Core Collection of Sorghum: II. Comparison of Three Random Sampling Strategies

C. Grenier,* P. Hamon, and P. J. Bramel-Cox

ABSTRACT

Since 1972, the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) has maintained a large collection of sorghum in India. The collection size has continuously increased, and the total number of accessions at present conserved in the gene bank has reached about 36 000 accessions. The need to help management was considered, and this study was conducted to establish core collections. This sorghum collection was earlier stratified into four clusters according to the photoperiod sensitivity. Then, considering the core collection strategy, we used three random sampling procedures to determine the specific accessions to be included in the core [i.e., a constant portion (Core C), a proportional (Core P), and a proportional to the logarithm (Core L)] of the photoperiod group size sampling strategy. Both the Core C and L were significantly different from the landrace collection with better representation of the smallest groups, such as landraces insensitive to photoperiod. Despite differences between the three core collections, estimates of global diversity through the Shannon-Weaver Diversity Indices were of the same magnitude as the landrace collection. When compared, the Core C and L were significantly different. Core L sampled better for the characters, the race, and the latitudinal classes that were related to the photoperiod-sensitive landraces. Thus, for establishing a core collection with the widest range of adaptation to photoperiod, we propose the use of a logarithmic sampling strategy, which identifies a broadly adapted set of genotypes.

A large collection, 36 719 accessions, of sorghum [*Sorghum bicolor* (L.) Moench] is conserved at ICRISAT. This collection is made up of breeding lines (14.5%), wild species (1.1%), and traditional varieties (i.e., landraces) (84.3%). The management of this collection involves acquisition of new accessions, conservation, characterization, evaluation, enhancement, and distribution of these germplasm (Brown, 1995). To facilitate the management and the use of genetic resources, the concept of a "core collection" was proposed by Frankel in 1984 to provide efficient access to the whole collection (Brown, 1989a). Thus, a core collection was defined as a limited set of accessions chosen to represent the genetic spectrum in the whole collection. Under the sampling theory of selectively neutral alleles, Brown (1989b) suggested a core collection size of about 10% of the entire collection.

To establish this core collection, a number of random sampling strategies were proposed by Brown (1989b). The initial step would be to stratify a collection. Then each stratum could be sampled by either a constant number for all the clusters (C strategy), a number proportional to the cluster size (P strategy), or a number proportional to the logarithm of the cluster size (L strategy). The P strategy has been the most commonly used to constitute a core collection (Holbrook et al., 1993; Basigalup et al., 1995; Prasada Rao and Ramanatha Rao, 1995; Ortiz et al., 1998). The C strategy has been used only by Vaughan (1991) on the IRRI rice collection, and the L strategy was used by Iguarta et al. (1998) on the Spanish barley (*Hordeum* L.) collection. From these studies, two questions have arisen on the use of these three possible sampling strategies. First, do the different strategies retain an equivalent level of diversity from the original collection? Second, how do the core collections identified from these various sampling strategies compare?

Previously, the world sorghum collection was stratified into four clusters according to photoperiod sensitivity (Grenier et al., 2001). The landrace collection held at ICRISAT was considered as 60 groups obtained from the cross tabulation table between the racial membership of the landraces (15 races) and their geographical origin (four latitudinal ranges). For flowering date and plant height, differences in the records between the two seasons of the semi-arid tropics (the rabi with shorter day length, and the kharif with longer day length) gave two continuous variables. Each one was translated into a qualitative variable by splitting into four frequency classes using the quartiles. Then, from the frequency distribution of landraces within race-latitudinal groups (60 groups) and the eight photoperiod classifications (four for flowering date and four for plant height), a K-means clustering procedure was used to get four clusters that pertained to the reaction to photoperiod. Thus, the four clusters obtained described the adaptations of the accessions due to both natural selection and farmers' selection. In this study, we established three core collections from the three random sampling strategies applied to this stratified landrace collection. The objectives of this study were to (i) compare the three core collections with each other and to the original landrace collection for their representation of diversity, and (ii) to determine the best strategy to sample the diversity of the entire landrace collection.

MATERIALS AND METHODS

From the entire collection of cultivated sorghum maintained at ICRISAT, only landraces assembled from the range of latitude 40° N to 40° S with complete data for all the characteristics (passport, qualitative, and quantitative) were considered. This collection represents 22 473 accessions.

