Genetic Resources and Crop Evolution (2006) 53: 1243–1253
DOI 10.1007/s10722-005-5678-7

Genetic and phenotypic diversity in downy-mildew-resistant sorghum (*Sorghum bicolor* (L.) Moench) germplasm

V. Kamala[1,*], P.J. Bramel[2], S. Sivaramakrishnan[3], S. Chandra[4], Seetha Kannan[4], S. Harikrishna[4] and D. Manohar Rao[5]

*[1]National Bureau of Plant Genetic Resources (NBPGR), Regional Station, Rajendra Nagar, Hyderabad – 500 030 AP, India; [2]International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria; [3]Department of Agricultural Biotechnology, Acharya NG Ranga Agricultural University (ANGRAU), Rajendra Nagar, Hyderabad 500 030 AP, India; [4]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, 502 324 AP, India; [5]Department of Genetics, Osmania University, Hyderabad 500 007 AP, India; *Author for correspondence (e-mail: kgksvp@yahoo.com; phone: +91-40-24015478; fax: +91-40-24014072)*

Received 12 October 2004; accepted in revised form 8 March 2005

Key words: Downy mildew, Genetic diversity, Microsatellite markers, Phenotypic diversity, *Sorghum bicolor*

Abstract

Genetic and phenotypic diversity among randomly selected 36 downy-mildew-resistant sorghum accessions were assessed, the former using 10 simple sequence repeat (SSR) marker loci and the latter using 20 phenotypic traits. The number of alleles (a_j) at individual loci varied from five to 14 with an average of 8.8 alleles per locus. Nei's gene diversity (H_j) varied from 0.59 to 0.92 with an average of 0.81 per locus. High gene diversity and allelic richness were observed in races *durra caudatum* (H_j = 0.76, a_j = 4.3) and *guinea caudatum* (H_j = 0.76, a_j = 3.8) and in east Africa (H_j = 0.78, a_j = 7.2). The regions were genetically more differentiated than the races as indicated by Wright's F_{st}. The pattern of SSR-based clustering of accessions was more in accordance with their geographic proximity than with their racial likeness. This clustering pattern matched little with that obtained from phenotypic traits. The inter-accession genetic distance varied from 0.30 to 1.00 with an average of 0.78. Inter-accession phenotypic distance varied from 0.01 to 0.55 with an average of 0.33. Eleven accession-pairs had phenotypic distance of more than 0.50 and genetic distance of more than 0.70. These could be used as potential parents in a sorghum downy mildew resistance-breeding program.

Introduction

Availability of adequate genetic variation is a fundamental prerequisite for genetic improvement of any crop species. An accurate assessment of this variation in a gene pool of potential breeding materials provides an objective basis to design efficient and cost-effective crop improvement strategies for sustainable long-term selection gains. Also, an assessment of the degree and distribution of this variation allows a better understanding of evolutionary relationships and permits an objectively targeted utilization of crop genetic resources for breeding and conservation.

Advances in sorghum improvement have resulted in the development of high yielding vari-

eties for diverse agro-climatic conditions. Sorghum productivity however continues to be seriously constrained by many pests and pathogens. Sorghum downy mildew (SDM), caused by the fungus *Peronosclerospora sorghi* [Weston et Uphall (Shaw)], is in particular a destructive disease prevalent throughout the sorghum growing areas of the world. Although some sources of resistance have been used to breed SDM resistant lines and hybrids (Pande et al. 1997), the presence of pathogen variability, and the risks associated with a narrow genetic base, underscore the need to identify diverse sources of resistance. The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), having a large *ex situ* sorghum collection (~35,000), has identified 130 accessions (out of ~16,000 screened) with a high level of resistance (0–5% mean disease incidence of the total plants systemically infected) to the ICRISAT isolate of the SDM pathogen (http://www.icrisat.org).

These 130 accessions, originating from various geographic regions and belonging to a number of races, represent a potentially useful gene pool for SDM resistance breeding. Their effective utilization is however constrained by a lack of knowledge about their genetic inter-relationships and the degree and distribution of variation at the phenotypic and genotypic levels. The objective of this paper is to assess and compare the genetic and phenotypic diversity in a sample of 36 accessions randomly selected from the 130 SDM resistant accessions.

