Indian J. Genet., 77(3): 329-340 (2017)
DOI: 10.5958/0975-6906.2017.00045.1

Review Paper

A critique of selection methodology of core collections and their use in crop improvement

Anurudh K. Singh* and S. N. Nigam[1]

2924, Sector 23, Gurugram 122 017, Haryana; [1]Plot No. 125, Road No. 74, Jubilee Hills, Hyderabad 500 033, Telangana

(Received: April 2017; Revised: July 2017; Accepted: August 2017)

Abstract

Growth in large germplasm collections in most important crops has led to the development of the concept of core collection or a smaller set of germplasm representing most spectrum of variability of total collections to facilitate their easy management, evaluation and use in crop improvement. In the last two decades, significant efforts have been made in this direction, following nearly identical sampling or selection strategies in most crops. The present article has tried to analyze critically the selection methodologies followed to assess how far core collections in different crops have succeeded in meeting the objectives, particularly those of crop genetic improvement. An attempt has also been made to address the possible ways for improvement in the selection strategy with additional steps to overcome the lacunas. Groundnut cores have been critically analyzed as a case study. This has revealed that cores have been able to capture only around 70% of variability, which has limited the value of core with non-capture of rare alleles. To overcome present lacunas, a modified stratification method has been suggested based on biogeographical distribution, and integration of gene pools (sets) of various desirable traits in the selection of accessions to formulate the core set. Greater emphasis needs to be given to genomics and characterization of accessions, particularly in relation to desired traits, using molecular markers associated with these traits to avoid masking effect caused by environment factors.

Key words: Gene pool, collection, core collection, mini-core collection, plant genetic resources, use in crop breeding

Introduction

Recognizing the importance of plant genetic resources (PGR) in crop improvement, Vavilov (1926), best known for identification of centers of origin of cultivated plant species, initiated a drive for collection of genetic diversity in crop species, existing in the form of diverse cultigens and their wild relatives. This initiative got further fillip with the establishment of Consultative Group on International Agricultural Research (CGIAR) in 1972, and International Board of Plant Genetic Resources (IBPGR) under it in 1974. The establishment of commodity based International Agricultural Research Centers (IARCs) under this system revolutionized these efforts with participation of international community in collection, evaluation, and conservation of the available genetic diversity of the mandated crops of the centers. These participatory efforts resulted in assembly of large collections in most of the food crops through joint collection and contribution of diverse accessions maintained by the National Agricultural Research Systems (NARSs). However, soon this enthusiasm created a problem of plenty of PGR in most crops, towards management of large numbers of accessions for curators and in selection of appropriate genetic resources, for efficient use to meet their specific requirements for breeders.

Frankel (1984) proposed the concept of core collection (CC), which was further developed by Frankel and Brown (1984) to create a manageable sample of germplasm, representing total spectrum of variability, particularly to facilitate genetic resources management (Brown 1988, 1989a & b). A CC was

*Corresponding author's e-mail: anurudhksingh@gmail.com

Published by the Indian Society of Genetics & Plant Breeding, F2, First Floor, NASC Complex, PB#11312, IARI, New Delhi 110 012
Online management by indianjournals.com; http://epubs.icar.org.in/journal/index.php/IJGPB

supposed to consist of a limited number of accessions derived from the total collection (about 10% of the full collection), representing the genetic diversity of a species and its relatives, within minimum number of accessions avoiding repetitiveness. Owing to the reduced size in CC, it was suggested that it can be studied extensively and the derived information can be used to guide more efficient utilization of the much larger 'reserve (base/total) collection', not included in the CC. With this understanding, there has been a spree for developing CCs in most major crops, particularly in those involving curators facing management problem. Consequently, in the last decade, CCs have been established in most important field and horticultural crops, using nearly identical stratification/selection strategies based on biometric principles, particularly in crops that are mandated to IARCs. However, the size of these CCs still appeared large to allow their extensive study and consequently their efficient use in crop improvement. Therefore, in response to this, Upadhyaya and Ortiz (2001) postulated the concept of the "mini core collection (MCC)", where the number of accession to represent the total spectrum of variability was reduced to 1% of the total collection without any formal genetic basis.

There is no documented information on the role of these CCs in facilitating greater usage and/or providing new and diverse sources used in breeding programs to obtain desired genetic enhancement/gains for all specific (geographic/agroecological regions and markets) requirements. The percent use of genetic resources has remained almost the same before and after CC development. For example, Duvick (1984) reported use of only around 1.5 percent of the total collections in the USA (across crops), and here too, the elite germplasm pool evolved through genetic enhancement efforts, possessed greater diversity, and provided more useful genetic resources than was usually supposed to be. Similarly, in 2012, from International Rice Research Institute (IRRI) only around 10,000 rice accessions, including duplicates and repeats, were distributed from the total holding of 1,27,000, which amounts to only 1.3 percent.

The CC development of their mandated crops at the CGIAR Centers started from 2003 onwards. Figure 1 clearly demonstrates that the number of germplasm samples distributed has nearly remained static and there has been no increase in number of samples distributed from 2003 onwards. In fact, it has declined and stabilized between 20,000 to 27,500 on an annual basis from the initial high of 32,000 in 1986 to 47,000 in 1988 (Noriega et al. 2013). Partly, this could be because most of the NARSs had already obtained the PGRs meeting their requirements in the initial years of conservation or slump was caused due to development of core. However, there is no data to demonstrate increased sharing and/or use of specific genotype(s) of the CC after development of CC in any crop. There may be an increase in sharing of the members of CC in most crops, particularly from ICRISAT than before, but to our knowledge most core accessions were predominantly distributed/shared for multilocation evaluation of core accessions rather than use in breeding programs. Moreover, from the time of