C. Grenier, Purdue Univ., 1150 Lilly Hall of Life Sciences, West Lafayette, IN 47907; P. Hamon, Université Montpellier III/IRD, BP5045, 34 032 Montpellier Cedex 1 France; and P.J. Bramel-Cox, ICRISAT-GREP Patancheru, 502 324 A.P., India. Received 30 Jul. 1999. *Corresponding author (grenier@purdue.edu).

Published in Crop Sci. 41:241–246 (2001).

Abbreviations: Core C, L, P, constant portion, logarithm, proportional portion; C, P, L strategy, sampling strategy in which each stratum could be samples by either a constant number for all the clusters (C), a number proportional to the cluster size (P), or a number proportional to the logarithm of the cluster size (L); FL, difference flowering (d) between the rainy and the dry seasons; PHT, difference in plant height (cm) between the rainy and the dry seasons.

Table 1. Descriptive statistics for the ten quantitative characters: Mean (± standard error), range, and variance for the sorghum landrace collection and for the core collections. Levene's test of homogeneity of variances and Newman-Keuls' test (N-K test) for means comparison between each core and the sorghum landrace collection.

	Days to flowering in dry season	Days to flowering in rainy season	Number of basal tillers	Plant height in dry season	Plant height in rainy season	Peduncle exertion	Panicle length	Panicle width	Grain size	100-seed weight
Landrace collection	d			cm					mm	g
Mean ± se	72.64 ± 0.10	97.05 ± 0.19	2.08 ± 0.01	219.02 ± 0.37	342.50 ± 0.59	16.55 ± 0.06	22.17 ± 0.06	9.42 ± 0.03	3.03 ± 0.00	3.08 ± 0.01
Range	36–136	35–199	1–20	55–465	80–655	0–71	2.5–86	1–80	0.8–5.6	0.32–8.92
Variance	216.56	845.34	0.89	3045.29	7921.58	86.54	73.84	21.89	0.27	1.06
Core P										
Mean ± se	72.70 ± 0.31	97.08 ± 0.61	2.07 ± 0.02	220.37 ± 1.15	344.34 ± 1.85	16.71 ± 0.19	22.35 ± 0.19	9.43 ± 0.10	3.04 ± 0.01	3.09 ± 0.02
Range	36–132	46–165	1–10	70–465	85–655	0–60	4–86	1.3–41	1–5.5	0.71–7.51
Variance	216.11	832.44	0.86	2986.47	7723.48	83.65	78.01	21.66	0.27	1.03
Levene's test	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
N-K test	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Core C										
Mean ± se	70.47 ± 0.29	83.09 ± 0.62	2.11 ± 0.02	207.80 ± 1.13	290.05 ± 2.09	16.17 ± 0.19	23.44 ± 0.16	9.52 ± 0.10	2.96 ± 0.01	2.87 ± 0.02
Range	36–132	40–164	1–13	65–465	90–610	0–60	5–57	1.3–41	0.8–5.5	0.32–6.84
Variance	193.52	854.64	0.82	2874.97	9812.20	77.10	58.04	21.54	0.25	0.77
Levene's test	**	NS	*	*	**	**	**	NS	**	**
N-K test	–	**	–	–	–	–	–	NS	–	–
Core L										
Mean ± se	71.11 ± 0.29	86.62 ± 0.63	2.07 ± 0.02	213.32 ± 1.12	303.23 ± 2.09	16.48 ± 0.19	23.52 ± 0.16	9.63 ± 0.10	2.98 ± 0.01	2.92 ± 0.02
Range	36–132	41–159	1–10	75–410	90–600	0–49	2.5–57	1–51	1–5	0.36–7.96
Variance	193.85	895.42	0.72	2937.14	9806.50	78.76	59.71	22.84	0.23	0.79
Levene's test	**	NS	**	*	**	**	**	NS	**	**
N-K test	–	**	–	–	–	–	–	*	–	–

NS, non-significant test at $P > 0.05$ level.

*, ** Significant at the $P < 0.05$ and $P < 0.01$ probability levels, respectively. N-K test could not be performed because of heterogeneous variances.

