The Pharma Innovation Journal 2022; 11(12): 3301-3306

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; 11(12): 3301-3306

www.thepharmajournal.com
Received: 02-09-2022
Accepted: 05-10-2022

Kaveri Chawan
Department of Genetics and Plant Breeding, College of Agriculture, UAS, GKVK, Bengaluru, Karnataka, India

P Ravishankar
Plant Scientist, Oilseeds Scheme, ZARS, College of Agriculture, UAS, GKVK, Bengaluru, Karnataka, India

S Ramesh
Professor, Department of Genetics and Plant Breeding, College of Agriculture, UAS, GKVK, Bengaluru, Karnataka, India

T Onkarappa
Aicrp on Soybean, ZARS, College Of Agriculture, UAS, GKVK, Bengaluru, Karnataka, India

Corresponding Author:
Kaveri Chawan
Department of Genetics and Plant Breeding, College of Agriculture, UAS, GKVK, Bengaluru, Karnataka, India

Development and validation of core sets in soybean (G*lycine max* L. Merrill) germplasm

Kaveri Chawan, P Ravishankar, S Ramesh and T Onkarappa

Abstract
In India, soybean is the most important protein-rich oilseed crop; it has a 40% protein content and 20% oil content. The use of genetic resources in breeding programmes will be managed and accelerated by core collection. The Standard Stratified Clustering (SSC) approach was used in the current study to develop core sets of soybean from a base collection of 2000 accessions based on data on 13 qualitative and 7 quantitative traits. The SSC approach with a combination of two core sizes (10% and 15%), two sampling strategies (proportional and logarithmic), and two allocation strategies (random and preferred) were used to develop eight core sets. Similarity of classes on the basis of qualitative traits of 8 core sets with the base collection was examined using Chi-square test, Shannon-Weaver diversity index, and 'class coverage' statistics. Univariate statistics, based on quantitative traits, such as mean and variance and multivariate statistics, standardized mean difference (SMD %), coincidence ratio (CR %), variance difference (VD %), and variable rate (VR %) were also used to assess the representativeness of core sets. Core set of 15% size developed using logarithmic sampling with preferred allocation strategy retained higher CR%, VD%, and VR% based on quantitative traits.

Keywords: Coincidence ratio, standard stratified clustering, standardized mean difference, variance difference

1. Introduction
A "wonder crop," soybean (*Glycine max* L. Merrill) has a protein content of 40% and an oil content of 20%. It is cultivated extensively in the USA, Brazil, China, India, and Argentina and is a major source of protein and edible oil worldwide. In contrast to Asian nations where soybean has traditionally been utilized as a staple food and consumed as soymilk, tofu, soy sprouts, fermented soy dishes, and soy sauce, soybean production in western nations predominantly concentrates on generating high-protein meals for livestock and vegetable oils. In addition, soybean is a significant source of polysaccharides, soluble fibres, phytosterols, lecithins, saponins, and phytochemicals, isoflavones, which either individually or collaboratively promotes health by reducing the prevalence of diseases like cancer, obesity, hyperglycemia, hypertension, dyslipidemia, and inflammation.

Abundant germplasm accessions have been collected by gene banks around the world, yet breeders still struggle to manage and utilize such a large collection. Development of core collections enables greater use of genetic resources in crop improvement programmes. To develop a manageable sample, or so-called "core collection," Frankel and Brown initially suggested sampling the collections (Brown 1989, Frankel and Brown 1984) [2, 11]. The genetic diversity of a species and its relatives is best represented by a core collection (CC) with the least amount of repetition. Due to its smaller size, CC can be thoroughly researched, and the knowledge gained can be applied to make better use of the much larger reserve collection. Core collections have been developed in many crops, including rice (Li *et al*. 2003) [20], wheat (Dong *et al*. 2003) [7], soybean (Wang *et al*. 2006) [38], cotton (Xu *et al*. 2006) [39], and peanut (Holbrook and Anderson 1995) [14]. Recently, core collection has evolved into a potent tool for assessing germplasm, identifying trait-specific accessions, discovering new genes through association mapping, allele mining, genomic research, marker development, and molecular breeding (Qiu *et al*. 2003) [25].

