Plant Genetic Resource Management

Editors
B S Dhillon
R K Tyagi
Arjun Lal
S Saxena

Narosa Publishing House
New Delhi • Chennai • Mumbai • Kolkata

Editors
B S Dhillon
R K Tyagi
Arjun Lal
S Saxena
National Bureau of Plant Genetic Resources
New Delhi 110 012, INDIA

NAROSA PUBLISHING HOUSE PVT. LTD.
22 Daryaganj, Delhi Medical Association Road, New Delhi 110 002
35-36 Greams Road, Thousand Lights, Chennai 600 006
306 Shiv Centre, DBC Sector – 17, K.U. Bazar P.O., Navi Mumbai 400 705
2F-2G Shivam Chambers, 53 Syed Amir Ali Avenue, Kolkata 700 019

www.narosa.com

ISBN 81-7319-579-X

Published by N K Mehra for Narosa Publishing House Pvt. Ltd.
22 Daryaganj, Delhi Medical Association Road, New Delhi 110 002
for Indian Society of Plant Genetic Resources, NBPGR, Pusa Campus, New Delhi 110 012
Printed at Aakriti Graphics, New Delhi

22

Core Collections for Efficient Management and Enhanced Utilization of Plant Genetic Resources

Hari D Upadhyaya
Genetic Resources and Enhancement Programme, International Crops Research Institute for the Semi-Arid Tropics, P.O. Patancheru-502 324, Andhra Pradesh

Plant diversity is vital to the development and welfare of human society. Plant genetic resources (PGR) contribute enormously towards achieving the global objectives of food security, poverty alleviation, environmental protection, and sustainable development. These are critical components of plant breeding efforts aimed towards increasing food security – both for short-term gains as well as for long-term increase in productivity. Much of the recent interest in plant genetic diversity developed from the experience in USA with southern corn leaf blight in 1969-1970. A new or previously undetected race of *Helminthosporium maydis* was first observed in Florida from where it moved rapidly northward, reaching the Corn Belt in 1970 (Tatum, 1971). The problem arose because of cytoplasmic uniformity of a large proportion of the maize grown at that time. This led to recognition of the need to diversify sources in plant breeding and broaden the genetic base of cultivars.

The establishment of *ex situ* germplasm collections has been the result of several decades of global efforts to conserve plant biodiversity. The need for large variability in scientific plant breeding and concern about potential loss of this variability, and non-availability of low cost tools to identify similarities and differences among accessions have led genebanks to hold large germplasm collections (Table 1, Fig. 1). This has resulted from the belief that the representativeness of collections can be achieved through large collection sizes (Frankel and Bennett, 1970). As collections rapidly grew beyond easily-manageable sizes, the task of quantifying diversity became daunting. Also, with increase in size of collections, the realization that they are little used by breeders also grew (Duvick, 1984). The large

Table 1. Total number of accessions in selected crops and the six major holders (CGIAR centres and regional genebanks) of *ex situ* germplasm collections

Crop	Total accessions	Major holders (% of accessions)					
Wheat	784 500	CIMMYT (13)	USA (7)	Russia (6)	India (6)	Germany (6)	Italy (5)
Barley	485 000	Canada (14)	USA (11)	UK (6)	ICARDA (5)	Brazil (5)	Russia (5)
Rice	420 500	IRRI (19)	China (13)	India (12)	USA (8)	Japan (5)	WARDA (4)
Maize	277 000	Mexico (12)	India (10)	USA (10)	Russia (7)	CIMMYT (5)	Colombia (4)
Phaseolus	268 500	CIAT (15)	USA (13)	Mexico (11)	Brazil (10)	Germany (3)	Russia (3)
Soybean	174 500	China (15)	USA (14)	AVRDC (10)	Brazil (5)	Ukraine (4)	Russia (3)
Sorghum	168 500	ICRISAT (21)	USA (20)	Russia (6)	Brazil (6)	Ethiopia (4)	Australia (4)
Brassica	109 000	India (16)	UK (10)	Germany (9)	USA (8)	China (6)	Rep. of Korea (3)
Cowpea	85 500	IITA (19)	Philippines (12)	USA (11)	AVRDC (7)	India (6)	Indonesia (5)
Groundnut	81 000	USA (27)	India (20)	ICRISAT (18)	China (8)	Argentina (6)	Zambia (2)
Tomato	78 000	USA (30)	AVRDC (9)	Philippines (6)	Russia (4)	Germany (4)	Colombia (3)
Chickpea	67 500	ICRISAT (26)	ICARDA (15)	Pakistan (9)	USA (9)	Iran (8)	Russia (4)
Cotton	49 000	India (34)	France (13)	Russia (12)	USA (6)	Pakistan (5)	China (3)
Sweet potato	32 000	CIP (21)	Japan (12)	USA (8)	Peru (6)	Philippines (5)	Several (4)
Potato	31 000	CIP (20)	Colombia (13	Germany (13)	USA (8)	Argentina (4)	Czech Rep. (4)
Faba bean	29 500	ICARDA (33)	Germany (18)	Italy (13)	Spain (6)	Russia (6)	France (6)
Cassava	28 000	CIAT (21)	Brazil (12)	IITA (8)	Uganda (6)	India (5)	Malawi (4)
Rubber	27 500	Malaysia (76)	Brazil (6)	Côte d'Ivoire (5)	Liberia (4)	Vietnam (4)	Indonesia (2)
Lentil	26 000	ICARDA (30)	USA (10)	Russia (8)	Iran (7)	Pakistan (4)	India (3)
Garlic/onion	25 500	Germany (18)	UK (10)	India (8)	Russia (5)	Hungary (6)	France (4)
Sugarbeet	24 000	Germany (25)	France (12)	Netherlands (9)	Yugoslavia (9)	Russia (7)	Japan (50
Oil palm	21 000	Zaire (83)	Malaysia (7)	Brazil (30)	Ecuador (1)	Colombia (1)	Indonesia (1)
Coffee	21 000	Côte d'Ivoire (35)	France (20)	Cameroon (7)	Costa Rica (7)	Ethiopia (6)	Colombia (5)
Sugarcane	19 000	Brazil (26)	India (22)	USA (11)	Domin. Rep.(11)	Cuba (8)	Venezuela (5)
Yam	11 500	IITA (25)	Côte d'Ivoire (20)	India (8)	Philippines (5)	Sri Lanka (4)	Solomon Is. (4)
Banana/plantain	10 500	INIBAP (10)	France (9)	Honduras (9)	Philippines (6)	Papua New Guinea (5)	Cameroon (5)
Tobacco	9 705	USA (19)	Poland (19)	India (15)	Cuba (7)	Greece (7)	Albania (5)
Cocoa beans	9 500	Brazil (24)	Trinidad/Tobago (22)	Venezuela (17)	France (7)	Costa Rica (6)	Colombia (5)
Taro	6 000	Malaysia (22)	Papua New Guinea (13)	India (11)	USA (8)	Indonesia (7)	Philippines (6)
Coconut	1 000	Sierra Leone (22)	Venezuela (20)	France (17)	India (13)	Colombia (11)	Philippines (9)

