CSIRO PUBLISHING
www.publish.csiro.au/journals/cp *Crop & Pasture Science*, 2011, **62**, 383–389

Developing proso millet (*Panicum miliaceum* L.) core collection using geographic and morpho-agronomic data

H. D. Upadhyaya[A,B], *Shivali Sharma*[A], *C. L. L. Gowda*[A], *V. Gopal Reddy*[A], *and Sube Singh*[A]

[A]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad, PO 502324, AP, India.
[B]Corresponding author. Email: h.upadhyaya@cgiar.org

Abstract. Proso millet (*Panicum miliaceum* L.) is a rich source of protein, minerals, and vitamins, and is an important cereal crop of Asia and Africa. Due to its lowest water and nutrient requirement, it has the potential for agriculture diversification. The development of a core collection would assist in efficient management and enhanced utilisation of proso millet genetic resources. The present investigation was conducted to develop a core collection of proso millet based on geographic information and 20 qualitative and quantitative traits recorded on 833 accessions conserved in the International Crops Research Institute for the Semi-Arid Tropics genebank. The entire germplasm collection was stratified into five groups based on races and data on 20 morpho-agronomic traits were used for clustering following Ward's method. About 10% (or at least one accession) was randomly selected from each of 101 clusters to constitute a core collection of 106 accessions. Comparisons of means, variances, frequency distribution, diversity indices, and correlation studies indicated that the variation in the entire collection has been preserved in the core collection. This core collection provides a gateway to identify diverse trait-specific germplasm accessions for important agronomic traits and for abiotic and biotic stresses for use in crop improvement research and in crop diversification programs.

Additional keywords: core collection, entire collection, genebank, mini core collection, proso millet, trait-specific germplasm accessions.

Introduction

Proso millet (*Panicum miliaceum* L.) is the oldest used cereal by humans besides wheat and barley. It was domesticated in Manchuria and introduced to Europe ~3000 years ago, followed by its introduction in the Near East and India. Because of its short growing seasons (60–90 days after planting), low water and nutrient requirements, it grows across wide environments up to 54° N/S latitude and also adapts well to plateau conditions and high elevations (Theisen *et al*. 1978; Matz 1986). Proso millet is found in high mountains in the former USSR up to 1200 m a.s.l. and in India up to 3500 m a.s.l. (Roshevits 1980). It has been grown in many areas of the world, including Russia, China, Romania, Afghanistan, Turkey, and India as a human food source (Boland 2003). The cultivation of proso millet can play an important role in the economy of many less-developed countries of the Old World because proso millet is consumed directly for human food. FAO data on area, production, and productivity of all the millets are given together under the general heading of millets and therefore exact statistical data about proso millet cultivation and production are unavailable. Proso millet is essentially a crop of temperate region, but is also grown in the subtropics and on high ground in tropical winters. Under drought and poor soil conditions, proso millet gives a yield which surpasses the yield of all other crops and has the potential to produce food where other grain crops would fail (Magness *et al*. 1971; Jiaju 1986). Thus, proso millet cultivation would contribute to crop diversifications besides better use of land, diet diversification, and better economic potential. Hence, genetic improvement of proso millet needs to be taken on priority at more institutions.

Proso millet is rich in protein, minerals, vitamins and micro-nutrients such as iron, zinc, copper and manganese and its nutritive parameters are comparable or better than common cereals (Kalinova and Moudry 2006). Proso millet is considered a self-pollinated crop, but natural cross-pollination may exceed 10%. On the basis of inflorescence morphology, the species *Panicum miliaceum* consists of subspecies *miliaceum*, which is subdivided into five races Miliaceum, Patentissimum, Contractum, Compactum and Ovatum.

The success of any breeding program depends upon the availability and exploitation of genetic diversity of crop plants for the traits of economic importance. At present, ~7.4 million accessions of different crops are stored in ~1750 genebanks (FAO 2010) around the world. Most of this genetic diversity (~80%) belongs to major crops and their relatives. Keeping in view the importance of proso millet in agricultural diversification, the genebank at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) has assembled and conserved 833 proso millet germplasm accessions from 30 countries for use in research and development programs globally. The effective

10.1071/CP10294 1836-0947/11/050383

utilisation of these germplasm accessions depends upon the identification of diverse trait-specific germplasm accessions. Therefore, these accessions need to be evaluated for important agronomic traits exhibiting genotype × environment interactions across multi-locations to identify trait-specific germplasm lines. However, evaluating the entire collection of this size is cumbersome, time consuming and resource demanding. In order to utilise and manage the germplasm collection more effectively and easily, Frankel (1984) proposed the concept of core collection. A core collection is defined as a representative sample of the entire collection with minimum repetitiveness and rich genetic diversity of a crop. The core collection, usually 10% of the entire collection, has a reduced size and can be evaluated extensively at a relatively low cost across multi-locations. Further, the information derived could be used as a guide towards more efficient utilisation of the entire collection (Brown 1989).