Drawing representative samples from whole collection for the constitution of core collection is the heart of core collection that determines its quality. Brown (1989a) [3] developed a number of methods for core collection, such as a random sampling strategy of 10% of the base collection that represents more than 70% of genetic variation, and suggested that the ideal core collection size should be between 10% and 20% of the entire collection (EC) (Brown 1989b)

[4]. The principle component score (PCS) method was developed by Noirot *et al.,* (1996) [23] and uses principal component analysis (PCA) to eliminate collinearity between variables while selecting individuals based on their cumulative relative contribution. Coffee (Hamon *et al.,* 1995) [13], mungbean (Bisht *et al.,* 1998) [1], groundnut (Upadhyaya *et al.,* 2003) [33], ragi (Upadhyaya *et al.,* 2006) [34], and sesame (Mahajan *et al.,* 2007) 21[] have all successfully established core collections using the PCS technique. The power core method of core collection, which utilizes the advanced M (maximisation strategy) implemented through the modified heuristic algorithm for the development of core collection, was created by (Kim *et al.,* in 2007) [18]. The power core programme is used to develop core collection in barnyard millet (Jayarame Gowda *et al.* 2009) [17] and ragi (Chandrashekhar *et al.* 2012) [5].

Other methods have been used to develop core sets in various crops, like Core Hunter (Thachuk *et al.* 2009) [31], M-Strat (Gouesnard and Bataillon *et al.* 2001) [12], genetic distance sampling (Jansen and Van Hintum 2007) [16], and standard stratified clustering (SSC) approach (Brown 1989a) [3]. Several studies have compared the effectiveness of various methods and tactics for creating core sets in various crops (Spagnoletti-Zeuli and Qualset 1993; Hu *et al*.. 2000; Franco *et al*.. 2005; Studnicki *et al*.. 2013) [15, 10, 28, 29]. The majority of researchers choose the SSC strategy for creating core sets to ensure the selection of common alleles (Crossa *et al*.. 1995) [6]. The effectiveness of SSC approach depends on a hierarchical classification of base collection accessions into genetically uniform groups. Many researchers have also suggested using a hierarchical approach to divide the base collections into smaller, homogenous groups (Spagnoletti-Zeuli and Qualset 1993; Holbrook *et al*.. 1993; Van Hintumn *et al*., 1995; Skinner *et al*.. 1999; Upadhyaya *et al*.. 2001) [28, 14, 32, 17]. The objective of the present study is to develop core sets using SSC approach-based combination of two core sizes (10% and 15%), two sampling strategies (proportional and logarithmic), and two allocation strategies (random and preferred).

2. Materials and Methods

The material for the study comprised of 2000 germplasm accessions of soybean (*Glycine max* L. Merrill) obtained from IISR (Indian Institute of Soybean Research), Indore, Madhya Pradesh and three check entries (KBS 23, JS 335 and DSB 21). The 2000 germplasm accessions along with 3 check varieties were characterized for 13 qualitative traits and evaluated for 7 quantitative traits following Augmented design during summer 2021 at ZARS, UAS, GKVK, Bangalore. Each block contained 100 germplasm accessions as well as three checks (replicated twice). Each entry's seeds were dibbled in a single row of 1.5-meter length with 45 X 10 cm row spacing. A base dose of 25:50:25 Kg NPK ha^{-1} was applied to the experimental plot at the time of sowing. Recommended agronomic and plant protection practices were followed to raise a healthy crop.

