Theor Appl Genet (2000) 100:918–925

Y. Djè · M. Heuertz · C. Lefèbvre · X. Vekemans

Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers

Received: 10 August 1999 / Accepted: 27 August 1999

Abstract Microsatellite markers are increasingly being used in crop plants to discriminate among genotypes and as tools in marker-assisted selection. Here we evaluated the use of microsatellite markers to quantify the genetic diversity within as well as among accessions sampled from the world germplasm collection of sorghum. Considerable variation was found at the five microsatellite loci analysed, with an average number of alleles per locus equal to 2.4 within accessions and 19.2 in the overall sample of 25 accessions. The collection of sorghum appeared highly structured genetically with about 70% of the total genetic diversity occurring among accessions. However, differentiation among morphologically defined races of sorghum, or among geographic origins, accounted for less than 15% of the total genetic diversity. Our results are in global agreement with those obtained previously with allozyme markers. We were also able to show that microsatellite data are useful in identifying individual accessions with a high relative contribution to the overall allelic diversity of the collection.

Key words Core collection · Genetic diversity · Germplasm · Microsatellites · Sorghum · SSR

Introduction

Methods of measuring genetic diversity have an important role within conservation programmes for genetic resources of crop plants (Newbury and Ford-Lloyd 1997). In the context of ex situ conservation of germplasm collections, methods based on genetic markers are used to address questions of identity, for instance to identify putative duplicate accessions, and questions of relationship and structure, for instance to determine how variation is distributed between individuals, accessions and races (Westman and Kresovich 1997). Information provided by the genetic markers is then used to design optimal procedures to manage extensive germplasm collections, in particular to highlight priorities in further sampling missions, to design germplasm regeneration programmes and to construct core collections. Many studies have aimed at assessing the genetic diversity in germplasm collections of crops using allozyme markers (e.g. Morden et al. 1989; Maquet et al. 1997) or various types of molecular markers such as restriction fragment length polymorphism (RFLP, Cui et al. 1995; Dubreuil and Charcosset 1998), random amplified polymorphic DNA (RAPD, Fofana et al. 1997), amplified fragment length polymorphism (AFLP, Tohme et al. 1996; Zhu et al. 1998). Recently, microsatellite or SSR (simple sequence repeat) loci, which correspond to tandemly repeated DNA with a very short repeat unit, have been introduced as powerful genetic markers in plants (Morgante and Olivieri 1993; Powell et al. 1996a). Comparative studies in crop plants have shown that microsatellite markers are more variable than most other molecular markers (Powell et al. 1996b; Taramino and Tingey 1996; Pejic et al. 1998) and provide a powerful methodology for discriminating between genotypes (Yang et al. 1994; Russell et al. 1997; Bredemeijer et al. 1998). Despite an explosive increase in the use of microsatellite markers in investigations on crop plants in the last 3 years, the number of individuals screened in any study remains low due to the low throughput of data acquisition techniques (Donini et al. 1998). In particular, we are aware of no single study that has assessed the partition of genetic diversity within and among germplasm accessions of cultivated crops based on microsatellites.

Our global aim is to evaluate the potential of the microsatellite technique to investigate the population genetic

Communicated by P.M.A. Tigerstedt

Y. Djè · C. Lefèbvre · X. Vekemans (✉)
Université Libre de Bruxelles,
Laboratoire de Génétique et d'Ecologie végétales,
1850 chaussée de Wavre, B-1160 Brussels, Belgium
e-mail: xvekema@ulb.ac.be

M. Heuertz
CREBS Research Unit
Centre de recherche public,
162a avenue de la faïencerie, L-1511 Luxembourg, Luxemburg

Table 1 List of sorghum germplasm accessions provided by ICRISAT

	Race	Code	Pedigree	Country	Geographic origin
b1	bicolor	11010	PI 329252	Ethiopia	Eastern Africa
b2	bicolor	21351		Kenya	Eastern Africa
b3	bicolor	21862	HDW 306	Philippines	Asia
b4	bicolor	22132	PG 122	India	Asia
b5	bicolor	24503		South Africa	Southern Africa
c1	caudatum	12123	PI 330909	Ethiopia	Eastern Africa
c2	caudatum	17589	Scheihe	Yemen	Asia
c3	caudatum	21361		Kenya	Eastern Africa
c4	caudatum	24792	Kubari	Nigeria	Western Africa
c5	caudatum	31504	US 661, Katome	Uganda	Eastern Africa
d1	durra	11114	PI 329357	Ethiopia	Eastern Africa
d2	durra	17803	KEP 144, Palapu Jonna	India	Asia
d3[a]	durra	20830	Irudan (Fodder)	USA	America[a]
d4	durra	23087	CER 19	Tanzania	Eastern Africa
d5	durra	27182	TGR 1126	Zimbabwe	Southern Africa
g1	guinea	7772	Doro Zabo	Nigeria	Western Africa
g2	guinea	21530		Malawi	Southern Africa
g3	guinea	27731	PCI 100	Sierra Leone	Western Africa
g4	guinea	30843		Cameroon	Western Africa
g5	guinea	35180		Togo	Western Africa
k1	kafir	14285		Botswana	Southern Africa
k2	kafir	20851	Kraspje	Madagascar	Southern Africa
k3	kafir	26696	Kwazulu Engini: 1963	South Africa	Southern Africa
k4	kafir	27076	TGR 210 C	Zimbabwe	Southern Africa
k5	kafir	30166	AMM 1154	Zimbabwe	Southern Africa