For developing core collections, several strategies have been proposed: constant (C) sampling strategy, proportional (P) sampling strategy and logarithmic (L) sampling strategy (Brown 1989). In C strategy, equal number of accessions are chosen from each cluster irrespective of cluster size to constitute core collection. In P strategy, several accessions proportional to the cluster size (usually 10%) is taken, whereas in L strategy, a number of accessions proportional to the logarithm of the cluster size is selected. The least distance stepwise sampling (LDSS) strategy was proposed for constructing core collection based on genotypic values (Wang *et al*. 2007). In LDSS strategy, the genetic distances between the accessions are calculated, which are used for grouping the accessions following hierarchical clustering. Stepwise sampling is performed to obtain a desired percentage of sampling. Core collection constituted using LDSS are representative of the entire collection (Wang *et al*. 2007). The choice of sampling strategy depends greatly on the objective of the research. In the situations where collections are large and the main aim is to manage collection efficiently and have a set representing diversity of collection, P strategy has been suggested as an optimum strategy (Grenier *et al*. 2001*a*). The main advantage of P strategy is that it is more efficient, includes more alleles, and often has lower variance (Cochran 1977). Brown (1989) also recommended P strategy in case of undifferentiated loci. The P strategy has been used in several crops to constitute core collections (Holbrook *et al*. 1993; Rao and Rao 1995; Ortiz *et al*. 1998; Upadhyaya *et al*. 2009*a*). In sorghum, Grenier *et al*. (2001*a*) found that core collection based on P sampling gave a better picture of the diversity of the entire collection and represented the landrace collection adequately in comparison to the core based on C or L strategies.

The core collections of important cereal crops such as wheat (Spagnoletti Zeuli and Qualset 1993; Martynov *et al*. 2003), rice (Li *et al*. 2002), maize (Li *et al*. 2005; Coimbra *et al*. 2009), barley (Knupffer and van Hintum Th 2003; Chabane and Valkoun 2004), cotton (Wang *et al*. 2007), pearl millet (Bhattacharjee *et al*. 2007; Upadhyaya *et al*. 2009*a*), sorghum (Grenier *et al*. 2001*b*; Dahlberg *et al*. 2004), finger millet (Upadhyaya *et al*. 2006) and foxtail millet (Upadhyaya *et al*. 2008) have already been developed. The present investigation was carried out to constitute a core collection of proso millet representing diversity of the entire collection for use by breeders and plant scientists in crop improvement.

Materials and methods

Plant material

The genebank at ICRISAT, Patancheru, India holds 833 accessions of the cultivated proso millet originating from 30 countries. Of these, 821 accessions were assembled through donations from 13 institutes in nine countries and 12 accessions were collected from three countries during collection missions. These accessions assembled in the ICRISAT genebank since 1977 were characterised for nine qualitative and 11 quantitative traits in eight lots. The characterisation site, Patancheru is located at 18°N and 78°E, at an altitude of 545 m a.s.l. The average annual rainfall at this location is ~850 mm, which normally occurs during June to September. The accessions were grown in Alfisols, in a single row of 4 m length, with inter-row spacing of 75 cm, and the plant spacing of 10 cm. Diammonium phosphate was applied at the rate of 100 kg ha^{-1} as a basal dose to supply nitrogen and phosphorus. In addition, 100 kg ha^{-1} of urea was applied as top dressing. Sowing was done in the last week of July. All cultural practices and data recordings were similar for all the years of characterisation during this period.

Data recording

The data were recorded on nine qualitative and 11 quantitative traits following the descriptors of *Panicum miliaceum* and *Panicum sumatrense* (IBPGR 1985). Of these, qualitative traits are not affected by environments, whereas the quantitative traits are highly influenced by environment and show genotype × environment interactions. The data on all nine qualitative traits (growth habit, culm branching, sheath pubescence, ligule pubescence, leaf pubescence, inflorescence shape, fruit colour, apiculus colour and overall plant aspect) and one quantitative trait (days to 50% flowering) were recorded on plot basis. Data on basal tiller number were taken on five representative plants of the plot. Data for the remaining nine quantitative traits [plant height (cm), flag leaf blade length (mm), flag leaf blade width (mm), flag leaf sheath length (mm), peduncle length (mm), panicle exsertion (mm), inflorescence length (mm), number of nodes, and inflorescence primary branches number] were recorded on main culms of the five representative plants of the plot.