Data were collected visually on 13 qualitative traits (hilum color, seed coat color, early plant vigour, hypocotyls color, flower color, leaf shape, leaflet color, plant pubescence, plant pubescence color, plant pubescence density, plant pubescence type, stem determination, pod color) using five randomly tagged plants. Data were also collected on seven quantitative traits based on counting/measurement (days to 50% flowering, plant height (cm), number of secondary branches $plant^{-1}$, number of pods $plant^{-1}$, days to 80% maturity, 100 seed weight (g), seed yield $plant^{-1}$ (g). Quantitative-trait data recorded on five randomly tagged plants were subjected to analysis of variance (ANOVA). Quantitative-trait means of germplasm accessions were adjusted by subtracting trait means from effects of respective blocks in which the accessions were evaluated. The adjusted quantitative-trait means of 2000 accessions were used for the statistical analysis described in the following sections for developing core sets of sizes 10% and 15% following SSC (Brown 1989a) [3].

3. SSC Approach

The process of forming core subsets starts with stepwise grouping of accessions into meaningful clusters followed by accession selection using pre-determined sampling and allocation strategies (Franco *et al*. 2005) [10]. The method used in the SSC approach for creating core sets is discussed below.

3.1 Stratification of the Germplasm Accessions

Accessions were classified into 10 clusters following Ward's hierarchical clustering algorithm based on adjusted means for 7 quantitative traits. Clusters were merged at each step by minimizing the variance within clusters and thus maximizing variance among clusters based on the adjusted means of 7 quantitative traits.

3.2 Sampling Strategies

Two sampling strategies, viz., proportional (P), a number proportional to cluster size and logarithmic (L), a number proportional to logarithm of cluster size (Brown 1989) [2] in relation to the chosen core set size of 10% and 15%, were followed to determine the number of accessions to be selected from each cluster for inclusion in core sets.

3.3 Allocation Strategy

Once the number of accessions to be selected from each cluster was determined, the accessions were chosen from each cluster following random and preferred sampling methods for inclusion in the core sets. Criteria such as number of pods per plant, 100 seed weight and seed yield per plant were used for selection in preferred sampling method. Thus, a total of 8 core sets were developed following SSC approach.

4. Validation of core sets

Validation is the process of examining the extent to which core collection represents the base collection. The base collection and eight core sets were compared and tested for homogeneity in quantitative traits mean (two-sample 't' test) and variances [Levene's (Levene, 1960) [19] test]. The homogeneity of accessions for qualitative traits frequency distribution of base and core collections was tested following chi-square statistic. Retention of qualitative trait classes by the core collection was determined using 'Shannon-Weaver diversity index' (Shannon and Weaver, 1949) [26] and 'class coverage' (Kim *et al.* 2007) [18] statistics. Non-significant differences between base and core sets for quantitative traits means and variances and homogeneity of frequency distribution of qualitative traits classes were considered as evidences for representativeness of core sets.

Quantitative trait-based composite and standardized validation statistics, such as standardized mean difference

(SMD %), variance difference (VD %), coincidence rate (CR %), and variable rate (VR %) (Hu *et al.* 2000) [15] were also used to assess the representativeness of 8 core sets.
The SMD (%) was estimated as,

$$\text{SMD (\%)} = \frac{1}{m}\sum_{j=1}^{m}\frac{Mb-Mc}{Mc} \times 100$$

The VD (%) was estimated as,

$$\text{VD (\%)} = \frac{1}{m}\sum_{j=1}^{m}\frac{Vb-Vc}{Vc} \times 100$$

The CR (%) was estimated as,

$$\text{CR (\%)} = \frac{1}{m}\sum_{j=1}^{m}\frac{Rc}{Rb} \times 100$$

The VR (%) was estimated as,

$$\text{VR (\%)} = \frac{1}{m}\sum_{j=1}^{m}\frac{CVc}{CVb} \times 100$$

Where,
m= Number of traits
Mb= Trait mean of base collection
Mc= Trait mean of core collection
Vb= Trait variance of base collection
Vc=Trait variance of core collection
Rb= Trait range of the base collection
Rc= Trait range of the core collection
CVb= Trait CV of the base collection
CVc= Trait CV of core collection

