Indian J. Pl. Genet. Resources 12(1): 1-9, 1999

SAMPLING STRATEGIES FOR DEVELOPING INDIAN SESAME CORE COLLECTION

R. K. MAHAJAN, I. S. BISHT AND P. L. GAUTAM, National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi 110 012

Various stratified random sampling strategies were compared for developing a core collection of indigenous sesame *Sesamum indicum* L accessions. Twenty diversity groups (strata) representing 3,129 Indian sesame accessions were used for the study. The Shannon Diversity Index (SDI) pooled over 19 descriptors was used as the measure of the diversity. Simple random sampling alongwith six stratified random sampling strategies were compared for varying sample fractions ranging from 5 to 30%. Stratified random sampling was always superior to simple random sampling. Of the six stratified sampling strategies, a genetic diversity dependent strategy with sampled accessions from various strata directly proportional to the product of the diversity and logarithm of the size of the strata (GL strategy) appeared to be the best strategy. A 10% sample fraction could be determined as the suitable sample fraction, invariably for all sampling strategies, serving both objectives of germplasm management and use.

Key words : Sesame, *Sesamum indicum* L genetic diversity, shannon diversity index, stratified sampling

Large assemblages of plant germplasm have accumulated in genebanks with time and have led to a serious concern among the scientific community as to whether the full range of genetic diversity they contain can be effectively managed and utilized. These collections are often poorly described due to lack of quality information on passport, characterisation and evaluation data. It has been estimated that about 65 per cent of the accessions in world collections have no passport data, 80 per cent are not characterised and only 1 per cent have been extensively evaluated (Peeters and William, 1984). Lack of easy access, resource constraints and the large number of accessions make proper evaluation of germplasm material a difficult task. Developing procedures for creating a collection of manageable and accessible size (a core collection) is among the most important issues in the management and utilization of plant germplasm collections (Frankel, 1984; Brown, 1989a; Marshall, 1990). The work carried out so far clearly indicates that it is possible to develop a core collection using existing data on the accessions in a collection. The basic issues that are to be addressed for developing a core collection are i) stratification of collection into homogenous (diversity) groups; ii) optimum sample size i.e., number of accessions to be sampled from the whole collection; and iii) the sampling strategies i.e., how to select core entries from groups.

The grouping approaches described and used in most of the core collection studies are mainly hierarchical. Taxonomic, geographical, agro-ecological, morpho-agronomic, biochemical and molecular markers have often been used individually or in combinations for hierarchical grouping of the accessions. Data from genetic markers could also be used in a variety of ways to create a representative (core) collection (Brown and Schoen, 1994; Bataillon et al., 1996; Gepts, 1995). The issue on optimum sample fraction was addressed by Brown (1989a,b) and a sample size of 10 per cent of the whole collection was suggested based on a theoretical model put forward by Ewens (1972). But in practice, the situations may differ markedly from the assumptions of the model. For sampling within groups, two techniques are mainly employed, i) random sampling and ii) use of principal components score strategy. Brown (1989a,b), Charmet and Balfourier (1995) and Yonezawa *et al.* (1995) proposed stratified random sampling strategy. Different strategies viz., Constant (C), Proportional (P), Logarithmic (L) by Brown (1989b), Genetic diversity dependent (G) by Yonezawa *et al.* (1995), M and H by Schoen and Brown (1995) were proposed for random selection of entries from groups. Principal components score strategy have been used in common bean (Singh *et al.*, 1991), okra (Hamon and van Sloten, 1989; Mahajan *et al.*, 1995), coffee (Hamon *et* al., 1995) and mungbean (Bisht *et al.*, 1998a).

The diversity in Indian sesame accessions has already been studied for a set of descriptors and the collection has been classified in different groups making combined use of passport information on origin of accessions and morpho-agronomic characterisation data (Bisht *et al.*, 1998b). In the present investigation, estimates of diversity using qualitative and quantitative descriptors were compared for various stratified random sampling allocation strategies and the optimum sample size determined for developing a core collection of Indian sesame accessions.

