjal germplasm comprising indigenous and exotic collections for 24 descriptors.

The essential features of a core collection are minimum redundancy and capturing maximum genetic diversity from entire collection of a crop species/ wild relatives. Some of the core collections in other crops have been developed under the Indian National Programme (Mahajan *et al.,* 1996; Bisht *et al.,* 1998) at NBPGR.

For the development of core set, several strategies have been proposed since Frankel first suggested the concept of a core collection in 1984 (Brown, 1989a). Brown (1989b) proposed the use of random sampling strategies among a stratified collection. He assumed that over 70% of the alleles would be retained in 10% of the total collection based on the theory of selectively neutral alleles. With the objective of retaining the maximum genetic diversity from the whole base collection in a manageable working collection, non random Principal Component Score Strategy (PCSS) has been suggested. This has been applied to identify a core set that will maximize the representation of the phenotypic variability of the base collection (Noirot *et al.,* 1996; Mahajan *et al.,* 1996; Sapra *et al.,* 2006). Sapra *et al.,* (2006) further discussed the issue of minimum sample size for capturing diversity. They suggested a clustering cum inertia score strategy which selects single entry from each group (cluster) having highest inertia score in the group and ensured higher diversity in terms of allelic evenness and richness in the sample. In the past the core sets were developed on the basis of inertia score computed from entire set of quantitative descriptors. The question here arises, can we take a subset of these descriptors for

Author for correspondence: mahajan@nbpgr.ernet.in

computing inertia score without compromising the genetic diversity in the core set. Such a situation may arise due to either non availability of the information for all the descriptors or to save the cost and time in recording the entire set of descriptors. It is quite logical that the highly variable descriptors play a significant role in computing inertia score in the PCSS in comparison to the low variable descriptors. Thus, the present study examines the merit of highly variable descriptors for computing inertia scores and compares the results with other strategies based on the entire set of descriptors.

Materials and Methods

The plant material consisted of 622 accessions of brinjal germplasm comprising 543 indigenous and 79 exotic accessions held at National Bureau of Plant Genetic Resources (NBPGR), New Delhi, India. The region/ country wise grouping of the accessions is presented in Table 1. The accessions were grown in 6 m row plots, with 75 cm row to row and 60 cm plant to plant spacing. These accessions were evaluated for 24 morphological descriptors (13 quantitative and 11 qualitative) during kharif 2003 cropping season (July to December) and the quantitative traits were transformed into qualitative ones (Table 2).

Allelic Evenness was measured by Shannon-Weiner Diversity Index (H′) (Weaver and Shannon, 1949) while allelic richness was measured by counting the descriptor states without considering their individual frequencies for different descriptors.

$$H' = -\Sigma\, p_{ij} \log_e p_i$$

where p_i is the proportion of the accessions for the i^{th} descriptor state of the qualitative character. In order to keep the value of H′ in the range of 0-1 each value of H′ was divided by the maximum value, $\log_e n$ where n is the number of descriptor states. The pooled H′ was obtained by summing the individual H′ over the entire set of descriptors. The average diversity was computed as the total diversity divided by the number of qualitative descriptors. Three strategies namely, Simple Random Sampling without replacement (SRSWOR), Stratified Random Sampling (SRS) and Principal Component Score Strategy were used for selecting the accessions with varying sample sizes i.e. 5, 10, 20 and 30 per cent of the entire collection having 31, 62, 124 and 187 accessions respectively. Due to the complex nature of H′, the bootstrapping was performed to estimate the confidence intervals (CI). The technique resamples the original data large number of times before drawing sample. In our case, 500 samples were drawn to estimate the expected mean and variance of the pooled H′. The selection of accessions and the estimation of CI for the above mentioned three strategies are described below.

a) Simple Random Sampling without Replacement (SRSWOR)

A random sample of size n was drawn by SRSWOR from N accessions. Let p_{ij} (i=1,2,..,d; j=1,2,…,k) denote the sample proportion of accessions for j^{th} state of i^{th} descriptor. Then an estimate of pooled H' is

$$H' = \Sigma_i \Sigma_j\, p_{ij} \log p_{ij}$$

For computation of 95% CI of H′, 500 independent random samples of a given size were drawn from the population by SRSWOR and mean [E(H′)] and variance [V(H′)] of H′ were computed. The 95% CI was computed as $(x - 1.96s/\sqrt{m}) < H' < (x + 1.96s/\sqrt{m})$, where x = E (H′); s = V (H′) and m = 500.

