Tobacco Research 47(1): 24-34, 2021

CHARACTERIZATION OF A SUB-SET OF TOBACCO GERMPLASM AND ITS CORE COLLECTION

SARALA K.[1], K. BAGHYALAKSHMI[2], K. PRABHAKARA RAO[1], D. DAMODAR REDDY[1], P. VINAY[1], P. SONIA[1], G. KIRAN[1] AND K. SANTINANDIVELU[1]

[1]ICAR-Central Tobacco Research Institute, Rajahmundry-533 105, India
[2]ICAR-Central Institute for Cotton Research, Research Station, Coimbatore, Tamil Nadu-641003

(Received on 4th February, 2021 and accepted on 4th June, 2021)

In order to effectively utilize the germplasm maintained at ICAR-Central Tobacco Research Institute, a core collection of 65 accessions (21 *bidi*, 19 country cheroot, 8 cigar filler and 17 chewing) constituted from the base collection of 647 accessions (205 *bidi*, 175 chewing, 185 country cheroot and 82 cigar filler) and compared with the diversity of source collection. Out of 24 morphological traits studied, six leaf and three reproductive characters found to contribute to diversity in base collection. Based on the 't' test, means of most of the morphological characters in base collections in each sub-group found be non-significant compared to their cores. Shannon-Weaver diversity index indicated that 93% of the diversity in the base collection was represented in the total core. In the Chi-square analysis, the frequency distribution of entries for 24 morphological characters in total base collection and its core found to be non-significant. Thus, all the tests indicated that the diversity in base collections was captured in respective core collections and hence, constituted cores can effectively be evaluated and utilized with minimum effort in breeding programmes.

INTRODUCTION

The crop germplasm conserved in genebanks around the world represents a total variability of the species and are maintained for further use by researchers in scientific advancements. Plant germplasm collections which include landraces, exotic cultivars, old varieties, related species etc. are the key resources that researchers use in breeding crop varieties with new traits viz., yield, stress resistance, chemical quality, improved handling characters etc. Utilization of diverse germplasm materials in breeding crop cultivars is essential to broaden the genetic base so as to reduce the genetic vulnerability of the cultivated varieties to biotic and abiotic stresses. However, conservation of large number accessions in germplasm collections is making it practically difficult to evaluate and effectively utilize the existing desirable variability in breeding programmes. For fruitfully utilizing large germplasm collections, Frankel (1984) proposed the concept of 'core collection'. The core collection represents a limited size of germplasm that has chosen to epitomize the genetic diversity of a large collection. Basic knowledge about the entire germplasm collection in a species is, however, essential to establish a core collection. The procedure for making a core collection involves five steps, they are: identifying the material, deciding the size of collection, dividing the material into distinct groups, deciding the number of entry per group and finally choosing the entries from each group to construct the core set (Van et al. 2000). Thus the constituted collection will represent the diversity of total collection and can effectively be utilized by crop breeders. Researchers are using different methods viz., random sampling (Van *et al.,* 2000; Schafleitner *et al.,* 2015), principal component scoring (Noirot *et al.,* 1996), stepwise clustering (Hu *et al.,* 2000) etc. using morphological and molecular data (Gouesnard *et al.,* 2001) in establishing core collections. Core collections are being constituted and evaluated for their successful utilization in Wheat (Zhang *et al.,* 2019), Melilotus sp. (Prasada *et al.,* 1995), mung bean (Schafleitner *et al.,* 2015), sorghum (Prasada *et al.,* 1995; Grenier *et al.,* 2001; Upadhyaya *et al.,* 2019), pearl millet (Bhattacharjee *et al.,* 2007; Upadhyaya

Key words: Core collection- diversity-germplasm-tobacco-*bidi*- chewing-country cheroot-cigar filler

et al., 2009a), chickpea (Upadhyaya *et al.,* 2001), groundnut (Upadhyaya *et al.,* 2003), pigeonpea (1290 accessions) (Reddy *et al.,* 2005), finger millet (Upadhyaya *et al.,* 2006) and foxtail millet (Upadhyaya *et al.,* 2009b).

Tobacco is a high value commercial crop cultivated in India generating livelihood security to millions of people and revenue to government in the form of huge excise duty and foreign exchange. Diverse types of tobacco including Flue-cured Virginia, Burley, Oriental, *Bidi*, Country Cheroot, Chewing, Cigar Wrapper and Filler, Hookah etc. are cultivated in the country for both domestic and export purposes. ICAR-CTRI as a national active germplasm site for tobacco maintains around 3380 germplasm accessions representing various tobacco types and *Nicotiana* Species in its genebank. There are 205 *bidi*, 175 chewing, 185 country cheroot and 82 cigar filler accessions available in the tobacco germplasm. *Bidi*, Chewing, Country Cheroot (CC) and Cigar Filler (CF) are the indigenous tobacco types cultivated in India mainly for smoking purposes. In order to efficiently characterize and utilize the diversity of these four germplasm groups, an effort made to constitute core collections for each type. Once, it is proved that the established cores epitomize the base collections; they can be fruitfully evaluated and utilized in tobacco breeding and other studies. Hence, in the present investigation, comparative analysis was made between constituted cores of *Bidi*, Country Cheroot, Chewing, Cigar Filler and their total with their base collections for measuring their resemblance for their effective utilisation.

