Theor Appl Genet (2000) 101:190–196

C. Grenier · P.J. Bramel-Cox · M. Noirot
K.E. Prasada Rao · P. Hamon

Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures

A. Using morpho-agronomical and passport data

Received: 5 October 1999 / Accepted: 3 November 1999

Abstract A large collection, such as the sorghum [*Sorghum bicolor* (L.) Moench] landrace collection held at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), represents a challenge for the maintenance of both the accessions of and the information documented for the germplasm collection. The accessibility and knowledge of the landrace collection are the essential factors for an efficient utilization of the genetic resources by both breeders and farmers. Different sampling strategies, either random or non-random, were proposed to obtain subsets of reduced size (core collection). Three subsets were established; a random sampling within a stratified collection (logarithmic strategy: L); a sample based upon morpho-agronomic diversity (principal component score strategy: PCS); and a sample based upon an empirical knowledge of sorghum (taxonomic strategy: T). Comparisons of these three samples for morpho-agronomic characterization and passport information were assessed to determine their impact on phenotypic diversity. For their overall diversity, the three subsets did not differ, as shown with the two-dimensional representation of the morpho-agronomic diversity and the Shannon-Weaver diversity indices. When comparisons for morpho-agronomic and passport data were considered, the PCS subset looked similar to the entire landrace collection. The L subset showed differences for characters associated with the photoperiod reaction that was considered in the stratification of the collection. The T subset was the most distinct from the entire landrace collection as it over-represented the landraces selected by farmers for specific uses and covered the widest range of geographical adaptation and morpho-agronomic characteristics.

Key words Core collection · Sorghum · Morpho-agronomic diversity · Sampling strategies

Introduction

Ex-situ conservation of biodiversity can result in large collections that are difficult to characterize, evaluate, utilize and maintain. An important task for curators is to find a way to preserve the widest range of genetic diversity within crop species as well as to improve the knowledge and utilization of the genetic resources. To alleviate management difficulties, the identification and use of core collections has been suggested. One of the objectives of this subset is to reduce the collection to a manageable size that will capture the widest range of genetic diversity. Different sampling strategies have been proposed since Frankel first suggested the concept of a core collection in 1984 (see Brown 1989a). Brown (1989b) proposed the use of random sampling strategies among a stratified collection. He assumed over 70% of the alleles present in the whole collection would be retained in 10% of the total collection's size based on the theory of selectively neutral alleles. For several plant species, core collections have been established and compared for their effectiveness to maintain the diversity of the base collection (Spagnoletti Zeuli and Qualset 1993; Bataillon 1994; Cordeiro et al. 1995; Diwan et al. 1995; Galwey 1995; van Hintum et al. 1995; Igartua 1998). For sorghum, Grenier et al. (2000) found that proportional sam-

Communicated by P.M.A. Tigerstedt

C. Grenier (✉)
CIRAD, TA 70/03, Avenue Agropolis,
F-34398 Montpellier cedex 5, France
Fax: +33 (0) 4 67 61 55 09
e-mail: grenier@cirad.fr

P.J. Bramel-Cox
International Crops Research Institute for the Semi-Arid Tropics,
Patancheru, 502 324 A.P., India

M. Noirot
Institut de Recherche pour le Developpement (IRD),
BP 5045, 34032 Montpellier Cedex 1, France

K.E. Prasada Rao
Academy of Biological Sciences, Visakhpatnam, 530 020 A.P.,
India

P. Hamon
Université Montpellier III/IRD, route de Mende,
34199 Montpellier Cedex 5, France

pling gave a better picture of the diversity of the base collection than constant or logarithmic sampling.

With the objective of retaining the maximum genetic diversity from the whole base collection in a manageable working collection, several non-random sampling strategies have been suggested. The principal component score (PCS) strategy, which is based on phenotypic characters, has been applied to identify a subset that will maximize the representation of the phenotypic variability of the base collection (Mahajan et al. 1996; Noirot et al. 1996; Balfourier et al. 1998; Bisht et al. 1998; Hamon et al. 1998). Alternatively, a subset could be identified based upon an empirical knowledge of the collection. This latter subset would be established with an understanding of the biology of the crop, its history, origins, the specific uses of its landraces, and their relative importance. Thus, both the phenotypic diversity as well as the crops' evolution is taken into account when this subset is selected.