Table 1. Region-wise number of accessions studied and their diversity indices

Group	Origin	No. of accessions	Average SDI (H′)
I	Zone I (North-western Himalaya)	17	0.56
II	Zone II (West Bengal and Assam)	141	0.59
III	Zone III (North-eastern region, Andaman & Nicobar islands)	28	0.51
IV	Zone IV (Indo-Gangetic plains)	49	0.58
V	Zone V (Eastern peninsular region)	160	0.58
VI	Zone VI (North-western plain and arid region)	33	0.61
VII	Zone VII (Central plateau region)	35	0.56
VIII	Zone VIII (Southern peninsular region)	80	0.66
IX	Bangladesh	22	0.55
X	Japan	7	0.45
XI	Sri Lanka	42	0.63
XII	Taiwan	8	0.44
	Total	622	

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.224.50 on dated 9-Feb-2023

Table 2. List of descriptors studied and alongwith their diversity indices

Quantitative descriptors	Frequency class	SDI
1. Number of primary branches	1=<4; 2=4-8; 3=8-12; 4=12-16; 5=>16	0.55
2. Plant height	1=Small (<50 cm); 2=Medium (50-100 cm); 3=Tall (>100cm)	0.45
3. Plant spread	1=Very narrow (<30 cm); 2=Narrow (30-40 cm); 3=Intermediate (40-60 cm) 4=Broad (60-90 cm) 5=Very broad (>90 cm)	0.59
4. Petiole length	1=<4 cm; 2=4-8 cm; 3=8-12 cm; 4=12-16 cm; 5=16-20 cm; 6=>20 cm	0.41
5. Leaf blade length	1=<8 cm; 2=8-12 cm; 3=12-16 cm; 4=16-20 cm; 5=20-24 cm; 6=>24 cm	0.63
6. Leaf blade width	1=<10 cm; 2=10-15 cm; 3>15 cm	0.69
7. Days to 50% flowering	1=<40; 2=40-55; 3>55-70; 4=>70	0.48
8. Fruit peduncle length	1=<4 cm; 2=4-8 cm; 3>8 cm	
9. Fruit length	1=<8 cm; 2=8-13 cm; 3=13-18 cm; 4=18-23 cm; 5=23-28 cm; 6=28-33 cm; 7=>33 cm	0.61
10. Fruit width	1=<5 cm; 2=5-10 cm; 3>10 cm	0.61
11. Number of fruits/plant	1=<10; 2=10-20; 3=20-30; 4=30-40; 5=40-50; 6=50-60; 7=60-70;8=70-80; 9=80-90; 10=90-100; 11=100-110	0.48
12. Fruit weight	1=<100g; 2=100-200 g; 3=200-300 g; 4=300-400 g; 5=400-500 g; 6=500-600 g; 7=600-700 g; 8=>700 g	0.48
13. Days to first fruit set	1=<30; 2=30-45; 3>45-60; 4=60-75; 5=>75	0.43
Qualitative descriptors		
14. Plant growth habit	1=Upright; 2=Intermediate; 3=Prostrate	0.29
15. Petiole color	1=Green; 2=Greenish violet; 3=Violet; 4=Dark violet; 5=Dark brown	0.45
16. Leaf blade lobing	1=Very Weak; 2=Weak; 3=Intermediate; 4=Strong; 5=Very strong	0.72
17. Leaf blade color	1= Light green; 2= Green; 3= Dark green; 4= Greenish violet and 5= Violet	0.45
Calyx colour	1=Green; 2=Light Purple; 3=Dark purple	0.82
18. Calyx spininess	3= Smooth; 5= Medium thorny; 7= High thorny	0.54
19. Corolla color	1=White; 2=Greenish white; 3=Pale violet; 4=Light violet; 5=Bluish Violet	0.65
20. Fruit length/breadth ratio	1=Broader than long; 3=As long as broad; 5=Slightly longer than broad; 7=Twice as long as broad; 9=Several times as long as broad	0.73
21. Fruit shape	1=Long; 2=Round; 3=Oblong; 4=Oval	0.71
22. Fruit color	1=Milky white; 2=Green; 3=Deep yellow; 4=Purple; 5=Purple black; 6=Black; 7=Light purple	0.58
23. Fruit flesh density	1=Very loose (spongy); 2= Loose (crumbly); 3=Medium compact;4=Compact; 5=Very compact	0.57
24. Seediness	3= Low; 5= Medium; 7= High	0.88

www.IndianJournals.com Members Copy, Not for Commercial Sale Downloaded From IP - 14.139.224.50 on dated 9-Feb-2023

b) Stratified Random Sampling (SRS)

A sample of size n was allocated to different groups in proportion to the diversity in the group. Let n_s be the sample size selected from s^{th} (s=1, 2,…, 12) group such that ($\sum n_s$=n). The samples from each group were selected by SRSWOR and an estimate of H´ was computed over the pooled set of accessions. CI was computed as mentioned in SRSWOR. Samples from different groups were retained for use in PCSS.

c) Principal Component Score Strategy (PCSS)

The entries from each group were selected by PCSS (Noirot *et al.,* 1996). The accessions were arranged in descending order in terms of inertia score and the top accessions having higher inertia scores were selected from each group. Two strategies of PCSS were designed, *i.e.* (i) entire set of 13 quantitative descriptors [PCSS(13)] and (ii) a sub set of 5 highly variable descriptors, namely, fruit weight, number of fruits/plant, fruit length, fruit width and peduncle length selected on the basis of high co-efficient of variation [PCSS(5)].

