Genetic Resources and Crop Evolution **43:** 249–255, 1996.

Studies on South Asian okra collection: Methodology for establishing a representative core set using characterization data

R.K. Mahajan, I.S. Bisht, R.C. Agrawal & R.S. Rana
National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi-110 012, India

Received 2 June 1994; Accepted 7 June 1995

Key words: Abelmoschus esculentus, core collection, South Asian okra collection

Summary

Proposed is a technique for establishing a representative core set of the South Asian okra germplasm collection maintained at the National Bureau of Plant Genetic Resources. The validity of the technique was confirmed by applying it to a well characterised sample data set of 260 accessions of diverse geographical origin utilizing qualitative and quantitative descriptors. A sample core set of 53 accessions was extracted.

Abbreviations: PCA – principal component analysis; SDI – Shannon diversity index

Introduction

The huge size of germplasm assemblages of crop species and their wild relatives at several national and international agricultural research centres has led to increasing demand among the scientific community for effective measures to improve accessibility for these vast collections, and thus promote better utilization for crop improvement. In the past, only small portions of these germplasm collections have been effectively utilized by research workers. Recognizing this, Frankel (1984) proposed the 'core collection' concept to denote a set which would confirm germplasm accessions with minimum repetitiveness, at the same time retaining most of the genetic diversity of the entire collection. This concept was further developed by Frankel and Brown (1984) and Brown (1989a, 1989b, 1992). This concept attracted considerable attention and debate at the global level.

Proper understanding of genetic diversity among the constituents of germplasm collection and their proper documentation is essential for developing a representative set. Several national and international groups have developed/or are developing core collections of various crops such as soybean (Brown et al., 1987), okra (Hamon & van Sloten, 1989), winter wheat (Mackay, 1990), barley (von Bothmer et al., 1990), common bean (Tohme et al., 1992), sorghum (Prasada Rao, 1992), coffee (Hamon et al., 1992) and tulip (Loos & van Duin, 1989).

Most of the approaches developed so far classify the material on the basis of their origin, but due to the genetic drift over time and also because of adaptation of germplasm material over different environments, classification of accessions on this basis appears to be a difficult task.

An okra collection of 189 accessions was established in 1985 by ORSTOM/IBPGR okra in Ivory Coast. It was selected with the objective of:

1. to have a manageable collection scaled down to the needs of users/breeders and
2. include the widest range of variability described by the passport, characterisation and evaluation data, also including rare gene types (Hamon & van Sloten, 1989).

It will be worthwhile to propose a technique for establishing a core collection of South Asian okra collection maintained at the NBPGR.

The National Bureau of Plant Genetic Resources (NBPGR) holds over 2,000 accessions of okra (*Abelmoschus esculentus* (L.) Moench) and related wild species. These include collections made from various regions of India and parts of neighbouring countries, viz, Nepal, Sri Lanka and Bangladesh, and also a

few early introductions from Brazil, Nigeria and other countries. Information on these germplasm accessions have been documented in the form of crop catalogues using International Board for Plant Genetic Resources (IBPGR) descriptors with a little modification (Thomas et al., 1990, 1991; Bisht et al., 1993). An attempt has also been made to measure and classify the extent of diversity in these accessions (Bisht et al., 1995).

In the present investigation a methodology for selection of accessions for establishing a core collection in okra is proposed. The technique will be used in establishing a core collection for South Asian okra maintained at NBPGR.

Materials and methods

A total of 260 represenative germplasm accessions with diverse geographical background was selected for the study. Along with two check cultivars, 'Pusa Sawani' and 'Sel-2', the accessions were planted at the NBPGR Experimental Farm, Issapur, New Delhi during the 1992 cropping season in an augmented design with 13 blocks. The crop was raised in two-row plots with row to row distance of 1 m and plant to plant distances of 50 cm. Usual agronomic practices were applied throughout the various stages of crop growth. Data on distinct morphological and agronomic characters were recorded (Table 1).

Characterization data on these 260 accessions were documented in the form of a catalogue (Bisht et al., 1993). In our earlier investigation (Bisht et al., 1995) data on 9 distinct quantitative descriptors viz. days to flowering, ridges/fruit, plant height, number of internodes, first flowering node, fruits on main stem, fruit length, fruit width and fruits/plant were used for non-hierarchical cluster analysis. Clustering was performed using Euclidean distance on the basis of these quantitative descriptors after standardization. The incremental sum of square was used as clustering criterion through the SPAR1 statistical package developed at the Indian Agricultural Statistics Research Institute, New Delhi. The software also uses the Beale (1969) criterion as stopping criterion. This resulted in 8 well-characterised clusters. We did not find much association between genetic diversity and geographical origin of the accessions under study. Number of days to flowering, plant height and various fruit parameters among quantitative descriptors, and pubescence and pigmentation of various plant parts among qualitative characters were found to be most significant in discriminating accessions and also in the analysis of variability. These clusters as such were utilized in sampling accessions for establishing a core collection in the present study.

