Indian J. Pl. Genet. Resources 9(1): 11-30, 1996

# DEVELOPING A CORE SET OR SUBSETS OF LARGE GERMPLASM COLLECTIONS : AN OVERVIEW

**R.K. Mahajan, Sudhir Kochhar, R.C. Agrawal and K.P.S. Chandel**

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

The concept of developing core germplasm set as a minimal yet diverse set, representing diversity of the germplasm collection has attracted wide attention in recent years with the expectation that it will provide an effective window to access the crucial genetic diversity. Procedures for selection of core entries have been considered in view of the available information on core collections for *Glycine* spp., wheat, barley, okra and others. An example of constituting a subset in okra and its proposed extension in *Sesamum* is given.

**Key Words :** Core collection, germplasm

Huge germplasm collections representing diverse genetic variability of crop species and their wild relatives have been accumulated by several International Agricultural Research Centres (IARCs) and many national programmes through continuing explorations and exchanges. The volume of these collections has outgrown the management resources. Scientific community is increasingly interested in assessing potential of the existing genetic diversity contained in these collections enabling promotion and use of plant genetic resources. Suggestions for ways of managing large, diverse germplasm collections to facilitate their efficient use despite their size have been proposed (Chang, 1991). Conventionally, the accessions were short listed after characterization and evaluation only by few morpho-agronomic attributes, identifying as the promising accession(s), for single or multiple traits, for promoting their utilization. Further, the idea of developing core collections has been proposed and debated for a long time leading to suggestions whether breeders need one core or many of them depending upon their requirements.

Frankel (1984) first proposed that a huge population could be condensed to a 'core collection' which would contain, with a minimum of repetitiveness, the genetic diversity of crop species and its wild relatives in the whole collection. The core subsets should include as much as possible of its genetic diversity. Accessions not included in the core would not be jettisoned, according to the proposal, but retained as the back up 'reserve collection'. Alternatively

defined, a core collection is a diverse, representative yet severely limited sample (Peacock, 1989; Palmer, 1989). The definition was modified by Frankel and Brown (1984), Brown (1989a, b) and elaborated by CENARGEN (1992) to state that the core collection is a part of the existing collection and does not replace it. According to IBPGR, core collection is to denote the part of a gene bank collection freely available on request without restriction (IBPGR, 1985). The alternative part, or 'reserve', is the one unavailable for general distribution because of restrictions by plant breeders rights, narcotic regulations, etc. This usage, however, has neither precedence nor semantics in its favour.

## CONCEPT OF A CORE COLLECTION

### Genesis

Collection and collation of characterization and evaluation data describing the extent of variation in the germplasm and a better understanding of distribution of the genetic diversity, in the base collection is required for effective management and utilization of PGRs (Hodgkin, 1991; Crossa *et al.*; 1992; Gawley, 1992; Jaradat, 1992; Kresovich *et al.*, 1992 and Rana, 1992 b). The 'core' concept revolves round the investigations on the extent and distribution of genetic diversity in a cultivated species and its relatives followed by the delineation of accessions that optimally represent this available genetic diversity.

For setting up of a core collection, a few assumptions should be fulfilled such as: the base collection includes (i) accessions from the entire range of adaptability of the species, (ii) all the congeners, to ensure representative inclusion of alleles, (iii) above 3000 individuals, to ensure inclusion of upto 70 per cent of selectively neutral alleles, (iv) related wild species, (iv) passport data, to ensure description of core entries with respect to geographical origin, status of the germplasm i.e. landrace, old or obsolete cultivar, breeders stock etc., and homogeneous or heterogeneous nature of core entries.

The word 'core' means the most central, the innermost part or the heart. Brown (1989a) opined that core collection was most important component of the Plant Genetic Resources (PGR) dogma. The term 'core' has been frequently used in scientific literature and fiction in different contexts, but its proposal in PGR was a radical departure (Frankel, 1984). Until then, efforts were concentrated on an open ended task, irrespective of continuing cost and use, for collecting as many samples as possible and securing them in gene banks. Elimination of duplicates was suggested to reduce the maintenance costs yet it was unlikely to have a significant impact due to already huge size of the collections and a high number of unique types (Chang, 1991). Besides, Frankel and Brown (1984) introduced the notion of adequacy of sampling of the species range. This implied stratification of accessions into non-overlapping groups

based on analysis of the passport information on eco-geographical parameters. The proposal included selecting a proportion of the germplasm accessions from these groups of base collection, on the basis of well defined sampling technique(s) to represent major kind of diversity. However, setting up a core in smaller collections, particularly in cleanly propagated crops, by eliminating genetical redundancies, was also recommended to promote resource utilization and effective management (CENARGEN, 1992). Two modifications to the core concepts have been proposed. Due emphasis was given by national PGR programmes to develop situation specific subsets of the base germplasm in order to promote utilization. Thus instead of developing a single core the germplasm would be categorised into 'usable' sets of collection which adequately represented specific traits like disease and pest resistance, tolerance to stress conditions and adaptability to specific ecogeographic range. Secondly, core subsets, alteration of the core sets, is normally used for a set of designated accessions within an existing collection. This approach is more suitable for situations where population size is no consideration and entries of the core need not be physically separable from the entire collection (Brown, 1992).

The needs of molecular plant breeders have been given added impetus to the designation of core collection. Identification of single genes using recombinant DNA technology have, of late, focussed the attention of plant molecular markers may be used for documentation of germplasm and detection of variability in collections. Efforts are also being made to develop trait-specific DNA probes for screening of germplasm accessions to detect genes for resistance to biotic and abiotic stresses as well as for desirable agronomic traits.