At present, the collection of cultivated sorghum maintained at ICRISAT contains more than 36 000 accessions. The objective of this study was to identify a subset or core collection that retained the widest range of phenotypic diversity from this large collection. Three different sampling strategies, random and non-random, were used to establish the subsets. The three subsets were compared with each other and to the entire landrace collection based upon the morpho-agronomic and passport data.

Materials and methods

Subset constitution

The entire sorghum landrace collection maintained at ICRISAT has been completely described for morphological and passport data. This collection contains accessions acquired from a latitudinal range of 40°N to 40°S. The sorghum landrace collection of 22 473 accessions was used to establish two of the three subsets. The third subset was based upon the landrace collection plus some very recent additions to the gene bank.

Characterization data was recorded at Patancheru, India (17°25′N latitude and 78°E longitude), for 18 morpho-agronomic characters from 1975 to 1996 during the two typical seasons in India, i.e. the rabi season and the kharif season (Appa Rao et al. 1996). Rabi is the dry season with short day lengths (September to April), while kharif is the rainy season with longer day duration (June to December). Ten quantitative characters were recorded. Days to 50% flowering were recorded as the mean emergence date to the date when 50% plants have started flowering. Plant height was observed at maturity and recorded as the length of the main stalk (cm). These two characters have been recorded in both rabi and kharif seasons, FLR, PHTR and FLK, PHTK, respectively. Peduncle exertion, PEDEX (cm), panicle length, EHLG (cm), panicle width, EHWD (cm) and the number of basal tillers, BT, were measured on plants in the field. After harvest, grain size GRS (mm) and 100-seed weight SWT (g) were measured. The eight qualitative characters noted were midrib pigmentation, shape of the inflorescence, glume and grain color, glume covering, type of grain (corneous or lustrous), and presence or absence of a subcoat. Passport information used included race classification into the five basic races (bicolor, caudatum, durra, guinea and kafir) and ten intermediate forms (hybrid combinations of the major races), according to Harlan and de Wet (1972), as well as the continent and latitudinal class of the country of origin.

The logarithmic subset (L) was sampled from a previously established logarithmic core collection that contained 2247 accessions (Grenier et al. 2000) which were previously distributed into four clusters for photoperiod sensitivity. A 10% random sampling [using the random generator in Statistica (StatSoft 1997)] was taken based on the logarithm of the number of accessions within each cluster. This sampling gave a subset of 225 accessions.

From the entire landrace collection, 2000 randomly identified accessions were used to define the PCS subset from the principal component scoring strategy (Noirot et al. 1996) (PCS strategy 'corequant' obtained from htpp://www.mpl.ird.fr/genetrop/rg/pcss/index.htm). Thus, a principal component analysis was performed from the ten quantitative characters plus two ordinal characters, i.e. glume covering and endosperm texture, after standardization. The first 229 accessions with the highest relative contribution to the generalized sum of squares were then selected. They represented 31% of the phenotypic diversity found in the 2000 accessions.

The taxonomic subset (T) was selected based on empirical knowledge of the sorghum collection held by K.E. Prasada Rao. He was formerly the senior botanist at ICRISAT who had been involved in the acquisition of the sorghum collection as well as its characterization, evaluation, and conservation. He selected landraces that had a particular character of interest or which originated from a specific region. Thus, a subset of 246 accessions was selected.

Data analyses

Principal component analysis was applied using the ten standardized quantitative characters from the landrace collection. The first and second principal component scores of the accessions identified in each subset were plotted in this reduced two-dimensional space.

The Shannon-Weaver Index was calculated from the eight qualitative characters plus the ten quantitative ones after their transformation into classes. Comparisons between the three subsets were made with a *t*-test that used the approximated variance as defined in Grenier et al. (2000). For the ten quantitative characters, the entire landrace collection and the three subsets were compared through Levene's test of the homogeneity of variance and Newman and Keuls's test for post-hoc mean comparisons.

Qualitative characters and passport data were compared between the three subsets and to the entire landrace collection with a chi-square test. When possible, classes with less than five individuals were grouped together. For racial distribution, the durra-bicolor, guinea-bicolor and guinea-durra races were grouped together, as well as the kafir intermediate forms. In contrast, when the grouping did not reach the minimum value (American plus European plus Oceanic landraces) the classes were not considered.