[Source: Food and Agriculture Organization of the United Nations (1996). Report on the State of the World's Plant Genetic Resources for Food and Agriculture]

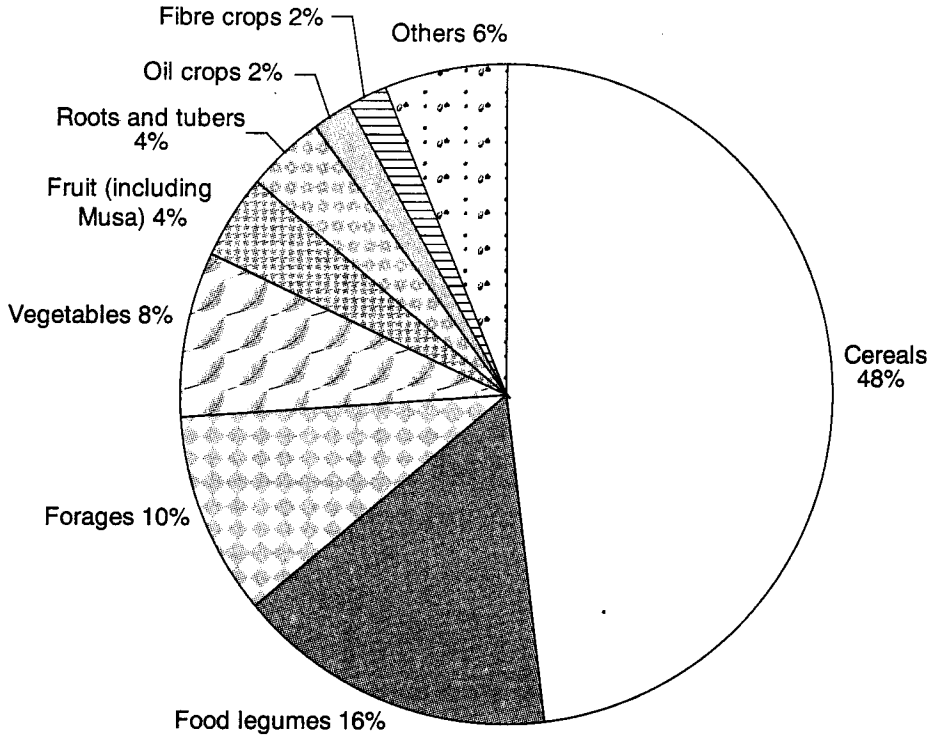

Fig. 1. Contribution of major crop groups to total ex situ collections. (Source: Food and Agriculture Organization of the United Nations 1996. Report on the State of the World's Plant Genetic Resources for Food and Agriculture)

variability within the genebank, rather than prompting its enhanced utilization, creates the "problem of plenty", that is not knowing what germplasm to begin with, in the genetic enhancement of crop breeding pool(s).

Core Collection Concept

In order to deal with the burgeoning number and size of germplasm collections, Otto Frankel suggested development of core collections (Frankel, 1984). Frankel's suggestion came at a biotechnology symposium where it was clear that molecular biology will have a significant impact on germplasm collection and utilization. Frankel was concerned that large germplasm collections might be stifled by their own apparent success. Thus, at a time when many were clamouring for more collection, he put forward the radical alternative that fewer, smaller collections were better. This was perhaps because growing accessions and the lag in corresponding data leads to collections not being used nearly as much as they should be (Marshall, 1989).