Statistical analysis and constitution of core collection

The entire germplasm collection of proso millet was stratified into five groups based on races: Compactum (98 accessions), Contractum (92 accessions), Miliaceum (539 accessions), Ovatum (48 accessions), and Patentissimum (56 accessions). The data on 11 quantitative traits in all five groups were standardised using the range of each variable to eliminate scale differences (Milligan and Cooper 1985). Data on qualitative traits were transformed to numerical scale (IBPGR 1985) to calculate phenotypic distance matrix. For simultaneous use of qualitative and quantitative traits in estimating genetic diversity, Gower's distance measure (Gower 1971) was used due to its ability to accommodate mixed data types. A phenotypic distance matrix

was created using data on nine qualitative and 11 quantitative traits following this distance measure. This distance matrix was subjected to hierarchical cluster algorithm (Ward 1963) at an R^2 (squared multiple correlation value) of 0.75 for clustering of accessions in each five groups, separately. This method optimises an objective function because it minimises the sum of squares within groups and maximises the sum of squares between groups. The proportional sampling strategy was used, and from each cluster ~10% of the accessions were randomly selected to constitute the core subset. At least one accession was included from each cluster having less than 10 accessions.

Validation of core collection

The data on geographic origin and nine qualitative and 11 quantitative traits were used to validate the core collection composition. The 30 countries of origin were grouped into 10 regions (Africa, East Asia, Europe, Mediterranean, North America, Oceania, Russia and CIS, South Asia, South-East Asia and West Asia). The information on country of origin was not available for 358 accessions. They were grouped as 'Unknown' region. Frequencies of geographic regions, countries within regions, races and qualitative traits in the entire and core collections were tested by the χ^2 test. The expected frequencies of the accessions in different classes of a trait in the core collection were based on the proportion of core to entire collection. The expected frequencies were tested against observed frequencies in the core collection for goodness of fit using χ^2 test. Yates' correction (Yates 1934) was used when the number of accessions in the entire collection was less than five in a class. The means of entire and the constituted core collection were compared using the Newman–Keuls procedure (Newman 1939; Keuls 1952) for all 11 quantitative traits. The homogeneity of variance between the entire and core collection was tested by Levene's test (Levene 1960). The percentage of the significant difference between the core and entire collections was calculated for the mean difference percentage (*MD*%) and the variance difference percentage (*VD*%) of traits. The coincidence rate (*CR*%) and the variable rate (*VR*%) were calculated to compare the entire and core collection following Hu *et al*. (2000).

$$CR\% = \frac{1}{m}\sum_{j=1}^{m}\frac{R_C}{R_E} \times 100$$

$$VR\% = \frac{1}{m}\sum_{j=1}^{m}\frac{CV_C}{CV_E} \times 100$$

where R_C is range of core collection, R_E range of entire collection, CV_C coefficient of variation of core collection, CV_E coefficient of variation of entire collection, and m is the number of traits.

The diversity index (H') of Shannon and Weaver (1949) was used as a measure of phenotypic diversity of each trait and was calculated as follows:

$$H' = -\sum p_i \ln(p_i)$$

where H' is the Shannon–Weaver Diversity Index, and p_i is the relative abundance of each group of organisms. The index was calculated separately in both entire and core collections to determine whether the diversity for each trait was retained in the core collection. Phenotypic correlations among 11 quantitative traits in the entire and core collections were estimated separately to determine whether these associations, which may be under the same genetic control, were conserved in the constituted core collection.

Results

Constitution of core collection

The hierarchical cluster analysis resulted in grouping the 833 accessions in the entire collection into 101 clusters. Race-wise, the number of clusters was 11 (98 accessions) in Compactum, 13 (92 accessions) in Contractum, 61 (539 accessions) in Miliaceum, 8 (48 accessions) in Ovatum and 8 (56 accessions) in Patentissimum. The number of accessions in clusters ranged between 3 and 46. From each cluster, 10% or at least one accession was included in the core collection resulting in a selection of 106 accessions, which is 12.7% of the entire collection.

Geographic distribution

The entire collection of proso millet in the ICRISAT genebank (833 accessions) was represented by 30 countries belonging to 10 regions spread over Africa, Asia, Europe and North America. About 40% accessions (358) in the entire collection had no information about their origin. The maximum number of accessions were from Russia and CIS (14.5%) followed by South Asia (13.5%), Mediterranean region (10.2%) and East Asia (9.1%). Africa, Oceania and South-East Asia were represented only by 0.2% accessions from each region in the entire collection (Table 1). The composition of core collection reflected the prevalence of accessions from East Asia (13.2%) followed by South Asia (12.3%), Russia and CIS (11.3%) and the Mediterranean (10.4%). About 5.7% accessions in the core collection were from Europe and West Asia, 3.8% from North America, and ~1% each from Africa, Oceania and South-East Asia. The χ^2-value and the heterogeneity were non-significant for all the regions. The countries within the region and the unknown group also showed non-significant χ^2-values, indicating that all the countries were optimally represented in the core collection (Table 1). Hence, the proportion of accessions in the entire versus core collection compared favourably across all 10 regions and 30 countries (Table 1).