Materials and methods

Plant material and DNA extraction

The 36 accessions, randomly selected from the 130 SDM resistant accessions, are listed in Table 1. Five accessions (IS 1054, IS 2462, IS 14387, IS 18683, IS 22231), known to be susceptible to the ICRISAT isolate of the pathogen, were also included for comparison. Nineteen (35%) of the 36 accessions originate from east Africa, 10 (26%) from west Africa, three (8%) from east Asia and Oceania (all three from Australia), one (3%) each from central Africa, southern Africa, north America and Indian subcontinent. Among the total 130 accessions, 54 (35%) originate from east Africa, 38 (26%) from west Africa, 13 (10%) from central and southern Africa, 10 (8%) from Indian subcontinent, nine (7%) from north America, five (4%) from east Asia and Oceania (all 5 from Australia) and one from Russia. The random sample of 36 accessions thus fairly represents the regional structure of the 130 accessions, with Africa adequately represented in the sample. Africa is known to be the major source of SDM resistance as well as the primary center of diversity of this crop and also probably of the SDM pathogen. The 130 accessions have 33 (25%) accessions belonging to race *caudatum*, 35 (27%) to race *durra*, 11 (8%) to race *durra caudatum*, 5 (4%) to race *guinea caudatum*, 12 (9%) to race *guinea*, 11 (8%) to race *caudatum bicolor*, 2 (2%) to race *guinea bicolor*, 13 (10%) to race *durra bicolor*, 7 (5%) to race *bicolor*, and 1 (1%) to race *kafir caudatum*. The random sample of 36 includes 8 (22%) accessions from race *caudatum*, 8 (22%) from race *durra*, 5 (14%) from race *durra caudatum*, 5 (14%) from race *guinea caudatum*, 4 (11%) from race *guinea*, 3 (8%) from race *caudatum bicolor*, 1 (2%) from race *guinea bicolor*, 2 (6%) from race *durra bicolor*, but no accessions from races *bicolor* and *kafir caudatum*. Except for the latter two races, the random sample adequately represents the racial structure of the 130 SDM resistant accessions. Any race and region with less than three accessions in the random sample were excluded from the allele-frequency-based diversity analysis reported in Table 3 and 4.

Genomic DNA of each accession was extracted from a bulked sample of 10 one-week-old etiolated seedlings (equal sample weight) using the DNA extraction protocol of Sivaramakrishnan et al. (1997) with two modifications: Mercaptoethanol was not included in the extraction buffer, and about 100 mg of poly vinyl pyrolidone was added while grinding the plant tissue with liquid nitrogen to avoid phenolic contamination.

Microsatellite analysis

Ten SSR loci (Brown et al. 1996) were used for genotyping (Table 2). They cover six of the 10 linkage groups. Markers Sb6-36 and Sb6-57 are from linkage group C, Sb1-10 and Sb4-121 from linkage group D, Sb4-32 and Sb4-15 from linkage group E, Sb6-84 from linkage group F, Sb5-236

Table 1. List of sorghum accessions used.

Accession	Code**	Geographic Region	Source Country
IS 22228	A-1c	East Asia & Oceania	Australia
IS 18757	A-2cb	East Asia & Oceania	Australia
IS 22231*	A-3cb	East Asia & Oceania	Australia
IS 22229	A-4gc	East Asia & Oceania	Australia
IS 7179	C-1gc	Central Africa	Malawi
IS 11201	E-1d	East Africa	Ethiopia
IS 11218	E-2d	East Africa	Ethiopia
IS 11980	E-3d	East Africa	Ethiopia
IS 11551	E-4db	East Africa	Ethiopia
IS 12646	E-5db	East Africa	Ethiopia
IS 11823	E-6dc	East Africa	Ethiopia
IS 8954	E-7c	East Africa	Kenya
IS 19018	E-8c	East Africa	Sudan
IS 19105	E-9dc	East Africa	Sudan
IS 19019	E-10dc	East Africa	Sudan
IS 3443	E-11gc	East Africa	Sudan
IS 19506	E-12gc	East Africa	Sudan
IS 19082	E-13gc	East Africa	Sudan
IS 8906	E-14c	East Africa	Uganda
IS 23855	E-15cb	East Africa	Yemen
IS 23948	E-16cb	East Africa	Yemen
IS 23836	E-17d	East Africa	Yemen
IS 23966	E-18dc	East Africa	Yemen
IS 23838	E-19gb	East Africa	Yemen
IS 18716	I-1d	Indian Subcontinent	India
IS 1054*	I-2d	Indian Subcontinent	India
IS 14332	S-1g	Southern Africa	South Africa
IS 14387*	S-2g	Southern Africa	Zimbabwe
IS 18683*	Usa-1c	North America	USA
IS 20665	Usa-2g	North America	USA
IS 2462*	Usa-3gb	North America	USA
IS 15070	W-1c	West Africa	Cameroon
IS 16059	W-2c	West Africa	Cameroon
IS 15141	W-3c	West Africa	Cameroon
IS 16038	W-4c	West Africa	Cameroon
IS 20452	W-5d	West Africa	Niger
IS 20405	W-6d	West Africa	Niger
IS 20205	W-7d	West Africa	Niger
IS 20478	W-8dc	West Africa	Niger
IS 19971	W-9g	West Africa	Senegal
IS 20049	W-10g	West Africa	Senegal