Results

Comparison of the subsets to the entire landrace collection

Multivariate analysis of the ten quantitative characters gives a representation of the morpho-agronomic diversity on a reduced space for the entire landrace collection and the three subsets (Fig. 1). The first axis accounts for 29.6% of the variance and the principal component scores are influenced by days to flowering and plant height in both seasons. The second axis accounts for 19.2% of the variance, and both grain size and 100-seed weight influence the principal component scores. Accessions sampled in the three subsets adequately represent

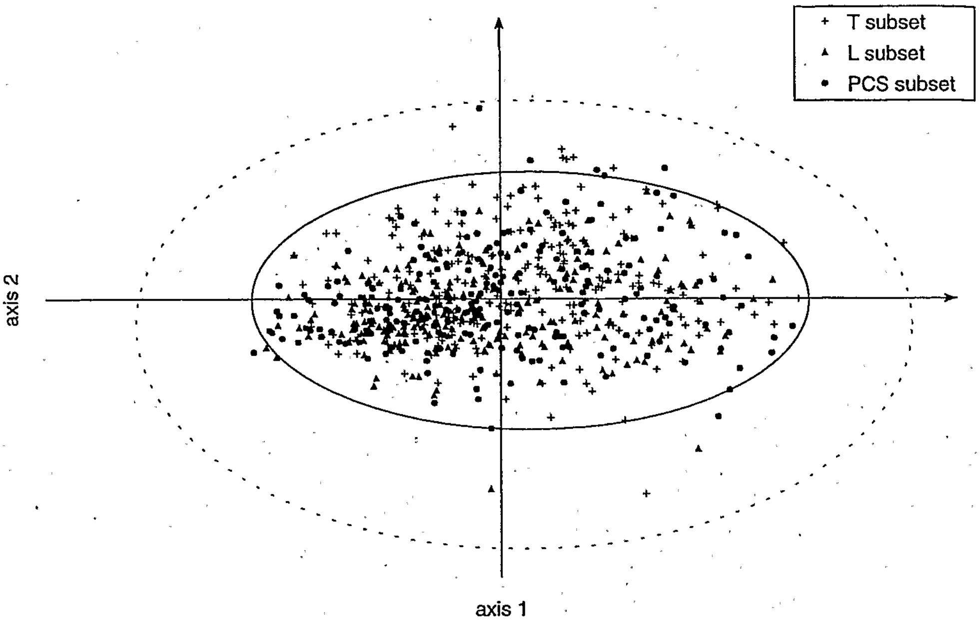

Fig. 1 Morpho-agronomic diversity in the two-dimensional space defined from the principal component analysis of the ten quantitative characters and for the 22 473 accessions. The *plain line* refers to the distribution of 90% of the entire landrace collection and the *dotted line* refers to the more external distribution of the entire landrace collection. Only the coordinates of the accessions sampled in the three subsets (principal Component Score strategy: PCS, logarithmic strategy: L, and taxonomic strategy: T) are plotted

the space that defined the main diversity of the entire landrace collection. Also, accessions from the subsets are plotted in the space where only a small proportion of the entire landrace collection is represented. When the three subsets are compared, the morpho-agronomic diversity of each subset looks similar.

In order to estimate the global phenotypic diversity levels, the Shannon-Weaver Index was calculated for the three subsets and the entire landrace collection. Statistically similar morpho-agronomic diversity levels were obtained (H′=1.261±0.111 for the entire landrace collection and 1.253±0.109, 1.214±0.106, and 1.242±0.113, for the PCS, the L and T subsets, respectively). High levels of diversity are also found with the standardized Shannon indices (SDI=0.853±0.053, 0.853±0.048, 0.823±0.052 and 0.826±0.051, for the entire landrace collection, and the subsets PCS, L and T, respectively).

The morpho-agronomic diversity based upon the mean and variance of the entire landrace collection and each of the three subsets was compared for the ten quantitative characters (Table 1). In the PCS subset, plant height in rabi and 100-seed weight have a larger variance, and the grain size is on average smaller than the entire collection. The L subset has a higher variance for plant height in kharif, and a smaller variance for basal tillers, panicle length, grain size and 100-seed weight. The mean value for days to flowering in kharif is also earlier in the L subset. In the T subset, the variance for days to flowering in kharif, panicle length, grain size and 100-seed weight is increased, while the variance for days to flowering and plant height in rabi is reduced. Also, the mean values for basal tillers and peduncle exertion are smaller in this T subset.

Morpho-agronomic diversity of the eight qualitative characters was also compared between the entire landrace collection and the three subsets (Table 2). For all characters, the PCS subset distribution does not differ significantly from the entire landrace collection. By contrast, the T subset has a significantly different distribution for the all characters. The L strategy differs significantly for panicle compactness, glume covering, grain color, and endosperm texture.