Racial pattern

The racial pattern of the entire collection reflected the prevalence of race Miliaceum (64.7% accessions) followed by Compactum (11.8%), Contractum (11%), Patentissimum (6.7%), and Ovatum (5.8%). In the core collection, race Miliaceum was the most common (61.3%) followed by Contractum (12.3%), Compactum (11.3%), Ovatum and Patentissimum (each 7.6%). The χ^2-values for the frequency distribution of accessions in entire and core collections were non-significant for all the five races ($\chi^2 = 0.018$, 0.143, 0.188, 0.586, and 0.107, respectively) (Table 2), revealing that the sampling technique to constitute the core collection was effective and adequate and the constituted core collection is the best representative of the entire collection.

Table 1. Chi-square test for the frequency distribution of proso millet entire and core collection accessions in different regions and countries within region

Region/country	Entire	Core	d.f.	χ^2	*P*
Africa	2	1	1	0.237	0.627
Kenya	1	0	1	2.000	0.157
Malawi	1	1	1	0.000	1.000
Heterogeneity			1	1.763	0.184
East Asia	76	14	1	1.938	0.164
China	2	1	1	0.047	0.828
Japan	1	1	1	0.541	0.462
Republic of Korea	73	12	1	0.156	0.693
Heterogeneity			2	1.194	1.000
Europe	28	6	1	1.667	0.197
Germany	12	2	1	0.127	0.722
Hungary	10	1	1	0.610	0.435
United Kingdom	4	1	1	0.149	0.700
Yugoslavia (former)	1	1	1	0.381	0.537
Romania	1	1	1	0.381	0.537
Heterogeneity			4	0.020	1.000
Mediterranean	85	11	1	0.003	0.956
Lebanon	1	0	1	3.061	0.080
Spain	1	1	1	1.061	0.303
Syria	34	5	1	0.082	0.775
Turkey	49	5	1	0.284	0.594
Heterogeneity			3	4.485	0.214
North America	14	4	1	2.763	0.097
Canada	1	1	1	0.161	0.689
Mexico	13	3	1	0.137	0.711
Heterogeneity			1	2.465	1.000
Oceania	2	1	1	0.237	0.627
Russia and CIS	121	12	1	0.750	0.387
South Asia	112	13	1	0.110	0.740
Bangladesh	2	0	1	2.309	0.129
India	68	7	1	0.101	0.751
Nepal	6	1	1	0.132	0.716
Pakistan	34	4	1	0.001	0.979
Sri Lanka	2	1	1	0.309	0.578
Heterogeneity			4	2.742	0.602
South-East Asia	2	1	1	0.237	0.627
Argentina	1	1	1	0.000	1.000
Kyrgyzstan	1	0	1	2.000	0.157
Heterogeneity			1	1.763	0.184
Unknown	358	37	1	1.607	0.205
West Asia	33	6	1	0.772	0.380
Afghanistan	16	2	1	0.284	0.594
Iran	9	1	1	0.247	0.619
Iraq	2	1	1	0.051	0.821
Kazakhstan	2	1	1	0.051	0.821
Ukraine	4	1	1	0.071	0.790
Heterogeneity			4	0.067	1.000
Overall	833	106		–	–

Table 2. Chi-square test for the frequency distribution of proso millet entire and core collection accessions in different races

Race	Entire	Core	d.f.	χ^2	*P*
Compactum	98	12	1	0.018	0.894
Contractum	92	13	1	0.143	0.706
Miliaceum	539	65	1	0.188	0.665
Ovatum	48	8	1	0.586	0.444
Patentissimum	56	8	1	0.107	0.743

Qualitative traits

The frequency distribution of classes in all nine qualitative traits using χ^2 analysis indicated homogeneity of distribution for all traits, thereby indicating that the sampling technique to constitute the core collection was adequate and appropriate (Table 3).

Quantitative traits

A comparison of the range for 11 quantitative traits across entire and core collections showed that 100% of the range available in the entire collection was included in the core collection for seven traits (basal tiller number, flag leaf sheath length, peduncle length, panicle exsertion, inflorescence length, number of nodes and inflorescence primary branches number) in the core collection (Table 4). For the remaining four traits (days to flowering, plant height, flag leaf blade length and width), the range included in the core collection varied from 91.7 to 98.3%. Differences between the means of the entire and core collections were non-significant for all the quantitative traits (Table 4). Variance of the entire and core collections were homogeneous for all traits. Phenotypic correlations were conducted between all 11 quantitative traits in the entire and core collections separately. The pattern of correlations was similar in the entire and core collections, demonstrating that associations observed in the entire collection were well preserved in the core collection. Only those traits with correlation coefficients greater than 0.707 and less than –0.707 are considered as biologically meaningful (Skinner *et al*. 1999) as more than 50% of the variation in one trait is predicted by the other (Snedecor and Cochran 1980). In our study, two such meaningful relationships between plant height and inflorescence length and panicle exsertion and peduncle length were found in the entire collection (Table 5). Both the relationships were retained in the core collection. In addition, the relationship between inflorescence length and flag leaf sheath length was also observed in the core collection ($r = 0.740$) and entire collection ($r = 0.604$). These relationships suggest that it is not necessary to measure both the related traits in future germplasm evaluations, and only easily measurable traits should be given priority.