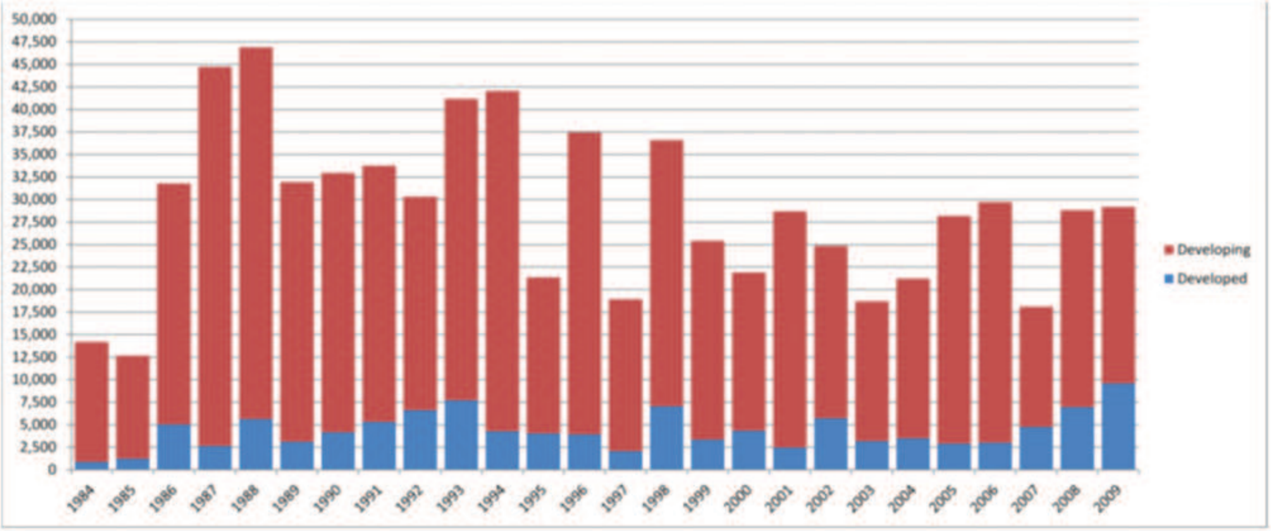

Fig. 1. Number of samples distributed by CGIAR gene banks to developed and developing countries from 1984 to 2009. Source: SINGER, Noriega et al. (2013)

over a decade of development of core sample sets, it is desirable to evaluate the contribution of CC in facilitating enhanced use of genetic diversity in crop breeding, identify gaps for the desired genetic variability, if any, and to search reasons/answers for questions listed below.

1. Does CC encompass total spectrum of variability, particularly the allelic variability of desired traits, to meet the demand of all agroecological regions and production systems?

2. Have the sources of desirable traits, particularly the most sought-after features such as resistance to biotic and abiotic stresses and adaptive traits such as earliness etc., changed after the advent of CC? Or the breeders are still relying on the same conventional (elite) sources?

3. Has there been an increase in the use of greater spectrum of variability in the form of more diverse accessions for a trait as per the need of agro-ecological situations and market?

4. What is the level of correlation between the mechanisms associated with phenotypic expression of desired features and quantitative morphological and agronomic traits used in selection of CC?

5. From the breeding point of view, is there a need for regular evaluation of established CCs to identify gaps in variability of alleles per locus covering all geographical/agro-ecological regions?

6. Lastly, but most importantly, the morphological characterization and evaluation data from single location with multivariate analysis of variability on total accessions have been the predominant basis of grouping, from where representative accessions for core are randomly selected. However, recognizing the influence of environmental and other conditions on DNA sequences/gene activity- gene activation and silencing through various processes (methylation), how reliable, or stable would be this data-set for selection of accessions to be used globally is a question? For example, whether some accessions with drought resistance will express all its associated features, when grown and characterized over generations under good experimental agronomic conditions.

7. Does this situation demand total genomic sequencing covering all chromosomes and greater dependence on molecular markers polymorphism (not affected by environment), particularly associated with loci of desired genes to identify unique accessions for stratification of gene pools of the desired traits, and integrate them in selection of CCs to capture total spectrum of variability for traits of significance in breeding programs?

Otherwise also, needs, preferences and procedures of germplasm users keep changing over time and it should be useful to evaluate the CCs from breeding perspective to modify/improve, and tailor them toward the needs of different groups.

Selection methodology used for development of core and mini-core collections

In most cases, the creation of a CC often starts with stratification of entire collection of accessions into small homogeneous groups. A group can be a collection of accessions with similar genotypic and phenotypic characteristics, taxonomically belonging to the same hierarchical entity (subspecies, botanical variety, etc.) and/or have common country/region of origin. Further sub-division within each group is based on strongly inherited genetic polymorphism in relation to both qualitative and predominantly quantitative morphological characters such as growth habit, branching pattern, etc. In this regard, abundant discriminating data are collected and accessions are classified into defined groups using multivariate analysis and clustering methods (Ward 1963). The clustering within the broad morphological or geographical group could be done to sort accessions out into clusters using standard hierarchical clustering methods. From each cluster 10% accessions are selected at random (1 accession in case the cluster contains <10 accessions) and pooled together to form a CC. The 10% is an arbitrary figure employed to all the crops, both autogmous and allogamous. The basic principle in arriving at the 10% criterion is explained as “a starting point” to retain a meaningful proportion of the total diversity, so that in theory, core may retain more than 70% of the alleles found in the original collection (Brown 1989 a and b), but short of total spectrum of variability.