For characterization, the accessions were grown on Vertisol soils at Patancheru, India (17°25′ N and 78° E). Field observations were made over several years from 1975 to 1996 in both the rainy (kharif) and the post-rainy (rabi) seasons. These two different environmental conditions are typical of the semi-arid tropics (Appa Rao et al., 1996). The rainy season occurs during long days (June–September) while the dry season occurs during short days (September–April). Data were recorded for ten quantitative characters (Table 1), as described by Appa Rao et al. (1996). Days to 50% flowering were recorded as the mean emergence date to the date when 50% of the plants had started flowering (i.e., ≈15 out of the 30 plants per row per accession). Plant height was observed at maturity and recorded as the length of the main stalk (cm) for ten plants. Both of these characters were measured in the dry and rainy seasons. Peduncle exertion (cm), panicle length (cm), panicle width (cm), and number of basal tillers were measured on ten plants in the field. After harvest, grain size (mm) and 100-seed weight (g) were measured. Two relative variables, FL and PHT, were defined as the differences in measurements obtained during the long day length of the rainy season and the shorter day length of the dry season. FL was calculated as the difference in days to flowering and PHT as the difference in plant height in the rainy and dry seasons. The nine qualitative characters (Table 2) recorded were leaf and midrib pigmentation, shape of the inflorescence, glume and grain color, glume covering, type of grain (corneous and lustrous), and presence or absence of subcoat.

The entire collection was stratified into four clusters by a K-means procedure (StatSoft, 1997) performed on a matrix of 60 groups and eight photoperiod classifications. The 60 groups were defined by the racial membership of the landraces for the 15 races defined according to the Harlan and de Wet racial classification (Harlan and de Wet, 1972) and the geographical origin was defined with four latitudinal ranges (Grenier et al., 2001). The eight photoperiod classifications were obtained from the two traits related to photoperiod response (FL and PHT), and split into four frequency classes according to the quartiles. From this matrix, a K-means clustering procedure was performed using an a priori number of four clusters. From this stratification, the first cluster consisted of 1160 accessions classified as photoperiod insensitive. The

Table 2. Chi-square test (χ^2) comparing the frequency distribution observed in each core collection (Core P, C, and L) with those expected in 10% of the sorghum landrace collection for the nine qualitative characters.

Qualitative Traits†	Number of character states	χ^2 Core P	χ^2 Core C	χ^2 Core L
Leaf pigmentation	2	0.61 NS	0.75 NS	0.03 NS
Midrib color	4	7.62 NS	79.13	71.82
Panicle compactness	10	7.70 NS	134.38	109.51
Glume color	13	8.01 NS	24.85	25.15
Glume covering	5	0.20 NS	174.11	151.13
Grain color	13	13.82 NS	145.26	101.00
Endosperm texture	5	2.93 NS	47.31	48.38
Grain lustre	2	0.66 NS	8.53	1.17 NS
Subcoat presence	2	2.16 NS	24.73	8.91

† Leaf pigmentation = tan, pigmented; Midrib color = white, dull, yellow, brown; Panicle compactness = semi-compact oval, compact oval, semi compact elliptic, compact elliptic, semi loose drooping branches, loose drooping branches, very loose drooping branches, semi loose stiff branches, loose stiff branches, very loose stiff branches; Glume color = white, straw, yellow, gray, light brown, brown, partly brown, purple, partly purple, black, light red, red, red-brown; Glume covering = uncovered (1) to fully covered (5); Grain color = white, chalky white, white and red, gray, straw, yellow, light red, light brown, red, brown, red-brown, purple, black; Endosperm texture = corneous (1) to floury (5); Grain lustre = lustrous, non-lustrous; Subcoat = presence, absence.

NS = non-significant difference at $P < 0.05$ probability level.

second cluster included 1062 accessions classified as mildly photoperiod sensitive. The third cluster consisted of 10 630 accessions classified as photoperiod-sensitive landraces. Finally, the fourth cluster consisted of 9621 accessions classified as highly photoperiod-sensitive landraces (Grenier et al., 2001).

Core collections that represented 10% of the landrace collection (2247 accessions) were established from three random sampling procedures. The C strategy sampled at random a constant number (562) of accessions from each cluster irrespective of its size. The P strategy sampled at random 10% of the number of accessions within each cluster (i.e., 116, 106, 1063, and 962 accessions). The L strategy sampled at random proportionally from the logarithm of the number of accessions within each cluster (i.e., 488, 482, 642, and 635 accessions).