Diversity in entire and core collection

The Shannon–Weaver diversity index (H') was calculated to compare diversity in the entire and core collections. The index

Table 3. Chi-square values and probability for frequency distribution of classes in nine qualitative traits in proso millet entire and core collection accessions

Traits	d.f.	χ^2	*P*
Growth habit	2	1.909	0.385
Culm branching	2	1.917	0.384
Sheath pubescence	2	2.649	0.266
Ligule pubescence	2	0.836	0.658
Leaf pubescence	2	0.089	0.956
Inflorescence shape	7	5.065	0.652
Fruit colour	10	9.829	0.456
Apiculus colour	1	0.071	0.791
Overall plant aspect	3	4.531	0.210

Table 4. Range, means and variances in entire and constituted core collections of proso millet

Traits	Range		Mean[A]		Variance[B]			
	Entire	Core	Entire	Core	Entire	Core	*F*-value	*F* probability
Days to flowering	26–50	28–50	34.5	34.9	11.9	14.0	0.68	0.409
Plant height (cm)	20–133	25–133	59.4	61.4	309.5	393.4	1.55	0.214
Basal tillers number	1–32	1–32	4.0	4.1	6.3	10.3	0.88	0.349
Flag leaf blade length (mm)	80–380	85–380	222.7	219.3	2867.3	3657.4	3.28	0.070
Flag leaf blade width (mm)	6–30	8–30	19.5	18.8	49.8	35.6	<0.01	0.984
Flag leaf sheath length (mm)	30–170	30–170	82.1	80.6	278.1	267.2	0.08	0.784
Peduncle length (mm)	15–400	15–400	181.3	179.7	4050.3	4605.4	0.64	0.422
Panicle exertion (mm)	0–320	0–320	100.1	102.8	3570.6	3881.4	0.25	0.619
Inflorescence length (mm)	22–400	22–400	193.1	193.8	3404.2	4222.9	2.34	0.127
Number of nodes	2–90	2–90	11.4	11.0	42.5	75.8	1.40	0.237
Inflorescence primary branches number	5–29	5–29	16.1	15.8	17.1	20.1	0.53	0.465

[A]Means compared using Newman–Keul test and were non-significant for all traits.
[B]Homogeneity of variances tested using Levene's procedure.

Table 5. Phenotypic correlation coefficients between 11 quantitative traits in the proso millet entire and constituted core collections
DTF, days to flowering; PLHT, plant height; BTN, basal tillers number; FLBL, flag leaf blade length; FLBW, flag leaf blade width; FLSL, flag leaf sheath length; PL, peduncle length; PAEX, panicle exertion; INFL, inflorescence length; NN, number of nodes; INFPBN, inflorescence primary branches number.
$*P \leq 0.05$; $**P \leq 0.01$; values in bold are >0.707 and <–0.707

Traits	Collection	DTF	PLHT	BTN	FLBL	FLBW	FLSL	PL	PAEX	INFL	NN
PLHT	Core	0.200*									
	Entire	0.261**									
BTN	Core	0.325**	0.296**								
	Entire	0.226**	0.261**								
FLBL	Core	0.078	0.634**	0.306**							
	Entire	0.123**	0.655**	0.324**							
FLBW	Core	–0.260**	0.286**	–0.006	0.562**						
	Entire	–0.261**	0.267**	–0.049	0.521**						
FLSL	Core	0.101	0.649**	0.059	0.586**	0.329**					
	Entire	0.038	0.600**	0.039	0.554**	0.414**					
PL	Core	–0.452**	0.368**	–0.181*	0.071	0.251**	0.275**				
	Entire	–0.354**	0.377**	–0.192**	0.083**	0.261**	0.341**				
PAEX	Core	–0.413**	0.196*	–0.243**	–0.128	0.107	0.016	**0.896****			
	Entire	–0.368**	0.219**	–0.220**	–0.081**	0.149**	0.086**	**0.937****			
INFL	Core	0.238**	**0.796****	0.250**	0.655**	0.128	**0.740****	0.215*	0.020		
	Entire	0.330**	**0.802****	0.280**	0.639**	0.107**	0.604**	0.147**	–0.025		
NN	Core	–0.095	0.263**	0.100	0.305**	0.231**	0.225**	–0.009	–0.122	0.223*	
	Entire	–0.059*	0.225**	0.125**	0.346**	0.268**	0.184**	–0.052	–0.122**	0.225**	
INFPBN	Core	–0.091	0.649**	0.169*	0.583**	0.377**	0.539**	0.213*	–0.009	0.617**	0.341**
	Entire	0.027	0.492**	0.150**	0.549**	0.389**	0.352**	–0.006	–0.128**	0.445**	0.394**