*Accessions susceptible to the ICRISAT isolate of the SDM pathogen; **A = Australia; C = central Africa; E = east Africa; I = Indian sub-continent; S = southern Africa; Usa = USA; W = west Africa, c = *caudatum*; d = *durra*; g = *guinea*; gb = *guinea bicolor*; dc = *durra caudatum*; cb = *caudatum bicolor*; gc = *guinea caudatum*; db = *durra bicolor*.

from linkage group G and Sb1-1 from linkage group H. Marker Sb4-22 is not yet mapped. Each 25 μl reaction contained 25 ng of genomic DNA, 1× PCR buffer (50 mM KCl, 20 mM Tris–HCl pH (8.4)), 10 pmol of each primer, 2 mM $MgCl_2$, 200 μM each of dCTP, dGTP, dTTP, 50 μM of dATP and 1 μ Ci of [α^{32}P]- dATP and 1 unit of Taq DNA polymerase (Amersham Pharmacia UK). PCR reactions were carried out in a PTC-100 Thermocycler (MJ Research Inc, USA) using a 65–55 °C Touchdown PCR cycle (Dean et al. 1999). Denaturation was carried out at 94 °C for 30 s and extension was carried out at 72 °C for 1 min. Annealing was carried out between the denaturation and extension steps using a touch down program: the first cycle at 65 °C for 30 s, followed by 63 °C for 3 cycles, 61 °C for 3 cycles, 59 °C for 5 cycles, 57 °C for 5 cycles and 55 °C for

Table 2. Allelic distribution and gene diversity.

Locus	a_j	H_j	a_j (ck)	H_j (ck)	a_j (min)	H_j (min)	a_j (s)	H_j (s)	Percent of accessions under	
									a_j (min)	a_j (s)
Sb 4-32	14	0.92	11	0.92	9	0.92	2	0.97	86	28
Sb 6-36	11	0.89	8	0.89	8	0.89	3	0.93	91	50
Sb 6-84	12	0.85	6	0.85	7	0.85	2	0.88	86	47
Sb 1-10	7	0.77	4	0.77	6	0.81	3	0.79	97	69
Sb 1-1	10	0.81	5	0.81	5	0.77	3	0.82	86	75
Sb 4-15	8	0.84	6	0.84	8	0.84	3	0.87	100	64
Sb 4-22	5	0.77	4	0.77	4	0.77	4	0.77	97	97
Sb 4-121	8	0.85	6	0.85	7	0.85	4	0.86	97	75
Sb 5-236	8	0.82	5	0.82	8	0.82	2	0.85	100	56
Sb 6-57	5	0.59	2	0.59	4	0.59	2	0.61	97	78
Total	88		57		66		28			
Mean	8.8	0.81	5.70	0.81	6.60	0.81	2.80	0.84		
SD		0.091		0.091		0.091	0.099			

a_j = total number of alleles observed; H_j = Gene diversity computed using all alleles;
a_j (ck) = Number of effective alleles computed using Crow and Kimura's (1970) formula;
H_j (ck) = Gene diversity computed using a (ck); a_j (s) = Number of SIGNALs at 5% level of significance; H_j (s) = Gene diversity computed using a(s); a_j (min) = Number of alleles with a frequency of 5% or greater; H_j (min) = Gene diversity computed using a(min); SD = Standard deviation.

14 cycles. In all, 31 cycles were carried out with a final extension at 72 °C for 5 min.

Electrophoresis and band scoring

PCR products were electrophoresed on a denaturing polyacrylamide gel (6% acrylamide, 7.5 M urea, 1 X TBE) at 1500 V for 2 h. The gel was then transferred to Whatman 3 filter paper, covered with Saran Wrap and dried under vacuum for 1 h at 80 °C. Autoradiograms were obtained by exposing the gel for varying periods (one to two days) in a cassette with intensifying screen using Kodak X-OMAT film. Fragment sizes were determined using end labeled AFLP marker (30–330 bp; Life Technologies, USA). The autoradiogram was manually scored for the presence (1) or absence (0) of a band for each locus across all the accessions.

Inter-accession phenotypic relationships

These were assessed using 10 qualitative and 10 quantitative traits. Qualitative traits were pigmentation, lustre, nodal tillering, presence of subcoat, midrib color, grain color, endosperm, ear head shape, glume color and grain covering. Quantitative traits were basal tillering, 50% flowering and plant height in rainy and post rainy seasons, panicle exertion, panicle length and width, seed size and 100 seed weight. Data on these traits were extracted from the ICRISAT Gene Bank database at the level of mean values of field observations made over several years (1975–1996) in both the rainy (*kharif*) and the post-rainy (*rabi*) seasons on vertisol soils at Patancheru, India.