According to Brown (1989b), a good CC should have no redundancy, should represent the whole collection with regards to species, subspecies and geographical regions and should be small enough to be easily managed while retaining the variability of entire collection. The best CC should contain relatively more material from the primary gene pool compared to the secondary or tertiary genepool, irrespective of the amount of diversity within the primary gene pool since there will be a strong preference by the breeders for material in an adapted genetic background. Recognizing the importance of adaptability gene(s), it would be relatively easy to use accessions with an adapted genetic background in a breeding program. However, it is important to note that irrespective of the type of CC, appropriate optimization and evaluation criteria should be used in creating and evaluating these selections. A CC should represent the whole spectrum of variability and diversity including the extremes, which is validated by comparing means, variances, the Shannon-Weaver diversity index (H') (Shannon and Weaver 1949) and the frequency distribution of traits between the entire collection and the core collections. Further, genetically controlled phenotypic correlations in the entire collection and the core collection are estimated and compared with each other to know if they are conserved in the latter.

The major concern and gaps

Most evaluation efforts using various statistical analyses confirm the abundance of genetic variation of the accessions in CC and MCC (Song et al. 2010; Upadhyaya et al. 2010). However, the percent recovery of variation over full collection varies (ca 75%) even at molecular level (Li et al. 2011). Moreover, being predominantly based on morphological traits, they have unequal distribution of different accessions representing the total spectrum of variability of desirable traits to satisfy the needs of geographic and agroecological regions. For these reasons, accessions with desirable agronomic or nutritional traits in a CC and MCC are limited and often rare alleles are missing for specific traits, which limit their value. This is owing to the effort of encompassing diversity for as many qualitative and quantitative traits as possible in a limited, or as minimum as possible accessions in a CC and exclusively based on morphological traits used for clustering of accessions (may or may not be associated with desirable traits). For this reason, the accessions identified in these CCs with specific traits at the most work as geographical indicators for directional identification of more elite accessions from CCs or full collections than actual accessions with desired variability meeting the requirements of wider global situations. For example, for studying the availability of resistance to diseases and other stress factors and other traits, accessions in MCC can first be used to characterize the geographical distribution of different traits, and then, many accessions may be evaluated from the same geographical area. This strategy has been used to identify elite accessions for salt tolerance and soybean cyst nematode (SCN) resistance (Yuan et al. 2012; Liu et al. 2011) in soybean.

Most CCs do have common widespread alleles and successful groups of common localized alleles. However, often recovery of rare alleles localized in diverse ecological niches go missing due to impracticality of conserving everything in CCs and MCCs (in 10 or 1% representation). This also happens because of juxtaposed objectives in development of CCs (management versus use) by curators, which are "conserving as much variation (phenotypic or genotypic) as possible in as few as possible accessions" against "optimizing the chance of finding a new/greater allele" to improve their use in breeding programs.

Further, for evaluation of utility of CCs, it will be advisable that whenever possible or appropriate, the evaluation of CCs should be based on data that have not been used for the selection of the accessions for the CC.

Groundnut core and mini-core: A case study

The groundnut CCs were developed with stratification of entire germplasm accessions by botanical varieties, country of origin and morphological characters. At ICRISAT, the entire collection of 14,454 accessions was stratified first by botanical variety within subspecies followed by their country of origin and then by 14 morphological traits(stem color, stem hair, branching pattern, leaf color, leaf shape, leaf hairs, flower color streak color, peg color, pod beak, pod constriction, pod reticulation, seeds per pod and seed color pattern), many of them having no or limited significance in breeding programs, thus limiting their utility in breeding programs. It was followed by principal component/ multivariate analysis and clustering, using data on predominantly quantitative morphological traits, which resulted in 75 groups. From these groups 10% of accessions were randomly selected to constitute a CC. Using similar or slightly modified methodology, CCs were developed on US germplasm (Holbrook et al. 1993); on global germplasm at ICRISAT (Upadhyaya et al. 2003) and in China (Jiang et al. 2007). MCCs by selecting 1% accessions of CCs were also developed at ICRISAT (Upadhyaya et al. 2002) and in China (Jiang et al. 2013). Comparison of CCs developed in different parts of the world showed different traits contributing to variability in different sets of collections, associated with the dominant subspecies and botanical varieties represented in a collection and the varied selection pressure (Jiang et al. 2008), indicating a need of a universal core collection for groundnut improvement, meeting everyone's needs.

Analyses of representation of diversity in core collections

Representation of taxonomic diversity

Based on analyses of data on 14 morphological characters, the CC, developed at ICRISAT, consists of 34.3% to 33.6% accessions belonging to ssp. *fastigiata* var. *vulgaris*, 17.5% to 17.9% to ssp. *fastigiata* var. *fastigiata*, 1.6% to ssp. *fastigiata* var. *peruviana*, 0.4% to 0.3% to ssp. *fastigiata* var. *aequatoriana*, and 46.0% or 46.4% [27.6 (bunch) + 18.8 (runner)] to ssp. *hypogaea* var. *hypogaea*, and 0.2% to ssp. *hypogaea* var. *hirsuta*. Whereas, the MCC created at ICRISAT consists of 32.6% accessions belonging to ssp. *fastigiata* var. *vulgaris*, 19% to ssp. *fastigiata* var. *fastigiata*, 1.1% to ssp. *fastigiata* var. *peruviana*, 0.5% to ssp. *fastigiata* var. *aequatoriana*, and 46.2% to ssp. *hypogaea* var. *hypogaea* (18.5% runner and 27.7% bunch types), and 1% to ssp. *hypogaea* var. *hirsuta* (Table 1). These figures suggest that the representation of subspecies and botanical varieties in the CC and the MCC nearly corresponds to the percent contribution of these taxonomic entities to the entire (full) collection (Upadhyaya et al. 2002; Upadhyaya et al. 2003). However, numerical representation does not necessarily mean the preservation of the same level of variability in CC and MCC as that occurring in entire collection. Botanical variety *hirsuta* of subsp. *hypogaea* and *peruviana* and *aequatoriana* of subsp. *fastigiata* are extremely under-represented in the entire collection of germplasm, both because of comparatively limited distribution and lesser collections efforts (Singh and Nigam 2016). Accordingly, they are also under-represented in the CC (variety *hirsuta* (0.2%), *peruviana* (1.6%) and *aequatoriana* (0.3 to 0,4%) and the MCC (variety *hirsuta* (0.5%), *peruviana* (1.1%) and *aequatoriana* (0.5%). This is resulting in (non-incorporation) non-availability of some useful variability for crop improvement. For example, var. *hirsuta* may possess significant variability for resistance against groundnut pests, because of its variable pubescence nature (leaf), a key feature of the botanical variety. Therefore, there is taxonomic genetic variability gap and need for assembly of more/new accessions of such varieties with appropriate initiative, either through collection missions or exchange of germplasm, to enhance their (genes) representation in the entire collection and hereby in CC and MCC produced by ICRISAT. Alternately, the selection methodology should be suitably modified to deliberately include some of these accessions in the CC and the MCC.