For the ten quantitative characters, comparisons between the core collections and the landrace collection were based upon Levene's test of homogeneity of variance as well as Newman and Keuls' test for post hoc comparison of means performed only when variances were comparable (Statsoft, 1997).

Because of the categorical nature of the qualitative characters, the passport data and the racial membership, a nonparametric statistical procedure was used to evaluate and compare the distributions. The χ^2 test was used to compare the distribution of accessions for the nine qualitative characters, passport data, and racial distribution in the landrace collection to those of the core collections. In the last case, the χ^2 test was obtained from the racial membership determined with the Harlan and de Wet's race classification. Passport data were compared with the frequency distributions for the geographical origin (i.e., continent and latitudinal ranges). The distributions observed in each core were compared to the distribution expected in the landrace collection adjusted for accession number (10%). The χ^2 test was also used to compare the distributions of the three core collections. Distributions for races and for passport data within each core collection were considered.

For each core, global diversity was estimated by the Shannon-Weaver Diversity Index as presented by Poole (1974; as cited by Li et al., 1996). This phenotypic diversity index is based upon the frequency distributions for the nine qualitative characters. This index also included the 10 quantitative characters transformed into four phenotypic classes defined by the four quartiles of the landrace collection. The Shannon-Weaver Diversity Index, H'_C, was estimated using

$$H'_C = -\sum_{i=1}^{n} p_i \log_e p_i \qquad [1]$$

where for a given character C, n was the number of phenotypic classes (for the qualitative characters $n = 2$ to 13 descriptor states, and for quantitative characters $n = 4$ frequency classes), and p_i was the proportion of the total number of entries in the ith class. Due to its additive property (Poole, 1974; as cited by Li et al., 1996), Shannon-Weaver Diversity Indices obtained for each character were pooled for each core collection and landrace collection. Means and standard errors were then calculated from the nonstandardized indices, and the variance was approximated by

$$\mathrm{Var}(H') = [\textstyle\sum p_i \log_e^2 p_i - (\sum p_i \log_e p_i)^2]/N + (n-1)/2N^2$$

where N was the number of observations (Hutcheson, 1970). The H' indices were compared by a t test

$$T = (H'_1 - H'_2)/[\mathrm{Var}(H'_1) + \mathrm{Var}(H'_2)]^{1/2}$$

with degrees of freedom

$$\mathrm{d.f.} = [\mathrm{Var}(H_1) + \mathrm{Var}(H_2)]^2/[\mathrm{Var}(H_1)^2/N_1 + \mathrm{Var}(H_2)^2/N_2]$$

where N_1 and N_2 were the numbers of entries in the two groups.

To compare indices obtained for different characters, a standardized index $\mathrm{SDI_C}$ (values range from 0–1) was calculated as

$$\mathrm{SDI_C} = H'_C/\log_e n$$

RESULTS

Considering the ten quantitative morphological characters, comparisons between the core collections and the landrace collection were based upon tests for homogeneity of variances and, when possible, means comparisons (Table 1). The results indicated that for Core Collection P, none of the ten traits had a significantly different variance from the landrace collection. Furthermore, means were not significantly different from those of the landrace collection. For both the Core C and the Core L, homogeneous variances were found for two traits, days to flowering in kharif and panicle width; however, compared with the landrace collection, only Core C did not have a significantly different mean for panicle width.

The χ^2 compared observed (in the core collection) vs. expected (obtained for 10% of the landrace collection) distributions calculated for the nine morphological qualitative characters (Table 2). For all these characters in the Core Collection P, the distribution was not significantly different from the observed in the landrace collection. Conversely, for nearly all characters, Core C and L were significantly different from the distributions expected in the landrace collection. Core C and Core L exhibited similar frequencies for leaf pigmentation, and only Core L did not differ from the landrace collection for grain lustre. Thus, when the qualitative and quantitative characters were considered, Core Collection P represented a similar pattern of diversity as the landrace collection while Core C and Core L exhibited different distributions.