A core collection consists of a limited set of accessions derived from a germplasm collection, which would "represent with a minimum of repetitiveness, the genetic diversity of a crop species and its relatives". The accessions excluded from the core collection are retained as the *reserve*

collection. The core collection, due to its reduced size, can be studied extensively and the information derived can be used to guide more efficient utilization of the much larger reserve collection (Brown, 1989b; Tohme *et al.*, 1995). Issues concerned with low use of collections new needs of molecular biologists, new techniques of germplasm conservation, and the core subset as a structured and efficient sample of collection, formed the agenda of an International Board of Plant Genetic Resources (IBPGR) symposium at Montepellier, France (Brown *et al.*, 1989b). Given the importance of the issue, another symposium was held more recently by Crop Science Society of America and International Plant Genetic Resources Institute, to discuss various issues on core collection (Johnson and Hodgkin, 1999).

There are four basic elements of the concept of core collection. These are (i) the original collection is large in size in view of management or use, but has taxonomic integrity; (ii) the core subset from this large original collection has a small size; (iii) the core subset is a representative sample of the collection; and (iv) it is diverse. For samples of collection which lack one or more of these four elements, it is not advisable to apply the term core collection to them (Brown and Spillane, 1999). However, the term core collection does not require that every part of whole collection be equally represented. Indeed unequal numbers from different classes of genetic resources like cultivated *versus* wild, or per subspecies, or per botanical varieties or geographical areas are to be expected. Also, core collection does not require the absolute maximum possible diversity because this will lead to a bias towards the large number of distant wild relatives. The diversity should be as high as possible, but keeping in view that the core is a representative genetic resources collection of practical utility for breeders and scientists.

Selecting the Core Collection

In setting up the core collection there are typically six steps as follows:

(i) Defining the collection to be represented in the core collection. Assembling all the relevant data on these accessions.

(ii) Deciding the size of core collection.

(iii) Grouping of accessions into groups that reflect the major genetic and ecological categories within the entire collection.

(iv) Selecting the core entries – how many from each group and which ones.

(v) Representativeness of core collection.

(vi) Managing the core collection.

Defining Core Collection and Assembling Data on Accessions

It is important to know very clearly what collection is being made a core collection. The guiding principles in taking this decision are that core should serve as many users and uses as possible, and that it should be comprehensive. The passport data on taxonomy, geographical origin, ecological adaptation for each accession in the collection should be assembled. Characterization data on morphological descriptor traits, genetic markers and evaluation data (if available) should be assembled. In fact, all the available information should be used to develop the most representative core collection (Diwan *et al.*, 1995; Skinner *et al.*, 1999).

Size of Core Collection

The first decision to make in setting up a core collection is regarding its size. On the basis of sampling theory of selectively neutral alleles, Brown (1989b) argued that the number of accessions in the core collection should be about 10% of the total collection, with a maximum of 3,000 per species. This level of sampling is effective in retaining in the core collection about 70% of alleles of the entire collection. Large increase in size of core collection has increasingly marginal effects on levels of diversity retained (Brown, 1989b). For example, in a population of 10,000, about 70% alleles were predicted to be retained in a core comprising 10% of accessions, but doubling the number of retained accessions to 20% increased the predicted diversity retention by about 5% only. This clearly indicated that a core collection of 10% size of entire collection is as efficient as much larger collection in representing allelic diversity, provided that the selection of accession is carried out in a manner likely to capture most of the diversity.

While diversity in a germplasm collection is not randomly distributed (Brown, 1989b), it is very difficult to predict where the most diverse accessions occur. Breeding system of a species is the primary determinant of how different populations are different from one another. Self-pollinated species show more intense population differences, and more uneven distribution of genetic diversity among populations, than the out-crossing species.

Grouping of Accessions

The grouping of accessions into categories of genetic similarity or commonality among accessions and determining groups in the entire collection is one of the most crucial steps. The hierarchy of grouping begins with the groupings suggested by taxonomy (species, subspecies, races) followed by assigning accessions to major geographic groups (country, state), climate or agro-ecological regions. Judgement is required to produce groups of comparable

ecological diversity (Brown, 1989a). The accessions from large countries like USA or India can be divided into ecological regions and those from small, adjacent and similar countries can be grouped together. The clustering within the broad geographical group could be done to sort accessions into clusters using standard hierarchical clustering methods. Franco *et al.* (1997) have reviewed various clustering strategy. In cluster analysis a researcher is faced with the problem of how to use different traits (continuous, discrete, ordinal, multi-state, binomial). These traits are measured on different scales. The distance measurement used in cluster analysis depends on types of variables and scale of measurements. The first thing in such situations is to eliminate scale differences by standardizing each variable by means of either standard deviation or its range. Milligan and Cooper (1988) found it better to standardize by range. In simulation studies of different hierarchical cluster algorithms and distance measurements, single linkage clustering algorithm was found to be worst clustering strategy, initially proposed by Florek *et al.* (1951), for recovering the true structure of the groups (Milligan, 1980; Milligan *et al.*, 1983; Milligan and Cooper, 1985). The Ward (1963) method was found the best clustering strategy when the sizes of the groups are similar and UPGMA (Unweighted pair group with arithmetic mean method) (Sokal and Michener, 1958) was appropriate when the groups are of different sizes. In most of our study at International Crops Research Institute for the Semi-arid Tropics (ICRISAT) we have used the Ward (1963) method for clustering. This method optimizes an objective function because it minimizes the sum of squares within groups and maximizes the sum of squares between groups. The agglomerative procedure starts with n groups, that is one observation in one group (maximum between group sum of squares) and proceeds by merging observations in groups so that the between-groups sum of squares increases and within-groups sum of squares decreases. In certain cases the within-groups sum of squares will remain the same.