Representation of geographical diversity

Groundnut is a tropical crop, which originated in tropical and sub-tropical regions of South America. It had

Table 1. Representation of subspecies and botanical varieties in total, core and mini-core of groundnut

Subspecies/botanical variety	Total collection[1]		Core collection[2]		Mini-core collection[2]	
	Total acc.	% share	Total acc.	% share	Total acc.	% share
Arachis hypogaea hypogaea (bunch)	6766	45.0%	784[1] revised 470[2]	46.0% 27.6%[2]	51	27.7%
Arachis hypogaea hypogaea (runner)			revised 320[2]	18.8%[2]	34	18.5%
Arachis hypogaea hirsuta	20	0.14%	4[1]	0.2%	1	0.5%
Arachis hypogaea fastigiata fastigiata	2302	16.1%	299[1] revised 305[2]	17.5% 17.9%[2]	35	19.0%
Arachis hypogaea fastigiata vulgaris	5102	35.7%	584[1] revised 573[2]	34.3% 33.6%[2]	60	32.6%
Arachis hypogaea fastigiata peruviana	249	1.72%	27[1]	1.6%	2	1.1%
Arachis hypogaea fastigiata aequatoriana	15	0.10%	6[1] revised 5[2]	0.4% 0.3%[2]	1	0.5%

Source: [1]Upadhyaya et al. (2001); [2]Revised lists of core and mini-core personal comm., ICRISAT subsequently spread to Africa, which has been considered its tertiary center of diversity as it offered a wide range of agro-climates for its cultivation and then tropical (India, etc.) and sub-tropical (China, etc.) Asia, and other countries (Central Asia, Southern Europe, USA, etc.). Genetic analysis done by ICRISAT and EMBRAPA (//ICRISAT// Groundnut Crop: www. icrisat.org/crop-groundnut-genebank.htm) revealed that accessions from the Americas had the highest number of unique alleles (109), while those from Africa and Asia had only six and nine, respectively. The greater allelic variability in accessions from the Americas is naturally expected, as the South America is the center of origin and diversity of cultivated *A. hypogaea*. However, the CC produced at ICRISAT consists of 317 accessions belonging to South and Central American countries (including island countries), 473 accessions of African countries, 429 accessions of tropical Asian countries and 77 accessions of temperate Asian countries (predominantly China). North America, where the crop has been introduced comparatively recently (ca. 200 years back), has been able to contribute 199 accessions (Table 2). This is perhaps because of strong genetic enhancement programs in the USA, using elite germplasm in breeding programs resulting in generating significantly higher level of variability for useful traits, a fact also observed by Duvick (1984) across crops, while assessing conservation and use of PGR in USA. This situation clearly reflects selection of proportionally lesser number of accessions from the primary biogeographical regions that are the center of origin/diversity of the crop, because of their comparatively lesser representation in total collections. Thus, indicating capture of lesser basic genetic and allelic variability in the CCs, particularly for the essential features, while capturing greater amount of variability of agronomic features, which have evolved in biogeographical areas, which are major groundnut growing areas, where crop was introduced. Many of the accessions from USA included in the CC might have been just introductions from other regions of diversity. In addition, the CC consists of 117 accessions (6.9%) of unknown origin raising the concern about authenticity of accessions selected, particularly from adaptability gene(s) point of view, which are of significance in specific breeding programs. Similarly, MCC produced by ICRISAT consists of 35 (15%) and 25 (14%) accessions of South American and North American origin, respectively, while 40 (24%) of African and 49 (26%) of Asian origin, respectively. Further corroborating that the CC and the MCC created at ICRISAT do not represent (captured) the actual genic and allelic variability (richness) pattern available in groundnut germplasm from different biogeographical regions, particularly from natural center of diversity. This is perhaps, because of proportional dominance of collections (accessions) from Africa (tertiary center of diversity) and Asia (particularly India), the present day major areas of cultivation, in the global collections assembled at ICRISAT. Principal Component analysis using 38 traits on world collections at ICRISAT broadly clustered the accessions into three clusters, representing South American, North American (USA), and African collections with a heterogeneous distribution of traits, which means differential selection pressure in different biogeographical regions contributing to genetic variability, particularly for agronomic features. The South American collections showed maximum variability for essential morphological features (//ICRISAT// Groundnut Crop: www. icrisat.org/crop-groundnut-genebank.htm). Therefore, for selecting accessions for CC representing genetic diversity of all countries of groundnut cultivation, the first stratification of full collection should be done based on biogeographical regions of groundnut distribution/cultivation, instead of country of origin, followed by stratification based on taxonomic diversity.