4.1 Criteria to examine representativeness of core set
The core set was considered to be representative of the base collection, if Core and base collections were significantly different for not more than 20% of the quantitative traits (SMD% ≤4).

a. The CR% retained by the core collection for quantitative traits is not less than 80 per cent (Hu *et al.* 2000) [15].
b. The differences between the actual numbers of accessions in each of defined classes in the core collection and those expected based on the core collection size is significant for not more than 20 per cent of the qualitative traits when tested using chi-square test. For example if the number of accessions in 3 different classes of a qualitative trait in the base collection are 30, 50 and 20, respectively, the expected number of accessions in the 10% sampled core collection were estimated as 10 per cent of 30=3, 10 per cent of 50=5 and 10 per cent of 20=2. These expected numbers of accessions in each of the 3 classes were compared for their deviation from those actually present in the core collection using chi-square test.
c. The core collections retained not less than 80 per cent 'class coverage' based on qualitative traits.
d. The average Shannon-Weaver diversity index across the qualitative traits of core collection is comparable to that of base collection.

5. Results and Discussion
5.1 Representativeness of Core Sets
Eight distinct core sets' classes of 13 qualitative traits were compared to those of the base collection. All core sets were comparable to those of the base collection for frequency distribution of qualitative traits, with the exception of proportional and logarithmic sampling with random allocation of sizes 10% and 15% (chi-square was significant for 4 traits), indicating their representativeness of the base collection for qualitative traits (Table 1).
The core set (s) with H' values similar to those of the base collection, according to Bisht *et al*. (1998) [1]; Upadhyaya (2003) [33]; Mahalakshmi *et al*. (2007) [22]; Dwivedi *et al*. (2008) [28] and Upadhyaya *et al*. (2009) [35], are thought to be representative of the base collection. The H' estimates of all the core sets in the current study were comparable to those of the base collection, indicating their representativeness according to the H' statistic. The 'class coverage' statistic was used to calculate the percentage retention of qualitative trait classes in the core sets (Kim *et al*. 2007) [18]. A good core collection, according to this criterion, should retain all of the classes of a given qualitative trait of the base collection. All 8 core sets covered more than 80% of the defined qualitative trait classes, indicating their representativeness.
To evaluate the representation of the core sets, the means, ranges, and variances of the quantitative traits of the core sets and those of the base collection are compared. For a representative core set, ranges should remain constant and means shouldn't fluctuate noticeably for less than 20% of the traits. However, due to effective stratification and fewer redundant accessions in the core set than in the base collection, the trait variances may rise in the core sets. In the current investigation, all of the core sets for quantitative trait means were comparable to those of the base collection ('t' test was significant for ≤4 traits).
To assess the representativeness of core collections, composite criteria such as standardized mean, variance, interquartile range, and coefficient of variation (Hu *et al*. 2000; Tai and Miller 2000) [15, 30] have been proposed. It involves conducting statistical tests to compare the core and base collections' means (SMD %), variances (VD %), range (CR %), and CV (VR %). The SMD% of all eight core sets was under 4, confirming their representativeness for quantitative trait means. In comparison to other approaches based core sets, logarithmic sampling with preferred allocation strategy (of 15% size) retained higher VD%, CR% and VR%.

5.2 Comparison of SSC Strategies
5.2.1 Core sizes
According to Brown (1989b) [4], a core set of 10% or less of the base collection is likely to hold at least 70% of the variety in the base collection. The amount of genetic redundancy among accessions, the resources available for maintaining core entries, and the frequency of entry regeneration all play a major role in determining the ideal size of a core set (Yonezawa *et al*. 1995) [40]. In the current study, a core size of 15% retained more CR% and VR%, indicating higher representativeness of the core sets (Table 2; Figure a).