## APPLICABILITY

Germplasm collections are reservoirs of biological diversity but search for desirable breeding stocks from these collections often takes very long time. Therefore, breeders depend heavily on limited strains evolved in various breeding programmes and tested in multilocational evaluation trials under different ecosystems. The core germplasm would assist in searching for desired traits more efficiently and authentically from the available core set and related passport data.

Designation of a core is needed for testing general combining ability of diverse representative germplasm with locally adapted strains for yield contributing traits (Frankel and Brown, 1984). Mating of a constant parent with core entries in a test cross can also provide fair estimates of combining ability with delineated set of germplasm (Zeuli, 1992). The combining ability of the standby subsets or reserve collections may be tested by 'Line x Tester' crossing with known lines out of the designated core subset. The in-depth studies

carried on various basic and applied aspects on accessions in the core would streamline their effective utilization.

Most of the germplasm related activities require making choice or set priorities among accessions. Difficulties encountered in decision making whether new additions in germplasm would add new information or they would be redundant, selecting a particular set of accessions for evaluation, testing for viability or regeneration or responding to breeders' requests are of great concern (Brown, 1989a). A set of well documented core collections will greatly assist in indexing new accessions into various categories and also provide a guide for conservation and exploration activities (Strauss *et al.*, 1988).

Ways and means for improving management or utilization of germplasm collections, the reservoir of candidate genes, through the application of powerful tools of biotechnology, which deserve sharp focus due to recent advances in techniques for rapid screening and gene manipulation, can be studied efficiently on a representative set of core collections than on a random lot out of base collection (Spagnoletti Zeuli, 1992; Vaughan and Jackson, 1992; Tohme *et al.*, 1992).

## CRITERIA FOR ESTABLISHING A CORE COLLECTION

Genetic diversity in plant populations is structured in a way that reflects the biological characteristics, distribution and ecology of the species examined (Hamrick and Godt, 1990; Nevo *et al.* 1988; Nevo and Beiles, 1989). A core subset should contain the breadth of such genetic diversity. The mode of reproduction alongwith the geographic origin of collections is significant in determining the observed distribution of quantitative characteristics and allozymes (Spagnoletti Zeuli and Qualset, 1987; Kahler and Allard, 1981). Outbreeding (Allogamous) species often possess higher diversity and less genetically differentiated populations than inbreeding (Autogamous) ones. Further, grouping of unique accessions into different sets/subsets require elaborate passport data, but many accessions may lack such information. Variable migration rates of germplasm material in the historical past and/or well structured selections in defined regions might have lead to uneven differentiation among groups of accessions. Such irregular phenomena make it difficult to select representative and yet non-redundant set of accessions. Methods for selecting a core subset still continue to be a matter of scientific debate (Brown, 1983a, b; Chapman, 1987; Strauss *et al.*, 1988).

Various statistical approaches have been advocated for core identification based alternatively on the sampling or the population genetic theory. Aspects like size of base population, size of 'core' sample, adequacy of data including diverse representation of collections from across the globe, skewness of evaluation data, and probability of upward or downward bias in sampling

while developing a core subset are a few important considerations underlying the statistical approach for core delineation. The case for a designated 'core' would strengthen if the information loss is minimal.

The criteria for set up of an efficient, representative and unbiased core which revolves around the genotype (population genetic theory) or the phenotype (sampling theory) may be reviewed separately in the relative contexts.

## POPULATION GENETICS THEORY

The germplasm accessions were broadly classifiable into four classes of alleles based upon the preponderance or rarity of different alleles into their genetic build up (Marshall and Brown, 1975, 1983). The accessions classified under Class I carry widespread and common allele, whereas those represented by Class II have widespread and rare allele; the Class III individuals may posses localized and common allele and class IV accessions have localized and rare allele in their genetic background. The priorities for inclusion of various classes of germplasm based on gene frequency may vary according to the preponderance and occurrence. Least priority is given to deliberate inclusion of class I as the widespread and common allele has a high probability of inclusion in any one of the indirect selection criteria. Class III variables may have a cross-adaptive value and thus be used sooner or later in breeding programmes and may be safely represented in the core by taking care of appropriate sample sizes. Inclusion of common-localized alleles poses certain problems for selecting neutral alleles. Brown (1989a) used the neutral allele model of Kimura and Crow (1964), and its sampling theory (Ewens, 1972; Nei, Maruyama and Chakravarty, 1975). They assumed random sampling in order to meet the baseline expectations. However, stratified random sampling based on ecological factors can strengthen the inclusion of widespread and rare alleles (class II) in core collections. Localized and rare allele in class IV is highly desirable for a 'core' due to the fear of its being lost during regeneration in routine. These localized rare alleles may also be new recurrent mutations which could be useful for mutation breeding programmes. This class has many a times been put forward in debates as a counter example to proposed strategies (Bogyo *et al.*, 1980). However, the accessions carrying localized rare alleles can be included in the flexible core accomplished on their subsequent detection during exhaustive statistical analysis of evaluation data followed by verification from rapid screening techniques.

Brown (1989a) showed that in large populations irrespective of different levels of polymorphism, a 10% sample (at least 3,000 accessions) of the entire population retained at least 70% of the 'rare -widespread' group of alleles with a frequency greater than 0.0001 and with a confidence limit of 95 per cent.