Selecting of Core Entries

The number of accessions in different groups are likely to vary greatly. The accessions allocated to a cluster will share genetic affinity. Once the decision on size of the core is taken, the decision on the number of accessions from each cluster will depend on the strategy to be adopted. The following three strategies have been suggested to decide on the number of accessions from each cluster.

Constant Strategy (C)

Equal number of accessions are sampled from each cluster into the core, irrespective of the total number of accessions in different groups. This strategy provides each cluster equal weightage.

Proportional Strategy (P)

A fixed proportion of each group selected to include into the core collection, so that the group is represented in proportion to its frequency in the entire collection.

Logarithmic Strategy (L)

The number of accessions included into core are in proportion to the logarithm of the number of accessions in that cluster.

Comparison of Strategies

The strategy C biases in favour of small clusters, whereas the P strategy biases in favour of large clusters. The main advantage of P strategy is that it is more efficient than simple random sampling. It includes more alleles and often has lower variance than simple random sampling (Cochran, 1977). The main advantage of C is that it avoids giving undue weightage to the very large clusters in which levels of redundancy is likely to be higher than in the smaller clusters. Strategy C is more efficient than P when the rare alleles occur in smaller clusters or when genetic variance is negatively related to group size. Strategy L is an intermediate strategy. Theoretically, L is the optimal strategy for fully differentiated loci, whereas P is optimal for undifferentiated loci (Brown, 1989a).

Representativeness of Core Collection

Of the four basic elements of the core collection concept, two elements, the original collection being large and the core collection being restrictive in size are easily met. The other two elements, core as representative of the entire collection, and its diversity, need to be assessed while setting up the core collection. Various parametric and non-parametric statistical methods can be used to compare the adequacy of core as a representative sample of the entire collection. In the case of quantitative traits, the means of the entire collection and the core collection can be compared using the Newman-Keuls procedure (Newman, 1939; Keuls, 1952), and variance homogeneity by Levene's test (Levene, 1960). Chi-square test can be used to test homogeneity of distribution of different classes in the entire and core collections. Similarly, the associations which may be under genetic control can be compared to know whether they were conserved in the core collection. The percentage of the significant differences for mean and variance of core and entire collection is calculated for the mean difference percentage (MD%) or the variance differences percentage (VD%) of traits (Hu *et al.*, 2000). The coincidence rate (CR%) for range and the variable rate (VR%) for coefficient of variation are also designed to evaluate the properties of core collection in terms of entire collection (Hu *et al.*, 2000).

$$CR\% = \frac{1}{m}\sum_{i=1}^{m}\frac{Rc}{\mathrm{Re}}\times 100$$

$$VR\% = \frac{1}{m}\sum_{i=1}^{m}\frac{CVc}{CVe}\times 100$$

Where Rc = range of the core collection, Re = range of the entire collection, CVc = coefficient of variation of the core collection and CVe = coefficient of variation of the entire collection and m = number of traits. The core collection is considered representative of the entire collection if no more than 20% traits have different means (significant at a = 0.05) between the core collection and the entire collection, and the CR% retained by the core collection is not less than 80% (Hu *et al.*, 2000).

The variances (high VD%) and the coefficient of variation (high CR%) in the core collection should be higher to represent higher genetic diversity. The diversity index (H') of Shannon and Weaver (1949) calculated independently for each trait in both entire and core collections, can be used as a measure of phenotypic diversity.

Managing the Core Collection

Managing the accessions included in a core collection is important so that it truly becomes a point of entry to the proper exploitation of genetic resources for crop improvement. The core accessions may be multiplied, made homogeneous, conserved separately, and evaluated further for important traits.

Core Selector

Core collections sometimes may not meet specific requests of users. For example, a user may want only one third the accessions of the core collection but nevertheless, a representative set, or the user may request material of a specific interest but with more accessions than included in the core. To meet these requests and to allow much more flexible use of the core collection concept, a system has been devised which allows selection of representative set meeting end-user needs. The number of entries is not fixed and the relative importance of specific parts can be adjusted according to needs. The system which allows this flexibility is called the core selector (van Hintum, 1999). It is a relatively simple system based on a formalization of the normal procedures to create a core collection. Depending upon the purpose, the size of the core subset can be set between 1% to 10% for the domain trait of importance. Domains can be divided into distinct groups depending upon the nature of traits for which a domain is defined. The

stepwise division of groups of qualitative nature can easily be accomplished as they are distinct by nature. Quantitative traits can be divided into groups based on the distribution and range value for the trait among accessions.

Progress in Developing Core Collections

Since the original proposal of Frankel (1984), core collections have been established for many species including common bean (*Phaseolus vulgaris* L.) by Tohme *et al.* (1995), barley (*Hordeum vulgare* L.) by Knupffer and van Hintum (1995), chickpea (*Cicer arietinum* L.) by Hannan *et al.* (1994), Upadhyaya *et al.* (2001a), annual and perennial *Medicago* species by Diwan *et al.* (1994), Basigalup *et al.* (1995), perennial *Glycine* by Brown *et al.* (1987), cassava (*Manihot esculenta* Crantz) by Cordeiro *et al.* (1995), coffee (*Coffea* spp.) by Dussert *et al.* (1997), lentil (*Lens culinaris* Medic.) by Erskine and Muehlbauer, (1991), okra (*Abelmoschus esculentus* L.) by (Mahajan *et al.* (1996), groundnut (*Arachis hypogaea* L.) by Holbrook *et al.* (1993); Upadhyaya *et al.* (2001b), potato (*Solanum tuberosum* L.) by Huaman *et al.* (2000), quinoa (*Chenopodium quinoa* Willd.) by Ortiz *et al.* (1998), *Saccharum spontaneum* by Tai and Miller (2001) and sweetpotato (*Ipomoea batatas* L.) by Huaman *et al.* (1999). The IPGRI conducted a global survey of genebanks to determine the extent of core collections. This survey indicated that at least 63 core collections covering 51 crop species have been formed (Brown and Spillane, 1999). The present status of core collections in different crops is presented in Table 2.