Results and Discussion

Among the total accessions, 88% were indigenous and 12% of exotic origin. The indigenous accessions were grouped into Zone I to Zone VIII whereas exotic accessions (Zone IX to XII) were from Bangladesh, Japan, Sri Lanka and Taiwan respectively. Zone V (160) followed by Zone II

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.224.50 on dated 9-Feb-2023

(141) had maximum number of accessions while Japan (7) represented the least number of accessions. Among the indigenous accessions, the average maximum diversity was observed in Zone VIII (0.66) followed by Zone VI (0.61) while in the exotic group it was maximum in accessions from Sri Lanka (0.63) followed by Bangladesh (0.55). The diversity for quantitative descriptors ranged between 0.41 for petiole length to 0.69 for leaf blade width and between 0.29 for plant growth habit to 0.88 for seediness among qualitative descriptors (Table 2).

Allelic evenness and allelic richness are the most commonly used parameters for measuring diversity. It would be worthwhile considering both the parameters simultaneously while discussing the issue of diversity. Table 3 in general shows a less pooled diversity as well as richness for all the three strategies when a sample size of 5% is considered. However, when the sample size is increased from 5 to 10% there is considerable increase both in diversity as well as in richness. Further increase in sample size to 20% showed marginal increase in diversity and it was negligible beyond 20% sample size in all the strategies. The size of core collection has long been under discussion. Brown (1989b) proposed that a fraction of about 10% is an appropriate sample size for sampling core entries and suggested a log proportion strategy (L) or absolute proportion strategy (P) when the whole collection is composed of several groups. He further made calculations based on the Ewens (1972) theory under certain assumptions that at least 70% of the existing alleles could be drawn with 95% certainty if 10% or more of the plants are sampled from the population. Our findings support his view and 10% sample size was found to be efficient for all the sampling strategies employed. Hence, increasing the sample size beyond 10% may not be much advantageous in terms of enhancing both evenness and richness.

Figure 1 shows the relationship between the sample size and the pooled SDI corresponding to different strategies. All the four curves rise sharply when the sample size is increased from 5 to 10% and after that the curves rise slowly upto 20% of the sample size. However, when the sample size increases from 20 to 30% the curves become almost parallel to the horizontal axis indicating it would not be advantageous in terms of increasing diversity. Table 3 and Figure 1 indicate that the random strategy (SRSWOR) is the most efficient one as the average SDI is the lowest among all the strategies. The stratified sampling is slightly better than that of the random one. The PCSS strategy based on 13 and 5 descriptors resulted in higher SDIs as compared to SRSWOR and Stratified strategies. However, the proposed strategy, PCSS (5) based on limited number of traits resulted in higher SDI for all sample sizes when the accessions were selected from the top after arranging the accessions in terms of decreasing inertia score.

Figure 2 represents more or less similar behaviour as that of Figure1 except for stratified sampling strategy where it has been found to be less efficient compared to SRSWOR in terms of pooled richness. The proposed strategy, PCSS (5) could capture richness of 92.16 % even with a smaller sample size of 10%. This value is even higher than 91.18% obtained through PCSS (13) strategy for a sample size of 30%. The proposed strategy, PCSS(5) based on limited number of traits resulted in the highest pooled H´ and allelic richness for all sample sizes. This may be due to the selection of five most variable descriptors in PCSS (5).

In general in previous studies, the stratified sampling has been found to be effective and efficient as compared