MATERIALS AND METHODS

In our earlier investigation, stratification of 3129 indigenous sesame accessions was done making combined use of agro-ecological passport information and morpho-agronomic characterisation data (Bisht *et al.*, 1998b). The morpho-agronomic characters were used for classifying the entire collection into 7 discrete clusters (main groups). The seven clusters, in combination with agro-ecological passport data, finally resulted in 20 diversity groups (Table 2). These diversity groups were used for comparing various sampling strategies in the present study.

Data on 18 characters, both qualitative and quantitative, were used for estimation of Shannon Diversity Index (Shannon and Weaver, 1963). The quantitative characters were converted into qualitative descriptors making frequency classes as described in Table 1.

Computation of Population Shannon Diversity Index (SDI)

Consider a population of N accessions spread over different descriptor states of each of the d qualitative descriptors. Let there be k_i ($i=1,2,3,....,d$) states for the i^{th} descriptor. Denote N_{ij} for the number of accessions associated with j^{th} descriptor state of the i^{th} descriptor such that $\Sigma_j N_{ij} = N \;\forall i$. Then $P_{ij} = N_{ij}/N$ denote the proportion of accessions belonging to the j^{th} descriptor state of the i^{th} descriptor. Thus the population Shannon Diversity Index (SDI_i) for the i^{th} descriptor is

$$SDI_i = -\Sigma_j P_{ij} \log_e P_{ij}$$

and the population diversity index is

$$SDI = \Sigma_i SDI_i$$

$$= \Sigma_i \Sigma_j P_{ij} \log_e P_{ij}.$$

Estimation of Shannon Diversity Index

i) Simple Random Sampling Without Replacement (SRSWOR)

A simple random sample of size n is drawn by SRSWOR from N accessions. The frequency

Table 1. List of descriptors recorded for the study

1.	Branching habit (1= un-branched; 2= low; 3= moderate; 4=high (bushy type))
2.	Stem hairiness (1= glabrous; 2= sparse; 3= hairy; 4= very hairy)
3.	Flower colour (1=deep violet/purple; 2=white with deep violet/purple shading; 3= white with light violet/purple shading; 4=wqhite)
4.	Corolla hairiness (1= glabrous to light hairy; 2= hairy; 3= very hairy)
5.	Internode length (l=low; 2=medium; 3=high)
6.	Capsule shape (1= narrow oblong; 2= broad oblong; 3= broad)
7.	Density of capsule hair (1=glabrous; 2=sparse, 3=high)
8.	Number of capsule per leaf axil (1=one; 2= two to three)
9.	Number of locules per capsule (1=four; 2=four and six; 3= four, six, eight and more than eight)
10.	Capsule length (1=small; 2=medium; 3=long)
11.	Seed colour (1=white; 2=black; 3=brown; 4= red; 5= grey; 6= others)
12.	Incidence of Phyllody disease (1= low susceptibility; 2=medium susceptibility; 3=high susceptibility)
13.	Days to 50% flowering (number of days to 50% plants having first flower open from date of planting; 1=30-35, 2=36-40, 3=41-45, 4=46-50, 5=51-55, 6=56-60, 7= >60 days)
14.	Days to 50% maturity (number of days to 50% plants reaching physiological maturity from date of planting; 1=65-75, 2=76-85, 3=86-95, 4=96-105, 5=106-115, 6= >115 days)
15.	Plant height (in cm, Mean of 5 random plants; 1=20-35, 2=36-50, 3=51-65, 4=66-80, 5=81-95, 6= >95)
16.	Capsules per plant (mean of 5 plants; 1=1-10, 2=11-20, 3= >20)
17.	Seeds per capsule (mean number of seeds from 5 capsules each per plant; 1=10-30, 2=31-50, 3=51-70, 4= >70)
18.	100-seed weight (g; 1=0.05-0.10, 2=0.11-0.15, 3=0.16-0.20, 4=0.21-0.25, 5=0.26-0.30, 6=0.31-0.35, 7= 0.35)

distribution of these n units over descriptor states for each of the descriptors is computed. Let p_{ij} ($i=1,2,3,\ldots,d$; $j=1,2,\ldots k_i$) denote the sample proportion of accessions for j^{th} state of the i^{th} descriptor. Then an estimate of SDI is