Uses of Core Collections

There are many important roles for core collections in the management and in utilization of genetic resources.

1. ***Management of Genetic Resources***

In genebank management, curators must decide on priorities among accessions. Core collections offer distinct advantages in addition to new accessions, conservation, characterization, germplasm distribution, and evaluation.

(i) ***Addition of new accessions***: The information on diversity preceding selection of a core collection helps to decide whether new accessions acquired by a genebank are worth adding to the collection or even to the core collection.

(ii) ***Conservation***: The core collection contains materials of the highest priority for conservation. It should, therefore, have first priority in monitoring viability, and in regeneration (depending upon viability and seed stock position). Core entries should be held in duplicate in other genebanks. Due to its representative nature, the core collection is suitable

for developing new methods of conservation such as ultra-dry seeds, *in vitro* or cryogenic storage.

(iii) ***Characterization***: The core collection, due to its reduced size, is the most suitable material to develop an adequate list of descriptors. A sufficient number of characters and states should be used to distinguish between core entries.

(iv) ***Evaluation***: The resources available to evaluate the germplasm are limited and dwindling steadily. Since evaluation of entire germplasm collections not possible, core collection due to its reduced size can provide a working collection that can be extensively examined for all economically important traits. This should result in identification of parents for different traits of economic importance for use by breeders. The core collection enables an efficient two-stage procedure in sampling the whole collection. Core entries can be evaluated first and then entries from reserve collection may be tested. The core provides a set of materials covering the range of variation in the whole collection. The core also assists the development of multivariate database to study the interrelationships between characters.

(v) ***Germplasm distribution***: Designation of the core collection helps to respond quickly since core entries can be multiplied, packaged and kept ready for dispatch. An important function of the core collection is to provide an opportunity to distribute the representative diversity of germplasm on a reduced scale and at a lower cost.

2. Utilization of Genetic Resources

The main aim of PGR is to use them in breeding programmes and to enhance productivity, quality and other desirable traits. The breeding of desirable traits from alien backgrounds into locally adapted stocks is a lengthy and expensive exercise. Core entries which form a reduced set have been suggested for use in testing for general combining ability (Frankel and Brown, 1984; Abel and Pollak, 1991; Spagnoletti Zeuli and Qualset, 1995). However, in situations where the number of core entries run into thousands, this seems a distant economic proposition. A more logical use of the core is in identifying sources for different traits and for their utilization in breeding work.

Now it is time to take stock of developments in the 17 years since Frankel (1984) proposed use of core collections. About 70 core collections in different crops have been developed but there are very few published reports indicating use of core collection in even identifying sources for use in crop improvement. Groundnut core collection developed by Holbrook

et al. (1993) has been evaluated for oil content (Holbrook *et al.*, 1998), fatty acid composition (Hammond *et al.* 1997), and resistance to tomato spotted wilt virus (Anderson *et al.*, 1996), early leaf spots (*Cercospora arachidicola* S Hori) and Cylindrocladium black rot [*Cylindrocladium crotalariae* (CA Loos) DK Bell & Sobers] (Isleib *et al.*, 1995), late leaf spots [(*Cercosporidium personata* (Berk. & MA Curtis)] Holbrook and Anderson, 1995), Rhizoctonia limb rot (*Rhizoctonia solani* (Kuhn) (Franke *et al.*, 1999), pre-harvest aflatoxin contamination (Holbrook *et al.*, 1997), groundnut root knot nematode [*Meloidogyne arenaria* (Neal) Chitwood] (Holbrook *et al.*, 2000). Similarly, the groundnut core (Upadhyaya *et al.*, 2001b) and chickpea core (Upadhyaya *et al.*, 2001a) have been evaluated, the former for protein and oil contents, shelling percentage, and 100-seed weight (Upadhyaya unpublished), early maturity (Upadhyaya *et al.*, 2001c,d), and tolerance to low temperature at germination (Upadhyaya *et al.*, 2001e) and the latter for early maturity, seed yield, seed size and other economic traits (Upadhyaya unpublished).

Mini Core Collections

The main purpose of a core collection is to improve the use of PGR in crop improvement programmes. In many crops the number of accessions contained in the genebank are several thousands (Table 1) and a core subset consisting of 10 per cent of total accessions would be an unwieldy proposition. Recognizing this, Upadhyaya and Oritz (2001) suggested a two-stage strategy to select mini core collection, consisting of only about 10 per cent of the entire collection held in the genebanks. The mini core collection subset still represents the diversity of the entire core collection. Of the two stages, the first stage involves developing a representative core subset (about 10

Table 2. Core collections developed in various crops

Crop group	Number of crops	Crops
Cereals	6	Amaranth, barley, maize (4), sorghum, wheat, quinoa
Pulses	6	Bean (2), chickpea (2), cowpea, mungbean (3), lentil, pea
Oilseeds	4	Safflower, sesame (2), soybean, groundnut (2)
Vegetables	5	Brassicas (2), capsicum, eggplant, lettuce (2), okra(2)
Fruits	12	Blueberries, citrus, currants, dates, grapes (2), hazelnut, persimmon, pear, pecan, plum, raspberries, strawberry
Forages	11	Alfalfa, annual medics, berseem clover, Kentucky bluegrass, red clover, ryegrass (3), shaftal clover, subclover, sweet clover, trefoil, white clover
Tuber crops	3	Cassava, sweet potato, potato (2)
Beverages, herbs, and spices	4	Coffee, garlic, mint, mountain mint
Industrial crops	3	Beet, hops, rubber