[a] Not used when classifying accessions according to geographic origin

structure of the world genebank collection of cultivated sorghum. Previous studies on sorghum based on allozyme markers have shown that genetic diversity is lower than in other cereals like maize and barley and that genetic differentiation is high among individual accessions but low among the five races defined on the basis of morphological characteristics (Morden et al. 1989; Ollitrault et al. 1989; Aldrich et al. 1992). Recently, several microsatellite loci have been identified in sorghum (Brown et al. 1996; Taramino et al. 1997). We have previously used three of these loci to assess genetic variation in sorghum landraces from North-western Morocco based on direct on-farm samples (Djè et al. 1999). Here we report on our study of samples from the world-wide collection of cultivated sorghum germplasm using five microsatellite loci that were characterised with an automated fluorescence-based detection method. We specifically attempted to quantify genetic diversity within and among accessions, races or geographic origins. We also evaluated the potential of microsatellites to identify individual accessions with a high relative contribution to the overall allelic diversity of the collection. We compared our results to those obtained by Morden et al. (1989) using allozyme markers.

Materials and methods

Plant material

Twenty-five accessions from the world germplasm collection of cultivated sorghum were obtained from ICRISAT (International Crops Research Institute for the Semi-Arid Tropics). The accessions were chosen to represent the five main races of cultivated sorghum, with 5 accessions sampled from each race, as well as a wide geographic distribution (Table 1). Ten seeds per accession were allowed to germinate in a dark chamber at 22°C, and seedlings were transferred to individual pots into the greenhouse. Leaves were taken for DNA extraction after about 4 weeks.

Microsatellite analysis

A total of 100—150 mg of young leaves was used to extract genomic DNA following a CTAB procedure (Doyle and Doyle 1990). Five primer pairs of microsatellite loci which had previously been shown to display a high degree of polymorphism in sorghum were used in the amplification reactions (Brown et al. 1996; Taramino et al. 1997; see Table 2). Polymerase chain reaction (PCR) conditions were optimised by adapting annealing temperature (T_m), $MgCl_2$ concentration and the amount of template DNA. The PCR reactions were performed in a mix containing 2.5 m*M* $MgCl_2$, 0.5 unit per reaction of AmpliTaq Gold™ DNA polymerase in buffer II from Perkin Elmer Applied Biosystems and approximately 10 ng template DNA (2.5 µl), in a total reaction volume of 25 µl. PCR cycling conditions were: a 10-min initial denaturation at 95°C; 34 cycles of amplification of 30 s at 94°C, 45 s at either 55°C (Sb4–22, Sb5–236 and SbAGH-04) or 53°C (SbAGA-01 and SbAGE-01) and a 1-min elongation at 72°C; final elongation of 5 min at 72°C. All PCR reactions were performed on a Mastercycler® Gradient 2331 version 1.2 (Eppendorf). The forward sequence of each primer pair was labelled with a fluorescent dye (from Perkin Elmer Applied Biosystems) at its 5' end: 6-FAM for Sb5–236 and Sb4–22; HEX for SbAGA-01 and SbAGE-01; and NED for SbAGH-04. The dyes were chosen in order to distinguish among PCR products of similar sizes from different loci. This allowed the pooling of PCR products from all loci in each well after appropriate adjustment of their relative concentration. An internal size standard (Genescan-350 with ROX dye) was also loaded in each well. Electrophoresis and detection of PCR products were carried out on denaturing polyacrylamide gels (5% Long Ranger; 36 cm) using an ABI PRISM® 377 DNA sequencer from Perkin Elmer with filter set D. Gels were run for 2 h at 3000 V in TBE buffer. The resulting electrophoregrams were analysed with the software GENOTYPER® 2.0 from Perkin Elmer.