5.2.2 Sampling strategies
Based on SMD% criteria, proportional sampling strategy-based core sets better represented the base collection than those based on logarithmic sampling strategy. However, the CR%, VR% and VD% of logarithmic sampling strategy based core sets were higher than those of proportional sampling strategy based core sets. (Table 2; Figure b). Many scholars, including Brown (1989a) [3], van Hintum *et al*. (1995) [36], and

others, have indicated that logarithmic and proportional sampling procedures are suitable for developing representative core sets.

5.2.3 Allocation strategies

Among the two allocation strategies, preferred allocation was superior to random allocation as indicated by least SMD%, and higher CR% and VR% (Table 2; Figure c).

5.3. Efficiency of SSC Approaches

Among the eight representative core sets identified, the base collection diversity was better represented by the logarithmic sampling with preferred allocation approach-based core set of 15% size than by the other core sets since it retained higher CR%, VD%, and VR% based on quantitative traits, H' estimates was comparable to those of the base collection, and "class coverage" statistics covered more than 80% of the defined qualitative trait classes.

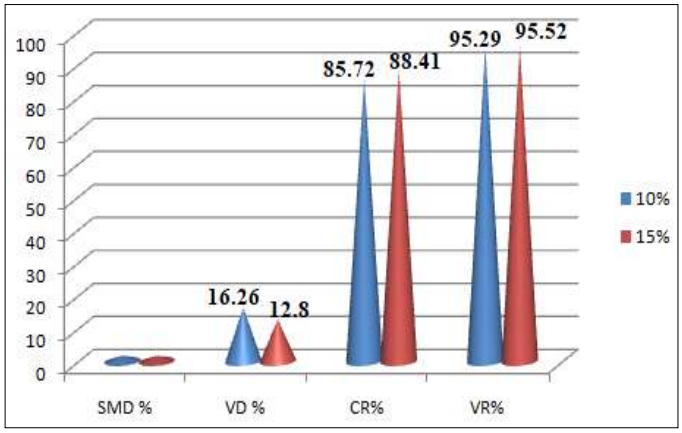
Table 1: Summary of validation statistics to identify representative and best core set (s) of soybean germplasm accessions

		Core size							
		10% of base collection				15% of base collection			
		Proportional sampling		Logarithmic sampling		Proportional sampling		Logarithmic sampling	
Sl. No	Random allocation	Random allocation	Preferred allocation	Random allocation	Preferred allocation	Random allocation	Preferred allocation	Random allocation	Preferred allocation
Qualitative traits									
1	#Significant Chi square	9	2	8	2	9	2	12	2
2	Shannon-Weaver diversity index (H^1)	1.00 ± 0.48	0.97 ± 0.611	1.00 ± 0.52	1.00 ± 0.50	0.99 ± 0.40	0.99 ± 0.48	1.02 ± 0.43	1.00 ± 0.63
3	Class coverage	96.7	96.7	97.8	96.7	97.8	96.7	97.8	96.7
Quantitative traits									
1	#Significant 't' test	4	4	4	4	3	4	4	4
2	#Significant 'F' test	7	6	7	6	7	6	7	6
3	SMD%	0.6	0.99	1.82	3.37	0.65	1.28	1.42	3.53
4	VD%	10.01	6.89	14.43	13.92	24.52	6.79	9.87	29.79
5	CR%	80.06	90.53	80.96	91.32	84.22	90.11	85.21	92.11
6	VR%	89.31	96.89	92.99	96.95	90.93	96.22	92.99	101.95