A plant breeder is usually interested in accessions which possess some desired qualitative characters for incorporation in breeding programmes. Consider m_j important heritable characters in the j^{th} cluster with more or less uniform distribution of variability (m_j need not be large). Since each cluster is well characterized by certain quantitative and qualitative descriptors, such heritable characters may vary from cluster to cluster. Let n_j be the frequency of accessions in the j^{th} cluster.

Situation I

$n_j \leq m$, where m is arbitrarily fixed (say, 20) depending on the number of quantitative characters considered. Then through subjective approach select accessions one by one, computing the Shannon Diversity Index (SDI) following Galwey (1992) at each stage on important qualitative characters until it is approximately equal to SDI for the entire cluster. Such entries constitute the sample core subset. Before adopting such a subjective approach it is necessary to have a good interaction with the crop specialist/breeder and curator.

Computation of SDI. SDI for i^{th} qualitative character

$$SDI_i = \sum_{j=1}^{d_i} p_{ij} \log p_{ij},$$

d_i being the descriptor states for i^{th} descriptor and p_{ij} the proportion of accessions for j^{th} descriptor state of i^{th} descriptor.

Situation II

$n_j > m$, split the cluster into s_j sub-clusters with s_j being the number of combinations of descriptor states of m_j important heritable characters, and let q_{ij} be the frequency of accessions in the i^{th} combination of j^{th} cluster. In the present study we consider two important heritable characters, stem and fruit pubescence with three descriptor states each; hence $m_j = 2$ and $s_j = 9$. If $q_{ij} \leq m$, follow the procedure as in Situation I. In the situation with $q_{ij} > m$, evaluation data (quantitative) on the accessions are utilized for identifying accessions explaining maximum variability

Table 1. Descriptors used for characterisation of the germplasm

1.	Growth habit (1 – erect, 2 – medium, 3 – procumbent)
2.	Branching (1 – profuse, 2 – low)
3.	Stem pubescence (1 – glabrous, 2 – slight, 3 – conspicuous)
4.	Stem colour (1 – green, 2 – green with red patches, 3 – purple)
5.	Leaf lobing (number of lobes above the sixth node)
6.	Lamina margin (1 – deeply incised, 2 – narrowly incised, 3 – serrate)
7.	Leaf colour (1 – green, 2 – green with red veins, 3 – red)
8.	Number of epicalyx segments (1 – from 5–7, 2 – from 8–10, 3 – more than 10)
9.	Shape of epicalyx segments (1 – linear, 2 – lanceolate, 3 – triangular)
10.	Persistence of epicalyx (1 – non-persistent 7 days after flowering, 2 - partially persistent upto 7 days, 3 – persistent)
11.	Petal colour (1 – green, 2 – yellow, 3 – golden)
12.	Red colouration at petal base (1 – inside only, 2 – both side)
13.	Position of fruit on main stem (1 – erect, 2 – horizontal, 3 – pendulous)
14.	Immature fruit colour (1 – green, 2 – dark green, 3 – yellowish green, 4 – red, 5 – dark red)
15.	Mature fruit colour (1 – yellowish green, 2 – green, 3 – dark green, 4 – green with red patches, 5 – dark red, 6 – others)
16.	Length of peduncle (1 – from 1–3 cm, 2 – more than 3 cm)
17.	Number of ridges/fruit (1 – 5-edged, 2 – more than 5-edged)
18.	Fruit pubescence (1 – downy, 2 – slightly rough, 3 – prickly)
19.	Seed shape (1 – round, 2 – reniform)
20.	Seed surface (1 – downy, 2 – glabrous)
21.	Maximum plant height (cm)
22.	Stem diameter at base (cm)
23.	Maximum number of internodes
24.	Leaf length (cm)
25.	Leaf width (cm)
26.	Days to 50% flowering
27.	First flowering node
28.	First fruit producing node
29.	Number of fruits on main stem
30.	Fruit length (cm)
31.	Fruit width (cm)
32.	Number of fruits/plant
33.	Days to initial maturity
34.	Number of seeds/fruit
35.	100-seed weight (g)
36.	Yellow vein mosaic virus (1 – low susceptibility, 2 – medium susceptibility, 3 – high susceptibility)
37.	Fruit and stem borer (1 – low susceptibility, 2 – medium susceptibility, 3 – high susceptibility)