Table 2. Representation of geographical region of origin in core and mini-core of groundnut

Geographical region of origin	Core collection		Mini-core collection	
	Total acc.	Percent share	Total acc.	Percent share
South and Central America	317	18.6%	35	19.0%
North America (USA)	199	11.7%	25	13.6%
Africa	473	27.7%	40	21.8%
Asia tropical (India etc.)	429	25.1%	49	26.6%
Asia sub temperate (China)	77	4.5%	11	6.0%
Asia Central & Mediterranean	26	1.5%	1	0.5%
Europe	19	1,1%	3	1.6%
Australia	22	1.3%	0	0.0%
Unknown	117	6.9%	17	9.2%

acc. = Accession

Representation of morphological diversity (qualitative and quantitative traits used)

An assessment of genetic diversity on world collection at ICRISAT for 16 morphological and 10 agronomic traits showed vast diversity in size and shape of pods and seeds. Principal Component analysis using 38 traits and clustering on first seven PC scores produced three clusters; first cluster consisting of accessions from North America, middle East, and East Asia, second cluster South America, and the third cluster West Africa, Europe, Central Africa, South Asia, Oceania, Southern Africa, Eastern Africa, Southeast Asia, Central America, and Caribbean. This means that agronomic traits differed significantly among regions as per the selection pressure, and the variances for all the traits among regions were heterogeneous. South American cluster showed 100 percent range of variation covering all possible classes for 12 of the 16 morphological traits, revealing highest range of variation.

Assessment of phenotypic diversity in core collection revealed significant variation. The average phenotypic diversity index was higher in the *fastigiata* group (0.146) than the *hypogaea* group (0.141). The *hypogaea* group showed significantly greater mean pod length, pod width, seed length, seed width, yield per plant, and 100-seed weight than the *fastigiata* group in both rainy and post-rainy seasons, whereas it was opposite for plant height, leaflet length, leaflet width, and shelling percentage with *fastigiata* group showing significantly greater means. Principal coordinate and principal component analyses showed that 12 morphological descriptors and 15 agronomic traits were important in explaining multivariate polymorphism. Leaflet shape and surface, color of standard petal markings, seed color pattern, seed width, and protein content did not significantly account for variation in the first five principal coordinates or components of *fastigiata* and *hypogaea* types, indicating their relatively low importance. The average phenotypic diversity index was similar in both subspecies. Further, the Shannon–Weaver diversity index varied among traits between the two subspecies and the diversity within a subspecies/group depended upon the seasons, and traits recorded, suggesting significant role of environmental factors in influencing variability. Molecular profiling of joint composite collection, developed by ICRISAT and EMBRAPA using 21 SSRs, showed rich allelic diversity, group-specific unique alleles, and common alleles sharing between subspecies and geographical groups. Gene diversity ranged from 0.559 to 0.926, with an average of 0.819. Group-specific unique alleles were 101 in wild *Arachis* spp., 50 in subsp. *fastigiata*, and only 11 in subsp. *hypogaea*. Accessions from Americas revealed the highest number of unique alleles (109), while Africa and Asia had only six and nine, respectively. But in the CC and the MCC, representation of accessions from Americas is less despite greater allelic variability, than that of Africa and Asia. The two subsp. *hypogaea* and *fastigiata* shared 70 alleles. In contrast, the wild *Arachis* shared only 15 alleles with *hypogaea* and 32 alleles with *fastigiata* (//ICRISAT// Groundnut Crop: www. icrisat.org/crop-groundnut-genebank.htm). This situation demands inclusion of wild *Arachis* species accessions in selection of CC to capture greater allelic diversity and improve the value of CC inbreeding programs. This can be achieved either through creation of wild *Arachis* species core or creation of subset from which representative accession of wild species with desirable variability can be randomly selected for inclusion into cores.

***Representation of desirable traits diversity* (biotic and abiotic stresses and nutritional traits)**

A CC is a 'true representative of the entire collection' and is 'nearly as diverse as the entire collection'. These principles guide the constitution of a CC. The entire range of variability for characters of significance in breeding, such as resistance to biotic stresses, like rust (*Puccinia arachidis* Speig.) and rosette, was covered by the accessions selected for constitution of the CC and the MCC. For example, the CC represented 100% range of entire collection for resistance to rust, early leaf spot (*Cercospora arachidicola* Hori.), and rosette virus disease. However, strong and weak correlations were observed between various morphological traits and traits of breeding significance (Upadhyaya et al. 2003; Upadhyaya et al. 2002). Upadhyaya et al. (2014) reported identification of multiple resistance and nutritionally dense germplasm from the MCC of groundnut. Subsequent evaluations have also resulted in identification of new sources of variation; for example, germplasm with improved oil quality, as determined by variation in oleic and linoleic fatty acids. Many of these accessions were agronomically at par or even superior over controls and showed specific and wide adaptation (Upadhyaya 2015). However, there are traits with limited or no variability in core sets, such as early leaf spot resistance in the case of MCC (Singh et al. 1997).

Comparison of CCs developed in different parts of the world showed different traits contributing to variability in different sets of collections, associated to the dominance of subspecies and botanical varieties in a collection and selection pressure (Jiang et al. 2008), indicating lack of a universal CC for groundnut improvement meeting everyone's needs.

It is clear from the procedure followed to develop CC and MCC in groundnut that they do not capture entire variability of the total collection. The level of uncaptured variability may vary from 0-30%. This unrepresented variability in CC and MCC may contain very useful alleles that may be missed out following the procedure described by Upadhyaya et al. (2012). Further, the sampling of accessions for inclusion in the CC and MCC, if it were random, was probably okay for quantitative traits. But for qualitative traits, because of their skewed distribution, random sampling is not going to be very helpful. For qualitative traits such as resistance to diseases and insect pests and other biotic and abiotic stresses etc., where prior knowledge is existing in the form of breeder's working collections for different desirable traits, choice sampling can be done to ensure their representation in CC and MCC. In such a scenario, would it not be desirable to formulate gene pools of trait-specific collections and select for desired alleles rather than taking the route of CC and MCC, which may not have captured all the desired alleles.