Schoen and Brown (1992) discussed assemblage of core collection of wild crop relatives in terms of maximisation of allelic diversity. Such issues are helpful for theoretical core collections but information about individual alleles in wild relatives is difficult to obtain due to limited availability of base primers to detect alternate alleles and also because alleles don't occur independently (Gawley, 1992). The representativeness of core collections in terms of available genetic diversity in the base population can be validated at different levels viz., allelic representation in terms of allozymes, RFLP analysis etc. (Brown, 1989b), phenotypic representation in terms of clustering diversity for quantitative traits (Mackay, 1992; Crossa *et al.* 1992) and then delineating core entries. Yet another level of representativeness ie. the genetic level which would be most useful for efficient PGR utilization and which may be expressed in terms of elaborate multilocational evaluation traits, combining ability analysis and even inheritance studies is difficult to attain due to extremely high level of input resources required to meet such objectives. On the other side, the definition of base population would also have to be viewed at two different planes depending upon existing strategies of core delineaion (Mackay, 1992). Accordingly, international genepool based core collection in terms of efficient management under networking of genebanks and situation specific subsets (Rana, 1992 b) or alternatively terms as 'specific attribute based subset' (Mackay, 1992) in terms of efficient utilization of PGR's strongly fulfil the criteria for core setup population genetic theory.

## SAMPLING THEORY

Sampling, drawing of a representative sample from the population, is bound to result in some loss of information in terms of adequate representation and remainders. The sampling technique e.g. stratified sampling are found adequately efficient for grouping collections into distinct groups (Brown, 1989 b) and better choice over simple random sampling, a default option.

The hierarchy of grouping is based on elaborate pass port and characterization data including taxonomy of the crop. At first step the taxa may be subgrouped or split according to the origin of accessions, as major regions based upon political maps i.e., continents, countries, states, or ecogeographic regions. Further, the collections subgrouped within a region may be clustered into groups of similar accessions for important inherited characters like resistance to major pests and diseases, pathogen races or physiological stresses etc. through multivariate clustering analysis (Spagnoletti Zeuli and Qualset, 1987). The hierarchical cluster analysis could, however, be used as a tool to classify germplasm collections even when no passport data is available (Peeters and Martinelli, 1989). Ideally, the extent of diversity within each cluster should be same as far as possible otherwise the selection strategies need be suitably modified. Lack of authentic information on evaluation data

or missing data, thereby making uneven distribution among accessions, leads to problems in hierarchial cluster analysis which generally require a complete rectangular data matrix. Exclusion of accessions with incomplete data is likely to reduce diversity of the core collection to a considerable degree (Gawley, 1992). Information on genetic diversity, from cytological studies, marker loci, or quantitative characters, if available, can also be used to ensure that each group has a substantial level of diversity.

Sample selection of accessions within each group for constituting the core may be guided by different strategies, viz., (i) Constant strategy (C): selection of equal sample size from each group, (ii) Proportional strategy (P): number of accessions being proportional to the total number of accessions in the particular group, and (iii) Logarithmic strategy (L): sample size being proportional to logarithm of population size in each group, have been widely acclaimed (Groth and Roelfs, 1987, Brown, 1989a, b). In addition, a few more sampling strategies have been proposed by various workers *viz.*, (i) Genetic-multiplicity-dependent strategy (G): sample size proportional to the amount of genetic multiplicity available in groups (Yonezama and Nomura, 1992). (ii) H-strategy; takes into account Neis' genetic diversity index (H) or panmictic heterozygosity to estimate variation among geographical regions using marker locus data in effective population size (Ne) or indices correlated with Ne (Schoen and Brown, 1992). It assumes selective neutrality of the marker locus variation and also that the populations are isolated. (iii) M-strategy: also suggested by above authors uses non-linear programming methods to pinpoint; accessions for maximum diversity within each geographical region. (iv) R-strategy: involves random sampling of accessions irrespective of stratification. Further, Gawley (1992) used C, L and P strategies for evaluation of effectiveness of the Shannon index H′ to account for the diversity of a discrete descriptor based on proportion of accessions having the ith descriptor state ($P_i$) where $H^l = -\sum p_i \log p_i$. Thus, a situation specific selection of particular sampling strategy would help developing a more meaningful, representative and effective core subset.

A well elaborated, systematic order for hierarchical (stratified) sampling has been reported by Chang (1991) in preference to random sampling. According to author the prioritis for stratification (continents or regions); species (*sativa, glaberrima*); ecogeographical race (indica, japonica, javanica, hybrid); geo-political information (country of origin); cultural type and hydroedaphic regime (lowland, upland, deepwaer, tidal wetland etc.); maturity, plant stature, other morphoagronomic traits; grain type (dimensions and shape, pericarp colour, endosperm type); known economic attributes (pest and stress tolerances); special types with genetic information; isozymes, seed proteins, RFLP's data, if available. The same can be suitably adapted, in general, subjected to modifications as per specific requirements of more or different strata in other crops as well.

Marshall (1990), on the other side, advocated that random, stratified-random proportional to geographic frequencies and based on statistical analysis of evaluation data as most optimal sampling strategies to ensure alleic inclusion from the four classes of their prevalence.

For delineation of core sets in structured populations, sampling proceaures from reserve collections were investigated on the principal of maintenance of largest genetic multiplicity within a given amount of resources (time, labour, facilities etc.). Empirical studies indicate that the P-strategy would have the widest range of application over C or L, although strategy G is expected to be superior to others in situations with known degree of genetic multiplicity within groups whereas the strategies H or M were most useful in situation where genetic diversity among loci within accessions was positively correlated (Yonezawa and Nomura, 1992). Brown (1989 b), however, demonstrated that strategy L was better placed for sampling accessions in *Glycine tomentella* and USDA's Barley core collections

## EXISTING CORE COLLECTIONS

A survey was initiated at the end of 1989 by IBPGR to look into the ongoing work on core collections by scientific community throughout the world. Based on the overall interest shown by various national and international organizations, over 20 projects involving cereals, legumes, forage crop species, fruits and vegetable crops were identified (T. Hodgkins, Pers. Comm.). But till date the number of core collections with published information is very limited. At the same time, the absence of a well defined approach in the published literature on core germplasm does not substantiate any definite criteria for future guidelines. A few representative projects related to core collections are reviewed hereunder.