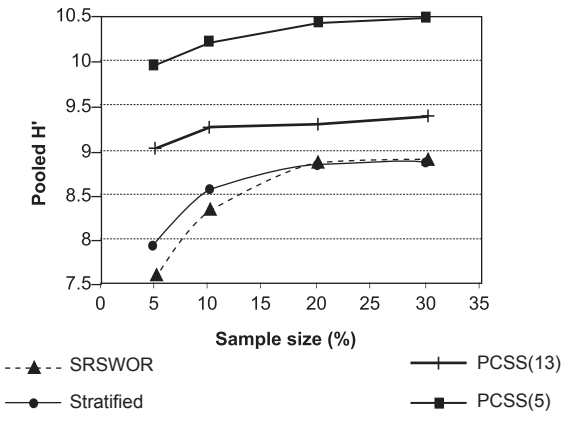

Fig. 1: Relationship between sample size and pooled H' for various strategies

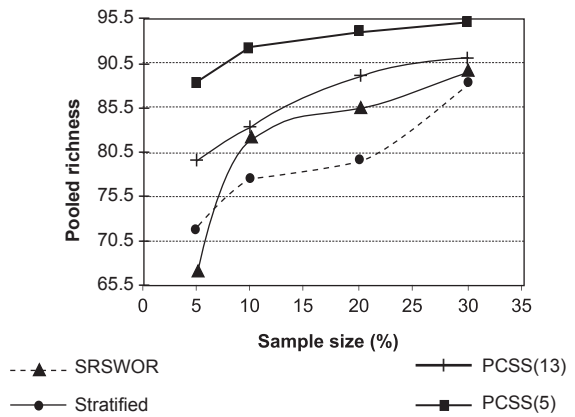

Fig. 2: Relationship between pooled richness and sample size for various strategies

Table 3. Comparison of Shannon-Weiner diversity index and pooled richness of selected accessions through various strategies

Sample size (%)	Number of accessions	SRSWOR (SDI)	Stratified (SDI)	PCSS (SDI)		SRSWOR (richness)	Stratified (richness)	PCSS (Richness)	
				13 Characters	5 Characters			13 Characters	5 Characters
5	31	7.606 [4.922-10.390]	7.920 [4.940-10.900]	9.031	9.932	66.67	71.57	79.41	88.23
10	62	8.334 [6.134-10.534]	8.563 [6.151-10.975]	9.240	10.211	82.27	77.24	83.33	92.16
20	124	8.862 [6.861-10.863]	8.845 [6.640-11.150]	9.271	10.430	85.58	79.50	89.22	94.12
30	187	8.870 [7.600-10.141]	8.846 [7.645-10.047]	9.379	10.481	89.72	88.24	91.18	95.10

Figures within parenthesis indicate the 95% confidence interval

to the random sampling in many of the cases and has been extensively used in developing the core sets. However, in our case also it has proved to be efficient for all sample sizes except for 5% and 10% where SRSWOR showed better results. The reason for this unusual behavior could be attributed to a chance factor. The relationship between sample size and the allelic richness showed similar pattern as in case of diversity except for stratified random sampling where it was found to be less efficient as compared to SRSWOR.

The results clearly indicate that the proposed strategy which uses limited number of traits has proven to be the most efficient in capturing both the pooled diversity as well as allelic richness for a given sample size. Thus, the strategy can be effectively utilized in developing core sets where the information is lacking, particularly for less heterogeneous traits and still ensuring diversity in the core set.

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.224.50 on dated 9-Feb-2023

References

Bisht IS, RK Mahajan, TR Loknathan and RC Agrawal (1998) Diversity in Indian sesame collection and stratification of germplasm accessions in different diversity groups. *Genet. Resour. Crop Evol*. **45**: 325-335.

Brown, AHD (1989a) Core collection: a practical approach to genetic resources management. *Genome* **31**: 818-824.

Brown AHD (1989b) The case for core collections. In: AHD Brown, OH Frankel, DR Marshall and JT William (eds.) *The Use of Plant Genetic Resources*. Cambridge University Press, pp. 135-156.

Daunay MC, RN Lester and G Ano (2001) Cultivated eggplants. In: A Charrier, M Jacquat, S Haman and D Nicolas (eds.) *Tropical Plant Breeding*, Oxford University Press, Oxford: 200-225.

Ewens WJ (1972) The sampling theory of neutral alleles. *Theoretical Population Biol.* **3**: 87-112.

FAOSTAT (2007) HYPERLINK "http://faostat.fao.org/site/567/default.aspx"

Kalloo G (1988) *Vegetable Breeding,* Vol III, Boca Raton, FL, CRC Press.

Khan R (1979) Solanum melongena and its ancesteral forms. In: JG Hawkes, RN Lester and AD Skelding (eds.) *The Biology and Taxonomy of Solanaceae.* Linnean Society Symposium, Series '7', Academic Press, London (GBR), pp. 629-636.

Kumar G, BL Meena, R Kar, SK Tiwari, KK Gangopadhyay, IS Bisht and RK Mahajan (2008) Morphological diversity in brinjal (*Solanum melongena* L.) germplasm accessions. *Plant Genet. Resour. Characterisation and Utilization* **6(2):** 232-236.

Mahajan RK, IS Bisht, RC Agrawal and RS Rana (1996) Studies on south Asian okra collection: Methodology for establishing a representative core set using characterization data. *Genet. Resour. 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. Resour. Crop Evol.* **43**: 1-6.

Sapra RL, SK Lal, A Talukdar and KP Singh (2006) Selecting accessions in soybean collections with high diversity. *Indian J. Plant Genet. Resour*. **19(2):** 283-284.

Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants. *Chron. Bot.* **13**: 1-364.

Weaver W and CE Shannon (1949) *The Mathematical Theory of Communications.* University of Illinois Press, Urbana.