Phenotypic relationships were assessed using simple matching coefficient for binary and multistate qualitative traits and Euclidean distance for continuous quantitative traits. A phenotypic distance matrix $P = \{P_{uv}\}$ was computed for the 36 accessions using Gower's (1971) method and subjected to sequential agglomerative hierarchical non-overlapping (SAHN) cluster analysis using the average-linkage (UPGMA) clustering algorithm and non-metric multi-dimensional scaling (MDS). Cophenetic correlation coefficients were estimated to assess the degree of agreement between the observed proximity matrices and the resultant dendrograms/MDS plots.

Inter-accession genetic relationships

Genetic similarity between accessions (u,v), $u,v = 1,\ldots,N$ was estimated as (Bowcock et al. 1994)

$$S_{uv} = [1/(2n_l)]\Sigma_j a_{juv}) \quad (1)$$

where a_{juv} is the number of shared alleles between individuals (u,v) at locus j. S_{uv} was estimated from the $a \times N$ binary allelic data matrix where $a = \Sigma_j a_j$ is the total number of alleles across the $n_l = 10$ loci. For inbreeding species, such as sorghum, S_{uv} is equivalent to the genetic similarity coefficient of Nei and Li (1979) for co-dominant markers (Powell et al. 1996). The genetic distance between individuals (u,v) was calculated as $d_{uv} = (1-S_{uv})$, resulting in an $N \times N$ distance matrix $D = \{d_{uv}\}$. The distance matrix was subjected to non-metric multi-dimensional scaling (MDS) and SAHN average linkage clustering (UPGMA).

Gene diversity

Gene diversity at locus j (H_{ej}) was estimated as follows (Nei 1987) for the entire sample of 36 accessions as well as for populations defined by races and regions

$$H_{ej} = [2N/(2N-1)](1 - \Sigma_i p_{ij}^2) \quad i = 1, \ldots, a_j \quad (2)$$

where $N = 36$ is the sample size, p_{ij} is the frequency of allele i at locus j, and a_j is the number of alleles at locus j. The average gene diversity (H_j) was estimated as $H = \Sigma_j H_j/n_l$ where $n_l = 10$ is the number of loci.

Effective number of alleles

Effective number of alleles at locus j was estimated as (Crow and Kimura 1970)

$$a_j(ck) = 1/\Sigma p_{ij}^2 \quad (3)$$

It is however useful to know which specific alleles significantly contribute to gene diversity, which Equation (3) does not tell. One approach would be to set a specified minimum limit $p_{\min}$, say 0.05 (Marshall and Brown 1975) on frequency p_{ij} of an allele in order for it to be considered as a major contributor to gene diversity. This is a subjective procedure. An objective procedure is to estimate the (1-α)% confidence interval (CI) for each estimated multinomial allele frequency p_{ij}. Using a Type 1 error probability of $\alpha = 0.05$, the 95% CI was estimated as (Snedecor and Cochran 1967)

$$p_{ij} \pm [1.96\sqrt{(p_{ij}q_{ij}/N)} + 1/2N] \quad (4)$$

where $q_{ij} = 1-p_{ij}$. The alleles for which the 95% CI did not include 0 were taken as significantly contributing to gene diversity. We term these alleles as significantly effective number of alleles (SIGNALs).

Genetic differentiation among racial and regional populations

Genetic differentiation among racial/regional populations with respect to allele frequencies was estimated using Wright's *F*-statistics (Weir and Cockerham 1984). Fisher's exact test for racial and regional differentiation was performed to determine if significant differences in allele frequencies existed among races/regions. The Markov Chain Monte Carlo (MCMC) simulations were used to estimate the exact probability of observed differences in allele frequencies. Inter-population genetic distances were computed using Rogers' modified distance (Rogers 1972).

Computations were done using GenStat version 6 (Payne 2002), NTSYSpc version 2 (Rohlf 1994) and TFPGA version 1.3 (Miller 1997).

Results and discussion

Gene diversity

High levels of polymorphism were revealed at each of the 10 SSR loci, with all accessions being clearly differentiated. A representative picture of the microsatellite amplification is given in Figure 1. Eighty-eight alleles with 5-14 alleles per locus were observed. Gene diversity had a range of 0.59–0.92 with a mean value of 0.81 ± 0.092 (0.092 is the standard error) (Table 2). Among the races, the *durra caudatums* and the *guinea caudatums* were most diverse with a gene diversity of 0.76 followed by the *caudatums* (0.75), the *guineas* (0.73) and the *durras* (0.70) (Table 3). All loci were polymorphic for all races except for the *durras* that were monomorphic at locus Sb 6-57. The number of observed alleles (allelic richness) was highest for the *caudatums* (49%) and the *durras* (49%) fol-