## GLYCINE Spp.

A core collection of 111 accessions of perennial *Glycine* spp was developed in Canberra, Australia (Brown *et al.*, 1987) from a collection of around 1400 accessions of 12 species. The criteria in choosing the accessions for developing a dynamic core were (i) more than one accession per *Glycine* spp. selected to provide replication for generalization of results from any basic or applied research carried out from these accessions at species level, (ii) geographical coverage in respect of political states as well as scatter and range of all possible habitats in Australia, (iii) inclusion of morphological, cytological and isozymes groups with known intraspecific variation and (iv) priority inclusion of accessions used for research in the past and the first hand collections. The core thus constituted was much different from a random set and achieved an overall target of about 10 per cent sampling. proportion from each category

i.e. species, geographical region or political state etc. varied drastically (Brown, 1989a).

Further, Brown (1989b) intensively demonstrated the effect of sampling strategy on specific inclusion of accessions in the core on *Glycine tomentella*, a species collected extensively from all across its natural range in Far East Asia and Oceania and which comprised 5 diploid (2n = 38,40) and tetraploid (2n = 78,80) groups. In this example the choice of alternate strategies (C, P or L) showed marked effects on the chance of including rare types. For the diploid strategy P gave the best recovery whereas for the tetraploids strategy C was the best option. The strategy L, however, emerged as a good compromise and was thus favoured by Brown (1989b) in cases where the representative set of germplasm showed skewed distribution for rarest variants, e.g. in diploids, occurring in the largest group and tetraploids, available in the smallest group, in *Glycine tomentella.*

## OKRA

Hamon and van Sloten (1989) established a core collection on okra based on the variability in passport, characterization and evaluation data from a single trial. A core comprising 189 accessions out of 2,283 accessions was developed with specific inclusion of rare types. This core collection on core was set up with a slightly varied approach in order to have a manageable collection scaled down to the needs of the breeders and other users. In order to include widest possible range of variability it was constituted primarily on the basis of rigorous statistical analysis of the available data on agromorphological characters. It suffered major limitations in terms of missing passport data to the tune of 20 percent for collectors' number, 30 per cent for town/province, 33 per cent for latitude/longitude 60 per cent for vernacular names and 99 percent for altitude thereby making it difficult to follow a stratified random sampling based on population genetics approach.

The statistical analyses resorted to were (i) univariate analysis *viz.*, for quantitative descriptors and some basic statistics like mean, range, standard deviation, coefficient of variation and frequency distribution for taxonomical or qualitative morphological descriptors, (ii) bivariate analysis e.g. correlations among quantitative traits, (iii) multivariate analysis like factor analysis, principal component analysis and hierarchical clustering methods elucidating the euclidean distances among various quantitative characters. Three distinct groups were earmarked at the euclidian distance level of 0.80 and a choice at level 0.30, which authors availed in this situation, allowed elimination of four descriptors.

Some of the interesting observations for setting up core in okra included using the cultivation system as an important source of information wherein

the graphical representation of variability by means of factorial analysis corresponded closely to the farming systems used, long tradition of okra cultivation in a particular country/agroecological zone and the choice of morphological descriptors. Generalization of such an indirect approach is, however, difficult to make due to the non- availability of descriptors like Farming System in the routine data recorded by collectors.

Other significant features, emerged from the core collection were that the reduction in number of descriptors obtained through multivariate analysis emerged as another significant feature in the okra core collection. A stepwise approach consisted of running a principal component/factor analysis projecting the variability in limited dimensions. Then a clustering analysis was done which could be later tested by a discriminant analysis. In such an approach the choice made by the curator/database manager leads to a deliberate but controlled loss of information. Nonetheless, the base collection was effectively reduced in size to a manageable representative set for further use.

## WINTER WHEAT

A 'core' on Australian winter wheat collection was developed by Mackay (1990) on the basis of passport, morphological and agronomic data. This core collection was built up as a specific attribute based subset rather than genebank based core collection. It puts considerable emphasis on the use of ecogeographic data for selecting accessions that were likely to include all the diversity for a particular desired character. This approach could be easily extended to developing situation specific subsets in the national PGR systems of the developing countries as proposed by Rana (1992a).

## BARLEY

Since 1989, an approach to develop a barley core collection was under consideration by the barley working group of the European Cooperative Programme for Conservaion and Exchange of Genetic Resources (von Bothmer *et al.*, 1990). It is based on World's barley holdings, and comprises the whole genus *Hordeum*, including the secondary and tertiary genepools as well as genetic stocks (Kunpffer, 1992). This is an obvious alternative to the built up situation specific or specific attribute based subsets. It was further mentioned that the accessions in the core would be maintained as homozygous lines. To ensure the maintenance of core by genebanks, different phases in the establishment and operation of core collection including networking of international barley genetic resources were described. It was further proposed that genetic and cytogenetic marker stocks be included.

Brown (1989b) reviewed the core set up of worked barley collections of USDA (Kahler and Allard, 1981) for the effect of various sampling strategies

(strategies C, L and P and simple random sampling R) on esterase diversity found in the accessions using geometric probability for sampling with replacement (Emigh, 1983; Brown, 1989b). About 75 per cent of the alleles were likely to be included in a 10 per cent core, as predicted. The sampling variance was expected to be greater than that for independent loci although the total would not be affected by the intense linkage disequilibrium among the loci. There was a meagre effect of the different strategies on allele inclusion for any one of the 4 esterase loci, eventhough strategy L was best for polyallelic esterase 2 locus. This was ascribed to coarse level of stratification and the former exchange of barley germplasm around the world.