The distribution of landraces among the continents is not significantly different between the three subsets and the entire landrace collection (Table 3). For the latitudinal distribution, the PCS subset and the entire landrace collection does not differ, while in the L subset the representation in the four latitudinal classes was different. The T subset differs from the entire landrace collection for all the latitude classes except the tropical latitude (latitude class 2).

Considering the racial distribution, only two races, guinea race and CB forms, are not distributed differently between the entire landrace collection and the three subsets. The DB-GB-GD group was significantly reduced in the PCS subset while the representation of the GC form is increased. The L subset differs significantly from the entire landrace collection; the representation of bicolor and kafir races as well as the GC forms and the kafir intermediates is increased, while the representation of durra race is decreased. The representation of bicolor, durra and kafir races, as for those of kafir intermediates, increase in the T subset but the representation of the caudatum race and DC form decrease.

Comparison between the subsets

The three subsets were compared for the ten quantitative characters using the variance and mean comparison tests (Table 1). When compared to the L subset, the PCS subset has a larger variance for plant height in rabi and seed

Table 1 Descriptive statistics for the ten quantitative characters; mean (±SE) and variance for the entire landrace collection (ELC) and the three subsets (principal component score strategy: PCS, logarithmic strategy: L, and taxonomic strategy: T). Levene's test of homogeneity of variances and Newman and Keuls' test for a means comparison (A) between each subset and the entire landrace collection, and (B) between the subsets

Sets	Flowering date in rabi (days)	Flowering date in kharif (days)	Basal tillers (no.)	Plant height in rabi (cm)	Plant height in kharif (cm)	Peduncle exertion (cm)	Panicle length (cm)	Panicle width (cm)	Grain size (mm)	100-seed weight (g)
(A)										
ELC 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
var.	216.56	845.34	0.89	3045.29	7921.58	86.54	73.84	21.89	0.27	1.06
PCS mean±SE	72.43±1.01	93.45±1.90	2.10±0.06	220.33±4.13	334.50±5.99	15.41±0.58	21.47±0.54	9.21±0.28	2.95±0.03	3.18±0.08
var.	234.04	824.24	0.73	3905.92	8216.08	77.51	66.20	18.07	0.24	1.38
Levene	NS	NS	NS	*	NS	NS	NS	NS	NS	*
N–K	NS	NS	NS	–	NS	NS	NS	NS	*	–
L mean±SE	71.24±0.92	84.98±1.98	2.08±0.05	223.89±3.52	306.31±6.35	16.40±0.61	24.47±0.51	9.92±0.34	2.96±0.03	2.93±0.06
var.	191.15	882.62	0.63	2786.48	9068.47	84.89	57.67	26.41	0.20	0.68
Levene	NS	NS	*	NS	*	NS	**	NS	**	**
N–K	NS	**	–	NS	–	NS	–	NS	–	–
T mean±SE	75.13±0.80	100.62±1.96	1.92±0.05	221.30±3.23	334.65±6.17	14.19±0.55	23.25±0.60	9.79±0.31	3.26±0.04	3.33±0.08
var.	156.39	948.29	0.70	2566.46	9372.84	73.97	88.17	23.55	0.37	1.60
Levene	*	*	NS	*	NS	NS	*	NS	**	**
N–K	–	–	**	–	NS	**	–	NS	–	–
(B)										
PCS vs L										
Levene	NS	NS	NS	*	NS	NS	NS	NS	NS	**
N–K	NS	**	NS	–	**	NS	**	NS	NS	–
PCS vs T										
Levene	*	*	NS	**	NS	NS	*	NS	**	NS
N–K	–	–	*	–	NS	NS	–	NS	–	NS
L vs T										
Levene	NS	NS	NS	NS	NS	NS	**	NS	**	**
N–K	**	**	NS	NS	**	*	–	NS	–	–

NS: non-significant test at $P>5\%$ level; *, ** significant at the $P<0.05$ and $P<0.01$ levels, respectively

Table 2 Comparison for the frequency distribution for the qualitative characters through a χ^2 test (A) between the three subsets (principal component score strategy: PCS, logarithmic strategy: L, and taxonomic strategy: T) and the entire landrace collection (ELC), and (B) between the three subsets