(accessions × descriptors) in the collection set through principal component analysis (PCA). This technique is applied after ensuring that correlation among these quantitative descriptors are significant. If the correlation for sizeable number of pairs are not significant, adopt the procedure as in situation I. Principal component technique takes care of the multicollinearity between descriptors. The information on (p × t) scores obtained from PCA is used for extracting the accessions accounting for maximum variability through the Hamon and Noirot (1990) approach described below:

Selection of accessions through PCA. Let y_{ij} be the score through PCA for i^{th} ($i = 1, 2, \ldots p$) accession and j^{th} ($j = 1, 2, \ldots t$) principal component. The inertia of i^{th} accession is

$$P_i = \sum_{j=1}^{t} y^2{}_{ij}.$$

Then the relative contribution of this i^{th} accession is given by $CR_i = P_i/(p \times t)$. First we search two accessions with the highest relative contribution (CR) and simultaneously compute the SDI on them for important heritable characters. Subsequently we go on adding accessions one by one with the next highest CR until the SDI on them approximately equates the corresponding values for the entire cluster.

The accessions sampled from various sub-clusters within a cluster constitute the sample core subset. The accessions from various sample core sub-sets are pooled to form a sample core set. An equal number of accessions are drawn by applying PCA on the data on 260 accessions through Hamon and Noirots' (1990) approach. Both approaches were compared to study their relative efficacy judged by the better representation of diversity in the selected accessions. This was assessed by computing the diversity index for important qualitative characters through SDI for the sampled core set.

It has been well established that a core collection need not necessarily include extremely rare genes. A small portion may get included in the core set due to chance factors alone and many may need not be present at all. The moderately/extremely rare genes, if available, in the collection set should be directly added to the sample core set.

Results

Significant correlation among important descriptors *viz.* plant height, days to flowering, number of internodes and various fruiting parameters were observed in various clusters. The principal components with eigen values greater than 1 were retained. The components together explained more than 55% variation in different clusters (Table 2). It also presents the relative contribution made by certain characters in the analysis of variability.

Since the variability in stem and fruit pubescence had a fairly good distribution in different clusters, further sub-grouping was done on the basis of these two qualitative descriptors in cluster I alone. The four identified sub-groups in cluster I were as follows:

Group I – stem pubescence = 1 and fruit pubescence = 1 with 24 accessions;

Group II – stem pubescence > 1 and fruit pubescence = 1 with 12 accessions;

Group III – stem pubescence = 1 and fruit pubescence > 1 with 20 accessions, and

Group IV – stem pubescence > 1 and fruit pubescence > 1 with 15 accessions.

Four accessions each from Group I and Group III through PCA and SDI, and 3 accessions each from Groups II and IV through subjective approach explaining maximum diversity were identified in cluster I. Thus the core sub-set from the first cluster consisted of 14 accessions. No sub-grouping was possible in other clusters as there were less than 20 accessions in their sub-clusters or groups. The PCA and SDI was, therefore, applied directly on the total number of accessions in each of these clusters. Application of PCA extracted 4,6,3,8,8 and 5 accessions from clusters II, III, IV, V, VI and VIII respectively (Table 3) thus constituting 6 sample core sub-sets. The two distinct accessions of cluster VII formed a separate core sub-set. The accessions from various core subsets were pooled to form a sample core set. Besides, 3 particular accessions were added to the selected 50 accessions making a total of 53 selected accessions for the sample core set. Another set of 50 accessions was also directly selected from the 260 accessions through the Hamon and Noirot approach. A comparison of variability obtained by both approaches is given in Table 3.

Discussion

The proposed procedure describes the statistical selection of accessions accounting for maximum variability in a given set, using characterization data. Initially, the clustering technique grouped the accessions into 8 distinct clusters on the basis of similarity in morphological characters. A critical examination revealed that qualitative and quantitative characters were well represented in each cluster, mainly pubescence of various plant parts, and days to flowering and other fruiting characters, respectively.