## COMMON BEAN

A core collection of 1400 accessions in *Phaseolus vulgaris* (common bean) from about 24,000 accessions stored at CIAT using evolutionary, agromorphology and agroecological data was established (Tohme *et al.* 1992). The core set comprised of 1200 accessions from primary centres, 200 from secondary centres and about 25 a priori selected bean lines as reference. A weighted stratified model was used. First an arbitrary weight as assigned to geographical areas delineated by similar subjective ecological classification. In corollary to the length of adoption of landraces in different farming systems in case of okra, the germplasm from the primary centers, in common bean, was assigned a higher grade than from area of recent introduction. The second level of stratification was based on morphological data.

## SORGHUM

A core collection of sorghum was planned by the Germplasm Resources Unit of ICRISAT which holds over 33,000 accessions from 88 countries. The size of core collection was worked out to be 3,475 i.e. nearly 10 per cent of the whole collection (Prasad Rao, 1992). For indentifying core accessions usage of three types of data was proposed (i) country of origin/geographical diversity, (ii) taxonomic group in numerical data of agronomic traits for cluster analysis. In this analysis wild sorghum were excluded as they need different descriptors.

## COFFEE

Out of nearly 20,000 genotypes representing about 75 taxons and raised under live conditions, 88 diversity groups corresponding to 73 taxons were taken into account for the hierarchical scheme of future coffee core collection (Hamon *et al.*, 1992). The core entries were proposed to be selected using principal component analysis to remove multicollinearity between descriptors and euclidian distances were weighted by standard deviation for measuring the basic distance. Genotype most distant from centroid (higher relative

contribution) was selected as the first genotype which maximizes the selected variability. Genotypes x descriptors interaction was involved to guide retention of accessions for the core or selection of descriptors for the analysis.

## ALFALFA (*MEDICAGO SATIVA* L. SENSU *LATO*)

A core set of 200 entries was extracted from about 1100 perennial *Medicago* P.I.s collections from 47 countries (Basigalup *et al*, 1995). The entire collection was classified into 18 geographical groups based on the available passport information on Germplasm Resource Information Network (GRIN) system. Within each group, entire were selected by eight methods (multivariate procedures, random and/or direct selection, and a totally random selected core) and comparison was made by sign test (a nonparametric statistical test). The method based on direct selection of entries within each group was the best one and was ultimately used for the selection of core entries. Also, the entries extracted by this method retained the greatest variability for 38 of 50 agronomic, forage quality, root and crown marphology, pest resistance and stress tolerance traits. Further, the methology uses incomplete data sets.

In another study, a core collection of 200 accessions (Diwan *et al*, 1994) was delinneated from a subset of 1240 germplasm accessions of 36 annual *medicago* species on the basis of 16 agronomic and morphological traits. Cluster analysis was followed to classify the accessions in each of the species separately. Within species, accessions were chosen proportionally to the country of the origin in the germplasm collection. Species with more clusters were considered more diverse than those with fewer clusters. One accession per cluster was selected. The core collection when grown for 2 years showed stability for the evaluated traits.

In addition to the above published evidences scattered information is available from various national programmes on the ongoing work in relation to core set up. In USA, a core of around 505 lines of chickpea has been delineated out of a germplasm collection of about 3491 accessions maintained at WRPIS (w-6), Pullman WA. The core germplasm was set up based on a few determinant characters in the descending priority order viz. 'country of origin > seed colour > seed size > seed shape > seed surface etc.' (Hannon, R.; Pers. Comm.*). This core collection was intensively screened for blight and other foliar diseases for effective utilization of representative germplasm. The pea core collection developed at Pullman, USA included 505 accessions (17.5%) in the core from a total of 2886 accessions. To efficiently represent genetic diversity, their first step was to include all *P. sativum* spp. *elatius* members based on high level isozyme variation in this group. In the second step, accessions were selected from the material of various countries based on number of accessions representing a particular country. Due to low seed

inventory, seed is only available for 84% of the core accessions (Hannon R., pers. Comm.). Loos and van Duin (1989) identified a set of 19 characterization traits and an additional set of 11 evaluation traits important for breeding in order to establish a core collection representing available genetic variation in tulip.

## CROP NETWORKING IN RELATION TO CORE

In the development of the core set or subsets accessions are selected from the existing holdings of genebank/s. These holdings are primarily available in base and/or active collections. The conceptual core collection may be safely treated as a curator guided working collection for the particular crop species. The plant improvement programmes, on the other side, revolve round the selection criteria and selection response on deliberately set up base population 'also referred to as working collection' which fulfills the criteria to meet the breeding objectives. A single working collection, leading to the set up of a base population for a defined breeding programme after a thorough uni/bi/multivariate analyses of the available data on quantitative traits, may be derived from a single base collection or may be developed from more than one base collections from different genebanks. Herein, the bulk population may be kept separate as a conservation alternative rather than an alternative for core set up or for active use in breeding programmes. Further, the base population/s may have be set up for individual breeding objectives, for incorporation of single or multiple traits from one or more of the working collections, core collections, base collections or collections over different genebanks depending upon the degree of diversity available.

The degree of diversity captured in a core set will be a reflection of diversity among accessions available in the genebank whereas the narrowness in apparent diversity for the adaptability factors may be specifically attributed to missing passport data as pointed out by Hamon and van Stolen (1989) in ORSTOM/IBPGR Okra collection, yet the real test of diversity could be provided by the biochemical or the genetical markers. Diversity is also restricted due to the closeness among cultivars originated from single breeding programme, co-adaptiveness of the landraces, or co-evolution of primitive cultivars etc. Various PGR programmes have duly recognized the need of conserving duplicate sets of germplasm accessions across the genebanks. There is a need of international collaboration to devise appropriate strategies likely to be of importance in maximizing the benefits of investigations on establishing core collections. The set up of international networks for effective exchange of database alongwith duplicate sets would help in appropriate statistical analysis, reduce the chance of availability of missing data for representative geographical sections of accessions and increase the probability of inclusion

of fullest range of diversity of the crop species in its core. The research workers' interest lies in exploiting maximum diversity in the genebanks for the core set up and, if the need be, to go in for further explorations in specific regions either unaccounted for or sparsely represented in the 'core'. It is thus desirable to have collaboration between genebanks, users and researchers in order to meet the objectives of maximum availability as well as diversity among the accessions. In other words, the development of crop-networks is closely linked to an effective core collection.