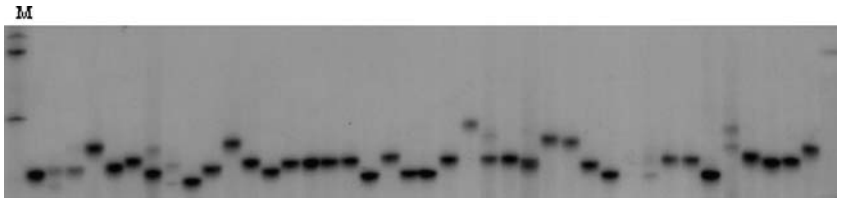

Figure 1. Allelic polymorphism as revealed by microsatellite marker Sb 1-1 across 41 accessions of *Sorghum*.

Table 3. Gene diversity (H_j) and number of alleles (a_j) in racial populations.

Locus	Overall (N = 36)	*Caudatum* (N = 8)	*Durra* (N = 8)	*Durra Caudatum* (N = 5)	*Guinea caudatum* (N = 5)	*Guinea* (N = 4)	*Caudatum bicolor* (N = 3)
	H_j (a_j)						
Sb 4-32	0.92(14)	0.80(5)	0.87(6)	0.80(4)	0.89(5)	0.86(4)	0.80(3)
Sb 6-36	0.89(11)	0.77(4)	0.87(6)	0.80(4)	0.71(3)	0.86(4)	0.80(3)
Sb 6-84	0.85(12)	0.73(4)	0.73(5)	0.80(4)	0.71(3)	0.71(3)	0.53(2)
Sb 1-10	0.77(7)	0.77(4)	0.67(3)	0.62(3)	0.80(4)	0.57(2)	0.53(2)
Sb 1-1	0.81(10)	0.57(3)	0.73(4)	0.89(5)	0.71(3)	0.86(4)	0.53(2)
Sb 4-15	0.84(8)	0.83(6)	0.80(5)	0.80(4)	0.62(3)	0.86(4)	0.53(2)
Sb 4-22	0.77(5)	0.70(3)	0.77(4)	0.62(3)	0.71(3)	0.71(3)	0.00(1)
Sb 4-121	0.85(8)	0.77(4)	0.80(4)	0.80(4)	0.80(4)	0.57(2)	0.53(2)
Sb 5-236	0.82(8)	0.83(6)	0.80(5)	0.80(4)	0.89(5)	0.71(3)	0.00(1)
Sb 6-57	0.59(5)	0.77(4)	0.00(1)	0.62(3)	0.71(3)	0.57(2)	0.53(2)
Mean	0.81(8.8)	0.75(4.3)	0.70(4.3)	0.76(3.8)	0.76(3.6)	0.73(3.1)	0.48(2.0)
SD	0.091	0.076	0.255	0.097	0.087	0.127	0.275

N = Number of accessions

lowed by *durra caudatums* (43%), *guinea caudatums* (41%) and *guineas* (35%). The *durras*, despite having a larger number of alleles (43), had mean gene diversity (0.70) comparable with that of the *guineas* (0.73) that had fewer alleles (31).

Grouping of the 36 accessions by geographical region gave gene diversity values of 0.78 for eastern Africa, 0.75 for western Africa and 0.51 for Australia (Table 4). Allelic richness was higher for the eastern Africa accessions (82% with 72 alleles) as compared to the western Africa accessions (61% with 55 alleles) though mean gene diversity values were not significantly different. About 47% of the alleles were common to the two regions and of the 47 unique alleles, eastern and western Africa accounted for 31 and 14 respectively. The remaining two alleles were found one each in accessions from southern and central Africa. All alleles in the three breeding lines from Australia were present in accessions from eastern, western or southern Africa.

The regions were observed to be genetically more differentiated than the races based on Wright's F_{st}, which provides an estimate of inter-population genetic differentiation. The F_{st} value was lower for races (0.03 ± 0.01) than for regions (0.06 ± 0.03). More diversity is present among accessions within a population (97% for races, 94% for regions) than among populations (3% for races, 6% for regions). This reflects high level of geneflow due to the absence of mating barriers between the races and regions on the one hand and human migration and agricultural trade between regions on the other.

The number of effective alleles identified by the three methods differed considerably, with the 95% CI method yielding the least number of effective alleles (Table 2). The three methods however gave quite similar value of gene diversity indicating that alleles with lower frequencies do not significantly contribute to the estimate of gene diversity. For example, for the most diverse locus Sb4-32 (H_j = 0.92), only 2 out of 14 alleles were SIGNALs, which gave a very similar value of 0.97 for gene diversity.

Table 4. Gene diversity (H_j) and number of alleles (a_j) in regional populations.