The solutions and the course correction for selection of a comprehensive core

Although CC development has clearly defined objectives and protocols, but most are not followed. A CC should not be developed from a single (site) evaluation data, but from data amassed from different evaluations at different geographical and cultural conditions, particularly in geographical region of origin. This may perhaps help to arrive at a better CC, which may accommodate rare alleles too. The blanket 10% or 1% of the cluster criterion requires redefinition in case of smaller clusters and larger clusters.

Improved selection with integration of stratification based on biogeographical region and gene pool of accessions with desirable traits, encompassing allelic variability

Most static CCs, selected by the gene-bank curators on priority, are often based on stratification by country of origin and taxonomic groups followed by clustering based on multivariate analysis performed over qualitative and quantitative morphological traits, and logarithmic random selections of accessions from each cluster to capture total spectrum morphological variability to facilitate conservation. For this reason, CCs are often limited in variability for a specific trait(s), particularly for those that are conferred by multiple genes or/are poorly associated with morphological traits (though desirable for improvement).Therefore, they are not represented in the encompassed spectrum of variability, for choice as per the need. This conflict arises because of the preference for allelic richness rather than ensuring allelic representativeness in the core subset. Therefore, an interactive core selection methodology, particularly involving breeders, is the need of the hour. This will ensure meeting wider and specific needs of the users and may further enhance germplasm utilization.

A comprehensive approach for selection of CCs meeting the wider adaptational needs and consisting of accessions representing the variability of desirable traits, such as tolerance to abiotic stresses, resistance to diseases and insect pests, and nutritional contents such as high protein or fat contents, aroma etc. with defined accessions containing different desirable levels of tolerance and/or resistance to stresses, agronomic and nutritional traits can fulfill the most interest of genetic research and breeding programs. Therefore, stratification of germplasm collection by biogeographical region instead of country of origin, and grouping of accessions sharing common characteristics, specifically (predominantly) those of significance in breeding program (stresses, earliness, dwarfness, high oil content, etc.), following a broader hierarchical structure of the gene pool must be considered in selection/sampling of accessions for the development of a CC. This shall help in encompassing wider genetic diversity and improve the scope of CCs from conservation to use. Presently, selection of accessions is confined to country of origin and taxonomy, and neutral or non-neutral descriptors, leaving aside the desirable characteristics required to overcome the factors/stresses responsible for yield reduction and those that can contribute to value addition in crop improvement. It is desired that the stratified selection methodology must integrate biogeographical distribution, and gene pool concept for covering full range of variability for traits of significance in breeding programs. It should be followed by pooling of *logarithm* representation from each group. This will allow encompassing total spectrum of variability for most ecologies (including agro-ecologies) and full range of variability of useful traits to enable users to choose the domain of their interest. Academically, at molecular level this can promote further mining of alleles, using known sequences associated with gene(s) conferring desirable features.

Algorithms such as Core Hunter (Thachuk et al. 2009) may assist defining core subsets based on user preference and having enough genetic diversity and appropriate average genetic distance among accessions. Core Hunter can also find small core subsets that keep all unique alleles found in the reference germplasm collection (total/full Collection).

Unique materials could be lost or discarded due to the inability of proper assessment of genetic diversity in the total collection, whereas certain traits may not be phenotypically expressed, because of inactivation/silencing of genes under certain environments. Therefore, DNA markers, which are not influenced by the environmental conditions, should be able to determine genetic diversity within a population and identify distinct accessions possessing maximum genetic diversity with greater accuracy. Furthermore, some assessments facilitated by DNA markers are revealing the impacts of plant breeding on improved crop gene pools, which may either narrow or widen their genetic base, and shift their genetic background.

Singh (2009) proposed for stratification of subsets/gene pool of accessions with similarity in desirable traits in addition to subsets based on taxonomic and geographic similarities and include them in selection of core accessions along with selection from subsets created on the basis of PCA of morphological qualitative and quantitative traits to enable representation of complete genetic diversity, including useful one across taxonomic and geographic boundaries. Using a similar approach, Guo et al. (2014) developed an integrated applied core collection (IACC) of soybean based on evaluation data for desirable agronomic and nutritional traits of available germplasm resources by including accessions with cold tolerance, drought tolerance, salt tolerance, soybean cyst nematode resistance, soybean mosaic virus resistance, high protein content, and high fat content. They found newly formed CC encompassed accessions with high genetic diversity and desirable agronomic traits. The genetic diversity of the newly formed IACC was compared with that of the established MCC of soybean with the aid of simple sequence repeat (SSR) markers and phenotypic traits. The results showed that at the molecular level, soybean IACC harbored a similar level of genetic diversity as the established MCC, and that at the phenotypic level, the IACC encompassed more accessions with desirable traits than did the established MCC. IACC laid foundation for establishment of a core that may function as a set of active collections to meet different objectives in different eco-regions.

Since most of the genetic resources collected and conserved are basically for use in breeding programs of the present and future, the most pragmatic way for development and evaluation of a CC from utility point of view should be based on the genetic criteria, i.e., gene/allelic diversity both in terms of genetic (sub-specific) and biogeographical diversity. A program for gene search and use in plant breeding should, therefore, consist of the following steps: (i) characterizing and evaluating genetic and phenotypic diversity available in the gene bank for a better understanding of the available variation in relation to taxonomy and biogeography (agroecology) for further use, (ii) screening/evaluating the gene bank accessions for desired traits, discerning allelic variation [the process may be shortened with screening of core subsets (morphology-based), if available- for desirable traits and allele diversity], (iii) creation of gene pools (subsets) of desired traits, such as earliness (maturity), dwarfism, resistance to biotic and abiotic stresses, nutritional (oil-, protein-rich, etc.) and other traits of interest, (iv) integrate these subsets with other subsets created to represent the taxonomic and biogeographical groups, and cluster/group created on the basis of multivariate analysis of morphological traits, in selection of accessions, (v) *logarithm* selection of accessions representing total spectrum of genetic and geographical diversity, and incorporating the desired trait(s) and their allelic diversity, and (vi) promote use of such core into the breeding of genetically enhanced populations for use in a crop improvement program as per the need.