$$sd_i = -\Sigma_i\Sigma_j \, P_{ij} \log_e P_{ij}$$

Due to the complex nature of sdi, it is difficult to judge its unbiasedness. However, sd_i is a consistent estimate of SDI. For computation of E(sdi) and V(sdi), 100 repeated independent random samples of size n by SRSWOR were drawn from the population and sdi, average sd_i and variance among sdi values were computed. Then mean square error (MSE) of sd_i is

$$MSE\ (sd_i) = V(sd_i) + bias^2\ (sd_i)$$

where bias $(sd_i) = E(sd_i) - SDI$

and then 95% confidence interval (CI) can be computed.

ii) Stratified Random Sampling

Let the population of N units be grouped into k strata with size $\{N_s\}$ $s = 1,2,\ldots,k$ for the s^{th} stratum such that $\Sigma_s N_s = N$ Samples of sizes $\{n_s\}$ are drawn independently by SRSWOR from the sth strata such that $\Sigma_s n_s = n$. Assuming the accessions in a stratum constituting the entire population, SDI is computed for the s^{th} stratum. Then SDIst, the pooled SDI_s over different strata, is defined as

$$SDIst = \Sigma_s \, W_s \, SDI_s$$

Where W_s is the weight attached to the s^{th} stratum and SDI_s is the SDI for the s^{th} stratum.

The weights W_s are to be selected such that $SDIst \cong SDI$. The weights Ws are chosen by the researcher on the basis of past experience. In the absence of any idea about the weights, the following three set of weights were considered:

a) *Equal weights*: Equal weights being attached to each of the k strata, i.e., $W_s = 1/k \; \forall \; s$.

b) *Proportional Weights*: Weights being proportional to the size of the strata, i.e., $W_s \propto N_s \Rightarrow W_s = N_s/N$ for the s^{th} stratum.

c) *Proportional to Stratum Diversity* : Weights being proportional to diversity in the s^{th} stratum., i.e.,

$$W_s \propto SDI_s \Rightarrow W_s = SDI_s/\Sigma_s SDI_s$$

The set of weights for which $SDIst \cong SDI$ was considered for estimation of population diversity. Let it be W'_s for s^{th} stratum. Let $(sdi)_{sts}$ be the sample diversity index for the sth stratum, then the estimator of SDI is

$$(sdi)_{st} = \Sigma_s W'_s (sdi)_{sts}$$

The researcher may select another set of weights on the basis of his past experience about the importance to be attached to different strata. In situations where it is not possible to choose W_s, $(sdi)_{st}$ will be an unbiased estimate of SDIst but a biased estimate of SDI.

Allocation Strategies

Assuming the sampling fraction (5, 10, 15, 20, 30 %) being fixed in advance, the following six allocation strategies were compared.

Strategy I (P strategy): The number of accessions sampled from different strata are directly proportional to the size of the strata i.e. $n_s \propto N_s$;

Strategy II (C strategy): Equal number of accessions i.e., $n_1 = n_2 = n_3 \ldots = n_s = n/k$ are selected from each stratum;

Strategy III (L strategy): The number of accessions sampled from different strata are directly proportional to the logarithm of the size of the strata i.e., $n_s \propto \log_e N_s$;

Strategy IV (G strategy): The number of accessions sampled from different strata are directly proportional to the diversity in the strata i.e., $n_s \propto SDI_s$;

Strategy V (GP strategy): The number of accessions sampled from different strata are directly proportional to the product of size and diversity in the strata i.e., $n_s \propto N_s \, SDI_s$;

Strategy VI (GL strategy): The accessions from different strata are directly proportional to the product of diversity and logarithm of the size of strata i.e., $(SDI)_{sts} \log_e N_s$.