Locus	Overall (N = 36)	Eastern Africa (N = 19)	Western Africa (N = 10)	Australia (N = 3)
		H_j (a_j)		
Sb 4-32	0.92(14)	0.92(11)	0.88(7)	0.53(2)
Sb 6-36	0.89(11)	0.87(9)	0.80(6)	0.53(2)
Sb 6-84	0.85(12)	0.90(11)	0.67(4)	0.53(2)
Sb 1-10	0.77(7)	0.80(6)	0.80(5)	0.00(1)
Sb 1-1	0.81(10)	0.79(7)	0.86(7)	0.53(2)
Sb 4-15	0.84(8)	0.83(6)	0.88(8)	0.53(2)
Sb 4-22	0.77(5)	0.77(5)	0.69(3)	0.53(2)
Sb 4-121	0.85(8)	0.79(6)	0.86(7)	0.80(3)
Sb 5-236	0.82(8)	0.65(7)	0.72(5)	0.53(2)
Sb 6-57	0.59(5)	0.51(4)	0.36(3)	0.53(2)
Mean	0.81(8.8)	0.78(7.2)	0.75(5.5)	0.51(2.0)
SD	0.091	0.122	0.158	0.203

N = Number of accessions

The high gene diversity values were also indicative of the number and distribution of the most frequent and rare alleles. When those alleles with a frequency of 5% or more were considered common, there were 66 alleles with the proportional representation varying from 86% (Sb 4-32,Sb 6-84, Sb 1-1) to 100% (Sb 4-15, Sb 5-236) across the 10 loci. Considering the 28 SIGNALs, the proportional representation ranged from 28% at locus Sb 4-32 to 97% at locus Sb 4-22 (Table 2). The distribution of SIGNALs among regions ranged from 96% in eastern Africa, through 82% in western Africa, to 46% in the Australian accessions. Among the races, the distribution of SIGNALs ranged from 46% to 89%. Information on allele frequencies however has practical implications for effective conservation of those that are less frequent, especially during seed regeneration. As Allard pointed out in 1992, frequently occurring alleles probably have a selective advantage over infrequent alleles (Allard 1992). Hence, while the most frequent alleles may be present in most collections, conservation of the rare alleles is important as possible sources of QTLs for various biotic and abiotic stresses and other agronomic traits (Schoen and Brown 1993).

These results on allelic richness and diversity are in accordance with SSR-based studies in sorghum by Grenier et al. (2000) and Kong et al. (2000). The degree of polymorphism revealed by SSRs in this study is much higher in contrast to that found with RFLP and RAPD markers (Deu et al. 1994; Oliviera et al. 1996). The high level of polymorphism associated with SSRs is expected because of the unique mechanism responsible for generating SSR allelic diversity by replication slippage (Tautz et al. 1986) rather than by simple mutations or insertions/deletions.

Phenotypic diversity

Ninety seven percent of the 36 accessions were predominantly pigmented, 89% displayed nodal tillering, 64% had lustrous grain and 75% did not possess a testa/subcoat. The most common grain color was brown (36%), followed by straw (19%), red (17%) and white (16%) colors. Yellow colored grains were less represented (5%). The principal glume color was purple/black (47%), followed by brown (22%), red (14%) and straw (14%), with only 3% having gray glumes. The mid rib was mainly white (81%) with dull green types accounting for 14% and white/yellow types being rare (0.6%). Forty-four percent of the accessions showed partly corneous endosperm with mostly corneous and completely corneous types accounting for 17% and 3% respectively. Accessions with mostly starchy and completely starchy endosperm types were represented to the extent of 25% and 11% respectively. Shape of earhead was primarily semi-compact elliptic (39%) with compact elliptic

types accounting for 8%. Earheads with semi-loose stiff branches accounted for 31%, while those with loose stiff branches, loose drooping branches, and semi-loose drooping branches accounted for 14%, 5% and 3% respectively. Grains were either half covered with glumes (44%) or one-fourth covered (42%), with only 11% being three-fourth covered and 3% completely uncovered. Grain color is an important trait in farmers' selection and consumer acceptance. Colored glumes and grains are known to play a significant role in grain mold resistance (Audilakshmi et al. 1999). However, their association with downy mildew needs to be explored. Wide variation was observed for quantitative traits: plant height in *rabi* (110–290 cm) and *kharif* (115–470 cm), days to 50% flowering in kharif (55–158) and rabi (52–107), number of basal tillers (1–4), peduncle exertion (4–48 cm), panicle length (8–35 cm) and panicle width (3–20 cm), and seed size (1.8–4.0 mm) and seed weight (2–6 g).