The Shannon-Weaver Diversity Indices for the landrace collection and the three core collections were calculated from the nine qualitative characters and the ten quantitative characters. The variance of the Shannon-Weaver Diversity Indices was used to compare these diversity estimates. Statistically similar morpho-agronomic diversity levels were obtained in the comparison of the landrace collection and the core collections ($H' = 1.261 \pm 0.111$ for the landrace collection, and 1.258 ± 0.111, 1.237 ± 0.106, and 1.231 ± 0.106 for the core collections P, L and C, respectively). Based upon the standardized Shannon-Weaver Diversity Indices, relatively high levels of diversity were found ($\mathrm{SDI_C} = 0.853 \pm 0.053$, 0.851 ± 0.052, 0.840 ± 0.050, and 0.835 ± 0.050, for landrace collection, Core P, L and C, respectively). While Core C and Core L had a different pattern of distribution from the landrace collection for most of the morpho-agronomic characters, the overall diversity was maintained in relation to the landrace collection.

The diversity of core collections was also compared for their deviation from the original distribution of accessions for the continent of origin, the latitudinal class, and the race classification through a χ^2 test (Table 3). Classes in the whole collection at a frequency lower than 1% were grouped together; such was the case for the American, European and Oceanic continents of origin, and kafir intermediates in the race classification. Passport information and racial distribution in the Core P and the landrace collection were not significantly different. As with the morpho-agronomic traits, the Core P gave a faithful image of the landrace collection for geographical and racial distribution. Conversely, Core C and Core L differed significantly from the landrace collection for latitudinal, continental (except Asia for the Core L) and racial distributions (except caudatum, guinea–bicolor, and guinea–durra for Core C, and guinea–bicolor, guinea–durra, and durra–caudatum for Core L).

The three core collections were compared for qualitative characters, quantitative characters, passport data and racial distribution (Table 4). From the previous comparison of Core P and the landrace collection, no significant difference was found. Thus, Core P is assumed to adequately represent the landrace collection in its comparisons with Core C and L. As both Core C and Core L differed from the landrace collection, an assessment was made only on the differences between these two core collections. Core C and Core L did not differ significantly for the qualitative characters. Core C and Core L differed significantly for days to flowering in kharif, plant height in the two seasons, and 100-seed weight (Table 4). Core C and Core L were not statistically different in their distribution over the continents; however, significant differences were observed for the distribution of landraces in the tropical and temperate latitudinal classes (LR-2 and LR-3) and for the distribution of the race guinea.

Table 3. Chi-square test (χ^2) comparing the frequency distribution observed in the core collections with those expected in 10% of the sorghum landrace collection for the passport data [continents of origin and latitudinal range (LR)] and the race classification (the five basic races plus the intermediates).

	χ^2 Core P	χ^2 Core C	χ^2 Core L
Africa	0.50 NS	5.25	6.14
Asia	0.41 NS	8.21	1.50 NS
America + Oceania + Europe	2.83 NS	538.43	337.26
LR-1: 10°S–10°N	0.03 NS	144.60	116.83
LR-2: 20°S–10°S, 10°N–20°N	0.19 NS	214.59	121.23
LR-3: 30°S–20°S, 20°N–30°N	0.41 NS	1024.72	589.98
LR-4: 40°S–30°S, 30°N–40°N	0.93 NS	1426.27	1093.69
Bicolor	≅ 0.00 NS	131.48	74.66
Caudatum	0.92 NS	2.43 NS	8.17
Durra	0.09 NS	104.19	80.65
Guinea	1.40 NS	75.54	37.88
Kafir	1.92 NS	693.90	549.37
Caudatum-bicolor	1.60 NS	41.22	35.60
Durra-bicolor	0.30 NS	13.47	10.96
Guinea-bicolor	0.41 NS	2.91 NS	1.63 NS
Durra-caudatum	0.10 NS	13.44	3.84 NS
Guinea-caudatum	0.43 NS	21.62	22.30
Guinea-durra	0.35 NS	3.75 NS	0.74 NS
Intermediate kafir	1.53 NS	310.52	165.71

NS = non-significant difference at $P < 0.05$ probability level.