Table 2 Characteristics of the microsatellite loci

Locus	Original publications			This study	
	Repeat motif	Number of alleles	Range of sizes (bp)	Number of alleles	Range of sizes (bp)
SbAGA-01	(AG)33	6	88–106	18	82–116
SbAGH-04[a]	(AG)39	8	110–170	24	111–161
Sb5—236[b]	(AG)20	3	165–185	17	157–197
SbAGE-01[a]	(AG)30	9	208–240	23	188–272
Sb4—22[b]	(ACGAC)4/(AG)6	5	270–300	14	293–333

[a] From Taramino et al. (1997)
[b] From Brown et al. (1996)

Estimation of genetic diversity

Analyses of genotypic data were performed using the computer programme GEN-SURVEY (Vekemans and Lefèbvre 1997). The genotypic structure of the overall sample of accessions was analysed with *F*-statistics computed according to Weir and Cockerham (1984), with standard deviations estimated by jack-knife (Weir 1990). Genetic diversity within accessions was estimated using four statistics averaged over loci: the percentage of polymorphic loci at the 5% level, *P*; the mean number of alleles per locus or allelic richness, *A*; the average observed heterozygosity, H_o; and the average gene diversity, H_e, computed according to Nei (1978). These statistics were averaged among accessions belonging either to the same race or having similar geographic origin. Two-level hierarchical gene diversity analyses were performed according to Nei and Chesser (1983) to estimate the partition of genetic diversity within and among accessions for each race and geographic origin. We also estimated differentiation among races and geographic origins using three-level hierarchical gene diversity analyses (within accession, among accessions within group and among groups). As a measure of genetic distance among pairs of accessions, we computed the standardised R_{ST} statistic (Slatkin 1995) with the programme MICROSAT ver. 1.5 d (distributed by Eric Minch, http://lotka.stanford.edu/microsat.html). The matrix of pairwise distances was used as input for a principal coordinate analysis using the package NTSYS-PC ver. 1.8 (Rohlf 1993).

Orientation of sampling strategies

Two methods were used to attempt identification of individual accessions with a high relative contribution to the overall allelic diversity in the collection. First, we evaluated the relative contribution (C^A_T) of each accession to total allelic richness and its partition into two components (one due to the level of allelic diversity within the accession; the other due to its divergence from all other accessions) according to Petit et al. (1998). To standardise the results of allelic richness across accessions we used the technique of rarefaction, which evaluates the expected number of different alleles among equal-sized samples drawn from the accessions (El Mousadik and Petit 1996). Analyses were done with a sample size set to ten gene copies per accession. The statistic C^A_T for accession *k* is computed as $(A_T\text{-}A_{T\backslash k})/A_T$, where A_T is the total allelic richness of the collection and $A_{T\backslash k}$ is the overall allelic richness obtained when removing accession *k* from the collection. The contribution to total allelic richness due to allelic diversity within accession *k* is computed as $(A_k\text{-}A_{S\backslash k})/(n.A_T)$, where A_k is the allelic richness within accession *k*, $A_{S\backslash k}$ is the mean over all accessions excluding *k* of the within accession allelic richness and *n* is the number of accessions (Petit et al. 1998). The contribution due to divergence of accession *k* from the whole collection is obtained as C^A_T minus the within-accession component computed as described above. Analyses were performed with a software provided by R. Petit. The second approach was to apply a specific algorithm derived for devising core collection strategies, the M (for maximisation) strategy introduced by Schoen and Brown (1995), to select a subset of *s* highly variable and well-differentiated accessions that represent at best the allelic diversity in the overall collection. The method essentially tries every subset of *s* accessions from the collection, with the condition that at least 1 accession from each of several predefined groups is chosen, and determines that/those with the highest total number of alleles. Because we found little differentiation among either races or geographical origins (see Results), we modified the original algorithm by removing the use of predefined groups. In our analyses, we used a value of $s = 5$. A computer programme performing this analysis is available from the last author (xvekema@ulb.ac.be).

Results

Characterisation of microsatellite loci

Figure 1 shows a typical electrophoregram obtained with multiple sample loading. Optimal separation between the five loci was obtained by combining differences in fragment sizes and in fluorescent dyes. The total number of putative alleles at each locus and the size ranges of these alleles are given as observed in this study and as reported in the original publications that first described the corresponding microsatellite loci (Table 2). It appears that all five loci scored in this study are highly polymorphic with many alleles (from 14–24 putative alleles) and a wide range of product sizes (differences between the longest and shortest alleles ranging from 34 to 84 bp). In all cases the observed number of alleles is much higher than reported in the original publications (Brown et al. 1996; Taramino et al. 1997), which is due to the larger number and wider geographic origin of the accessions used here. Also, for most loci the size range of PCR products obtained here is substantially wider than that intially reported. A similar observation was reported in wheat and barley by Donini et al. (1998).