A unique feature of the proposed methodology is that the accessions from clusters/sub-clusters were selected using quantitative data (through PCA/subjective approach) and qualitative data (through SDI) in a sub-cluster. Accessions selected were judged for their distinctness and were approved by the crop specialist and the curator. Accessions selected through PCA and SDI from individual clusters proved to adequately represent the diversity present. Sub-grouping of accessions based on certain qualitative characters, followed by selecting accessions through PCA/subjective approach resulted in a still better representation of diversity. This could be confirmed

Table 2. Eigen values, variation explained and weightage for important characters in different clusters

PC Axes	Eigen values	Total variation explained		Weightage for
		%	Cumulative	
Cluster I				
I	2.02	22.50	22.5	First fruiting node (0.75)
II	1.64	18.30	40.8	Fruit width (0.51), Internodes (−0.52), Days to flowering (0.68)
III	1.36	15.20	56.0	Fruit length (−0.54)
Cluster II				
I	2.27	25.28	25.28	Plant height (0.82), Fruits on main stem (0.45)
II	1.76	19.63	45.91	First fruting node (−0.56), Fruit width (−0.47)
III	1.49	16.60	61.51	Fruits on main stem (−0.62), Days to flowering (0.44)
Cluster III				
I	2.83	31.50	31.50	Days to flowering (0.88)
II	1.67	18.60	50.10	Fruit length (−0.61) First fruiting node (0.68)
III	1.56	17.30	67.40	Plant height (−0.56)
Cluster IV				
I	3.00	33.38	33.38	Fruits per plant (−0.78)
II	1.69	18.76	52.14	Number of internodes (0.69)
III	1.30	14.44	66.58	Fruit width (−0.61), Plant height (0.66)
Cluster V				
I	2.07	23.04	23.04	Days to flowering (0.68)
II	1.60	17.80	40.84	Fruits per plant (0.66) Number of internodes (0.67)
III	1.40	15.49	56.33	Fruit length (0.45)
Cluster VI				
I	2.59	28.80	28.80	Number of internodes (−0.88)
II	1.86	20.70	49.50	First fruiting node (0.67), Plant height (0.48)
III	1.18	13.12	62.62	Fruit length (0.66), Plant height (0.54)
Cluster VIII				
I	2.62	29.12	29.12	Days to flowering (0.52), Plant height (−0.48)
II	1.49	16.58	45.70	Number of internodes (−0.57), Plant height (−0.55)
III	1.18	13.13	58.83	First fruiting node (0.67), Fruit width (−0.41)

by comparing the explained variability by selected core accessions from a total of 71 in cluster I and the one computed from an equal number of accessions selected from various sub-groups of cluster I. It was observed that core accessions constituted 20% of all accessions. This high percentage could be attributed to a relatively small set of 260 accessions used in the present study.

Table 3. Number of core accessions and variation explained by their contribution in different clusters

Cluster	Variation explained	Number of accessions	Selected accessions		Common accessions
			Through proposed technique*	Directly from the set of 260**	
I	45	71	14(45)	0	0
II	48	28	4(48)	5	3
III	57	28	6(55)	17	4
IV	49	17	3(48)	14	3
V	48	51	8(45)	1	0
VI	55	31	8(52)	11	2
VII	—	2	2	2	2
VIII	45	32	5(49)	0	0
Total		260	50	50	14

*The figures in parenthesis indicate the percentage variability explained by the core entries. Overall variability explained was 59% by these 50 accessions.
**48% variation was explained by 50 selected accessions.

Table 4. Shannon Diversity Index for important qualitative characters

Character	Diversity Index		
	Whole set of 260 accessions	Selected accessions from a set of 260 (Hamon & Noirot)	Selected accessions from various clusters (Suggested approach)
Branching	0.212	0.262	0.259
Stem pubescence	0.341	0.268	0.297
Stem colour	0.245	0.217	0.248
Fruit pubescence	0.338	0.254	0.312
Seed pubescence	0.302	0.278	0.300

The present approach was found to be better than that suggested by Hamon and Noirot. Its usefulness was shown by examining the SDI on certain qualitative characters for the core accessions (Table 4). The SDI was consistently higher for the suggested approach, particularly for stem and fruit pubescence. Further, the variability explained by the core set was found to be 59% in contrast to 48% through the Hamon & Noirot approach (Table 3).

So the proposed approach will be used to select 200–250 accessions for the South Asian okra core collection. The work on measuring diversity index considering all qualitative characters put together, has already been initiated and would be made use of in establishing the core collection. It was observed earlier that the geographical diversity contributed very little in the analysis of genetic variability (Bisht et al., 1995; Ario, 1987). This criterion, however, has been a major consideration in sampling genetical diversity for establishing core collections of most crops.

Acknowledgements

The International Board for Plant Genetic Resources provided financial support for the okra germplasm evaluation programme at NBPGR.

References

Ario, O.J., 1987. Measurement and classification of genetic diversity in okra (*Abelmoschus esculentus*). Annals of Applied Biology 116: 335–340.