The basic idea of establishing crop-networks in order to strengthen core collections could be viewed differently at the international and national levels. Various international agencies like CGIAR, IBPGR and IARCs share the global responsibilities for crop based PGR programmes. On the other side, the national programmes like USDA, NBPGR would tend to fulfil the requirements for farming systems' need based, agroecological situation specific subsets for their respective countries and also contribute to global reserves through collaborative exploration, exchange and duplicate sets etc.

The region specific smaller PGR units within the national set ups constitute significant and integrated part within the national PGR network (Rana, 1992b). Besides other factors, such a netowrking within a national PGR programme would be helpful in retrieving further evaluation data (FED) from the multi-locational tests at National Active Collection Sites (Rana *et al.*, 1993) which would ultimately provide desirable database for constituting core collections. The need based, situation specific subsets, of the developing countries, may be developed by following either of the population genetic approach under stratified sampling using C, P or L strategy as in Soybean (Brown, 1989 a, b), in case the passport data is complete, or biometrical approach using statistical analysis of the evaluation data on quantitative characters as done in Okra (Hamon and van Sloţen, 1989).

## DEVELOPING A CORE SUBSET PROGRAMME AT NBPGR

A flow diagram depicting the essential features in establishment of core subsets program is given in Fig. 1. Essential features of such a programme aimed at developing a minimal core subset representing the diversity of base collections, would be as follows :

At the onset, grouping of accessions in the base collection is carried out on the basis of passport data for ecological, geographical or climatic factors, plant types through appropriate multivariate statistical tools. In the next step, within group selection is carried out for accessions accounting for maximum variability in qualitative and quantitative descriptors. A group may invariably be split into sub-groups equaling total number of combinations for a few

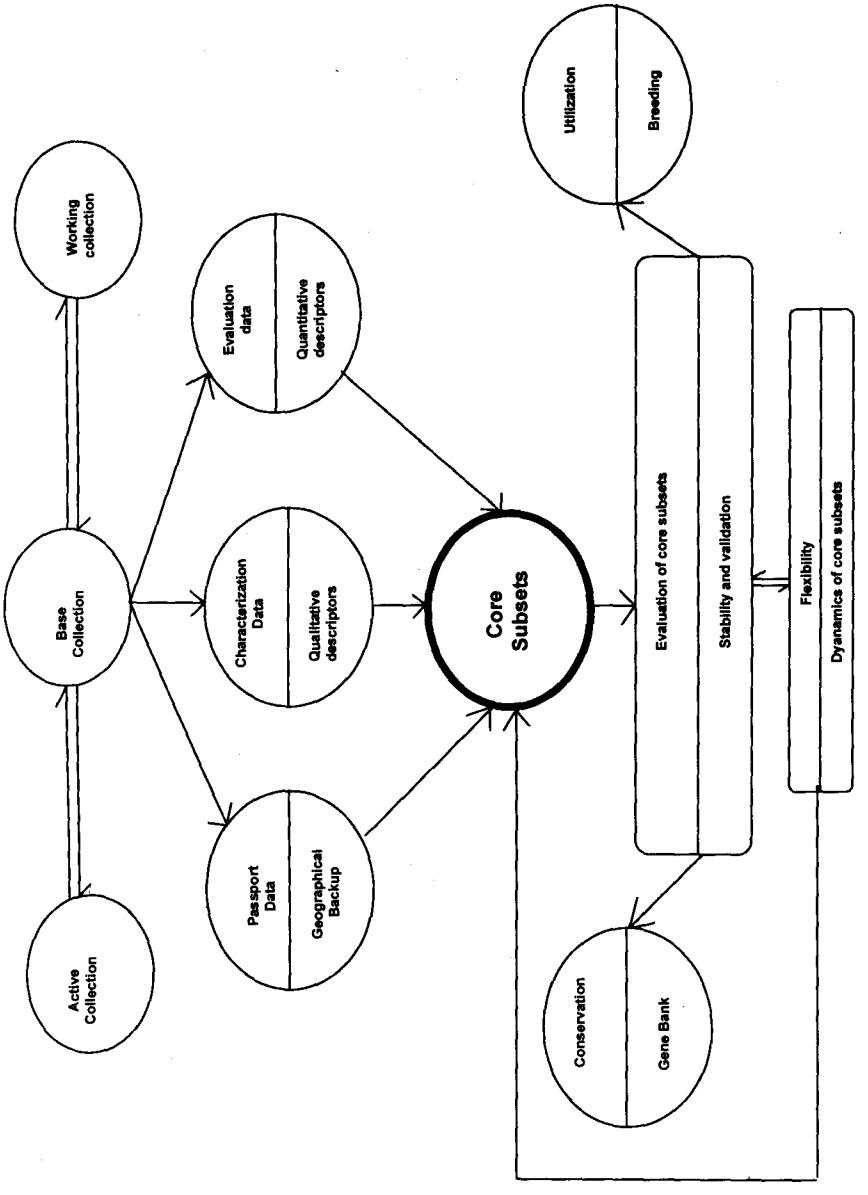


Fig. 1. Flow diagram for establishing core subsets

important qualitative descriptors at thistage, showing a fairly spread distribution order for accessions explaining maximum diversity within each subgroup through some appropriate multivariable technique (Mahajan *et al*, 1996). At the same time, Shannon Diversity Index (SDI) for these accessions is computed, for the important qualitative traits, on a cumulative scale. Accessions are retained upto a stage when the SDI of accessions becomes approximately equal to to greater than the corresponding value for the entire group/sub-group. The accessions thus selected from different sub-groups within a group would constitute a core subset. Situation specific subsets may be similarly obtained from other other groups.