DISCUSSION

As suggested by Brown (1989b), the stratification of the base collection has often been the first step to constitute core collections (Brown, 1989b; Erskine and Muehlbauer, 1991; Spagnoletti Zeuli and Qualset, 1993; Basigalup et al., 1995; Diwan et al., 1995; Cordeiro et al., 1995; Crossa et al., 1995; van Hintum et al., 1995; Prasada Rao and Ramanatha Rao, 1995; Tohme et al., 1995; Balfourier et al., 1998; Iguarda et al., 1998; Ortiz et al., 1998; Huamán et al., 1999; Skinner et al., 1999). In sorghum, Prasada Rao and Ramanatha Rao (1995) stratified the base collection from a principal component analysis based on seven morphological traits. In our study, the base collection was stratified according to photoperiod sensitivity and such clustering was discussed in Grenier et al. (2001). In the second step, either only one sampling strategy was applied (Prasada Rao and Ramanatha Rao, 1995), or comparisons between random sampling strategies were reported (Brown, 1989b; Erskine and Muehlbauer, 1991; Spagnoletti Zeuli and Qualset, 1993; Cordeiro et al., 1995; Diwan et al.,

Table 4. Differences between the three random stratified sorghum core collections when compared for the ten quantitative characters, the nine qualitative characters, the passport data, and the racial distribution. Different letters were used when the comparisons between the core collections were significantly different at $P < 0.05$ probability level. Letter A indicates a similar distribution as the sorghum landrace collection. Letters B and C indicate significantly different distributions than the landrace collection.

Traits	Core P	Core C	Core L
Days to flowering in rabi	A	B	B
Days to flowering in kharif	A	B	C
Basal tillers	A	B	B
Plant height in rabi	A	B	C
Plant height in kharif	A	B	C
Peduncle exertion	A	B	B
Panicle length	A	B	B
Panicle width	A	A, B	B
Grain size	A	B	B
100-seed weight	A	B	C
Leaves pigmentation	A	A	A
Midrib color	A	B	B
Panicle compactness	A	B	B
Glume color	A	B	B
Glume covering	A	B	B
Grain color	A	B	B
Endosperm texture	A	B	B
Grain lustre	A	B	A, B
Subcoat presence	A	B	B
Africa	A	B	B
Asia	A	B	A, B
America + Oceania + Europe	A	B	B
LR-1: 10°S–10°N	A	B	B
LR-2: 20°S–10°S, 10°N–20°N	A	B	C
LR-3: 30°S–20°S, 20°N–30°N	A	B	C
LR-4: 40°S–30°S, 30°N–40°N	A	B	B
Bicolor	A	B	B
Caudatum	A	A, B	B
Durra	A	B	B
Guinea	A	B	C
Kafir	A	B	B
Caudatum-bicolor	A	B	B
Durra-bicolor	A	B	B
Guinea-bicolor	A	A	A
Durra-caudatum	A	B	A, B
Guinea-caudatum	A	B	B
Guinea-durra	A	A	A
Intermediate kafir	A	B	B

1995; Galwey, 1995; van Hintum et al., 1995). In the latter case, the main conclusions drawn from these studies were that a better representation of the pattern of variation present in the total collection was obtained with the proportional sampling. The range of variation present in the whole collection was maximized when constant or logarithmic sampling were used. In the simulation studies of Schoen and Brown (1993), Schoen and Brown (1995), and Bataillon et al. (1996), and in the hypothetical populations of Yonezawa et al. (1995), proportional sampling appeared as the optimal strategy, although differences between it and logarithmic sampling were small.

Differences in representation between core collections could result from differences in the stratification of the base collection as it partitioned the total diversity. Stratification can lead to near equal or unequal cluster sizes. When group sizes are equal, the three random sampling procedures will probably be equivalent. For unequal group size, Brown (1989b) found in the *Glycine tomentella* (Hayata) core collection that when the rarest variants were classified in the smaller groups, then the constant strategy sampled the highest number of types or alleles per locus. Conversely, if the rarest variants occurred in the largest groups, the proportional sampling identified the best core. Brown (1989b) concluded that the logarithmic strategy was a good compromise that both lowered the representation from the largest clusters and sampled the rare variants. In addition, for the Brazilian cassava collection, where areas of high diversity were poorly represented in the whole collection, the logarithmic strategy was recommended (Cordeiro et al., 1995). In a similar manner, the logarithmic strategy better sampled the underrepresented countries in the durum wheat (*Triticum durum* Desf.) collection (Spagnoletti Zeuli and Qualset, 1993). The logarithmic strategy seems to reduce the weight given to the larger groups, which probably have a high level of genetic redundancy. Also, it could be used to increase the probability of sampling rare alleles that confer wide or local adaptation.