Beale, E.M.L., 1969. Cluster Analysis. Scientific Control Systems Ltd., Berners Street, London.

Bisht, I.S., D.P. Patel, R.L. Sapra, T.A. Thomas, M.N. Koppar & R.S. Rana, 1993. Catalogue on okra (*Abelmoschus esculentus*) germplasm Part II, 77 pp. NBPGR, New Delhi.

Bisht, I.S., R.K. Mahajan & R.S. Rana, 1995. Genetic diversity in South Asian okra (*Abelmoschus esculentus*) germplasm collection. Annals of Applied Biology 126: (In Press).

Brown, A.H.D., 1989a. The case for core collections. In: A.H.O. Brown, O.H. Frankel, D.R. Marshall & J.T. Williams (eds.), The use of plant genetic resources, pp. 136–156, Cambridge University Press, Cambridge.

Brown, A.H.D., 1989b. Core collections: A practical approach to genetic resources management. Genome 31: 818–824.

Brown, A.H.D., 1992. The core collections at the crossroads. Paper presented in International Workshop on Core Collections of Plant Germplasm held at CENARGEN, Brasilia, Brazil in September.

Brown, A.H.D., J.P. Grace & S.S. Speer, 1987. Designation of a "core" collection of perennial *Glycine*. Soybean Genetics Newsletter. 14: 59–70.

Frankel, O.H., 1984. Genetic perspectiveness of germplasm conservation. In: W.K. Arber, K. Llimensee, W.J. Peacock & P. Starlinger (Eds.), Genetic manipulation: Impact on man and Society, pp. 161–170, Cambridge University Press, Cambridge.

Frankel, O.H. & A.H.D. Brown, 1984. Current plant genetic resources — a critical appraisal. In Genetics: new frontiers. 4, pp. 1–11, Oxford & IBH Publishing Co., New Delhi.

Galwey, N.W., 1992. Veryfying and validating the representatives of a core collection. Paper presented in International Workshop on Core Collection of Plant Germplasm held at CENARGEN, Brasilia, Brazil in September 1992.

Hamon, S. & M. Noirot, 1990. Some proposed procedures for obtaining a core collection using quantitative plant characterization data. Paper presented at International Workshop on Okra held in NBPGR, New Delhi, India, October. pp 8–12

Hamon, S., M. Noirot & F. Anthony, 1992. A statistical procedure for selecting a core collection with quantitative data-application to coffee. Paper presented in International Workshop on core collections of plant germplasm held at CENARGEN, Brasilea, Brazil in September.

Hamon, S. & D.H. van Sloten, 1989. Characterization and evaluation of okra. In: A.H.D. Brown, O.H. Frankel, D.R. Marshall & J.T. Williams (Eds.), The use of plant genetic resources, pp. 173–196, Cambridge University Press, Cambridge.

Kendall, S.M., 1980. Multivariate Analysis. Charles Griffin & Co., Second edition, London. pp. 13–45.

Loos, B.P. & P.J.W. van Duin, 1989. Establishing a core collection representing genetic variation in tulip. Plant Genetic Newsletter. June-September. pp. 11–12.

Mackay, M.C., 1990. Strategic planning for effective evaluation of plant germplasm. In: J.P. Srivastava & D.S. Damania (Eds.), Wheat genetic resources: meeting diverse needs, pp. 21–25, John Wiley & Sons, Chichester.

Prasada Rao, K.E., 1992. The use of characterization data in developing a core collection. Paper presented in International Workshop on Core Collections of Plant Germplasm held at CENARGEN, Brasilia, Brazil in September.

Thomas, T.A., I.S. Bisht, D.P. Patel, R.L. Sapra & R.S. Rana, 1990. Catalogue on okra (*Abelmoschus esculentus*) (L.) Moench) germplasm: Part I, NBPGR, New Delhi. 51 p.

Thomas, T.A., I.S. Bisht, D.P. Patel, R.C. Agrawal & R.S. Rana, 1991. Catalogue on okra (*Abelmoschus esculentus*) (L.) Moench) germplasm: Part II, NBPGR, New Delhi. 100 p.

Tohme, J., P. Jones, S. Beebe & M. Iwanaga, 1992. The CIAT *Phaseolus vulgaris* core collection established by the combined use of evolutionary, characterization and agroecological data. Paper presented in International Workshop on Core Collection of Plant Germplasm held at CENARGEN, Brazil in September.

von Bothmer, R., G. Fischbeck, Th. van Hintum, T. Hodgkin & H. Knüpffer, 1990. The Barley core collections, Final report of BBC working group, Weihenstephan. 14 p.