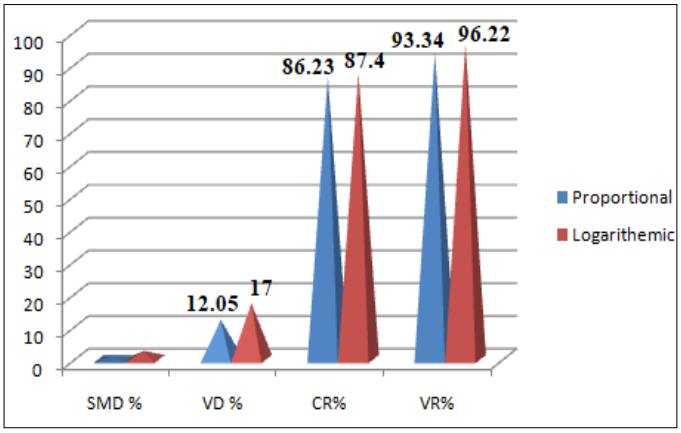
(SMD) standardized mean difference; (VD) variance difference; (CR) coincidence ratio; (VR) variable rate; (#) representative core sets

Table 2: Comparison of core sizes, sampling strategies, and allocation strategies of developing core sets in Soybean

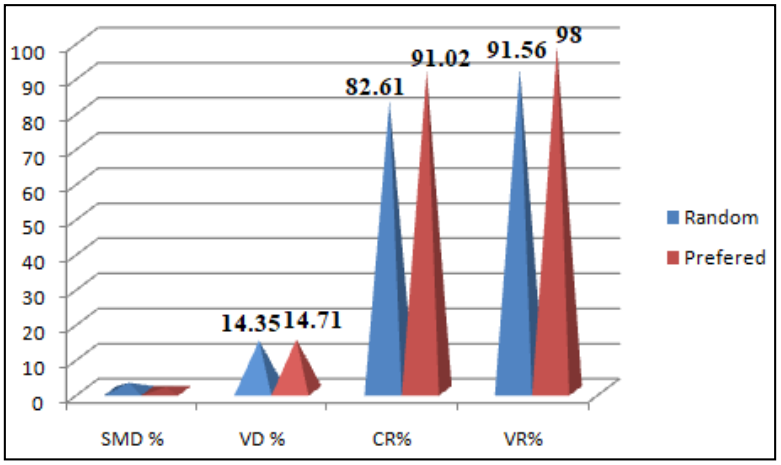
Statistics	Core sizes		Sampling strategy		Allocation strategy	
	10%	15%	Proportional	Logarithmic	Random	Preferred
SMD %	1.72	1.7	0.88	2.54	2.29	1.12
VD %	16.26	12.8	12.05	17	14.35	14.71
CR%	85.72	88.41	86.23	87.4	82.61	91.02
VR%	95.29	95.52	93.34	96.22	91.56	98



(a) Core sizes



(b) Sampling strategies



(c) Allocation strategies

Fig 1: Graph depicting comparison of (a) core sizes, (b) sampling strategies, and (c) allocation strategies for developing core sets in soybean (SMD) standardized mean difference; (VD) variance difference; (CR) coincidence ratio; (VR) variable rate

6. Conclusions

The soybean core collection created in this study will be valuable genetic resources for soybean breeders and researchers for screening soybean germplasm and identifying desirable genotypes for economically important traits. The development of core collections will also aid in addressing challenges posed by climate change because core collections represent the genetic variability of the entire collection and desirable genotypes can be easily identified. Owing to limited available resources, evaluating the entire collection may not be practically feasible; therefore, core collection can act as a working collection for breeders to be used in evaluation and breeding programmes. The soybean core collection created in the current study can also be used in association mapping studies to identify the genes and QTLs linked to numerous economically significant features. As new accessions of soybean germplasm are gathered and new data are developed, the current core collection of soybeans needs to be periodically revised.