MATERIALS AND METHODS

A total tobacco germplasm of 647 indigenous tobacco accessions (base collection) including 205 *bidi*, 175 chewing, 185 country cheroot (CC) and 82 cigar filler (CF) collected from various tobacco growing areas and countries were raised under field condition at Katheru Farm, Rajahmundry, India following standard production practices. All the entries were studied for 24 morphological characters under field condition as per the proforma developed for characterization of germplasm lines at the Institute (Table 1). All the quantitative and qualitative values recorded for different characters were converted to scores (1-9) as indicated in the brackets next to each note at Table 1.

Cluster analysis: The diversity existing among the base collections of each type and the entire germplasm entries was computed based on the scores recorded for morphological characters using Computer Software Program–DARwin (Perrier, Jacquemond, 2006). Dissimilarly matrix for morphological observation was constructed using Rogers-Tanimoto coefficient of associations to find out genetic relationships. The matrix data were subjected to unweighted pair groups method with arithmetic mean (UPGMA) analysis to generate dendrogram using DARwin 5.0 Software and dissimilarity was estimated based on the respective morphological scoring.

Construction of core collection and characterization: Further, based on diversity and passport data of base collection, about approximately 10 per cent of entries were randomly selected from each tobacco sub-type to form core collections in *Bidi*, Country Cheroot, Chewing, Cigar Filler types. Core accessions from four groups combined together to form a tobacco core collection referred henceforth in the text as 'total core collection'. The variability available in the base collections of each sub-group, total germplasm (sum of four sub-groups) and their respective cores was studied through the comparison of means and means significance. Shannon diversity index and Chi-square test were used to compare total germplasm base collection and its core.

***Comparison of Mean of morphological traits*:** The means of base collections of each sub-group, total germplasm and their respective core collections were compared using t-test for all the morphological data. Non-significant t-test results indicates both the groups have similar diversities and vice versa.

***Shannon-Weaver diversity Index and Chi-square test*:** It is a commonly used diversity index (Clarke and Warwick, 2001) that takes into account both abundance and evenness of species present in the community. The index assumes that the individuals are sampled randomly from a large independent population. It is explained by the

Table 1: Details of morphological observations recorded on tobacco germplasm accessions

S. No.	Character	Notes and their assigned scores (given in parenthesis)
	Plant Characters	
1	Shape (PS)	Conical (1), Cylindrical (2), Elliptical (3)
2	Mean Stem Height (MSH)	Very Short (1), Short (3), Medium (5), Tall (7),Very Tall (9)
3	Habit (PH)	Open (1), Erect(3), SemiErect(5), Squattering(7), Bouquet (9)
4	Internodal Length (PI)	< 4 cm (1),<6 cm (2),>6cm (3)
	Leaf Characters	
5	Number of Leaves (PN)	Very Few (1),Few (3),Medium Many(7),Very Many (9),
6	Type (LT)	Sessile (1), Petiolate (2)
7	Angle of Insertion (LAI)	Acute(1),Moderately Acute(2),Right Angle(3), Moderately Right(4)
8	Length (LL)	Very Short (1),Short (3),Medium (5), Long (7),Very Long (9)
9	Width (LW)	Very Narrow (1),Narrow (3),Medium (5), Broad (7)
10	Midrib (LM)	Thin (1),Medium (2),Thick (3)
11	Blade Shape (LBS)	Lanceolate (1),Narrow Elliptic (2), Broad Elliptic (3), Ovate (4),Obovate (5),Cordate (6),Rounded (7)
12	Tip Shape (LTS)	Obtuse (1),Slightly Pointed (3), Medium Pointed (5), Strongly Pointed (7), Very Strongly Pointed (9)
13	Blistering of Blade (puckering) (LBB)	Absent or Very Weak (1),Weak (3), Medium (5), Strong (7),Very Strong (9)
14	Undulations of Margin (LUM)	Absent or Very Weak (1),Weak (3),Medium (5), Strong (7)
15	Development of Auricles (LDA)	Absent or Very Weak (1),Weak (3),Medium (5),Strong (7), Very Strong (9)
16	Color of Blade (LCB)	Yellow Green (1),White Green (2),Light Green (3), Medium Green (4),Dark Green (5)
17	Colour of midrib (LCM)	Whitish (1), White Greenish (2), Greenish (3)
	Inflorescence Characters	
18	Inflorescence Shape (IS)	Spherical (1),Flattened Spherical (2), Inverted Conical (3),Double Conical (4)
19	Inflorescence Compactness (IC)	Very Loose (1),Loose / Slightly loose(3), Medium (5), Slightly Dense / Dence (7), Very Dence (9)
	Flower Characters	
20	Expression -Tips of Corolla (FETC)	Absent (1),Weak (3), Medium (5),Strong (7), Very Strong (9)
21	Colour of Corolla (FCC)	White(1), Light Pink \sh white / whitish pink (2), Medium Pink (3), Dark Pink (4),Red (5), Variegated (6), Pinkish Yellow(7),Yellow(8)
22	Length of Pistil Relative to Stamens (FLPRS)	Shorter (1),Equal Length (2),Longer (3)
	Fruit & Seed characters	
23	Fruit Form (FF)	Rounded (1),Intermediate (2),Ovate (3),Conical (4), Elongated (5)
24	Seed Testa Colour (TC)	Light Brown (1),Dark Brown (2),Cream (3),Yellow (4)

formula :H = $-\sum_{i=1}^{T}(P_i * \ln P_i)$, where,H = the Shannon diversity index, P_i = fraction of the entire population made up of species 'i'. Comparable H values for core and its base population indicates core represent base population and vice versa. The Chi-square test was used to analyse the distribution homogeneity for all the characters in the total base collection and core collection. Non-significant Chi-squares demonstrate similar nature of the populations.