Qualitative characters	No. of modalities	(A) Comparison to the entire landrace collection			(B) Comparison between subsets		
		PCS vs ELC	L vs ELC	T vs ELC	PCS vs L	PCS vs T	L vs T
Midrib color	2	3.22 NS	0.34 NS	28.06**	3.04 NS	20.82**	8.10**
Panicle compactness	4	0.62 NS	15.39**	16.73**	7.88*	7.76 NS	27.22**
Glume color	9	5.43 NS	15.07 NS	24.96**	11.01 NS	14.76 NS	23.64**
Glume covering	5	1.92 NS	31.68**	31.66**	11.68*	14.67**	13.86**
Grain color	7	4.46 NS	16.96**	55.91**	9.13 NS	36.86**	47.80**
Endosperm texture	5	1.59 NS	13.21*	15.32**	11.16*	9.43 NS	7.37 NS
Grain lustre	2	0.00 NS	0.94 NS	30.59**	0.52 NS	16.83**	11.31**
Subcoat presence	2	0.00 NS	0.13 NS	28.33**	0.06 NS	16.26**	18.07**

NS: non-significant test at $P>5\%$ level; *, ** significant at the $P<0.05$ and $P<0.01$ levels, respectively

Table 3 Frequency for the passport data and comparison between the entire landrace collection (ELC) and the subsets (PCS, L and T), and between the three subsets. The frequency distributions were compared through a chi-square test. The same letters showed no significant differences at the $P<0.05$ level. The number of accessions is given in brackets

Set	ELC		PCS		L		T	
Africa	(15 652)	a	(162)	a	(151)	a	(171)	a
Asia	(6329)	a	(64)	a	(56)	a	(66)	a
Latitude 1: 10°N–10°S	(8919)	a	(102)	a	(47)	b	(75)	c
Latitude 2: 20°N–20°S	(10 583)	a	(96)	a	(69)	b	(102)	a, b
Latitude 3: 30°N–30°S	(1843)	a	(23)	a, c	(64)	b	(37)	c
Latitude 4: 40°N–40°S	(1128)	a	(8)	a	(45)	b	(32)	b
bicolor	(718)	a	(6)	a, c	(16)	b	(14)	b, c
caudatum	(4893)	a	(54)	a	(37)	a, b	(34)	b
durra	(4545)	a	(42)	a, b	(22)	c	(66)	b
guinea	(3907)	a	(33)	a	(27)	a	(43)	a
kafir	(520)	a	(7)	a	(23)	b	(21)	b
CB	(760)	a	(6)	a	(13)	a	(10)	a
DB–GB–GD	(1620)	a	(7)	b	(12)	a, b	(13)	a, b
DC	(2713)	a	(35)	a	(23)	a, b	(19)	b
GC	(2399)	a	(35)	b	(39)	b	(16)	a
Intermediate kafir	(398)	a	(4)	a, c	(13)	b	(10)	b, c

weight, a higher mean value for days to flowering and plant height in kharif, and a smaller mean value for the panicle length. Comparison between the PCS and the T subsets shows a greater variance for days to flowering and plant height in rabi in the PCS subset. The T subset has a greater variance for days to flowering in kharif, panicle length and grain size. When compared to the L subset, the T subset has a higher variance for panicle length, grain size and 100-seed weight; a higher mean value for days to flowering in both rabi and kharif, plant height in kharif, and a smaller mean value for peduncle exertion.

Qualitative characters were also compared between each subset by a chi-square test (Table 2). The distribution for three characters is significantly different between the PCS and the L subsets (panicle compactness, glume covering and endosperm texture). When the T and PCS subsets are compared, five characters have a different distribution (midrib color, glume covering, grain color, grain lustre and subcoat presence). Comparison between the T and the L subsets reveals a different distribution for all characters except endosperm texture. Only the glume covering has a significantly different distribution for every subset comparison. Also, the L subset differs from the two other subsets for panicle compactness, and the T subset differs significantly from the two other subsets for midrib color, grain color, grain lustre and subcoat presence.

The distribution of accessions within the continent of origin is similar between the three subsets, while the latitudinal distribution differs between each subset for the equatorial latitude (latitude class 1). Moreover, the PCS subset differs significantly from the two other subsets for the highly temperate latitudes (latitude class 4), and the L subset differs significantly from the two other subsets for the temperate latitude (latitude class 3).

The racial distribution is not significantly different between the three subsets for the guinea race, the DB-GB-GD group and the CB forms. However, when compared to the two other subsets, the representation of the kafir race is reduced in the PCS subset. Also the PCS subset has a smaller representation of the bicolor race and kafir intermediates than the L subset, as well as a higher representation of the caudatum race and the DC forms than in the T subset. The L subset has a smaller number of accessions classified as the durra race than the PCS and T subsets. The T subset contrasts from the two other subsets, as its representation of the GC forms is not increased.