7. References

1. Bisht IS, Mahajan RK, Patel DP. The use of characterisation data to establish the Indian mungbean core collection and assessment of genetic diversity. Genetic Resources and Crop Evolution. 1998a;45:127-133.
2. Brown AHD. A case for core collections. In: The Use of Plant Genetic Resources, edited by AHD Brown OH, Frankel Marshall DR and Williams JT. Cambridge, MA: Cambridge University Press; c1989. p. 136-156.
3. Brown AHD. Core collections: a practical approach to genetic resources management. Genome. 1989a;31:818-824.
4. Brown AHD. Core collections: a practical approach to genetic resources management. Genome. 1989b;31:818-824.
5. Chandrashekhar H, Gowda J, Ugalat J. Formation of core set in Indian and African finger millet (*Eleusine coracana* L. Gaertn) germplasm accessions. Indian Journal of Genetics and Plant Breeding. 2012;72:358-363.
6. Crossa J, Delacy IH, Taba S. The use of multivariate methods in developing a core collection. In Core Collections of Plant Genetic Resources (T. Hodgkin, A. H. D. Brown, Th. J. L. van Hintum and E. A. V. Morales, eds.). John Wiley and Sons, UK; c1995. p. 77-92.
7. Dong YC, Cao YS, Zhang XY, Liu SC, Wang LF, You GX, *et al*. Establishment of candidate core collections in Chinese common wheat germplasm. J Plant Genet. Resour. 2003;4:1-8
8. Dwivedi SL, Puppala N, Upadhyaya HD, Manivannan N and Singh S. Developing a core collection of peanut specific to Valencia market type. Crop Sci. 2008;48:625-632.
9. Federer WT. Augmented (or hoonuiaku) designs. Hawaii. Plant Res. 1956;2:191-208.
10. Franco J, Crossa J, Taba S, Shands H. A sampling strategy for conserving genetic diversity when forming core sub sets. Crop Sci. 2005;45:1035-1044.
11. Frankel OH, Brown ADH. Current plant genetic resourcesA critical appraisal. In: Genetics New Frontiers. Proc. of XV Int. Congress of Genetics. Oxford & IBH publishing Co. 1984;(4):3-13.
12. Gouesnard B, Bataillon TM. MSTRAT: An algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. J Hered. 2001;92(1):93-94.
13. Hamon S, Noirot D, Anthony F. Developing a coffee core collection using the principal components score strategy with quantitative data. In: Hodgkin, T., Brown, A. H. D., van Hintum, T. J. L., and Morales EAV (eds) Core Collections of Plant Genetic Resources. Chichester, UK: IPGRI-Wiley & Sons; c1995. p. 117-126.
14. Holbrook CC, Anderson WF. Evaluation of a core collection to identify resistance to late leafspot in peanut. Crop Sci. 1995;35:1700-1702.
15. Hu J, Zhu J, Xu HM. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. Theor. Appl. Genet. 2000;101:264-268.
16. Jansen Jand, Van Hintum JL. Genetic distance sampling: a novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. Theor. Appl. Genet. 2007;114(3):421-428.
17. Jayarame Gowda, Bharathi S, Somu G, Krishnappa M, Rekha D. Formation of core set in barnyard millet [*Echinochloa frumentacea* (Roxb.) Link] germplasm using data on twenty four morpho-agronomic traits.