RESULTS AND DISCUSSION

Variability found to exist in the base collections of all the four sub-groups and total collection for all the 24 morphological characters studied in the present study. Sarala *et al.,* (2018, 2019) and Baghyalakshmi *et al.,* (2018) also reported the availability of diversity in tobacco germplasm groups. Total tobacco germplasm base collection of 647 indigenous tobacco accessions including 205 *bidi*, 175 chewing, 185 country cheroot and 82 cigar filler were characterized for 24 morphological characters (Table 1). Variability found to exist in each of the base collections of germplasm types and their total (all four types put together) for all the morphological characters studied. Observations recorded varied from minimum of two variables in leaf type to eight in corolla colour. However, minimum variables actually recorded differs from two in plant shape, leaf type and seed testa colour to six in leaf blade shape.

In order to effectively utilize the germplasm, about 10 percent of genotypes were selected to establish core collection in bidi, CC, CF and chewing types and all these selections were combined together to form total core collection. In tobacco, Zhide *et al.,*. (2001) established a core collection with 411 accessions (~10%) from 3979 catalogued China tobacco germplasm collections based on cluster analyses with 7-11 qualitative traits combined with experience. Agostino *et al.,* (2012) produced a core collection merging five sets of tobacco accessions composed of 12 "Burley" (including 1 "Maryland"), 20 "Flue-cured", 20 "Oriental", 14 "Cigar", 10 "Primitive", 8 "Dark", and 5 "Other" based on the genetic diversity and potential for tobacco breeding from a selected set of 312 out of 1,900 accessions of *N. tabacum*. Core can be established to represent only a part of a collection for their effective utilisation (Bisht *et al.,* 1998, Erskine eta al. 1991). In most of the cases, variations in different characters would be proportionally represented in the cores constituted compared to their base collections. This primarily indicates the cores embody all the variation of their base collections. As per the definition of the core collection, the size should be much smaller compared to the original set. There are studies showing the core collections ranging from 5 to 20 per cent (Van *et al.,* 2000). Brown (1989) suggested that the entries to a core collection should be limited to ~10%, using the sampling theory of selectively neutral alleles, with a ceiling of 2000 per species. While analyzing the size and grouping strategies Charmet and Balfourier (1995) and Bisht *et al.,* (1998) found that sizes of 5–10% were optimal, capturing 75–90% of the diversity. In contrast, Noirot *et al.,* (1996) have suggested that higher percentages (20–30%) are needed, particularly where the objective is to capture the genetic diversity of quantitatively inherited characters. This level of sampling is effective in retaining 70% of alleles of entire collection. Brown (1989) further suggested that when there is no base for stratification, simple random sampling can be used for the formation of core collection. These can be thoroughly evaluated and the information so derived can be utilized for improving the efficiency of breeding programs. The accessions that constitute a subgroup would be more or less uniform and therefore ~10% of accessions are retained from each subgroup generally (Upadhyaya *et al.,* 2010). When the core developed from larger collection, the percentage size may be much smaller than 5%. The International Barley Core Collection (1600 accessions) is less than 0.3% of the world barley holding and the ICRISAT (International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India) sorghum core collection of approximately 600 accessions was formed from a collection of 40 000 accessions (1.5%).

Diversity analysis: Initially, clustering of all the 647 germplasm accessions was done based on diversity analysis of 24 morphological descriptors. Characters namely leaf angle of insertion, leaf length, leaf width, leaf tip shape, leaf blistering of blade, leaf development of auricles, inflorescence shape, flower color of corolla and length of pistil relative to stamens were found to contribute much to the diversity between the four tobacco types (data not presented). Clustering through UPGMA produced groups defining four distinct clusters each in *bidi*, chewing and country cheroot germplasm with 0 to 4 distance and three clusters in cigar filler genotypes with 0-2 distance (Fig. 1). The genotypes falling in different clusters with larger distances may be having huge diversity and genotypes within the clusters/sub-clusters with less diversity. The maximum dissimilarity value in bidi was found between bidi entries- line 693-4-29-23 and Akkol (4.25), Cigar filler accessions-

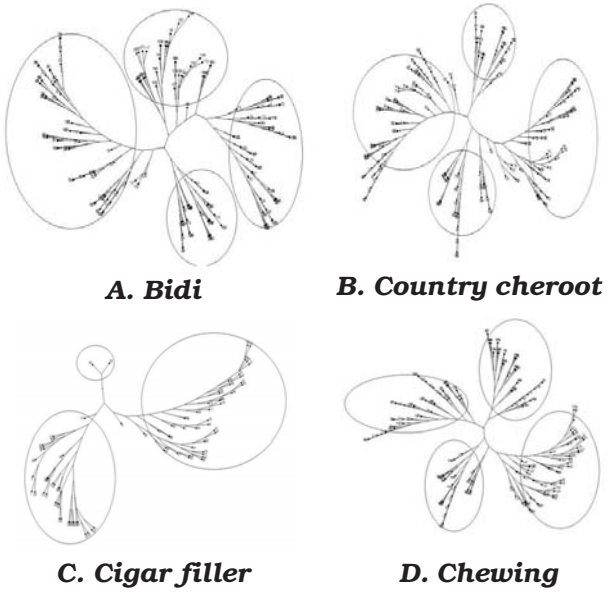

Fig. 1: Clustering pattern of *bidi*, country cheroot, cigar filler and chewing tobacco base collections using DarWIN. (Each germplasm entry is named with a prefix number).