Genotypic structure of the germplasm collection

F-statistics over the 25 accessions are given for each microsatellite locus as well as their averages over loci (Table 3). Altogether very little variation was observed among loci with respect to the values of these *F*-statistics. The overall inbreeding coefficient, F_{IT}, has a mean

Table 3 *F*-statistics over the 25 accessions of sorghum for each microsatellite locus; F_{IT}, overall inbreeding coefficient; F_{ST}, fixation index; F_{IS}, average inbreeding coefficient within accessions. Mean values are given ± standard deviation (SD) estimated by jack-knife

Locus	F_{IT}	F_{ST}	F_{IS}
SbAGA-01	0.868 ± 0.043	0.629 ± 0.066	0.645 ± 0.106
SbAGH-04	0.900 ± 0.029	0.677 ± 0.063	0.691 ± 0.076
Sb5–236	0.854 ± 0.034	0.744 ± 0.042	0.434 ± 0.114
SbAGE-01	0.813 ± 0.044	0.658 ± 0.059	0.454 ± 0.093
Sb4–22	0.857 ± 0.032	0.719 ± 0.045	0.496 ± 0.102
Mean ± SD over loci	0.858 ± 0.015	0.683 ± 0.020	0.553 ± 0.054

Table 4 Genetic diversity within and among accessions of sorghum classified either by race or by geographic origin, and for the overall sample of 25 accessions

Groups of accessions	P[a]	A[b]	H_o[c]	H_e[d]	H_T[e]	G_{ST}[f]
Classification by race						
Bicolor	64.0	2.40	0.113	0.286	0.847	0.656
Guinea	72.0	2.24	0.089	0.224	0.742	0.689
Durra	56.0	2.24	0.071	0.260	0.801	0.661
Caudatum	80.0	2.64	0.178	0.364	0.782	0.520
Kafir	68.0	2.64	0.218	0.313	0.677	0.540
Classification by geographic origin						
Eastern Africa	88.6	2.97	0.159	0.395	0.876	0.535
Western Africa	72.0	2.36	0.104	0.260	0.743	0.639
Southern Africa	57.5	2.25	0.152	0.244	0.791	0.684
Asia	50.0	2.00	0.113	0.237	0.745	0.706
Overall mean	68.0	2.43	0.134	0.290	0.897	0.668
SD[g]	28.3	0.74	0.092	0.142	0.033	0.051

[a] Percentage of polymorphic loci at the 5% level within accession
[b] Mean number of alleles per locus within accession
[c] Average observed heterozygosity within accession
[d] Average gene diversity within accession
[e] Total gene diversity
[f] Proportion of among-accessions differentiation
[g] Standard deviation

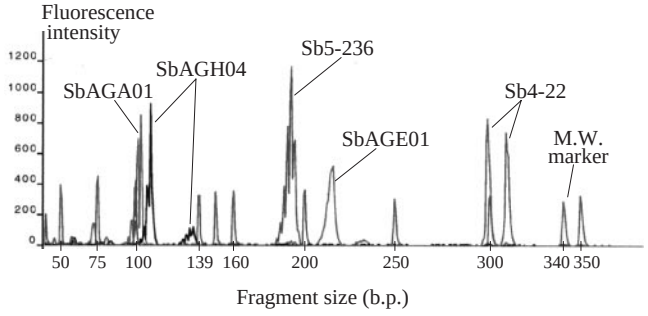

Fig. 1 Typical electrophoregram produced by the automated DNA sequencer after multiple sample loading

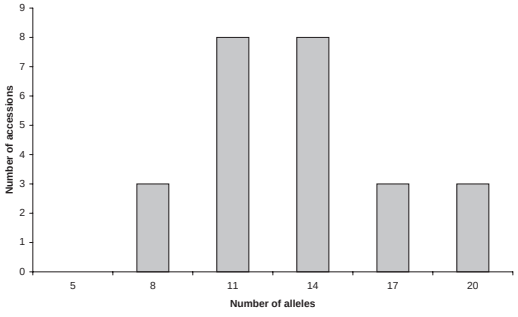

Fig. 2 Distribution among accessions of the number of alleles recorded within each accession, expressed as a sum of alleles over the five loci

(± SD) of 0.86 ± 0.02, showing that the sorghum germplasm collection is highly inbred overall. This inbreeding is due in part to inbreeding within individual accessions, with a mean value F_{IS} = 0.55 ± 0.05, and in part to strong genetic differentiation among accessions as shown by a mean fixation index of F_{ST} = 0.68 ± 0.02.