For obtaining the expectation, MSE of $(sdi)_{st}$ and the corresponding CI the procedure adopted was same as in SRSWOR with bias being given by

$$\mathbf{Bias((sdi)_{st}) = E((sdi)_{st}) - SDI_{st}}$$

RESULTS NAD DISCUSSION

The population SDI was recorded to be 14.62. The SDI from stratified sampling with weights proportional to (i) size of strata, (ii) equal weights and (iii) diversity in the strata was found to be 11.36, 11.59 and 11.82 respectively. SDI with weights proportional to diversity was the highest but not very close to the population SDI. But in the absence of any knowledge of Ws, the weights proportional to diversity were used for comparing alternative sampling allocation strategies in stratified sampling.

Shannon diversity indices of different strata are presented in Table 2. It is evident from the table that diversity in different strata is not uniform. It ranged from 8.58 to 14.43. Diversity was maximum in stratum 6 (14.43) followed by stratum 8 (13.83) and 15 (13.36). It varied from 10 to 12 in the remaining strata except 7, 10 and 20 where the SDI was observed to be 9.43, 9.26 and 8.58, respectively. SDI was comparatively high in small groups as compared to the large ones.

Table 2. Diversity group size and their Shannon Diversity Indices (SDI)

Diversity group	No. of accessions	SDI
1	333	11.65
2	59	12.00
3	60	11.76
4	142	12.00
5	50	13.82
6	66	14.43
7	211	9.43
8	65	13.83
9	100	12.31
10	235	9.26
11	470	10.81
12	200	11.92
13	65	10.78
14	37	10.07
15	66	13.36
16	124	10.62
17	395	11.65
18	402	11.89
19	28	11.75
20	22	8.58

Comparison of different sampling strategies is presented in Table 3. In strictly random sampling, the diversity varied from 68.23 to 93.72 per cent for a sample size of 5 to 30 per cent. The mean square error (MSE) varied from 1.00 to 29.60. The width of the confidence interval decreased with increase in sample size. There was considerable increase in %SDI for all sampling strategies from 5 to 10 per cent sample fraction. Thereafter, there was no substantial increase in the %SDI with increasing sample size. A comparison of MSE clearly indicates that the genetic diversity dependent strategy particularly IV and VI always performed better in comparison to all other strategies (Table 3). For equal allocation strategy, it was not possible to go beyond 10per cent sample size. Strategy VI as consistently superior to other strategies with least MSE and a narrow 95 per cent confidence interval for different sample sizes considered. Strategy I with % SDI ranging from 51.68 to 76.36 per cent was poor performer with highest MSE for different sampling fractions.

The sampling strategy to obtain a core sample should maximize the diversity in the sample while attempting to reduce the redundancy of identical genotypes (Frankel and Brown, 1984). Genetic diversity, however, is not randomly distributed over plant populations. It has a structure that can generally be summarised in a hierarchical model (van Hintum, 1995).

In the present study, both agroecological passport data on origin of accessions and clustering technique based on phenetic analysis of individual accessions has been used for stratification of accessions in diversity groups.

A commonly used measure of diversity is the Shannon-Weaver information theoretic expression (Shannon and Weaver, 1963), often known as the Shannon diversity index. The SDI, pooled over different descriptors, has been considered as a measure of diversity. It gives larger values when more descriptor states are present, and other things being equal, it gives larger values when descriptor states are equally common than when some are common and others are rare. However, it is rarely affected by inclusion or

Table 3. Shannon Diversity Index and its statistical parameters for simple random sampling and other allocation strategies