Strain 608 and S-98 (2.15), Country cheroot lines- Natu Pothavithanam and II-1876 (4.728) and chewing genotypes- NP 63 and MRP-64 (3.8). Similarly, minimum distance was observed between chewing entries- B. Kalizong and Bandi Karhor (0.19), *bidi* entries- line 3-2-21-33-81 and ABD-65 (0.058), CF entries- Havana-381 and Havana-307 (0.14) and CC entries- NG-72 and NG-76 (0.11).

In the current study, comparable clustering patterns were observed in total germplasm base collection and its core indicating their similarities in diversity. Further, comparison of means of total collection and its core were statistically non-significant for all plant, leaf, floral and seed characters indicating that the variation available in the base collections is represented in core collection. Zhide *et al.,* (2001) in their study compared the traits mean, standard variance etc. among core collections and catalogued entire China tobacco germplasm and proved that the core collection well represented the total collection for genetic diversity. Yanhua *et al.,* (2008) studied the levels of genetic variation within 81 tobacco core collection including four populations (*N. rustica* L. 43 landrace, *N. tabacum* L. 13 landrace, 10 bred and 15 introduced lines), using sequence related amplified polymorphism (SRAP).

Core collection construction: Core collection for each sub-group was constructed carefully choosing each entry representing the variability available in the respective base collections. Based on the diversity, passport data and data generated over period of time at the Institute, about 10 percent of genotypes viz., 21 *bidi*, 19 CC, 8 CF and 17 chewing entries were selected to establish core collection in each the types. The 't' test was calculated each time to reject the entries contributing to core when there was significant difference between the means of germplasm and its core collection. All the accessions thus selected in each sub-group combined together to form a total core collection of 65 entries. Total base collection and its core found to demonstrate comparable clustering patterns indicating their similarities in diversity (Fig. 2). In the total core, minimum distance of 1.41 was observed between chewing entry, NP-24 and bidi entry, ABD-58 and maximum of 8.32 between chewing entry, B.KALIZONG and bidi entry, ABD-78.

Comparison of base and core collections: Mean values of all the morphological traits in the base collections of each sub-group and total germplasm was compared with their respective core collections using't' test (Table 2). Means of the most of the characters in Bidi, country cheroot, cigar filler, chewing and total germplasm base collections and their respective core collections were non-significant indicating that the variation in the each

A. Total tobacco germplasm base collection

B. Total core collection Note: Red color ones- *Bidi* entries, Pink-Chewing, Blue-CC, Green-CF

Fig. 2: Clustering pattern of total tobacco germplasm and its core collectionusing DarWIN. (Each line is named with a prefix number).

Table 2: Comparison of Mean (± standard error) values for morphological traits among the tobacco germplasm base and core collections