Genetic diversity within and among accessions

Statistics of genetic diversity within and among accessions are given in Table 4 for the overall sample of 25 accessions as well as for each group of accessions classified either by race or by geographic origin. High genetic polymorphism for microsatellite loci is observed within accessions as, on average, more than two-thirds of the loci are polymorphic for a given accession (mean percentage of polymorphic loci, P=68.0%), more than 2 alleles per locus per accession are found on average (A=2.4) and the probability that 2 randomly sampled alleles in a given accession are different is higher than a quarter (H_e = 0.29). The distribution among accessions of the total number of alleles recorded within each accession over the five microsatellite loci is shown in Fig. 2. It

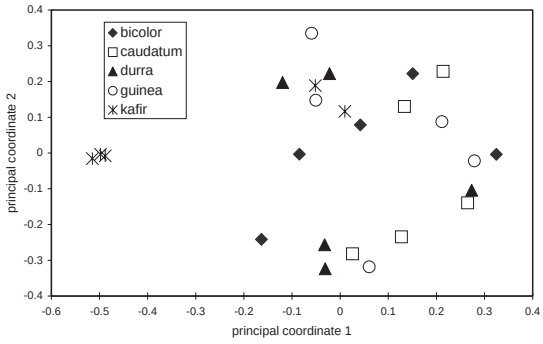

Fig. 3 Scatterplot of sorghum accessions against the two first principal coordinates computed from the genetic distance matrix

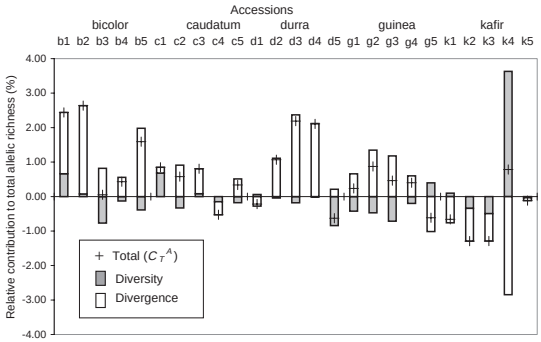

Fig. 4 Relative contribution to total allelic richness (C^A_T expressed in percentage) of each accession of sorghum, and its subdivision into a diversity and a divergence component

is shown that accessions with either low or very high polymorphism were rather rare within our sample, and none were completely fixed at the five loci. This is in contrast to what was found for wild relatives of sorghum (Morden et al. 1990; Schoen and Brown 1995) where the majority of the populations are highly depauperate in allozyme variation. It is not clear at present whether this discrepancy is due to the origin of the material (domesticated versus wild) or to the nature of the genetic markers (microsatellites versus allozymes). Despite this high level of genetic variation within accessions of cultivated sorghum, the average proportion of heterozygote individuals is low (H_o=0.13), in agreement with the high level of inbreeding noted above.

The overall gene diversity in the entire sample is very high (H_T=0.90), with two-thirds of the diversity due to differentiation among accessions (G_{ST}=0.67; Table 4). Including classification according to races as a new level in a hierarchical analysis of gene diversity showed that 53% of the overall gene diversity is due to differentiation among accessions but within races, whereas only 14% of diversity is due to differentiation among races. A similar analysis using classification by geographic origin showed that only 11% of the overall gene diversity is due to differentiation among origins.

When the accessions were classified according to races, the highest overall gene diversity occurred in race bicolor, whereas the lowest value of H_T was found for race kafir (Table 4). This seems to be due to a much higher genetic differentiation among accessions belonging to race bicolor than to race kafir (G_{ST} equal to 0.65 and 0.54, respectively) as accessions from the latter display a somewhat higher genetic variation within accessions than those from race bicolor. The classification according to geographic origin revealed that the Eastern African region has the highest overall level of gene diversity as well as the highest diversity within accessions. In contrast, accessions from Asia have a much lower diversity within, but a higher genetic differentiation among accessions, than those from other regions.

Classical statistics of genetic distances between pairs of accessions, as for instance Nei's D (Nei 1987) could not be used for microsatellite data because many accession pairs did not share a single allele at any locus as a result of the large number of alleles observed (96 distinct alleles in total). In order to estimate genetic relatedness between accessions we used the R_{ST} statistic (Slatkin 1995) that takes into account the difference in number of repeats among alleles. A principal coordinate analysis was performed based on the R_{ST} distance matrix, and a scatterplot of individual accessions against the two first principal coordinates (PC 1 and PC 2, that explain together 36.5% of the total variance) is shown in Fig. 3. Overall, accessions belonging to the same race or geographic origin are widely scattered, with the exception of 3 accessions from race kafir (k1, k3 and k4) which form a genetically distinct group.

Orientation of sampling strategies

The relative contribution to total allelic richness (C^A_T) of each accession and its subdivision into a diversity and a differentiation component are given in Fig. 4. All accessions from race bicolor have a positive value of (C^A_T), and thus contribute more than average to total allelic diversity. These positive values of C^A_T are due essentially to a higher than average divergence from other accessions, but also to high allelic richness within the accession (for accessions b1 and b2). Two accessions from race durra (d3 and d4) also show high positive values of C^A_T. In contrast, most accessions from race kafir have negative values of C^A_T, mostly as a consequence of low differentiation among accessions from race kafir. Independently of the morphological classification, all accessions originating from Eastern Africa have values of C^A_T higher than average, with the exception of accession d1 (Fig. 4).