Values in parentheses indicate the number of core collections for the crop

per cent) from the entire collection using all the available information on origin, geographical distribution, and characterization and evaluation data of accessions. The second stage involves evaluation of the core subset for various morphological, agronomic and quality traits, and selecting a further subset of about 10% accessions from the core subset. At both stages standard clustering procedure was used to separate groups of similar accessions. A mini core subset consisting 211 chickpea accessions from 1956 core collection accessions (total collection 16,991 accessions), using data on 22 morphological and agronomic traits was selected. The mini core subset, due to its drastically reduced size will prove a point of entry to the proper exploitation of chickpea genetic resources (Upadhyaya and Ortiz, 2001).

Have Core Collections Served their Objectives?

Many crop improvement scientists ask whether core collections have served their purpose for which they were proposed, and if they can serve the curator in so many ways, why have they not been implemented widely? Brown and Spillane (1999) have presented the summary of IPGRI survey on various issues/objections related to core collections. These can be grouped into four categories.

1. *Vulnerability of Reserve Collection*

In view of the economy of operations the core may threaten the reserve collection as excess to needs. This assumes that the core is an entity by itself, ignoring the fundamental role of a core as a guide to the use of whole collection (Brown, 1995). In fact, the appraisal of the entire collection that occurs in selecting a core collection can provide evidence on the need to increase the size of the entire collection through targeted collection. Our experience at ICRISAT with chickpea illustrates this very well. Ethiopia which is a secondary centre of diversity for chickpea, is under-represented by 928 accessions (5.5%) in the entire collection (16,991) and by 120 (6.1%) accessions in the core collection (Upadhyaya *et al.*, 2001a).

2. *Bias Towards Representing Diversity*

This concern originates from the fact that the core collection is chosen to represent diversity of the entire collection, and is, thus, presumed not to represent usefulness. In fact, a diverse core is more likely to contain adequate sources of many characters than that selected by other strategies. The groundnut core collection developed at ICRISAT represented 100 per cent range variation for reaction to three important diseases, rust (*Puccinia arachidis* Speig.), early leaf spot (*Cercospora arachidicola* Hori) and rosette virus disease (Upadhyaya *et al.*, 2001b). For rust and rosette diseases, the percentage of

accessions with a score in the core collection was similar to the entire collection (Upadhyaya *et al.*, 2001b). Acquaintance of the breeder with the phenotypic diversity of crop is an advantage of a diverse core collection.

3. ***Inflexibility of Core Entries***

At what rate the changes should be made in the composition of a core collection needs to be addressed more seriously. A balanced and flexible approach to accommodate needs for change and for stability is required.

4. ***Lack of General Validity in Sampling Variation***

This concern originates from reliability of information on genetic diversity on which the core is based, and from the fact that the character of interest for breeders may be rare. If a core is formed on incomplete or misleading information on patterns of diversity in the whole collection, it is possible that it could leave out important types. This is a fault not of the concept of the core collection, but of inadequacy of information used. In fact, a core selected by simple random sampling showed good retention statistics (Brown, 1989b) and might be better than the one based on poor quality data (Brown and Spillane, 1999). The other concern is regarding the absence from the core collection of a really rare variant, which may be only one in an entire collection. Resistance to the grassy stunt virus in one population of *Oryza nivara* is an often-cited example of a rare variant. However, to identify such variant from entire collection will depend on luck or the capacity to cope with sampling the entire collection. Evaluation of a core collection provides information on whether the variant is rare. It may indicate sources of acceptable expression or suggest that the accessions from a hot spot of diversity be evaluated or that the rare character may be searched in a core subset of wild related species (Brown, 1995).

Conclusions

The core collection concept is about 20 years old and has been under discussion since its inception. It has been implemented in some cases. The theory behind core collections and methods of selection have come of age. The arguments for and against the core collection concept have been put forth. I believe as a practicing plant breeder that representative core collection is an important asset to the plant breeder, and can be effectively used in enhancing use of PGR in crop improvement. The mini core concept takes care of unwieldy size of core collections. The mini core can also be used to study genetic variability using molecular markers. The core or mini core collections assist in dealing with a deluge of local materials, and their development will render the whole collection more workable to the users, which in turn should help in conservation for long term and effective

utilization. The synergy between breeder and curator is the key to successful and efficient exploitation of PGR for enhancing productivity and quality of different crop species.

References

Abel BC and LM Pollak (1991) Rank comparisons of unadapted maize populations by testers and *per se* evaluation. *Crop Sci.* **31**: 650-656.

Anderson WF, CC Holbrook and AK Culbreath (1996) Screening peanut core collection for resistance to tomato spotted wilt virus. *Peanut Sci.* **23**: 57-61.

Basigalup DH, DK Barnes and RE Stucker (1995) Development of a core collection for perennial *Medicago* plant introductions. *Crop Sci.* **35**: 1163-1168.

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

Brown AHD (1989b) The case for core collections. *In:* Brown AHD, OH Frankel, DR Marshall and JT Williams (eds) *The Use of Plant Genetic Resources*. Cambridge University Press, Cambridge, UK, pp 136-155.