S. No.	Character	Bidi			Country Cheroot			Cigar Filler			Chewing			Total Tobacco Germplasm		
	Plant	Base	Core	t test Signi. ficance	Base	Core	t test Signi. ficance	Base	Core	t test Signi. ficance	Base	Core	t test Signi. ficance	Base	Core	t test Signi. ficance
1	Shape	1.805 ± 0.028	1.882 ± 0.144	NS	1.681 ± 0.034	1.474 ± 0.118	S	1.725 ± 0.050	1.5 ± 0.258	S	1.806 ± 0.030	1.786 ± 0.114	NS	1.760 ± 0.017	1.685 ± 0.063	NS
2	Main stem height	3.541 ± 0.092	3.118 ± 0.352	NS	3.886 ± 0.133	3.737 ± 0.381	NS	3.707 ± 0.171	3.4 ± 0.400	NS	3.126 ± 0.109	3.143 ± 0.390	NS	3.554 ± 0.061	3.364 ± 0.180	NS
3	Habit	2.044 ± 0.120	1.588 ± 0.435	NS	2.708 ± 0.136	3.632 ± 0.406	NS	2.634 ± 0.206	2.6 ± 0.980	NS	2.406 ± 0.132	2.286 ± 0.496	NS	2.406 ± 0.071	2.564 ± 0.253	NS
4	Internodal Length	1.744 ± 0.046	1.647 ± 0.161	NS	1.714 ± 0.043	1.842 ± 0.138	NS	1.829 ± 0.048	1.4 ± 0.245	NS	1.754 ± 0.038	1.500 ± 0.139	S	1.749 ± 0.023	1.655 ± 0.079	NS
5	Number of Leaves	4.063 ± 0.088	3.588 ± 0.228	S	4.719 ± 0.109	4.474 ± 0.208	NS	4.561 ± 0.167	4.2 ± 0.490	NS	4.120 ± 0.085	3.857 ± 0.275	NS	4.329 ± 0.053	4.018 ± 0.136	NS
	Leaf															
6	Type	1.015 ± 0.008	1.059 ± 0.059	NS	1.043 ± 0.015	1.000 ± 0.000	NS	1.024 ± 0.017	1.2 ± 0.200	NS	1.063 ± 0.018	1.071 ± 0.071	NS	1.037 ± 0.007	1.055 ± 0.031	NS
7	Angle of insertion	2.024 ± 0.043	2.059 ± 0.181	NS	1.865 ± 0.030	1.947 ± 0.093	NS	1.829 ± 0.073	1.8 ± 0.200	NS	1.909 ± 0.041	1.929 ± 0.165	NS	1.923 ± 0.022	1.964 ± 0.078	NS
8	Length	6.298 ± 0.090	6.059 ± 0.388	NS	5.454 ± 0.136	4.895 ± 0.358	NS	5.390 ± 0.132	5 ± 0.632	NS	5.160 ± 0.080	5.000 ± 0.296	NS	5.634 ± 0.058	5.291 ± 0.204	NS
9	Width	2.824 ± 0.076	2.882 ± 0.208	NS	2.200 ± 0.090	2.158 ± 0.279	NS	2.927 ± 0.107	3 ± 0.000	NS	2.691 ± 0.073	2.857 ± 0.143	NS	2.623 ± 0.044	2.636 ± 0.128	NS
10	Midrib	1.610 ± 0.043	1.412 ± 0.150	NS	1.546 ± 0.048	1.579 ± 0.176	NS	1.451 ± 0.060	1.4 ± 0.245	NS	1.383 ± 0.038	1.214 ± 0.155	NS	1.510 ± 0.023	1.418 ± 0.089	NS
11	Blade Shape	1.456 ± 0.056	1.294 ± 0.143	NS	1.573 ± 0.053	1.579 ± 0.159	NS	1.951 ± 0.107	2 ± 0.316	NS	1.816 ± 0.064	2.000 ± 0.277	NS	1.650 ± 0.033	1.636 ± 0.108	NS
12	Tip Shape	6.093 ± 0.085	6.412 ± 0.228	NS	6.535 ± 0.118	6.579 ± 0.361	NS	5.829 ± 0.139	5.4 ± 0.400	NS	5.839 ± 0.088	5.714 ± 0.339	NS	6.118 ± 0.053	6.200 ± 0.177	NS
13	Blistering of Blade	4.337 ± 0.074	4.412 ± 0.228	NS	3.789 ± 0.077	3.632 ± 0.267	NS	3.829 ± 0.120	3 ± 0.000	S	4.269 ± 0.095	4.000 ± 0.277	NS	4.097 ± 0.045	3.909 ± 0.145	NS
14	Undulation of Margin	4.063 ± 0.075	3.941 ± 0.303	NS	4.092 ± 0.084	4.053 ± 0.235	NS	4.122 ± 0.144	4.2 ± 0.800	NS	4.121 ± 0.080	4.000 ± 0.348	NS	4.095 ± 0.044	4.018 ± 0.163	NS
15	Auricles Development	3.966 ± 0.088	3.588 ± 0.285	NS	3.935 ± 0.092	3.842 ± 0.233	NS	3.439 ± 0.092	3.4 ± 0.400	NS	4.051 ± 0.082	4.286 ± 0.339	NS	3.913 ± 0.046	3.836 ± 0.153	NS
16	Colour of Blade	4.844 ± 0.036	5.000 ± 0.000	NS	4.837 ± 0.027	4.895 ± 0.072	NS	4.890 ± 0.043	5 ± 0.000	NS	4.800 ± 0.038	4.857 ± 0.143	NS	4.836 ± 0.018	4.927 ± 0.044	NS
17	Colour of Midrib	2.946 ± 0.016	3.000 ± 0.000	NS	3.000 ± 0.000	3.000 ± 0.000	NS	3.000 ± 0.000	3 ± 0.000	NS	3.000 ± 0.000	3.000 ± 0.000	NS	2.983 ± 0.005	3.000 ± 0.000	NS
	Inflorescence															
18	Shape	2.166 ± 0.096	2.235 ± 0.369	NS	2.249 ± 0.102	2.263 ± 0.304	NS	2.244 ± 0.137	1.6 ± 0.245	S	1.954 ± 0.096	2.214 ± 0.381	S	2.142 ± 0.053	2.182 ± 0.182	NS
19	Compactness	4.815 ± 0.098	4.529 ± 0.322	NS	4.838 ± 0.104	4.895 ± 0.323	NS	5.146 ± 0.154	5.4 ± 0.748	NS	4.726 ± 0.111	4.857 ± 0.490	NS	4.839 ± 0.056	4.818 ± 0.203	NS
	Flower															
20	Expression Corolla Tips	6.854 ± 0.077	7.000 ± 0.000	NS	6.967 ± 0.045	7.000 ± 0.000	NS	6.963 ± 0.096	7 ± 0.000	NS	6.793 ± 0.073	6.143 ± 0.404	NS	6.884 ± 0.036	6.782 ± 0.112	NS
21	Colour of Corolla	2.668 ± 0.080	2.529 ± 0.244	NS	2.897 ± 0.104	2.474 ± 0.221	NS	2.878 ± 0.159	3.2 ± 0.735	NS	3.103 ± 0.122	3.357 ± 0.476	NS	2.878 ± 0.055	2.782 ± 0.177	NS
22	Length of Pistil Relative to Stamens	1.887 ± 0.051	1.941 ± 0.181	NS	1.951 ± 0.053	1.842 ± 0.158	NS	2.244 ± 0.073	2.4 ± 0.245	NS	1.983 ± 0.050	2.214 ± 0.155	NS	1.977 ± 0.028	2.018 ± 0.092	NS
23	Fruit Form	3.420 ± 0.057	3.118 ± 0.241	NS	3.378 ± 0.057	3.368 ± 0.175	NS	3.512 ± 0.070	3.2 ± 0.583	NS	3.554 ± 0.042	3.643 ± 0.133	NS	3.456 ± 0.029	3.345 ± 0.114	NS
24	Testa Colour	1.210 ± 0.028	1.235 ± 0.106	NS	1.243 ± 0.033	1.263 ± 0.104	NS	1.171 ± 0.042	1 ± 0.000	NS	1.097 ± 0.024	1.071 ± 0.071	NS	1.184 ± 0.016	1.182 ± 0.052	NS