International Journal of Plant Sciences. 2009;4:1-5.
18. Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwag JG, *et al.* Power Core: A program applying the advanced M strategy with a heuristic search for establishing core sets. Bioinformatics. 2007;23:2155-2162.
19. Levene H. Robust tests for equality of variances. In: Olkin, *et al.* (ed.). Contributions to probability and statistics: Essays in honour of Harold Hotelling. Stanford University Press, Stanford; c1960. p. 278-292.
20. Li ZC, Zhang HL, Cao YS, Qiu ZE, Wei XH, Tang SX, *et al.* Studies on the sampling strategy for primary core collection of Chinese ingenious rice. Acta Agron. Sin. 2003;29:20-24.
21. Mahajan RK, Bisht IS, Dhillon BS. Establishment of a core collection of world sesame (*Sesamum indicum* L.) germplasm accessions. SABRAO Journal of Breeding and Genetics. 2007;39:53-64.
22. Mahalakshmi V, Lawson M and Ortiz R. Cowpea (*Vigna unguiculata* L. Walp.) core collection defined by geographical, agronomical and botanical descriptors. Plant Genet. Res. 2007;5(3):113-119.
23. Noirot M, Hamon S, Anthony F. The principal component scoring: A new method of constituting a core collection using quantitative data. Genetic Resources and Crop Evolution. 1996;43:1-6.
24. Odong TL, Jansen J, Van Eeuwijk FA, Van Hintum TJL. Quality of core collections for effective utilization of genetic resources review, discussion, and interpretation. Theor. Appl. Genet. 2013;126:289-305.
25. Qiu, LJ, Cao YS, Chang RZ, Zhou XA, Wang GX, Sun JY, *et al.* Establishment of Chinese soybean (*G. max*) core collection: Sampling strategy. Sci Agri Sin. 2003;36:1442–1449.
26. Shannon CE, Weaver W. The mathematical theory of communication. University Illinois Press, Urbana, USA; c1949.
27. Skinner DZ, Bauchan GR, Auricht G, Hughes S. A method for the efficient management and utilization of large germplasm collections. Crop. Sci. 1999;39:1237-1242.
28. Spagnoletti Zeuli LW, Qualset CO. Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. Theor. Appl. Genet. 1993;87:295-304.
29. Studnicki M, Mądry W, Kociuba W. The efficiency and effectiveness of sampling strategies used to develop a core collection for the Polish spring triticale (×*Triticosecale* Wittm.) germplasm resources. Communications in Biometry and Crop Sci. 2010;5(2):127-137.
30. Tai P, Miller JD. A core collection for *Saccharum spontaneum* L. from the world collection of sugarcane. Crop Sci. 2000;41(3):879–885.
31. Thachuk C, Crossa J, Franco J, Dreisigacker S, Warburton M, Davenport GF. Core Hunter: an algorithm for sampling genetic resources based on multiple genetic measures. BMC Bioinformatics. 2009;10:243.
32. Upadhyaya HD, Bramel PJ, Singh S. Development of a chickpea core subset using geographic distribution and quantitative traits. Crop Sci. 2001;41:206–210.
33. Upadhyaya HD, Ortiz R, Bramel PJ, Singh S. Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. Genetic Resources and Crop Evolution. 2003;50:139-148.
34. Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CLL, Reddy VG, Singh S. Development of core subset of finger millet germplasm using geographical origin and data on 14 quantitative traits. Genetic Resources and Crop Evolution. 2006;53:679–685.
35. Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CLL, Reddy VG, Singh S. Developing a mini core collection of sorghum for diversified utilization of germplasm. Crop Sci. 2009;49:1769-1780.
36. Van Hintum THJL, Von Bothmer R, Visser DL. Sampling strategies for composing a core collection of cultivated barley (*Hordeum vulgare*) collected in China. Hereditas. 1995;122:7-15.
37. Van Hintum THJL, Brown AHD, Spillane C, Hodgkin T. Core collections of plant genetic resources. IPGRI Technical Bulletin 3. International Plant Genetic Resources Institute, Rome, Italy; c2000.
38. Wang LX, Guan Y, Guan RX, Li YH, Ma YS, Dong ZM, *et al.* Establishment of Chinese soybean (*Glycine max*) core collections with agronomic traits and SSR markers Euphytica. 2006;151:215-223.
39. Xu H, Mei Y, Hu J, Zhu J, Gong P. Sampling a core collection of island cotton (*Gossypium barbadense* L.) based on the genotypic values of fiber traits Genet. Resour. Crop. Evol. 2006;53:515-521.
40. Yonezawa K, Nomura T, Morishima H. Sampling strategies for use in stratified germplasm collections. In Core Collections of Plant Genetic Resources (Hodgkin T, Brown ADH, van Hintum JL and Morales EAV eds.). John Wiley and Sons, UK; c1995. p. 35-54.