In fact, the choice of the sampling strategy greatly depends on the objectives given to the core collection. When the objective of defining a core collection and the properties of the three random sampling strategies are considered, two options are possible. If the maintenance of a collection has become unmanageable because of its increasing size, the main objective of the core is to concentrate efforts on a smaller set that faithfully represents the diversity of the entire collection. In this situation, and when characterization data are not available, the P strategy would be the best to give a reliable image of the landrace collection. If the core is to be used by the breeders for genetic enhancement, it should contain the maximum amount of the diversity present in the base collection and facilitate access to specific traits underrepresented in the base collection. For this objective, the C or L strategy would be the best approach. In particular situations, L sampling is preferred to C sampling. For example, the logarithmic sampling is recommended to better sample the rarest variants from the smaller groups (Brown, 1989b on perennial wild relatives of soybean; Spagnoletti Zeuli and Qualset, 1993 on durum wheat; Cordeiro et al., 1995 on Brazilian cassava; Galwey, 1995 on beans).

In the present work, the constant and logarithmic core collections had different patterns of diversity when compared to the landrace collection. For sorghum breeders, photoperiod-insensitive landraces are important for short-term genetic enhancement in the temperate latitude. Furthermore, photoperiod sensitivity is an important trait for tropical breeding programs. Indeed, it ensures yield stability both for quantity and quality traits particularly for cultivars grown under climatic constraints such as in western Africa (Chantereau et al., 1998; Kouressy et al., 1998; Trouche et al., 1998). Thus, there is a need to keep the widest range of photoperiod sensitivity to retain a wide range of adaptation to specific ecogeographical zones in a core collection. The logarithmic core retained an increased number of the photoperiod-insensitive landraces and highly photoperiod-sensitive landraces that would serve as an important source of broadly adapted genetic backgrounds for breeding programs.

ACKNOWLEDGMENTS

We acknowledge M. Deu, for her contributive remarks and review of the manuscript, and the ICRISAT-GRD staff for providing data information.

REFERENCES

Appa Rao, S., K.E. Prasada Rao, M.H. Mengesha, and V. Gopal Reddy. 1996. Morphological diversity in sorghum germplasm from India. Genet. Resour. Crop Evol. 43:559–567.

Balfourier, F., G. Charmet, J.-M. Prosperi, M. Goulard, and P. Monestiez. 1998. Comparison of different spatial strategies for sampling a core collection of natural populations of fodder crops. Genet. Sel. Evol. 30 (Suppl. 1):215–235.

Basigalup, D.H., D.K. Barnes, and R.E. Stuker. 1995. Development of a core collection for perennial *Medicago* plant introduction. Crop Sci. 35:1163–1168.

Bataillon, T., J.L. David, and D.J. Schoen. 1996. Neutral genetic markers and conservation genetics: Simulated germplasm collections. Genetics 144:400–417.

Brown, A.H.D. 1995. The core collection at the crossroads. p. 3–19. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.

Brown, A.H.D. 1989a. The case of core collections. p. 135–156. *In* A.H.D. Brown et al. (ed.) The use of plant genetic resources. Cambridge Univ. Press, Cambridge, UK.

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

Chantereau, J., C. Luce, M.A.G. Hamada, and G. Trouche. 1998. L'utilisation des sorghos guinea en sélection dans le programme conjoint ICRISAT-CIRAD sur le sorgho. p. 65–72. *In* A. Ratnadass et al. (ed.) Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre. Actes de l'atelier de restitution du programme conjoint sur le sorgho ICRISAT-CIRAD. Bamako, Mali. Collection colloques, Montpellier, CIRAD-CA.

Cordeiro, C.M.T., E.A.V. Morales, P. Ferreira, D.M.S. Rocha, I.R.S. Costa, A.C.C. Valois, and S. Silva. 1995. Towards a Brazilian core collection of cassava. p. 155–168. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.