of germplasm groups was captured in their respective core collections. Nevertheless, for very few (1-2) characters viz., number of leaves for bidi, plant shape in CC, plant shape and inflorescence shape in CF and intermodal length and inflorescence shape in chewing tobacco the difference between cores and their base collections was significant. However, means of entire base collection and its core were statistically non-significant for all plant, leaf, floral and seed characters.

Shannon-Weaver diversity index ('H') was used to characterize and compare the diversity within total base collection and its core collection. The diversity index between both total germplasm and its core collection was almost equal for majority of the morphological traits except for flower expression of tips of corolla and seed testa colour (Table 3). The average H' within the entire set was 0.823 and within the core were 0.769 inferring that the 93% of the diversity in the entire germplasm was represented in the total core. Shannon-Weaver diversity index ('H') between both total collection and its core was almost equal for majority of the morphological traits in our study. Similar results were obtained in groundnut core collection where the difference in H' between the original and the core was negligible (Upadhyaya *et al.*, 2003). Non-significant p values in Chi-square analysis in the present study on frequency distribution of entries for 24 morphological characters among total germplasm and its core also proved that both the entries have similar distribution patterns.

The p value was found to be non-significant in the Chi-square analysis of total tobacco germplasm and its core collection on frequency

Table 3: Shannon-Weaver diversity index for different morphological descriptors in the total tobacco germplasm base collection and its core

S.No	Character	Total base collection	Core collection
1	Plant: shape (PS)	0.586	0.604
2	Plant: Height (PHT)	1.148	1.058
3	Plant: Habit (PH)	0.904	0.907
4	Plant: Internodal Length (PIL)	0.857	0.860
5	Leaf: Number of Leaves (LNL)	1.001	0.692
6	Leaf: Type (LT)	0.159	0.187
7	Leaf: Angle of Insertion (LAI)	0.821	0.763
8	Leaf: length (LL)	1.106	1.090
9	Leaf: Width of Blade (WB)	0.827	0.754
10	Leaf: Midrib (LM)	0.853	0.835
11	Leaf: Blade Shape (LBS)	1.061	1.016
12	Leaf: Tip Shape (LTS)	1.023	0.950
13	Leaf: Blistering of Blade (puckering) (LBB)	0.837	0.750
14	Leaf: Undulations of Margin (LUM)	0.832	0.876
15	Leaf: Development of Auricles (LDA)	0.863	0.810
16	Leaf: Color of Blade (LCB)	0.467	0.310
17	Leaf: Colour of Midrib (LCM)	0.086	0.000
18	Inflorescence: Shape (IS)	1.056	1.068
19	Inflorescence: Compactness (IC)	1.054	1.054
20	Flower: Expression of Tips of Corolla (FTC)	0.519	0.352
21	Flower: Colour of Corolla (FCC)	1.025	0.961
22	Flower: Length of Pistil Relative to Stamens (FLPS)	1.039	1.006
23	Fruit Form (FF)	0.981	1.032
24	Seed Testa Colour (TC)	0.654	0.521
	Average±SE	0.823±0.058	0.769±0.062

distribution of entries for 24 morphological characters (Table 4). Thus, all these studies clearly showed that the core collections (sub-group wise and total) constituted are the representative of the base collections and harboring the variation of source collections. Similarly, at ICRISAT, core collections capturing over 80% of variability in the entire collections of sorghum (3575 accessions; 2247 accessions) (Prasada *et al.,* 1995; Grenier *et al.,* 2001), pearl millet (1600 accessions; 2094 accessions) (Bhattacharjee *et al.,* 2007; Upadhyaya *et al.,* 2009a), chickpea (1956 accessions) (Upadhyaya *et al.,* 2001), groundnut (1704 accessions) (Upadhyaya *et al.,* 2003), pigeonpea (1290 accessions) (Reddy *et al.,* 2005), finger millet (622 accession) (Upadhyaya *et al.,*. 2006) and foxtail millet (155 accessions) (Upadhyaya *et al.,* 2009a) have been developed using passport information and characterization data generated over a period of time. Charmet and Balfourier (1995) and Bisht *et al.,* (1998) suggested that sizes of 5–10% were optimal, capturing 75–90% of the diversity.