Sample Size (%)	Statistical parameters	Simple random sampling	Stratified sampling allocation strategies					
			I	II	III	IV	V	VI
5%	SDI	9.66	7.55	9.60	9.59	9.62	7.83	9.87
	%SDI*	68.23	51.68	65.66	65.59	65.82	53.61	67.51
	MSE	29.60	8.20	4.90	4.94	4.78	7.62	4.12
	CI	1.00-20.32	1.94-13.16	5.26-13.93	5.23-13.94	5.33-13.90	2.42-13.24	5.88-13.86
10%	SDI%	11.34	9.66	10.75	10.65	10.99	9.71	11.05
	SDI	80.15	66.14	73.58	72.82	75.24	66.47	75.58
	MSE	12.86	4.59	1.12	1.35	1.02	4.37	0.95
	CI	4.31-18.37	5.47-13.86	8.68-12.83	8.38-12.93	9.02-12.97	5.61-13.81	9.14-12.96
15%	SDI	12.59	10.35	-	11.12	11.18	10.41	11.28
	%SDI	86.12	70.80	-	76.07	76.49	71.24	77.15
	MSE	5.88	2.13	-	0.47	0.39	1.96	0.29
	CI	7.84-17.34	7.49-13.21	-	9.77-12.47	9.92-12.41	7.66-13.15	10.23-12.33
20%	SDI	12.88	10.71		11.38	11.37	10.79	11.58
	%SDI	88.12	73.96	-	77.90	77.83	73.81	79.21
	MSE	3.13	1.21	-	0.18	0.19	1.04	0.14
	CI	11.15-18.09	8.55-12.86	-	10.55-12.22	10.52-12.23	8.78-12.79	10.85-12.31
30%	SDI	13.70	11.16	-	11.55	11.56	11.19	11.62
	%SDI	93.72	76.36	-	79.03	79.10	76.62	79.48
	MSE	1.00	0.42	-	0.07	0.06	0.37	0.05
	CI	11.74-1566	8.82-12.43	-	11.04-12.06	11.08-12.03	10.00-12.39	11.18-12.06

*SDI expressed as percentage of population SDI

exclusion of rare descriptor states in the sample (Emlen, 1973). This is a desirable attribute in some context, but in a core collection rare descriptor states are of considerable importance, and need to be included with subjective approach, wherever possible.

In stratified sampling, weights are required to be attached to different strata in view of their relative importance over one another based on the degree of diversity in the different strata. All the weights are equal when the strata have equal variability and then stratification is as good as random sampling. But stratification is done when sampling units are heterogeneous with regards to characters of interest. Pooling of results over different strata is done such that the resulting SDI is approximately equal to the population SDI. In development of SDI for stratified sampling, unweighted estimates are considered. Due consideration for suitable weights is a practical necessity. Therefore, we consider here weights on the basis of closeness of pooled estimates to the population value. Since for weights proportional to diversity, the SDI^{st} is closer to population SDI, therefore these weights formed the basis of comparison of different sampling strategies in stratified sampling.

In the present study, sampling through stratification based on well identified diverse groups were more efficient efficient over simple random sampling. To overcome the problem of giving undue higher weights to large groups, likely to contain a higher level of genetic redundancy, Growth and Roelfs (1987) and Brown (1989b) suggested that the number of accessions to be drawn should be proportional to the logarithm of the frequency in each group (L strategy). The C strategy gives each group equal weighting and biases the core in favour of small groups. The P strategy considers a fixed proportion of each group so that the group is represented in the core in proportion to the frequency in the group. Clearly this strategy is biased towards large groups (Brown, 1989b). The P strategy is likely to perform better when diversity is distributed according to the size of groups, that in actual situation does not hold. The C strategy is likely to perform better than the P strategy when diversity is concentrated in smaller groups. In other situations, constant allocation can be inferior to simple random sampling. The genetic diversity dependent (G) strategy proposed by Yonezawa *et al.* (1995) takes into consideration the diversity in different groups irrespective of their sizes and it has been emphasized that this strategy is always better when the diversity in different groups is known. This directed us towards the development of some other intermediate strategies. Therefore, in the present study two more probable diversity dependent strategies, GL and GP, were also considered that may take care both of group sizes and the level of diversity. We find that the GL strategy performed better than the G and GP strategies. The GL strategy is more appropriate when one encounters situations with different levels of diversity and sufficiently large strata sizes while GP is likely to perform better in situations when groups are of small sizes and diversity is unevenly distributed within groups. In the last three strategies, we come across situations when the diversity in certain strata is very large and the corresponding strata size are small enough. Then ns is likely to be greater than Ns. In such situations we take ns = Ns. Apart from the above suggested strategies, other sampling strategies might be applied to develop a core collection but their performance will depend upon the species, the composition and distribution of diversity in collection and the type of characters of interest (Spagnoletti Zeuli and Qualset, 1993).