A modification of the M, or maximisation, strategy derived by Schoen and Brown (1995) was used to select a subset of 5 well-differentiated accessions that represent at best the allelic diversity in the overall sample of 25 accessions. From the 53 130 possible combinations, we found a single best subset that conserves 62 (65%) of the 96 alleles recorded. This subset contains 3 accessions from race bicolor (b1, b2 and b4) and 1 from each of the

races caudatum (c1) and guinea (g5). Three of these accessions originate from Eastern Africa (b1, b2 and c1). Most of the subsequent subsets identified by the procedure contain 2 accessions from race bicolor (b1 and b2), 1 from race kafir (k4), and the last 2 belonging to any of the three remaining races. The accessions selected by the maximisation algorithm correspond in general to those showing a higher than average diversity component of C^A_T (Fig.4).

Discussion

Assessment of genetic diversity with allozymes versus microsatellites

To our knowledge this study is the first to report estimates of genetic diversity within and among accessions of a domesticated plant species based on microsatellite markers. A previous detailed study of genetic structure within the germplasm collection of sorghum was performed based on allozyme markers using 8 plants from each of 83 accessions (Morden et al. 1989). These authors found a low level of allozyme variation within accessions with an average percentage of polymorphic loci (P) of 2.1%, and a mean gene diversity (H_e) of 0.008; this is of an order of 30 times lower than our estimates for microsatellite variation. Assuming that accessions correspond to populations at equilibrium between genetic drift and mutation in the absence of selection, the expected gene diversity is given by $4N\mu/(1+4N\mu)$ where N is the population size and μ is the mutation rate per locus per generation (Nei 1987). Based on this equation it appears that the mutation rate of the microsatellite loci surveyed should be about 40 times higher than that of allozymes. From their results, Morden et al. (1989) concluded that sorghum accessions were depauperate for allozymic variation as compared to other cereals like maize or barley. As similar studies with microsatellites in other cereals are lacking, it is not possible yet to compare among species. Studies based on nuclear RFLP have shown that genetic variation is substantially higher within wild than within cultivated sorghum (Aldrich and Doebley 1992; Cui et al. 1995), suggesting the occurrence of a genetic bottleneck during domestication.

Morden et al. (1989) also reported that 91% of the total variation present in the data was found among as opposed to within accessions, which is higher than our estimate (67%). This discrepancy could be due to the different evolutionary properties of microsatellite and allozyme loci (Estoup et al. 1998). For instance, the standard deviation of G_{ST} estimates over loci is four times higher for allozymes (unweighted G_{ST} = 0.87 ± 0.20, mean ± SD) than for microsatellites (G_{ST} = 0.67 ± 0.05), which may suggest that some allozyme loci are subject to either disruptive or homogenising selection. Also, theoretical considerations suggest that the higher mutation rate of microsatellites, as compared to allozymes, are expected to bias downwards the values of the parameters F_{ST} or G_{ST} (Slatkin 1995). Using a presumably unbiased statistic, R_{ST} (Slatkin 1995), we estimated that 71% of total variation is occurring among accessions (data not shown), a value closer but still lower than that reported using allozyme variation.

Based on allozyme loci, Morden et al. (1989) also found that about 9% of the total variation occurred among races of sorghum, a value close to our estimate of 14% for microsatellite loci. Hence, the two studies agree in pointing out that the races of sorghum are not substantially differentiated genetically, a result which is well illustrated by our principal coordinate analysis (Fig. 3). In conclusion, neither allozymes nor microsatellites do substantiate the system of racial classification for sorghum introduced by Harlan and De Wet (1972) based on morphological variation. However, based on nuclear RFLP markers, Deu et al. (1994) showed some differentiation among races kafir, guinea, and caudatum, with accessions from race bicolor intermixed with all four. According to Morden et al. (1989), genetic variation was more closely associated with geographic origin than racial classification. Our results do not confirm this observation, as only 11% of total genetic variation at microsatellite loci is due to differentiation among geographic origins. A lack of correlation between genetic and geographic distances in cultivated sorghum was also reported by Aldrich and Doebley (1992) based on nuclear RFLP loci.

Studies based on allozyme and microsatellite markers also agree in pointing out that the highest overall gene diversity occurs in race bicolor, whereas the lowest is found in race kafir, with the latter displaying little genetic differentiation among accessions. This is consistent with race bicolor having more primitive morphological characters (Harlan and De Wet 1972; Doggett 1988) associated with a wide geographic distribution and race kafir being mostly restricted to Southern Africa.