Discussion

Global diversity, as depicted through multivariate analysis of the ten continuous characters recorded on the entire landrace collection, demonstrates that the global morpho-agronomic diversity is well represented in the three subsets and that the morpho-agronomic diversity is not influenced by the random or non-random sampling strategies. These results are supported by the high and similar values of SDI calculated from the entire landrace collection and the three subsets.

However, comparative analyses for each character between the three subsets and the entire landrace collection illustrate several differences. The PCS strategy results in a subset with no change in the variances or means for most of the quantitative characters from that of the entire landrace collection. This observation can be explained by the normal distribution of the characters in the entire landrace collection. However, the PCS sampling modifies the distribution patterns of the quantitative characters as the curves are flattened along the X-axis in comparison to the entire landrace collection. Nevertheless, it is noteworthy that this sampling strategy does not affect the distribution of the characters that are not considered in the statistical procedure. In our study, this result is verified irrespective of the type of character, i.e. qualitative morpho-agronomic character or passport data. Few reports on the use of the PCS strategy have been published. The application of this strategy to two perennial trees (rubber tree and coffea) and two annual crops (rice and sorghum) found that the distributions of the qualitative characters in the subsets were not modified to any significant degree while the quantitative characters demonstrated changes in the variances and means (Hamon et al. 1998). When the distribution was considered, for example in coffee, whatever the shape of the curve in the base collection, the effect of the PCS sampling was characterized by a more or less flattened curve. This change in the pattern of distribution was most likely due to less redundancy in this core. Thus, the PCS strategy permitted both the elimination of redundancies and the retention of the diversity of the entire landrace collection in a subset of manageable size (1% of the entire landrace collection in our work).

Second, when the logarithmic (L subset) or empirical (T subset) are compared, several or all characters (morpho-agronomic characters, either quantitative or qualitative, as well as passport data) have a modified distribution from the entire landrace collection. These discrepancies between the three subsets could be explained by two hypotheses. The first is that the differences are due to the sampling strategy itself and reflect a bias in the sampling. In this case, it would be possible to observe differences for any other set of data considered (resistance to biotic or abiotic stress, biochemical or molecular markers). The other hypothesis is that these differences are due to the method used for the constitution of these particular subsets. This second hypothesis looks the most likely especially when the L sampling strategy is considered. Indeed, the L subset was constituted from the landrace collection after it had been stratified into four clusters according to photoperiod sensitivity. The characters with the modified statistical parameters are those associated with the photoperiod sensitivity. For example, variance for plant height recorded during the kharif season is increased while the mean value is decreased. The mean value for days to flowering recorded during the kharif season is also reduced. Thus, the representation of early and short plant types was increased in the L subset. The sampling was also biased toward landraces with loose panicles and stiff branches which are the main characteristics for the varieties classified as either bicolor or kafir races. The proportions of the bicolor, kafir and the kafir intermediates, as well as the proportion of the temperate material, are higher in the L subset than in the entire landrace collection. These races and latitudinal class are mainly composed of photoperiod-insensitive landraces and represent only a small portion among the entire landrace collection (Grenier et al. 2000). Thus, the representation of photoperiod-insensitive landraces is clearly increased in the L subset in relation to the entire landrace collection.

Differences between the subset and the entire landrace collection in the representation of the morpho-agronomic diversity, the geographical distribution and the race classification are even more marked in the T subset. This subset was defined from empirical knowledge based on many years of fieldwork. The criteria integrated a very wide set of characteristics, i.e. phenotypic characters, eco-environmental adaptation, geographic origin and race classification. As all these traits were considered together in the selective sampling, this subset was expected to differ from the entire landrace collection. This expectation is indeed observed since, when compared to the entire landrace collection, most of the quantitative characters have a heterogeneous variance and all the qualitative characters are differently distributed. In addition, racial and geographical distributions are greatly modified in this subset.

In conclusion, a random strategy (L subset) and a non-random strategy (PCS and T subsets) seem to retain a good representation of the overall diversity from the entire landrace collection. However, dependent upon the sampling procedure, the statistical parameters of some or all characters are modified. Two hypotheses, bias introduced during the sampling and bias due to the specific constitution of the subsets, are proposed to explain our results. A further assessment to compare the genetic diversity of the three subsets with the use of molecular markers, such as microsatellites, will be undertaken. Furthermore, it would be worthwhile to consider non-neutral characters for the sampling of core collections and neutral characters (such as molecular markers) to estimate the effect of the sampling.