is used in genetic studies as a convenient measure of both allelic richness and allelic evenness. A low H' indicates an extremely unbalanced frequency of classes for an individual trait and a lack of genetic diversity. The average H' for the seven qualitative and 11 quantitative traits and a high H' for two qualitative traits in the core collection was similar to the entire collection (Table 6) indicating that the diversity of the entire collection for qualitative and quantitative traits was adequately represented in the core collection.

Coincidence rate and variable rate

The coefficients of variations or variable rate for most of the traits were higher in the core collection than in the entire collection, resulting in 109.97% *VR* for quantitative traits (Table 6). The variances and coefficients of variation in the selected collection should be higher than in the entire collection (Hu *et al*. 2000). High range (91.7–100%) variation or *CR*% was captured for all quantitative traits in the core collection. The higher *CR*% captured for quantitative (97.93%) traits (Table 6) in the core collection confirmed that the core collection is the best representative of entire collection.

Discussion

The availability of sufficient variability is a prerequisite for any crop improvement program. Keeping this in view, plant genetic resources comprising landraces, obsolete varieties, advanced breeding lines, and wild relatives of crop species are being conserved in various genebanks. For proso millet, 833

Table 6. Shannon–Weaver diversity (*H'*) index, coincidence rate (*CR*%) and variable rate (*VR*%) in proso millet entire and constituted core collection

Traits	*VR*%	*CR*%	*H'*	
			Entire	Core
Growth habit	–	–	0.315	0.355
Culm branching	–	–	0.453	0.439
Sheath pubescence	–	–	0.472	0.471
Ligule pubescence	–	–	0.437	0.435
Leaf pubescence	–	–	0.426	0.422
Inflorescence shape	–	–	0.769	0.807
Fruit colour	–	–	0.717	0.733
Apiculus colour	–	–	0.286	0.283
Overall plant aspect	–	–	0.415	0.394
Mean	–	–	0.477	0.482
s.e.±	–	–	0.0545	0.0577
Days to flowering	106.9	91.7	0.586	0.600
Plant height (cm)	109.2	95.6	0.576	0.559
Basal tillers number	122.2	100.0	0.424	0.306
Flag leaf blade length (mm)	114.7	98.3	0.618	0.604
Flag leaf blade width (mm)	87.6	91.7	0.607	0.529
Flag leaf sheath length (mm)	99.8	100.0	0.601	0.624
Peduncle length (mm)	107.6	100.0	0.622	0.639
Panicle exertion (mm)	101.5	100.0	0.599	0.585
Inflorescence length (mm)	111.0	100.0	0.629	0.623
Number of nodes	138.4	100.0	0.337	0.319
Inflorescence primary branches number	110.7	100.0	0.649	0.651
Mean	109.97	97.93	0.568	0.549
s.e.±	3.904	1.017	0.0292	0.368

germplasm accessions from 30 countries have been conserved in the ICRISAT genebank. However, effective utilisation of this diversity depends upon the identification of diverse trait-specific germplasm accessions. Extensive evaluation of the entire germplasm collection consisting of 833 accessions is expensive and uneconomical particularly in view of limited crop improvement efforts in this crop. Therefore, the constituted core collection, consisting of 106 accessions (12.7% of the entire collection) representing the total diversity of the entire collection can be used very efficiently for the identification of new sources of diversity for important traits following extensive multi-location evaluation. The accessions for core collection were selected following clustering of five racial groups in the entire collection using hierarchical cluster analysis. Selection of 10% accessions from each cluster or at least one accession from the cluster having less than 10 accessions resulted in a selection of 106 accessions.

A good core collection should represent diversity of entire collection. In the present study, the validation of this core collection using different measures such as data on geographic origin and nine qualitative and 11 quantitative traits revealed that the constituted core collection is the representative of the entire collection. The accessions from different countries and regions as well as the distribution of five races in the entire collection were well represented in the core collection. The non-significant differences between the means of entire and core collection, homogeneity of variance for all 11 quantitative traits, Shannon–Weaver diversity index, *CR*% and *VR*% also suggest that the sampling technique to constitute core collection was appropriate and the core collection has captured adequate diversity from the entire collection. Ortiz *et al*. (1998) advocated that an adequate and proper sampling, essential in developing a representative core collection, should consider the conservation of phenotypic associations arising out of co-adapted gene complexes. In the present study, all the associations observed in the entire collection were retained in the core collection indicating the effectiveness of the technique in constituting the core collection. Therefore, the entire collection is well represented by the constituted core collection.