Brown AHD and C Spillane (1999) Implementing core collections–principles, procedures, progress, problems and promise. *In:* Johanson RC and T Hodgkin (eds) *Core Collections for Today and Tomorrow.* International Plant Genetic Resources Institute, Rome, Italy, pp 1-9.

Brown AHD, JP Grace and SS Speer (1987) Designation of a core collection of perennial *Glycine*. *Soybean Genet. Newsl.* **14**: 59-70.

Brown AHD (1995) The core collection at the crossroads. *In:* Hodgkin T, AHD Brown, TJL van Hintum, and EAV Morales (eds) *Core Collection of Plant Genetic Resources*. John Wiley and Sons, Chicester, UK, pp 3-19.

Cochran WG (1977) *Sampling Techniques.* 3rd Edition, John Wiley and Sons, New York, USA.

Cordeiro CMT, EAV Morales, P Ferreira, DMS Rocha, IRS Costa, ACC Valois and S Silva (1995) Towards a Brazilian core collection of cassava. *In:* Hodgkin T, AHD Brown, Th JL van Hintum and BAV Morales (eds) *Core Collections of Plant Genetic Resources*. International Plant Genetic Resources Institute (IPGRI), John Wiley and Sons, New York, USA. pp 155-168.

Diwan N, GR Bauchan and MS McIntosh (1994) A core collection for the United States annual *Medicago* germplasm collection. *Crop Sci.* **34**: 279-285.

Diwan N, MS McIntosh and GR Bauchan (1995) Methods of developing a core collection of annual *Medicago* species. *Thoer. Appl. Genet.* **90**: 755-761.

Dussert S, N Chabrillange, F Anthony, F Engelmann, C Recalt and S Hamon (1997) Variability in storage response within a coffee (*Coffea* spp.) core collection under slow growth conditions. *Plant Cell Rep.* **16**: 344-348.

Duvick D (1984) Genetic diversity in major farm crops on the farm and in reserve. *Eco. Bot.* **38**: 161-178.

Erskine W and FJ Muehlbauer (1991) Allozyme and morphological variability, outcrossing rate and core collection formation in lentil germplasm. *Theor. Appl. Genet.* **83**: 119-125.

Florek K, J Lukaszewicz, J Perkal, H Steinhaus, and S Zubrzycki (1951) Sur la liason et la division des points d'un ensemble fini. *Colloq. Math.* **2**: 282-285.

Franco J, J Crossa, J Villasenor, S Taba, and SA Eberhart (1997) Classifying Mexican maize accession using hierarchical and density search methods. *Crop Sci.* **37**: 972-980.

Franke MD, TB Brenneman and CC Holbrook (1999) Identification of resistance to *Rhizoctonia* limb rot in a core collection of peanut germplasm. *Plant Dis.* **83**: 944-948.

Frankel OH (1984) Genetic perspectives of germplasm conservation. *In:* Arber WK, K Llimensee, WJ Peacock, and P Starlinger (eds) *Genetic Manipulation: Impact on Man and Society.* Cambridge University Press, Cambridge, UK, pp 161-170.

Frankel O and E Bennett (eds) (1970). *Genetic Resources in Plants: Their Exploration and Conservation.* Blackwell, Oxford, UK.

Frankel O and AHD Brown (1984) Current plant genetic resources- a critical appraisal. *In:* Chopra VL, BC Joshi, RP Sharma and HC Bansal (eds) *Genetics: New Frontiers* (Vol IV), Oxford and IBH Publishing, New Delhi, pp 3-13.

Hammond EG, D Duvick, T Wang, H Dodo and RN Pitman (1997) Survey of fatty acid composition of peanut (*Arachis hypogaea*) germplasm and characterization of their epoxy and eicosenoic acids. *J. Am. Oil Chem. Soc.* **74**: 1235-1239.

Hannan RM, WJ Kaiser and FJ Muehlbauer (1994) Development and utilization of the USDA chickpea germplasm core collection. *In: Agronomy Abstracts* 1994. ASA, Madison, Wisconsin, USA.

Holbrook CC and WF Anderson (1995) Evaluation of a core collection to identify resistance to late leaf spot in peanut. *Crop Sci.* **35**: 1700-1702.

Holbrook CC, WF Anderson and RN Pitman (1993) Selection of a core collection from the US germplasm collection of peanut. *Crop Sci.* **33**: 859-861.

Holbrook CC, MG Stephenson and AW Johnson (2000) Level and geographical distribution of resistance to *Meloidogyne arenaria* in the US peanut germplasm collection. *Crop Sci.* **40**: 1168-1171.

Holbrook CC, J Bruniard, KM Moore and DA Knauft (1998) Evaluation of the peanut core collection for oil content. *In:* Agronomy Abstracts. ASA, Madison, Wisconsin, USA. 159 p.

Holbrook CC, DW Wilson and ME Matheron (1997) Results from screening the peanut core collection for resistance to preharvest aflatoxin contamination. *In:* Robens J and J Dorner (eds) *Proc. Aflatoxin Elimination Workshop,* Memphis, TN. 27-28 Oct. 1997. USDA-ARS, Beltsville, MD, USA.

Hu J, J Zhu and HM Xu (2000) Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Genet.* **101**: 264-268.

Huaman Z, C Aguilar and R Ortiz (1999) Selecting a Peruvian sweet potato core collection on the basis of morphological, eco-geographical, and disease and pest reaction data. *Theor. Appl. Genet.* **98**: 840-844.