While comparing various sampling strategies on Durum wheat, Spagnoletti Zeuli and Qualset (1993) found that the three strategies i) random stratified by geographic origin with P strategy, ii) random stratified by geographic origin with L strategy and iii) random stratified with canonical variables produced the desired effects of increasing frequencies from the less representative countries of origin. Diwan *et al.* (1994) proposed random sampling and classified the accessions of annual *Medicago* species-wise and accessions within species country-wise. Sampling within species was done by cluster analysis on agronomic and morphological traits. Basigalup *et al.* (1995) classified accessions of perennial *Medicago* plant introductions according to country of origin and eight sampling strategies were compared. Two methods, combined cluster analysis based on principal components within each geographical group with random selection of entries within each cluster and direct selection of entries within each geographical group were adjudged to be the best strategies for designating the core collection. Charmet and Balfourier (1995) compared Shannon Diversity index averaged over 200 samples generated by computer simulations using various proportions for sampling methods. The clustering method based on agronomic traits with geographic contiguity constraints was better than others. A random sample of 5% of accessions maintained 86 per cent of the diversity while clustering method based on geostatistics gave the best results with 92 per cent of the variation being maintained in a 5 per cent of the core sample.

Considerable increase in %SDI was observed when sample fraction increased from 5 to 10 per cent. No substantial increase was observed beyond

10 per cent sample fraction. Brown (1989a) also proposed an optimum sample size of 10 per cent. Applying Ewens theory, Brown (1989a) showed that the fraction of alleles retained from a population increased rapidly as the sample fraction increased to 0.1, but rather slowly after it surpassed 0.1. Similar was the case in the present study in all the sampling strategies tested. On the basis of the theoretical model, Yonezawa *et al.* (1995) showed that the optimum sample fraction depends upon various genetic and resource parameters, primarily on the degree of genetic redundancy among accessions comprising the whole collection and the amount of resources available for maintenance of the core collection. The optimum sampling fraction was large, with a lower redundancy among accessions or a lower initial allelic diversity within accessions, indicating that a large sample fraction is better in species with a large selfing rate. A 20-30 per cent sample was estimated to be appropriate in situations where accessions in the collection are neither very heavily nor very lightly redundant in terms of diversity.

Based on the above discussion, it is evident that suitable choice of weights, selection of an appropriate sampling allocation strategy along with a optimum sampling fraction is essential for capturing the maximum diversity. In the present study, the weights proportional to diversity in various strata, sampling allocation strategy with sample size directly proportional to the product of diversity and logarithm of the strata size and 10% sample fraction is appropriate for the development of sesame core set and it is likely to serve both the objectives of germplasm management and utilization.

ACKNOWLEDGEMENTS

We are thankful to Dr. Toby Hodgkin, Principal Scientist, Genetic Resources Science and Technology Group and Dr. A.H.D. Brown, Sr. Consultant, IPGRI, Rome for critically going through the manuscript and valuable suggestions. Thanks are also due to IPGRI for financial support for the sesame core collection project operative at the NBPGR, New Delhi.

REFERENCES

Bataillon, T.M., J.L. David and D.J. Schoen. 1996. Neutral genetic markers and conservation genetics. *Genetica* **144**: 409-417.

Basigalup, D.H., D.K. Barnes and R.E. Stucker. 1995. Development of Core Collection for Perennial *Medicago* Plant Introductions. *Crop Science* **35**: 1163-68

Bisht, I.S., R.K. Mahajan and D.P. Patel. 1998a. The use of characterisation data to establish the Indian mungbean core collection and assessment of genetic diversity. *Genetic Resources and Crop Evolution* **45**: 127-133

Bisht, I.S., R.K. Mahajan, T. R. Loknathan and R.C.Agrawal. 1998b. Diversity in Indian sesame collection and stratification of germplasm accessions in different diversity groups. *Genetic Resources and Crop Evolution* **45**: 325-335.

Brown, A.H.D. and D.J. Schoen. 1994. Optimum sampling strategy for core collections of plant genetic resources. V. Loeschgke, J. Tomiuk and S.K. Jain (eds). *In*: Conservation Genetics p. 357-370 Birkhauser Verlag, Basel, Switzerland.