Acknowledgements The authors thank Dr. Rodomiro Ortiz at the ICRISAT Genetic Resources Enhancement Program, Patancheru, India, and Dr. Serge Hamon at the Institut de Recherche pour le Developpement Montpellier, France, for their valuable comments on the manuscript. The experiments comply with the current laws of the country in which the experiments were performed.

References

Appa Rao S, Prasada Rao KE, Mengesha MH, Gopal Reddy V (1996) Morphological diversity in sorghum germplasm from India. Genetic Resour Crop Evol 43:559–567

Balfourier F, Charmet G, Prosperi J-M, Goulard M, Monestiez P (1998) Comparison of different spatial strategies for sampling a core collection of natural populations of fodder crops. Genet Selec Evol 30 (Suppl 1):S215–S235

Bataillon T (1994) Comparaison de diverses stratégies d'échantillonnage pour la constitution de *core collections* de ressources génétiques végétales (études par simulation informatique). Mémoire de DEA, Institut National Agronomique, Paris

Bisht IS, Mahajan RK, Patel DP (1998) The use of characterisation data to establish the Indian mungbean core collection and assessment of genetic diversity. Genetic Resour Crop Evol 45:127–133

Brown AHD (1989a) The case for core collections. In: Brown AHD, Frankel OH, Marshall DR, Williams JT (eds) The use of plant genetic resources. Cambridge University Press, Cambridge, pp 135–156

Brown AHD (1989b) Core collections: a practical approach to genetic resources management. Genome 31:818–824

Cordeiro CMT, Morales EAV, Ferreira P, Rocha DMS, Costa IRS, Valois ACC, Silva S (1995) Towards a Brazilian core collection of cassava. In Hodgkin T, Brown ADH, van Hintum TJL, Morales EAV (eds) Core collections of plant genetic resources. John Wiley and Sons, Chichester, UK, pp 155–168

Diwan N, McIntosh MS, Bauchan GR (1995) Methods of developing a core collection of annual *Medicago* species. Theor Appl Genet 90:755–761

Galwey NW (1995) Verifying and validating the representativeness of a core collection. In Hodgkin T, Brown ADH, van Hintum TJL, Morales EAV (eds) Core collections of plant genetic resources. John Wiley and Sons, Chichester, UK, pp 187–198

Grenier C, Bramel-Cox PJ, Hamon P (2000a) Core collection of the genetic resources of sorghum. I. Stratification based on eco-geographical data. Crop Sci (in press)

Grenier C, Hamon P, Bramel-Cox PJ (2000b) Core collection of the genetic resources of sorghum. II. Comparison of three random sampling procedures. Crop Sci (in press)

Hamon S, Dussert S, Deu M, Hamon P, Seguin M, Glaszmann J-C, Grivet L, Chantereau J, Chevallier M-H, Flori A, Lashermes P, Legnate H, Noirot M (1998) Effects of quantitative and qualitative principal component score strategies on the structure of coffee, rubber tree, rice, and sorghum core collections. Genet Selec Evol 30 (Suppl 1):S237–S258

Harlan JR, de Wet JMJ (1972) A simplified classification of cultivated sorghum. Crop Sci 12:172–176

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

Igartua E, Gracia MP, Lasa JM, Medina B, Molina-Cano JL, Montoya JL, Romagosa I (1998) The Spanish barley core collection. Genet Resour Crop Evol 45:475–481

Mahajan RK, Bisht IS, Agrawal RC, Rana RS (1996) Studies on South Asian okra collection: methodology for establishing a representative core set using characterization data. Genet Resour Crop Evol 43:249–255

Noirot M, Hamon S, Anthony F (1996) The principal component scoring: a new method of constituing a core collection using quantitative data. Genetic Resourc Crop Evol 43:1–6

Spagnoletti Zeuli PL, Qualset CO (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. StatSoft, USA

THEORETICAL AND APPLIED GENETICS

International Journal of Plant Breeding Research

Volume 101 Number 1–2 July 2000

Aragao FJL, Sarokin L, Vianna GR, Rech EL: **Selection of transgenic meristematic cells utilizing a herbicidal molecule results in the recovery of fertile transgenic soybean [*Glycine max* (L.) Merril] plants at a high frequency** 1