Development of core collections with reduced size will not only assist in the more efficient conservation and management of crop genetic resources but also pave a way for their precise evaluation using replications and multi-locations to identify trait-specific germplasm accessions for the traits of economic importance that show high genotype × environment interactions. Such evaluations have already proved to be very useful in identifying new sources of resistance or tolerance to biotic and abiotic stresses and variation for agronomic and quality traits in chickpea, pigeonpea, groundnut, sorghum, pearl millet, foxtail millet and finger millet (summarised in Upadhyaya *et al*. 2009*b*). In a crop like proso millet in which breeding efforts are in their infancy, systematic evaluation of this core collection should prove very useful in selecting promising lines for large-scale evaluation and release as cultivars. Seeds of the proso millet core collection are available for research from ICRISAT genetic resources using the Standard Material Transfer Agreement.

References

Bhattacharjee R, Bramel PJ, Khairwal IS, Reddy KN (2007) Establishment of a pearl millet [*Pennisetum glaucum* (L.) R. Br.] core collection based on geographical distribution and quantitative traits. *Euphytica* **155**, 35–45. doi:10.1007/s10681-006-9298-x

Boland M (2003) Proso millet. Available at: http://test.agmrc.org/agmrc/commodity/grainsoilseeds/prosomillet [Retrieved 6 November 2005].

Brown AHD (1989) Core collections: a practical approach to genetic resources management. *Genome* **31**, 818–824. doi:10.1139/g89-144

Chabane JCK, Valkoun J (2004) Characterization of genetic diversity in ICARDA core collection of cultivated barley (*Hordeum vulgare* L.). *Czech Journal of Genetics and Plant Breeding* **40**, 134–136.

Cochran WG (1977) 'Sampling techniques.' 3rd edn (John Wiley & Sons: New York)

Coimbra R, Miranda GV, Cruz CD, Silva DJH, Vilela RA (2009) Development of a Brazilian maize core collection. *Genetics and Molecular Biology* **32**, 538–545. doi:10.1590/S1415-47572009005000059

Dahlberg JA, Burkee JJ, Rosenow DT (2004) Development of a sorghum core collection: refinement and evaluation of a subset from Sudan. *Economic Botany* **58**, 556–567. doi:10.1663/0013-0001(2004)058[0556:DOASCC]2.0.CO;2

FAO (Food and Agriculture Organization of the United Nations) (2010) The state of *ex-situ* conservation. In 'The second state of the world's plant genetic resources for food and agriculture'. pp. 52–90. (FAO: Rome) Available at: www.fao.org/agriculture/seed/sow2/en

Frankel OH (1984) Genetic perspective of germplasm conservation. In 'Genetic manipulations: impact on man and society'. (Eds W Arber, K Limensee, WJ Peacock, P Stralinger) pp. 161–170. (Cambridge University Press: Cambridge, UK)

Gower JC (1971) A general coefficient of similarity and some of its properties. *Biometrics* **27**, 857–874. doi:10.2307/2528823

Grenier C, Bramel PJ, Hamon P (2001*b*) Core collection of the genetic resources of sorghum: 1. Stratification based on eco-geographical data. *Crop Science* **41**, 234–240. doi:10.2135/cropsci2001.411234x

Grenier C, Hamon P, Bramel PJ (2001*a*) Core collection of the genetic resources of sorghum. II. Comparison of three random sampling procedures. *Crop Science* **41**, 241–246. doi:10.2135/cropsci2001.411241x

Holbrook CC, Anderson WF, Pittman RN (1993) Selection of a core collection from the US germplasm collection of peanut. *Crop Science* **33**, 859–861. doi:10.2135/cropsci1993.0011183X003300040044x

Hu J, Zhu J, Xu HM (2000) Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theoretical and Applied Genetics* **101**, 264–268. doi:10.1007/s001220051478

IBPGR (1985) 'Descriptors for *Panicum miliaceum* and *P. sumatrense*.' (International Board for Plant Genetic Resources: Rome)

Jiaju C (1986) Importance and genetic resources of small millets with emphasis on foxtail millet (*Setaria italica*) in China. In 'Small millets in global agriculture'. (Eds A Seetharam, KW Riley, G Harinarayana) pp. 93–100. (Oxford & IBH Publishing Co Pty Ltd: New Delhi)

Kalinova J, Moudry J (2006) Content and quality of protein in proso millet (*Panicum miliaceum* L.) varieties. *Plant Foods for Human Nutrition* **61**, 43–47. doi:10.1007/s11130-006-0013-9

Keuls M (1952) The use of the Studentized range in connection with an analysis of variance. *Euphytica* **1**, 112–122. doi:10.1007/BF01908269