Keep core collection dynamic with regular evaluation for gaps and inclusion of accession **(with desirable feature and rare alleles)**

For the desired success of CCs in breeding programs, they should be dynamic with continuous feedback from users and periodic revision with inclusion of distinctively new sources, revision of grouping, review of users' needs, and inclusion of better and authentic accessions for specific traits. The target of core should be to provide a working collection that can be extensively examined for all economically important traits and use in the breeding program. If certain accessions have not been used in breeding programs over time, they should be dropped from the core collection. Similarly, duplicates with similar alleles for desired traits should be eliminated from CCs. This can be facilitated with development of database on accessions that are part of different core and putting it in public domain.

Further, congener accessions may possess few useful genes, while others may have linkage with undesirable genes, therefore, their importance may be limited to genes of interest. The worth of CC (accessions) may be evaluated based on its usefulness. A CC is expected to represent total genetic variability; however, its constitution should be evaluated to see if all the useful traits are represented? As the accessions of core are representative of a larger set, a data base with related cultivars/ genotypes for each core accession shall facilitate their greater use.

The CC identified from different parts of the world showing differential contribution of variability due to the dominance of certain subspecies/botanical varieties and/or selection pressure, and those lagging in representation of accessions from primary centers of diversity, perhaps due to sociopolitical reasons, can be improved by combining CCs from different world centers for a crop to develop a global CC (GCC) for that crop. This may lead to development of a more effective CC. For example, pooling of rice CC of IRRI, WARDA, China and CIAT to make a GCC of rice.

Breeder's approach facilitating greater use of core

A dynamic crop improvement program should have two approaches: 1) long-term to generate new diversity in the superior agronomic background by accessing new/novel alleles from primitive landraces including wild relatives, and 2) short-term to address immediate threats to the crop and ever-changing demands from the industry and consumers using superior sources to access alleles for resistance/desirable traits. In the latter, mostly advanced breeding lines with desirable traits are intercrossed to develop a new variety with desirable combination of traits as they offer more useful diversity than general germplasm (Duvick1984). The new variety can deviate from the current variety only for the traits requiring improvement and the rest of the traits need to stay the same, as industry and consumers' specifications are very rigid. This results in exclusion of primitive/wild species germplasm from hybridization in the short-term program due to large linkage drag that they carry with them.

Breeding programs, which concentrate on genetic enhancement with desirable variability, can afford to have long-term approach, can use diverse sources including primitive and wild species of desired alleles in their programs. Once the new alleles are incorporated in desired agronomic background without any linkage drag, they are ready to be used in short-term programs. Most of the breeding programs serve a designated geographical area, thus, they look for traits/alleles which are adapted to their designated geographical area. This further restricts the choice of germplasm for the breeder. What breeders require is trait-specific gene pools or CCs and MCCs. Since, these can be evaluated and studied extensively, the information thus generated on them will help breeders to select the most appropriate parent(s) avoiding the duplication of alleles for the desired traits. Studies on combining abilities and stability of the parents included in CCs and MCCs will further help in selecting the parents, which are likely to give desired combinations in segregating generations. Over a period, successful breeders develop good knowledge of their working collection of germplasm and this allows them to make most appropriate choice of germplasm for their effective use in their breeding programs.

Conclusion

A rational stratification (sub-grouping) of full collection on the basis of biogeography and taxonomy and integration of collection group(s)/gene(s) pool with desirable traits of breeding value in random selection of accessions for core, along with selection of accessions from groups (cluster), created on the basis of genetic diversity of morphological traits can result in creation of more representative core sets, with equal emphasis on inclusion of ecological, and genic and allelic variability for traits of significance for genetic improvement of crops. This can be further strengthened with integration of genomics and molecular genetics in accession selection strategy, identifying molecular markers (sequences) and variability within, associated with desirable features. However, often molecular level variations are not comparable to phenotypic variations. Therefore, type of molecular marker used is of significance and one should go for a marker system that is 'truly' genomic in the sense of covering both coding and non-coding sequences. In this respect SNPs are the best. Currently, protocols for deciphering genetic diversity at sub-population level using model based approaches (Pritchard et al. 2000) and large data sets such as whole genome SNP variations are also available to aid effective development of CCs. This may help to incorporate stable traits and in specific breeding providing greater resilience to the crop species across ecologies with improved quality and productivity. Such core set will become the basis of genetic improvement to meet different objectives in different eco-regions, engineering cultigens to face various challenges, including climate change, and nutritional and productivity. This will also resolve the concerns regarding non-concentrating on useful genetic diversity in reducing sample size. Further, the strategy of integrating accessions of all biogeographic regions with various desirable traits shall also help in meeting global goal of genetic improvement. Accessions with more than one specific trait can be used directly for breeding elite varieties.

Declaration

The authors declare no conflict of interest.

References

Brown A. H. D. 1988. The case for core collection. In: The Use of Plant Genetic Resources (Eds. A.H.D. Brown et al.), Cambridge University Press, Cambridge. pp. 136-156.

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

Brown A. H. D. 1989b. The case of core collection. In: The use of plant genetic resources. (Eds. A.H.D. Brown et al.), Cambridge University Press, Cambridge. pp. 135-56.

Data Base of the System-Wide Information Network for Genetic Resources (SINGER). Available online: http://www.singer.cgiar.org/ (accessed on March 1, 2012).

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

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

Frankel O. H. and Brown A. H. D. 1984. Current plant genetic resources – a critical appraisal. In: Genetics: New Frontiers Vol. IV. Oxford IBH Publ. Co. New Delhi, pp. 1-11.