A sample core set in okra has been worked out with limited size of base collection originated from series of interrelated exploration programmes and covering the variability from South East Asia, including some SAT (South African Taxons). There is ample variability for a species/taxa and the within species accessions in this set. However, this limited set was exercised at the initial stage to suit the available softwares of non- hierarchical cluster analysis which could accommodate 500 accessions and 10 quantitative descriptors and also to confirm the validity of the above approach. The 260 okra accessions produced 8 homogeneous unequal groups on the basis of quantitative characters (Bisht *et al.* 1995). In a later assessment of the database, these groups were ascribed to plant types. Each group was split into subgroups depending upon its size, frequency distribution of accessions on qualitative characters and variability observed within groups. Accessions were then selected from within each group/subgroup to account for maximum variability in quantitative characters, using Principal Component Technique, and qualitative characters using Shannons' Diversity Index (SDI). The selected accessions from various subgroups within a group consisted of a sample core set. This core set comprised of 50 accessions to which 3 rare types were deliberately incorporated making it a final set of 53 accessions.

Sesamum collection, more than 4,550 accessions is maintained at NBPGR. To provide effective germplasm utilization, improved genebank management and also to reduce the effective size of genebank holdings for different crops, a collaborative project with IPGRI for establishing a core collection has been initiated. The development of the methodology is in progress.

Identitficaton of Core Collections is likely to help in greater utilization of germplasm holdings in breeding programmes since researchers will find it to be a more practical way to reach ultimately the primary collection which is really the target for identifying desired traits and also for diversifying the sources of resistances to pests and pathogens among other goals.

## REFERENCES

Basigalup D.M., D.K. Barnes and R.E. Stucker. 1995. Development of a core collection for Perennial *Medicago* Plant Introductions. *Crop Science*. **35** : 1163-1168.

Bisht I.S., R.K. Mahajan and R.S. Rana. 1995. Studies on South Asian Okra (*Abelmoschus esculentus*) germplasm Collection I. Measurement and Classification of Genetic Diversity. *Annals of Applied Biology* : 539-550.

Bogyo T.P., E. Porcedolu and P. Perrino. 1980. Analysis of sampling strategies for collecting genetic material. *Economic Botany*, **34** : 160-74.

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

Brown, A.H.D. 1989b. Core collections: a practicl 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. Sept., 92.

Brown A.H.D., J.P. Grace and S.S. Speer. 1987. Designation of a "core" collection of perennial Glycine. *Soybean Genet. Newsl.* **14** : 59-70.

Chang T.T. 1991. Guidelines on developing cores collections of rice cultigens. *In*: Rice Germplasm: Collecting, preservation, use. IRRI: 105-107.

Chapman C. 1987. Criteria for selective core collection: Basis passport data. International Board for Plant Genetic Resources, Rome, Italy.

Crossa J., S. Taba and I. Dlacy. 1992. The use of multivariate methods in the process of forming core subsets. Paper presented in International Workshop on Core Collections of Plant Germplasm. Held at CENARGEN, Brasilia, Brazil. Sept., 92.

Diwan, Noa, Bauchan, H. Gray, MacIntosh and S. Marla. 1994. A core collection for the United States Annual *Medicago* Germplsm Collection. *Crop. Science.* **34**: 279-285.

Emigh T.H. 1983. On the number of observed classes from a multinomial distribution. *Biometrics*, **39**: 485-491.

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

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

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

Gawley N.W. 1992. Verifying and validating the representatives of a core collection. Paper presented in International Workshop on Core Collections of Plant Germplasm. Held at CENARGEN, Brasilia, Brazil. Sept., 92.

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

Hamon S. and M. Noirot. 1992. Some proposed procedures for obtaining a core collection using quantitative plant characterization data. Paper presented at International Workshop on Core Collections of Plant Germplasm. Held at CENARGEN, Brasilia, Brazil, Sept., 92.

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

Hamrick J.L.L. and M.J. Godt. 1990. Allozyme diversity in plant species. In: Brown, A.H.D.; Clegg, M.T.; Kahler, A.L. and Weireds, B.S. (eds). Plant Population Genetics, breeding and genetic resources. Sinaur Association, Sunderland, Massachussets. pp. 43-63.

Hodgkin T. 1991. Improving utilization of plant genetic resources through core collections. In: Rice germplasm-Collecting, Preservation and Use., IRRI, Philippines. pp. 99-104.

IBPGR. 1985. Report of the III meeting of IBPGR Advisory Committee on Seed Storage, IBPGR Rome.

Jaradat A.A. 1992. The dynamics of core collections. Paper presented in International Workshop on Core Collections of Plant Germplasm. Held at CENARGEN, Brasilia, Brazil. Sept., 92.

Kahler A.L. and R.W. Allard. 1981. Worldwide patterns of genetic variation among four esterase loci in barley (*Hordeum vulgare* L.). *Theor. Appli. Genet.* **59**: 101-111.

Kimura M. and J.F. Crow. 1964. The number of alleles that can be maintained in a finite population. *Genetics*, **49**: 725-738.

Knupffer H. (1992). The barkey core collection- An international effort. Paper presented in International Workshop on Core Collections of Plant Germplasm. Held at CENARGEN, Barsilia, Brazil. Sept., 92.