Knupffer H, van Hintum Th JL (2003) Summarised diversity – The Barley Core Collection. In 'Diversity in barley (*Hordeum vulgare*)'. (Eds R von Bothmer, Th JL van Hintum, H Knupffer, K Sato) pp. 259–267. (Elsevier Science BV: Amsterdam, The Netherlands)

Levene H (1960) Robust tests for quality of variances. In 'Contribution to probability and statistics: essays in honour of Harold Hotelling'. (Ed. I Olkin) pp. 278–292. (Stanford University Press: Stanford, NJ)

Li Y, Shi Y, Cao Y, Wang T (2005) Establishment of a core collection for maize germplasm preserved in the Chinese National Genebank using geographic distribution and characterization data. *Genetic Resources and Crop Evolution* **51**, 845–852. doi:10.1007/s10722-005-8313-8

Li Z, Zhang H, Zeng Y, Yang Z, Shen S, Sun C, Wang X (2002) Studies on sampling schemes for establishment of core collection of rice landraces in Yunnan, China. *Genetic Resources and Crop Evolution* **49**, 67–74. doi:10.1023/A:1013855216410

Magness JR, Markle GM, Compton CC (1971) Food and feed crops of the United States. Interregional Research Project IR-4, Interregional Research Bulletin 1 (Bulletin 828, New Jersey Agricultural Experimental Station), Rutgers University, New Brunswick, NJ.

Martynov SP, Dobrotvorskaya TV, Dotlacil L, Stehno Z, Faberova I, Bares I (2003) Genealogical approach to the formation of the winter wheat core collection. *Genetika* **39**, 1091–1098.

Matz SA (1986) Millet, wild rice, adlay, and rice grass. In 'Cereal science'. pp. 225–229. (Avi Press: Westport, CT)

Milligan GW, Cooper M (1985) An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**, 159–179. doi:10.1007/BF02294245

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

Ortiz R, Ruiz-Tapia EN, Mujica-Sanchez A (1998) Sampling strategy for a core collection of Peruvian quinoa germplasm. *Theoretical and Applied Genetics* **96**, 475–483. doi:10.1007/s001220050764

Rao KEP, Rao VR (1995) The use of characterization data in developing a core collection of sorghum. In 'Core collections of plant genetic resources'. (Eds T Hodgkin, AHD Brown, Th JL van Hintum, BAV Morales) pp. 109–116. (John Wiley & Sons: Chichester, UK)

Roshevits RY (1980) 'Grasses: an introduction to the study of fodder and cereal grasses.' (Indian National Scientific Documentation Centre: New Delhi)

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

Skinner DZ, Bauchan GR, Auricht G, Hughes S (1999) A method for the efficient management and utilization of large germplasm collections. *Crop Science* **39**, 1237–1242. doi:10.2135/cropsci1999.0011183X003900040046x

Snedecor GW, Cochran WG (1980) 'Statistical methods.' 7th edn (Iowa State University Press: Ames, IA)

Spagnoletti Zeuli PL, Qualset CO (1993) Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. *Theoretical and Applied Genetics* **87**, 295–304. doi:10.1007/BF01184915

Theisen AA, Knox EG, Mann FL (1978) Feasibility of introducing food crops better adapted to environmental stress. In 'Individual crop reports. Vol. II'. pp. 168–172. (National Science Foundation: Arlington, VA)

Upadhyaya HD, Gowda CLL, Pundir RPS, Reddy VG, Singh S (2006) Development of core subset of finger millet germplasm using geographical origin and data on 14 quantitative traits. *Genetic Resources and Crop Evolution* **53**, 679–685. doi:10.1007/s10722-004-3228-3

Upadhyaya HD, Gowda CLL, Reddy KN, Singh S (2009*a*) Augmenting the pearl millet [*Pennisetum glaucum* (L.) R. Br.] core collection for enhancing germplasm utilization in crop improvement. *Crop Science* **49**, 573–580. doi:10.2135/cropsci2008.06.0378

Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CLL (2009*b*) Mini core collections for efficient utilization of plant genetic resources in crop improvement programs. Information Bulletin No. 78. International Crops Research Institute for the Semi Arid Tropics, Patancheru Andhra Pradesh, India. 52 pp.

Upadhyaya HD, Pundir RPS, Gowda CLL, Reddy VG, Singh S (2008) Establishing a core collection of foxtail millet to enhance utilization of germplasm of an underutilized crop. *Plant Genetic Resources* **7**, 177–184. doi:10.1017/S1479262108178042

Wang JC, Hu J, Xu HM, Zhang S (2007) A strategy on constructing core collections by least distance stepwise sampling. *Theoretical and Applied Genetics* **115**, 1–8. doi:10.1007/s00122-007-0533-1

Ward J (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* **38**, 80–83.

Yates F (1934) Contingency table involving small numbers and the χ^2 test. *Journal of the Royal Statistical Society* Suppl. 1, 217–235.

Manuscript received 3 September 2010, accepted 11 May 2011