However, these cores have to be revised periodically in order to add new genotypes which are acquired during the course of time (Perrier and Jacquemond-Collet, 2006). As the availability of time and resource to evaluate the entire germplasm each time to select a trait is limited, these core collections serve the purpose. The entries in the cores can be extensively evaluated considering its small size and the trait of interest could be studied without consuming huge resource. Tobacco being annual crop and the breeding programme takes more than ten years, evaluating core collection for target traits can thus be saved through. Hence, the constituted core collections can be utilized by tobacco researchers working in various disciplines

Table 4: Comparison of frequency distribution of entries among total tobacco germplasm base collection and its core for various morphological traits

S. No.	Characters	No. of Classes	Chi square	P value
1	Plant: shape (PS)	3	1.615	0.446
2	Plant: Height (PHT)	5	1.030	0.905
3	Plant: Habit (PH)	3	0.488	0.784
4	Plant: Internodal Length (PIL)	4	0.662	0.718
5	Plant: Number of Leaves (PNL)	5	6.798	0.147
6	Leaf: Type (LT)	2	0.151	0.697
7	Leaf: Angle of Insertion (LAI)	4	2.560	0.465
8	Leaf: length (LL)	5	2.220	0.695
9	Leaf: Width of Blade (WB)	4	1.630	0.653
10	Leaf: Midrib (LM)	3	4.676	0.097
11	Leaf: Blade Shape (LBS)	6	4.454	0.486
12	Leaf: Tip Shape (LTS)	4	0.842	0.839
13	Leaf: Blistering of Blade (puckering) (LBB)	5	2.845	0.584
14	Leaf: Undulations of Margin (LUM)	4	3.442	0.328
15	Leaf: Development of Auricles (LDA)	5	0.897	0.925
16	Leaf: Color of Blade (LCB)	4	1.721	0.632
17	Leaf: Color of Midrib (LCM)	2	1.131	0.288
18	Inflorescence: Shape (IS)	4	0.111	0.991
19	Inflorescence: Compactness (IC)	4	4.288	0.232
20	Flower: Expression of Tips of Corolla (FTC)	4	2.220	0.528
21	Flower: Colour of Corolla (FCC)	5	1.872	0.814
22	Flower: Length of Pistil Relative to Stamens (FLPS)	3	1.213	0.545
23	Fruit Form (FF)	4	3.532	0.317
24	Seed Testa Colour (TC)	3	4.207	0.122

in their studies for drawing valid conclusions on respective group of tobacco germplasm and total of them with minimum effort, time and resources. These constituted core collections are currently being genotyped using SSR markers and phenotyped for nutrient contents for further exploitation.

ACKNOWLEDGEMENT

The authors are thankful to ICAR-Central Tobacco Research Institute for providing the necessary facilities for undertaking the research presented in the article.

Reference

Agostino, F., B. Nicolas, D. C. Marcello, P. Pietro, D. Paolo, S. Alessandra, V.I. Nikolai and P. Carlo. 2012. Molecular diversity, population structure, and linkage disequilibrium in a worldwide collection of tobacco (*Nicotiana tabacum* L.) germplasm. **BMC Genet.** 13, 18. https://doi.org/ 10.1186/1471-2156-13-18.

Baghyalakshmi, K., K. Sarala, K. Prabhakara Rao, D. Damodar Reddy, C. C. S. Rao, P. Soni, P. Vinay, G. Kiran, and K. Santinandivelu. 2018. K Genetic Diversity in Tobacco Genotypes Evaluated under All India Network Project. **Multilogic in Sci.** 8: 30-34.

Bhattacharjee, R., P. J. Bramel, I. S. Khairwal and K. N. Reddy. 2007. Establishment of a pearl millet [*Pennisetum glaucum* (L) R. Br.] core collection based on geographical distribution and quantitative traits. **Euphytica** 155: 35-45.

Bisht, I. S., R. K. Mahajan, T. R. Lokknathan and R. C. Agrawal. 1998. Diversity in Indian sesame collection and stratification of germplasm accessions in different diversity groups. **Genet. Resour. Crop Evol.** 45: 325-335.

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

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

Clarke, K. R. and R. M. Warwick. 2001. *Changes in Marine Communities: An Approach to Statistical Analysis and Interpretation* (2nd ed.). Plymouth, PRIMER-E.

Dutta M., *et al.,* 2015. Development of Core Set of Wheat (*Triticum* spp.) Germplasm Conserved in the National Genebank in India. In: Ogihara Y., Takumi S., Handa H. (eds) *Advances in Wheat Genetics: From Genome to Field*. Springer, Tokyo. pp 33-45.

Erskine W. and F. J. Muehlbauer. 1991. Allozyme and morphological variability, outcrossing rate and core collection formation in lentil germplasm. **TAG** 83: 119-125.

Frankel, O. H. 1984. Genetic perspective of germplasm conservation. In: Arber, W., Llimensee, K.., Peacock, W. J., & Starlinger, P. (eds), *Genetic manipulations: impact on man and society*. Cambridge University Press, Cambridge, pp. 161–170.

Gouesnard, B., T. M. Bataillon, G. Decoux, C. Rozale, D. J. Schoen and J. L. David. 2001. MSTRAT: An Algorithm for Building Germ Plasm Core Collections by Maximizing Allelic or Phenotypic Richness. **J. Hered.** 92: 93–94. https://doi.org/10.1093/jhered/92.1.93.