Management of germplasm collections of sorghum

Our results confirm previous assessments that the germplasm collection of sorghum is highly structured genetically, with about 70% of the genetic diversity occurring among accessions. This level of differentiation is higher than the average reported for crop species (G_{ST}=0.34; Hamrick and Godt 1997) but is common among predominantly selfing species (average G_{ST} over 36 selfing crop species=0.59; Hamrick and Godt 1997). Within accessions we observe substantial inbreeding (F_{IS}=0.55), which would be expected under the mixed mating-system model as a consequence of self-fertilisation at a rate of $s=2\ F_{IS}/(1+\ F_{IS})=0.71$. Similar levels of selfing have been reported for sorghum by Ellstrand and Foster (1983). Slightly higher values of the inbreeding coefficient ($F_{IS}\approx0.70$) were obtained using both allozyme and microsatellite markers in cultivated sorghum sampled in situ in North-western Morocco (Djè et al. 1999).

For highly structured germplasm collections, it has been shown that molecular markers can be useful to opti-

mise sampling strategies in the face of limited resources for conservation (Schoen and Brown 1995). We found evidence for unequal diversity and differentiation among accessions of sorghum (see Table 4, Fig. 4). In particular, we found the highest genetic diversities to occur in accessions belonging to race bicolor and/or originating from Eastern Africa. Clearly, such accessions should receive a high priority in conservation. In close agreement with this result, individual accessions from race bicolor or originating from Eastern Africa were shown to exhibit a high relative contribution to overall allelic richness and were consistently chosen in a majority position within representative subsets of accessions determined with the maximisation strategy of Schoen and Brown (1995). Moreover, each accession from race bicolor shows higher than average genetic differentiation with respect to the whole collection, which may indicate that this group is more ancestral. Other studies using molecular markers also concluded that race bicolor is highly variable (nuclear RFLP, Deu et al 1994; mitochondrial RFLP, Deu et al. 1995) and that accessions from Eastern Africa are highly variable (Aldrich and Doebley 1992; Deu et al. 1994).

Accessions originating from Asia tend to show lower genetic diversity locally but higher inter-accession differentiation than those from other geographic groups. This could arise as a result of a multiple independent introduction of sorghum associated with local genetic bottlenecks. Molecular evidence for multiple introduction of sorghum in China has been presented recently (Yang et al. 1996).

In conclusion, it appears that microsatellite data pertaining to diversity within and among accessions of a crop species may be useful in devising strategies for managing highly structured germplasm collections. Their use in genetic conservation should be encouraged, provided that successful strategies improving the analysis throughput could be introduced (Donini et al. 1998).

Acknowledgements We thank M. El Jaziri and M. Milinkovitch for allowing us the use of their laboratory equipment, Dr. N.K. Rao, genebank curator at the ICRISAT, for providing our plant material, and R. Petit for comments on the manuscript and help in data analysis. We are also grateful to D. Monteyne for technical support. The scholarship of the first author was provided by the government of Ivory Coast.

References

Aldrich PR, Doebley J (1992) Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor.* Theor Appl Genet 85:293–302

Aldrich PR, Doebley J, Schertz K, Stec A (1992) Patterns of allozyme variation in cultivated and wild *Sorghum bicolor*. Theor Appl Genet 85:451–460

Bredemeijer GMM, Arens P, Wouters D, Visser D, Vosman B (1998) The use of semi-automated fluorescent microsatellite analysis for tomato cultivar identification. Theor Appl Genet 97:584–590

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S (1996) Multiple methods for the identification of polymorphic simple repeats (SSRs) in sorghum (*Sorghum bicolor* (L) Moench). Theor Appl Genet 93:190–198

Cui YX, Xu GW, Magill CW, Schertz KF, Hart GE (1995) RFLP-based assay of *Sorghum bicolor* (L) Moench. genetic diversity. Theor Appl Genet 90:787–796

Deu M, Gonzalez-de-Leon D, Glaszmann JC, Degremont I, Chantereau J, Lanaud C, Hamon P (1994) RFLP diversity in cultivated sorghum in relation to racial differentiation. Theor Appl Genet 88:838–844

Deu M, Hamon P, Chantereau J, Dufour P, D'Hont A, Lanaud C (1995) Mitochondrial DNA diversity in wild and cultivated sorghum. Genome 38:635–645

Djè Y, Forcioli D, Ater M, Lefèbvre C, Vekemans X (1999) Assessing population genetic structure of sorghum landraces from North-western Morocco using allozyme and microsatellite markers. Theor Appl Genet 99:157–163

Doggett H (1988) Sorghum, 2nd edn. Longman scientific and technical, London

Donini P, Stephenson P, Bryan GJ, Koebner RMD (1998) The potential of microsatellites for high throughput genetic diversity assessment in wheat and barley. Genet Res Crop Evol 45: 415–421