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

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

Charmet, G. and F. Balfourier. 1995. The use of geostatistics for sampling a core collection of perennial ryegrass populations. *Genetic Resources and Crop Evolution* **42**: 303-309.

Diwan, Noa, Gary R. Bauchan and Marla, S. McIntosh. 1994. A Core Collection for the United States annual *Medicago* germplasm collection. *Crop Science* **34**: 279-85.

Emlen, J.M. 1973. Ecology: An Evolutionary Approach. Reading, Massachusetts, USA: Addison-Wesley.

Ewens, W.J. 1972. The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**: 87-112.

Frankel, O.H. and A.H.D. Brown. 1984, Current plant genetic resources - a critical appraisal. *In*: Genetics: New frontiers. 4, p. 1-11, Oxford and IBH Publishing Co., New Delhi.

Frankel, O.H. 1984. Genetic perspectiveness of germplasm conservation. *In*: W.K. Arber, K. Llimensee, W.J. Peacock and P. Starlinger (eds.). Genetic manipulation: Impact on man and society, p. 161-170. Cambridge University Press, Cambridge.

Gepts, P. 1995. Genetic markers and core collections. *In*: Hodgkin T., A.H.D.Brown, Th.J.L.van Hintum and E.A.V. Morales (eds.), Core Collection of Plant Genetic Resources, p 127-146, John Wiley & Sons.

Growth, J.V. and A.D. Roelfs. 1987. The concept and measurement of phenotypic diversity in *Puccinia graminis* on wheat. *Phytopathology* 77:1395-1399.

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

Hamon, S., M. Noirot and F. Anthony. 1995. Developing coffee core collection using the principal components score strategy with quantitative data. *In*: Hodgkin T., A.H.D.Brown, Th.J.L.van Hintum and E.A.V.Morales (eds.), Core Collection of Plant Genetic Resources, p 117-126, John Wiley & Sons.

Mahajan, R.K., I.S. Bisht, R.C. Agrawal and R.S. Rana. 1996. Studies on South Asian okra collections: Methodology for establishing a representative core set using characterisation data. *Genetic Resources and Crop Evolution* 43: 249-255

Marshall, D.R. 1990. Crop genetic resources: Current and emerging issues. *In*: Brown, A.H.D., M.H. Clegg, A.L. Kahlor and B.S. Weir (eds). Breedings and Genetic Resources. Sunderland, Massachusetts, U.S.A., Sinauer.

Peeters, J.P. and Williams, J.T. 1984. Towards better use of gene banks with special reference to information. *Plant Genetic Resources Newsletter* 60: 22-32.

Schoen, D.J. and A.H.D. Brown. 1995. Maximising genetic diversity in core collections of wild relatives of crop species. *In*: Hodgkin T., A.H.D.Brown, Th.J.L.van Hintum and E.A.V.Morales (eds.), Core Collection of Plant Genetic Resources, p 55-76, John Wiley & Sons.

Shannon, C.E. and Weaver, W. 1963. The Mathematical Theory of Communication. Urbana, Illinois.

Singh, P.S., R. Nodari and P. Gepts. 1991. Genetic diversity in cultivated common bean: I. Allozymes. *Crop Science* 31: 19-23.

Spagnoletti Zeuli and C.O. Qualset. 1993. Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. *Theoretical and Applied. Genetics* 87: 295-304.

van Hintum, Th. J.L. 1995. Hierarchical approaches to the analysis of genetic diversity in crop plants. In: T.Hodgkin, A.H.D. Brown, Th. J.L.van Hintum and E.A.V.Morales (eds.), Core Collection of Plant Genetic Resources, p 23-34, John Wiley & Sons.

Yonezawa, K., T. Nomura and H. Morishima. 1995. Sampling strategies for use in stratified germplasm collections. *In*: Hodgkin T., A.H.D. Brown, Th.J.L.van Hintum and E.A.V. Morales (eds.), Core Collection of Plant Genetic Resources, p 35-54, John Wiley & Sons.