Huaman Z, R Ortiz, D Zhang and F Rodriguez (2000) Isozyme analysis of entire and core collections of *Solanum tuberosum* subsp. *andigena* potato cultivars. *Crop Sci.* **40**: 273-276.

Isleib TG, MK Beute, PW Rice and JE Hollowell (1995) Screening the core collection of peanut for resistance to *Cylindrocladium* black rot and early leaf spot. *Proc. Am. Peanut Res. Edu. Soc.* **27**:25.

Johnson RC and T Hodgkin (eds) (1999) *Core collections for Today and Tomorrow.* International Plant Genetic Resources Institute, Rome, Italy, 81p.

Keuls M (1952) The use of the "Studentized range" in connection with an analysis of variance. *Euphytica* **1**: 112-122.

Knupffer H and T JL van Hintum (1995) The barley core collection: an international effort. *In:* Hodgkin T, AHD Brown, TJ Lvan Hintum and BAV Morales (ed) *Core Collections of Plant Genetic Resources*. International Plant Genetic Resources Institute (IPGRI), John Wiley and Sons, New York, USA, pp 171-178.

Levene H (1960) Robust tests for equality of variances. *In:* Olkin I (ed) *Contributions to Probability and Statistics: Essays in Honour of Harold Hotelling*. Stanford University Press, Stanford, USA, pp 278-292.

Mahajan RK, IS Bisht, RC Agrawal and RS Rana (1996) Studies on south Asian okra collection: a methodology for establishing a representative core set using characterization data. *Genet. Resour. Crop Evol.* **43**: 244-255.

Marshall DR (1989) Limitations to the use of plant genetic resources. *In:* Brown AHD, OH Frankel, DR Marshall, and JT Williams (eds) *The Use of Plant Genetic Resources.* Cambridge Univ. Press, Cambridge, UK, pp 105-120.

Milligan GW (1980) An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika* **45**: 325-342.

Milligan GW, SC Soon and SL Sokol (1983) The effects of cluster size, dimensionality, and the number of clusters on recovery of true cluster structure. *IEEE Trans. Pattern Anal. Machine Intelligence* **5**: 40-47.

Milligan GW and M Cooper (1985) An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**: 159-179.

Milligan GW and M Cooper (1988) A study of standardization of variables in cluster analysis. *J. Classif.* **5**: 181-204.

Newman D (1939) The distribution of range in samples from a normal population expressed in terms of an independent estimate of standard deviation. *Biometrika* **31**: 20-30.

Ortiz R, EN Ruiz-Tapia and A Mujica-Sanchez (1998) Sampling strategy for a core collection of Peruvian quinoa germplasm. *Theor. Appl. Genet.* **96**: 475-483.

Shannon CE and W Weaver (1949) The mathematical theory of communication. University of Illinois Press, Urbana, IL, USA.

Sokal R and C Michener (1958) A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.* **38**: 1409-1438.

Skinner DZ, GR Bauchan, G Auricht and S Hughes (1999) A method for the efficient management and utilization of large germplasm collections. *Crop Sci.* **39**:1237-1242.

Spagnoletti Zeuli P and CO Qualset (1995) The durum wheat core collection and the plant breeder. *In:* Hodgkin T, AHD Brown, TJL Van Hintum and EAV Morales (eds) *Core Collection of Plant Genetic Resources.* John Wiley and Sons, Chicester, pp 213-228.

Tai PYP and JD Miller (2001) A core collection for *Saccharum spontaneum* L. from the world collection of sugarcane. *Crop Sci.* **41**: 879-885.

Tatum LA (1971) The southern corn leaf blight epidemic. *Science* **171**: 1113-1116.

Tohme J, P Jones, S Beebe and M Iwanaga (1995) The combined use of agroecological and characterization data to establish the CIAT *Phaseolus vulgaris* core collection. *In:* Hodgkin T, AHD Brown, T JL van Hintum and EAV Morales (eds) *Core collections of Plant Genetic Resources.* International Plant Genetic Resources Institute (IPGRI). John Wiley and Sons, New York, USA, pp 95-108.

Upadhyaya HD and R Ortiz (2001) A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theor. Appl. Genet.* **102**: 1292-1298.

Upadhyaya HD, PJ Bramel and S Singh (2001a) Development of a chickpea core subset using geographic distribution and quantitative traits. *Crop Sci.* **41**: 206-210.

Upadhyaya HD, R Ortiz, PJ Bramel and S Singh (2001b) Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. *Genetic Res. Crop Evo.* (in press).

Upadhyaya HD, ME Ferguson and PJ Bramel (2001c) Status of the *Arachis* germplasm collection at ICRISAT. *Peanut Sci.* (in press).

Upadhyaya HD, SN Nigam and S Singh (2001d) Identification of new sources of early-maturity in groundnut core collection. *Paper presented in the Diamond Jubilee Symposium on Hundred years of Post-Mendelian Genetics – Retrospect and Prospects.* 6-9 Nov 2001. Indian Agricultural Research Institute, New Delhi, India.

Upadhyaya HD, SN Nigam and S Singh (2001e) Evaluation of groundnut core collections to identify sources of tolerance to low temperature at germination. *Indian J. Plant Genet. Resour.* **14**: 165-167.

van Hintum TJL (1999) The core selector, a system to generate representative selections of germplasm accessions. *Plant Genet. Resour. Newsl.* **118**: 64-67.

Ward J (1963) Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **38**: 236-244.