Doyle JD, Doyle JL (1990) Isolation of plant DNA from fresh tissue. BRL Focus 12:13–15

Dubreuil P, Charcosset A (1998) Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. Theor Appl Genet 96:577–587

Ellstrand NC, Foster KW (1983) Impact of population structure on the apparent outcrossing rate of grain sorghum (*Sorghum bicolor*). Theor Appl Genet 66:323–327

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theor Appl Genet 92:832–839

Estoup A, Rousset F, Michalkis Y, Cornuet JM, Adriamanga M, Guyomard R (1998) Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). Mol Ecol 7:339–353

Fofana B, Vekemans X, Du Jardin P, Baudouin J-P (1997) Genetic diversity in Lima bean (*Phaseolus lunatus*) as revealed by RAPD markers. Euphytica 95:157–165

Hamrick JL, Godt MJW (1997) Allozyme diversity in cultivated crops. Crop Sci 37:26–30

Harlan JR, De Wet JMJ (1972) A simplified classification of cultivated sorghum. Crop Sci 12:127–176

Maquet A, Zoro Bi I, Delvaux M, Wathelet B, Baudoin J-P (1997) Genetic structure of a lima bean base collection using allozyme markers. Theor Appl Genet 95:980–991

Morden CW, Doebley J, Schertz KF (1989) Allozyme variation in old world races of *Sorghum bicolor* (Poaceae). Am J Bot 76: 247–255

Morden CW, Doebley K, Schertz KF (1990) Allozyme variation among the spontaneous species of *Sorghum* section Sorghum (Poaceae). Theor Appl Genet 80:296–304

Morgante CW, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583–590

Nei M (1987) Molecular evoluationary genetics. Columbia University Press, New York

Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversity. Ann Hum Genet 47:253–259

Newbury H, Ford-Lloyd BV (1997) Estimation of genetic diversity. In: Maxted N, Ford-Lloyd BV, Hawkes JG (eds) Plant genetic conservation-the in situ approach. Chapman and Hall, London, pp 192–206

Ollitrault P, Arnaud M, Chantereau J (1989) Polymorphisme enzymatique des sorghos. II-Organisation génétique et évolutive des sorghos cultivés. Agron Trop 44:211–222

Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. Theor Appl Genet 97:1248–1255

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 12:844–855

Powell W, Machray GC, Provan J (1996a) Polymorphism revealed by single sequence repeats. Trends Plant Sci 1:215–222

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996b) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Rolf RJ (1993) NTSYS-PC, numerical taxonomy and multivariate analysis system, version 1.8, Exeter Software, Setauket, N.Y.

Russell J, Fuller J, Young G, Thomas B, Taramino G, Macauley M, Waugh R, Powell W (1997) Discriminating between barley genotypes using microsatellite markers. Genome 40:442–450

Schoen DB, Brown AHD (1995) Maximising genetic diversity in core collections of wild relatives of crop species. In: Hodgkin T, Brown AHD, van Hintum ThJL, Morales EAV (eds) Core collections of plant genetic resources. John Wiley and Sons, Chichester, pp 55–76

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Taramino G, Tingey S (1996) Simple sequence repeats for germplasm analysis and mapping in maize. Genome 39:277–287

Taramino G, Tarchini R, Ferrario S, Lee M, Pé ME (1997) Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theor Appl Genet 95:66–72

Tohme J, Gonzalez DO, Beebe S, Duque MC (1996) AFLP analysis of gene pools of wild bean core collection. Crop Sci 36: 1375–1384

Vekemans X, Lefèbvre C (1997) On the evolution of heavy-metal tolerant populations in *Armeria maritima:* evidence from allozyme variation and reproduction barriers. J Evol Biol 10: 175–191

Weir BS (1990) Genetic data analysis. Sinauer Assoc, Publ, Sunderland, Mass.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Westman AL, Kresovich S (1997) Use of molecular marker techniques for description of plant genetic variation. In: Callow JA, Ford-Lloyd BV, Newbury HJ (eds) Plant genetic resources. Conservation and use. CAB Int, Wallingford, UK, pp 9–48

Yang GP, Saghai Maroof MA, Xu Qifa Zhang CG, Biyashev RM (1994) Comparative analysis of microsatellite DNA polymorphism in landraces cultivars of rice. Mol Gen Genet 245: 187–194

Yang W, de Oliveira AC, Goodwin I, Schertz K, Bennetzen JL (1996) Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums. Crop Sci 36:1669–1676

Zhu J, Gale MD, Quarrie S, Jackson MT, Bryan GJ (1998) AFLP markers for the study of rice biodiversity. Theor Appl Genet 96:602–611