Kresovich S., W.F. Lamboy, J.R. McFerson and P.L. Forsline. 1992. Utilizing different types of genetic information to develop useful and representative collections. Paper presented in International Workshop on Core Collections of Plant Germplasm. Held at CENAGEN, Brasilia, Brazil, Sept., 92.

Loos B.P. and P.J.W. van Duin. 1989. Establishing a core collection representing genetic variation in tulip. *Plant Genetic Newsletter*. June/September:11-12.

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

Mackay M. 1992. One core or many. Paper presented in International Workshop on Core Collections of Plant Germplsm. Held at CENARGEN, Brasilia, Brazil. Sept., 92.

Mahajan R.K., I.S. Bisht, R.C. Agrawal and R.S. Rana. 1996. Studies on South Asian Okra Collection: Methodology for establishing core collection using characterization data. *Gene. Resou.& Crop Evol.* (In press).

Marshall D.R. 1990. Crop Genetic Resources: Current and emerging issues. *In*: A.H.D. Brown, Clegg M.T., Kahler A.L. and Weireds B.S. (eds). Plant Population Genetics, breeding and genetic resources. Sinaur Association, Sunderland, Massachussets.pp. 384-385.

Marshall D.R. and A.H.D. Brown. 1975. Optimum sampling strategies in genetic conservation. *In*: O.H. Frankel and Hawkes J.G., Crop genetic resources for today and tomorrow. Cambridge University Press, Cambridge. pp. 53-80.

Marshall D.R. and A.H.D. Brown. 1983. Theory of forage collection. *In*: J.C. Melver, Bray R.A. (eds). Genetic resources in Forage Plants, CSIRO, Melbourne. pp. 135-48.

Nei M., R. Maruyama and R. Chakravarty. 1975. The bottleneck and genetic variability in populations. *Evolutions*, **29**: 1-10.

Nevo E. and A. Beiles. 1989. Genetic diversity of wild emmer wheat in Israel and Turkey; Structure evaluation and application in breeding. *Theor. Appl. Genet.* **76**: 737-52.

Nevo E., A. Beiles and T. Krugman. 1988. Natural Selection of allozyme polymorphisms: a microgeographical differentiation buedaphic, topographical and temporal factors in wild emmer wheat (*Triticum dicoccoides*). *Theor. Appl. Genet.* **77**: 421-55.

Palmer R.G. 1989. Germplasm collections and the experimental biologist. *In*: A.H.D. Brown, Frankel O.H., Marshall D.R. and Williams J.T. (eds). The use of plant genetic resources. Cambridge. pp. 32-45.

Peacock W.J. 1989. Molecular biology and genetic resources. *In*: A.H.D. Brown, Frankel O.H., Marshall D.R. and Williams J.T. (eds). The use of plant genetic resources. Cambridge. pp. 363-376.

Peeters J.P.L. and J.A. Martinelli. 1989. Hierarchical cluster analysis as tool to manage variation in germplasm collections. *Theo. Appl. Genet.* **78** : 42-48.

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. Sept., 92.

Rana R.S. 1992a. Indian National Plant Genetic Resources Systems. *In*: R.S. Rana, Sapra R.L., Agrawal R.C. and Gambhir Rajeev (eds). Plant Genetic Resources-Documentation and information management. NBPGR, New Delhi, India. pp. 1-16.

Rana R.S. 1992b. Core collections and priorities of national programmes-Indian perspective. Paper presented in International Workshop on Core Collections of Plant Germplasm. Held at CENARGEN, Brasilia, Brazil. Sept., 92.

Rana R.S. 1993. Indian Programme on conservation an use of Plant Genetic Resources. *In*: R.S. Rana, Ram Nath, Khetrapal R.K., Gokte, Nandini and Bisht I.S. (eds). Plant Quarantine and Genetic Resources Management. N.B.P.G.R., N.Delhi, India. pp. 1-16.

Schoen D.J. and A.H.D. Brown. 1992. Maximizing genetic diversity in core collection of wild crop relatives. Paper presented in International Workshop on Core Collections of Plant Germplasm. Held at CENARGEN, Brasilia, Brazil. Sept., 92.

Spagnoletti Zeuli P.L. 1992. The 'core' collection and the plant breeder. Paper presented in International Workshop on Core Collections of Plant Germplasm. Held at CENARGEN, brasilia, Brazil. Sept., 92.

Spagnoletii Zeuli P.L. and C.O. Qualset. 1987. Geographical diversity for qualitative spike characters in a world collection of durum wheat. *Crop Sci.* **27** : 235-241

Strauss M.S., J.A. Pinoja and J.I. Cohen. 1988. Quantification of diversity in *ex situ* plant collections. *Diversity.* **16** : 30-32.

Tohme J., P. Jones, S. Beebe and 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 Collections of Plant Germplasm. Held at CENARGEN, Brasilia, Brazil. Sept., 92.

Vaughan D.A. 1990. Core collections of wild relatives of rice In Rice Germplasm: Collection, Preservation and Use. IRRI, Philippines. pp. 109-110.

Vaughan A. Duncan and M.T. Jackson. 1992. The relationship between the whole collection and its core. Paper presented in International Workshop on Core Collections of Plant Germplasm. Held at CENARGEN, Brasilia, Brazil. Sept., 92.

Von Bothmer R., G. Fischbeck, van Hintum Th., T. Hodgkin and H. Kniipffer. 1990. The barley core collections, Final report of BBC working group, Weihenstapham: **14**.

Yonezawa K. and T. Nomura. 1992. Sampling strategies for use in structured germplasm collections. Paper presented in International Workshop on Core Collections of Plant Germplasm. held at CENARGEN, brasilia, Brazil, Sept., 92.