Guo Y., Li Y., Hong H. and Quie Li-Juan. 2014. Establishment of the integrated applied core collection and its comparison with mini core collection in soybean (*Glycine max*). Crop J., **2**: 38-45.

Holbrook C. C., Anderson W. F. and Pittman R. N. 1993. Selection of a core collection from the U.S. germplasm collection of peanut. Crop Sci., **33**: 859-861.

Jiang H., Ren X., Chen Y., Huang L., Zhou X., Huang J., Froenicke L., Yu J., Guo B and Liao B. 2013. Phenotypic evaluation of the Chinese mini-mini core collection of peanut (*Arachis hypogaea* L.) and assessment for resistance to bacterial wilt disease caused by *Ralstonia solanacearum*. Plant Genet. Resour. C, **11**(1): 77-83.

Jiang H., Ren X., Liao B., Huang J., Lei Y., Chen B., Guo B., Holbrook C. C. and Upadhyaya H. D. 2008. Peanut core collection established in China and compared with ICRISAT mini core collection. Acta Agron. Sin., **34**: 25-30.

Jiang H., Ren X., Shou Liao B., Huang J. and Chen Y. 2007. Establishment of peanut core collection in China. Wuhan Zhiwuxue Yanjiu., **25**(3): 289-293.

Li Jinjie, Zhang Hongliang, Wang Deping, Tang Bo, Chen Chao, Zhang Dongling, Zhang Minghui, Duan Junzhi, Xiong Haiyan and Li Zichao 2011. Rice Omics and biotechnology in China Plant Omics Journal, P.O.J., **4**(6): 302-317.

Liu G. Y., Guan R. X., Chang R. Z. and Qiu L. J. 2011.Correlation between Na^+ contents in different organs of soybean and salt tolerance at the seedling stage Acta Agron. Sin., **37**: 1266-1273. (in Chinese with English abstract).

Noriega I. López, Halewood M., Galluzzi G., Vernooy R., Bertacchini E., Gauchan D. and Welch E. 2013. How policies affect the use of plant genetic resources: The experience of the CGIAR. Resources, **2**: 231-269; doi:10.3390/resources2030231.

Pritchard J. K., Stephens, M. and Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics, **155**: 945-959.

Shannon C. E. and Weaver W. 1949. The Mathematical Theory of Communication. Urbana: University of Illinois Press.

Singh A. K., Mehan V. K. and Nigam S. N. 1997. Sources of resistance to groundnut fungal and bacterial diseases: an update and appraisal. Information Bulletin no. 50, Patancheru 502 324, Andhra Pradesh, India: ICRISAT: 42.

Singh Anurudh K. and Nigam S. N. 2016. *Arachis* gene pools and genetic improvement in groundnut. In: Gene Pool Diversity and Crop Improvement. (Eds. V.R. Rajpal, S.M. Rao and SN Raina), Springer International Publishing, Switzerland, pp 17-77.

Singh Anurudh K. 2009. Role of core collection and pre-breeding in management and use of genetic resources for designing crops under changing climate. Indian J. Genet., **69**(4): 1-6.

Song X. E., Li Y. H., Chang R. Z., Guo P. Y. and Qiu L. J.

2010. Population structure and genetic diversity of mini core collection of cultivated soybean (*Glycine max* (L.) Merr.) in China Sci. Agric. Sin., **43**: 2209-2219. (in Chinese with English abstract).

Thachuk C., Crossa J., Franco J., Dreisigacker S., Warburton M. and Davenport G. F. 2009. Core Hunter: An algorithm for sampling genetic resources based on multiple genetic measures. BMC Bioinf, **10**: 243 DOI 10.1186/1471-2105-10-243.

Upadhyaya H. D. 2015. Establishing core collections for enhanced use of germplasm in crop improvement. Ekin J. Crop Breed. Genet., **1**(1): 1-12. Retrieved from http: //dergipark.gov.tr/ekinjournal/issue/22784/243174.

Upadhyaya Hari D., Dronavalli Naresh, Gowda C. L. L. and Singh S. 2012. Mini Core Collections for Enhanced Utilization of Genetic Resources in Crop Improvement. Indian J. Plant Genet. Resour., **25**(1): 111-124.

Upadhyaya H. D., Bramel P. J., Ortiz R. and Singh S. 2002. Developing a mini core of peanut for utilization of genetic resources. Crop Sci., **42**: 2150-2156.

Upadhyaya H. D., Dwivedi S. L., Vadez V., Hamidou F., Singh S., Varshney R. K. and Liao B. 2014. Multiple resistance and nutritionally dense germplasm identified from mini core collection in groundnut. Crop Sci., **54**: 679-693.

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

Upadhyaya H. D., Ortiz R., Bramel P. J. and Singh S. 2003. Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. Genet. Resour. Crop. Evol., **50**: 139-148.

Upadhyaya H. D., Yadav D., Dronavalli N., Gowda C. L. L. and Singh S. (2010). Mini core germplasm collections for infusing genetic diversity in plant breeding programs. Electronic J. Plant Breed., **1**(4): 1294-1309.

Upadhyaya H. D., Ferguson M. E. and Bramel P. J. 2001. Status of *Arachis* germplasm collection at ICRISAT. Peanut, **28**: 89-96

Vavilov N. I. 1926. Centers of origin of cultivated plants. Trpo. Prikl. Bot. Genet. Sel. [Bull. Appl. Bot. & Genet. Sel.], **16**(2): 139-248 [in Russian].

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

Yuan C. P., Li Y. H., Liu Z. X., Guan R. X., Chang R. Z. and Qiu L. J. 2012. DNA sequence polymorphism of the *Rhg4* candidate gene conferring resistance to soybean cyst nematode in Chinese domesticated and wild soybeans. Mol. Breed., **30**: 1155-1162.