Crossa, J., I.H. DeLacy, and S. Taba. 1995. The use of multivariate methods in developing a core collection. p. 77–92. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.

Diwan, N., M.S. McIntosh, and G.R. Bauchan. 1995. Methods of developing a core collection of annual *Medicago* species. Theor. Appl. Genet. 90:755–761.

Erskine, W., and F.J. Muehlbauer. 1991. Allozyme and morphological variability, outcrossing rate and core collection formation in lentil germplasm. Theor. Appl. Genet. 83:119–125.

Galwey, N.W. 1995. Verifying and validating the representativeness of a core collection. p. 187–198. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.

Grenier, C., P.J. Bramel-Cox, and P. Hamon. 2001. Core collection of sorghum: I. Stratification based on eco-geographical data. Crop Sci. 41:234–239 (this issue).

Harlan, J.R., and J.M.J. de Wet. 1972. A simplified classification of cultivated sorghum. Crop Sci. 12:172-176.

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

Huamán Z., C. Aguilar, and R. Ortiz. 1999. Selecting a Peruvian sweetpotato core collection on the basis of morphological, eco-geographical, and disease and pest reaction data. Theor. Appl. Genet. 98: 840–844.

Hutcheson, K. 1970. A test for comparing diversities based on the Shannon formula. J. Theor. Biol. 29:151–154.

Igartua, E., M.P. Gracia, J.M. Lasa, B. Medina, J.L. Molina-Cano, J.L. Montoya, and I. Romagosa. 1998. The Spanish barley core collection. Genet. Resour. Crop Evol. 45:475–481.

Kouressy, M., M. Ouattara, and M. Vaksmann. 1998. Importance du photopériodisme chez les sorghos tropicaux, conséquences pour un programme de sélection. p. 49–54. *In* A. Ratnadass et al. (ed.) Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre. Actes de l'atelier de restitution du programme conjoint sur le sorgho ICRISAT-CIRAD. Bamako, Mali. Collection colloques, Montpellier, CIRAD-CA.

Li, Y., S. Wu, Y.S. Cao, and X. Zhang. 1996. A phenotypic diversity analysis of foxtail millet [*Setaria italica* (L.) P. Beauv.] landraces of Chinese origin. Genet. Resour. Crop Evol. 43:377–384.

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

Prasada Rao, K.E., and V. Ramanatha Rao. 1995. The use of characterisation data in developing a core collection of sorghum. p. 109–116. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.

Schoen, D.J., and A.H.D. Brown. 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc. Natl. Acad. Sci., (USA) 90:10 623–10 627.

Schoen, D.J., and A.H.D. Brown. 1995. Maximising allelic diversity in core collections of wild crop relatives: The role of genetic markers. p. 55–76. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.

Skinner D.Z., G.R. Bauchan, G. Auricht, and S. Hughes. 1999. A method for the efficient management and utilization of large germplsam collections. Crop Sci. 39:1237–1242.

Spagnoletti Zeuli, P.L., and C.O. Qualset. 1993. Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. Theor. Appl. Genet. 87:295–304.

StatSoft. 1997. Statistica for Windows. USA, StatSoft, Tulsa, OK.

Tohme, J., P. Jones, S. Beebe, and M. Iwanaga. 1995. The combined use of agroecological and characterisation data to establish the CIAT *Phaseolus vulgaris* core collection. p. 95–107. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.

Trouche, G., M. Vaksmann, J. Chantereau, M. Kouressy, H.D. Maiga, and C. Barro. 1998. Etude du déterminisme génétique du photopériodisme des sorghos guinea. p. 37–48. *In* A. Ratdnadass et al. (ed.) Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre. Actes de l'atelier de restitution du programme conjoint sur le sorgho ICRISAT-CIRAD, Bamako, Mali, Collection colloques, Montpellier, CIRAD-CA.

van Hintum, Th.J.L., R. Von Bothmer, and D.L. Visser. 1995. Sampling strategies for composing a core collection of cultivated barley (*Hordeum vulgare* s. lat.) collected in China. Hereditas 122:7–17.

Vaughan, D.A. 1991. Choosing rice germplasm for evaluation. Euphytica 54:147–154.

Yonezawa, K., T. Nomura, and H. Morishima. 1995. Sampling strategies for use in stratified germplasm collections. p. 35–53. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.