Schmidt A, Doudrick RL, Heslop-Harrison JS, Schmidt T: **The contribution of short repeats of low sequence complexity to large conifer genomes** 7

Tu J, Datta K, Khush GS, Zhang Q, Datta SK: **Field performance of *Xa21* transgenic indica rice (*Oryza sativa* L.), IR72** 15

Yamamori M, Fujita S, Hayakawa K, Matsuki J, Yasui T: **Genetic elimination of a starch granule protein, SGP-1, of wheat generates an altered starch with apparent high amylose** 21

Chen CC, Chen CM, Hsu FC, Wang CJ, Yang JT, Kao YY: **The pachytene chromosomes of maize as revealed by fluorescence in situ hybridization with repetitive DNA sequences** 30

Seo B-S, Peterson PA: **Transposition of the *En/Spm* transposable element system in maize (*Zea mays* L.): reciprocal crosses of *a1-m*(*Au*) and *a1-m*(r) alleles uncover developmental patterns** 37

Brutovska R, Cellarova E, Schubert I: **Cytogenetic characterization of three *Hypericum* species by in situ hybridization** 46

Friebe B, Qi LL, Nasuda S, Zhang P, Tuleen NA, Gill BS: **Development of a complete set of *Triticum aestivum-Aegilops speltoides* chromosome addition lines** 51

Hayano-Saito Y, Saito K, Nakamura S, Kawasaki S, Iwasaki M: **Fine physical mapping of the rice stripe resistance gene locus, *Stub-i*** 59

Sobir, Ohmori T, Murata M, Motoyoshi F: **Molecular characterization of the SCAR markers tightly linked to the *Tm-2* locus of the genus *Lycopersicon*** 64

Mengistu LW, Mueller-Warrant GW, Barker RE: **Genetic diversity of *Poa annua* in western Oregon grass seed crops** 70

Suzuki G, Watanabe M, Nishio T: **Physical distances between *S*-locus genes in various *S* haplotypes of *Brassica rapa* and *B. oleracea*** 80

Popovsky S, Paran I: **Molecular genetics of the *y* locus in pepp: its relation to capsanthin-capsorubin synthase and to fruit color** 86

Charmet G, Bert PF, Balfourier F: **A computerised algorithm for selecting a subset of multiplex molecular markers and optimising linkage map construction** 90

Pestsova E, Salina E, Börner A, Korzun V, Maystrenko OI, Röder MS: **Microsatellites confirm the authenticity of inter-varietal chromosome substitution lines of wheat (*Triticum aestivum* L.)** 95

Pestsova E, Korzun V, Goncharov NP, Hammer K, Ganal MW, Röder MS: **Microsatellite analysis of *Aegilops tauschii* germplasm** 100

Amand PCSt, Skinner DZ, Peaden RN: **Risk of alfalfa transgene dissemination and scale-dependent effects** 107

Varghese JP, Rudolph B, Uzunova MI, Ecke W: **Use of 5′-anchored primers for the enhanced recovery of specific microsatellite markers in *Brassica napus* L.** 115

Scotti C, Pupilli F, Salvi S, Arcioni S: **Variation in vigour and in RFLP-estimated heterozygosity by selfing tetraploid alfalfa: new perspectives for the use of selfing in alfalfa breeding** 120

Tai GCC, Seabrook JEA, Aziz AN: **Linkage analysis of anther-derived monoploids showing distorted segregation of molecular markers** 126

Balfourier F, Imbert C, Charmet G: **Evidence for phylogeographic structure in *Lolium* species related to the spread of agriculture in Europe. A cpDNA study** 131

Buso JA, Boiteux LS, Tai GCC, Peloquin SJ: **Direct estimation of the effects of meiotic recombination on potato traits via analysis of 4*x*-2*x* progenies from synaptic mutants with 2n-pollen formation by FDR without crossing-over** 139

Negi MS, Devic M, Delseny M, Lakshmikumaran M: **Identification of AFLP fragments linked to seed coat colour in *Brassica juncea* and conversion to a SCAR marker for rapid selection** 146

Fischer D, Bachmann K: **Onion microsatellites for germplasm analysis and their use in assessing intra- and interspecific relatedness within the subgenus *Rhizirideum*** 153

Diwan N, Bouton JH, Kochert G, Cregan PB: **Mapping of simple sequence repeat (SSR) DNA markers in diploid and tetraploid alfalfa** 165

(Continued on page A 3)

0040-5752(200007)101:1-2;1-P