Inter-accession genetic and phenotypic relationships

The 36 accessions were genetically very diverse, with inter-accession genetic distance varying between 0.30 and 1.00 with an average of 0.78. Similar results have been reported by Dean et al. (1999). Out of 630 accession pairs, 573 pairs had genetic distance of more than 0.60. Ninety-four pairs had a genetic distance of 1.00. While no clear racial or regional separation was visible in the MDS plot (Figure 2), the UPGMA dendrogram (not shown), with several sub-clusters, reflected some geographic associations. Inclusion of data from the five susceptible lines retained the geographic grouping. The two guineas from South Africa (one resistant and the other susceptible) clustered together, as did the two accessions from India (one resistant and the other susceptible). The four Australian breeding lines, of which one was susceptible, also grouped together. Three accessions from USA were observed to be closer to those from Africa than to each other (Figure 3). Two of these (IS 2562 and IS 18683) were subsequently traced to Somalia and Sudan respectively. This is consistent with the fact that sorghum in USA has primarily been introduced from Africa (Duncan et al. 1991). Similar associations between genetic variation and geographic origin have also been reported by Morden et al. (1989).

The inter-accession phenotypic distance varied from 0.01 to 0.55 with an average value of 0.33 ± 0.091. Only 11 accession pairs, out of 630, had a distance of more than 0.50. These were: (IS-7179, IS-22229), (IS-23948, IS-22229), (IS-19971, IS-22229), (IS-20049, IS-22229), (IS-11218, IS-7179), (IS-11823, IS-3443), (IS-15070, IS-3443), (IS-11218, IS-11201), (IS-19971, IS-16059), (IS-20405,

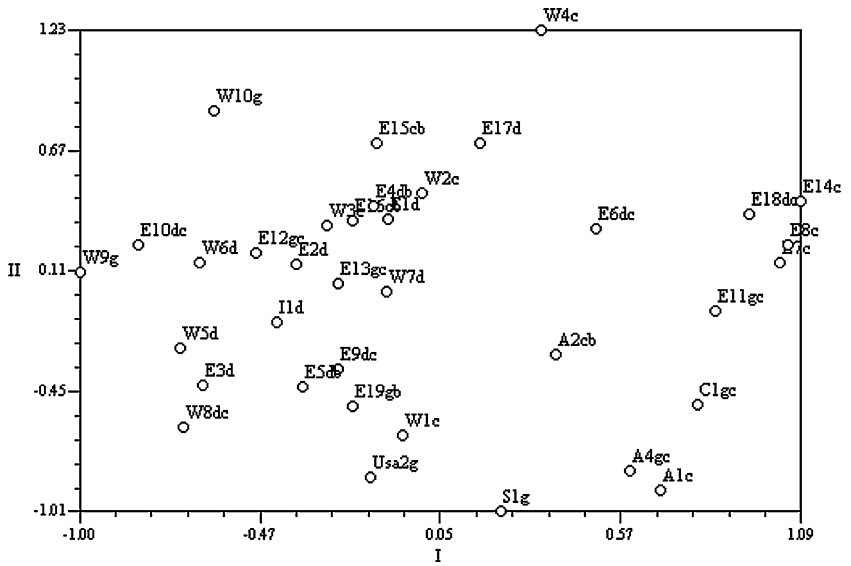

Figure 2. MDS plot of 36 *Sorghum* accessions with SSRs (Cophenetic correlation $r = 0.78$, $p < 0.05$). See Table 1 for accession codes.