Grenier, C., P. J. Bramel, and P. Hamon. 2001. Core collection of the genetic resources of sorghum: 1. Stratification based on eco-geographical data. **Crop Sci.** 41: 234-240.

Hu J., Zhu J. and H. Xu. 2000. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. **TAG** 101: 264–268. https://doi.org/10.1007/s001220051478.

Noirot, M., S. Hamon and F. Anthony. 1996. The principal component scoring: a new method of constituting a core collection using quantitative data. **Genet. Resour. Crop Evol.** 43: 1-6.

Perrier, X. and J. P. Jacquemond-Collet. 2006. *DARwin software*. http://darwin.cirad.fr/darwin.

Prasada R. K. E. and R. Ramanatha 1995. The use of characterization data in developing a core collection of sorghum. In. *Core collection of plant genetic resources* (Hodgkin T., Brown, A. H. D., van Hintum Th J. L., & Morales, B.A.V. (eds.). Chichester, UK: A Wiley-Sayee Publication Pages 109–116

Reddy, L. J., H. D. Upadhyaya, C. L. L. Gowda, and S. Singh. 2005. Development of core collection in pigeonpea [*Cajanus caja*n (L.) Millsp.] using geographic and qualitative morphological descriptors. **Genet. Resour. Crop Evol.** 52: 1049-1056.

Sarala, K., K. Baghyalakshmi, K. Prabhakara Rao, D. Damodar reddy, P. Vinay, G. Kiran and P.Sonia. 2018. Genetic diversity among mutant germplasm accessions of *Nicotiana Tabacum* as determined by morphological parameters. **Tob. Res.** 44: 47-53.

Sarala, K, K. Prabhakara Rao, K. Baghyalakshmi, G. Kiran and K. Sharavan kumar. 2019. Morphological diversity in burley tobacco germplasm. **Tob. Res.** 45: 39-47.

Schafleitner, R., M. N. Ramakrishnan, R. Abhishek, W. Yen-wei, L. Chen-yu, C. Shu-hui, L. Pin-yun, C. Jian-Cheng, and W.E. Andreas. 2015. The AVRDC – The World Vegetable Center mungbean (*Vigna radiata*) core and mini core collections. **BMC Genomics**, 16: 344. DOI 10.1186/s12864-015-1556-7.

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

Singh A. K. and S. N. Nigam. 1997. Groundnut. In: Fuccillo D., Sears L., Stapleton P. (eds). *Biodiversity in trust – conservation and use of plant genetic resources in CGIAR centers*. Cambridge Univ. Press, Cambridge, pp. 114–128.

Upadhyaya, H. D., P. J. Bramel and S. Singh. 2001. Development of a chickpea core collection using geographic distribution and quantitative traits. **Crop Sci.** 41: 206-210.

Upadhyaya, H. D., Y. Devvart, D. Naresh, C.L.L. Gowda and Singh S. 2010. Mini core germplasm collections for infusing genetic diversity in plant breeding programs. **EJPB.** 1(4): 1294-130.

Upadhyaya H. D., C. L. L. Gowda, R. P. S. Pundir, V. G. Reddy and S. Singh. 2006. Development of core subset of finger millet germplasm using geographical origin and data on 14 quantitative traits. **Genet. Resour. Crop Evol.** 53: 679-685.

Upadhyaya H. D., C. L. L. Gowda, K. N. Reddy and S. Singh. 2009a. Augmenting the pearl millet [Pennisetum glaucum (L.) R. Br.)] core collection for enhancing germplasm utilization in crop improvement. **Crop Sci.** 49: 573-580.

Upadhyaya H. D., V. Mani, M. A. Abdullah, C. R. A. Vania, C. S. Hari, S. Rajan, P. S. Suraj and Yi-Hong W. 2019. Multi-Trait Diverse Germplasm Sources from Mini Core Collection for Sorghum Improvement. **Agriculture**. https/ doi/10.3390/agriculture9060121.

Upadhyaya H. D., R. Ortiz, P. J. Bramel 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., R. P. S. Pundir, C. L. L. Gowda, V. G. Reddy and Singh S. 2009b. Establishing a core collection of foxtail millet to enhance utilization of germplasm of an underutilized crop. **Plant Genet. Resour.** https/doi/ 10.1017/S1479262108178042.

Van Hintum Th J. L., A. H. D. Brown, C. Spillane and Hodgkin T. 2000. Core collection of Plant Genetic Resources. IPGRI Technical Bulletin No. 3. Pg 48.

Yanhua L., W. Zhide, M. U. Jianmin, L. Jianfeng

and Peigang, D. 2008. Study on genetic diversity of part of tobacco core collection. CORESTA Congress, Shanghai, 2008, APPOST 10.

Zhang H., R. Bai, F. Wu, G. Wenli, Y. Zhuanzhuan, Y. Qi, Z. Yufei, M. Jinxing and J. Zhang. 2019. Genetic diversity, phylogenetic structure and development of core collections in *Melilotus* accessions from a Chinese gene bank. **Sci Rep.** https:// doi.org/10.1038/s41598-019-49355-y.

Zhide W., J. Yuen, D. Peigang and Jianmin M. 2001. Establishment and evaluation of tobacco core germplasm in China. CORESTA Meeting, Agronomy/Phytopathology, Cape Town, 2001, APPOST 04