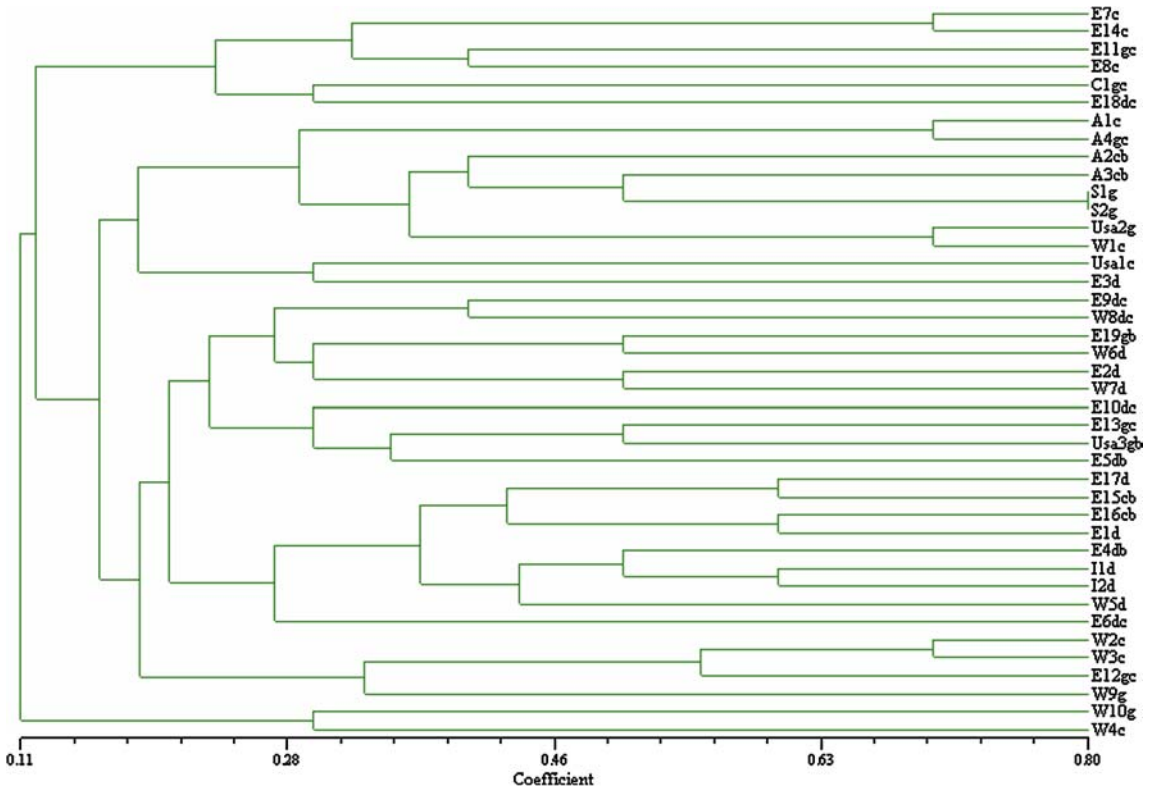

Figure 3. UPGMA dendrogram of 41 *Sorghum* accessions with SSRs ($r = 0.67$, $p < 0.05$). See Table 1 for accession codes.

IS-16059), and (IS-20205, IS-16059). These 11 accession pairs also had a genetic distance varying from 0.70 to 1.00. Clustering the accessions using the 20 traits produced an MDS plot (Figure 4) that separated the accessions into two groups principally reflecting grain lustre and color. Sub-clusters within

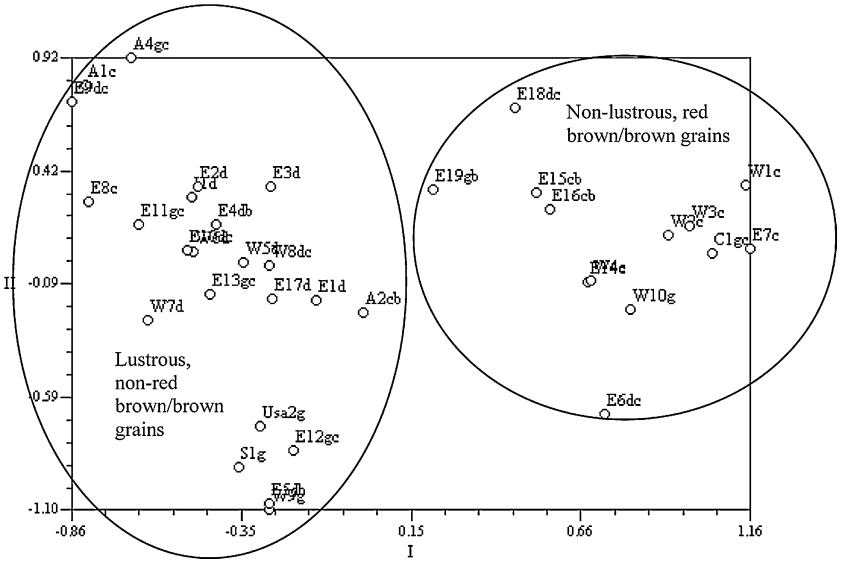

Figure 4. MDS plot of 36 *Sorghum* accessions using qualitative and quantitative traits ($r = 0.91$, $p < 0.05$). See Table 1 for accession codes.

each group were more suggestive of geographical rather than racial associations. The inclusion of the five selected SDM susceptible accessions did not alter the basic grouping into two clusters based on grain lustre and color. Mantel test showed little correlation between genetic and phenotypic distances ($r = 0.04$, $p > 0.05$).

The SSR markers allowed inferring the genetic structure of SDM resistant germplasm, their fingerprinting and identification of genetically diverse germplasm. The 11 accession pairs, identified on the basis of their high genetic and phenotypic distances, may be useful as parental lines in SDM resistance breeding and for generating mapping populations to identify markers linked to SDM resistance.

Acknowledgements

This is part of the first author's PhD work who would like to gratefully thank Director NBPGR New Delhi and the ICAR to permit and to financially support her to pursue this work, and ICRISAT for providing the necessary facilities and cooperation to carry out this study. The authors gratefully thank Dr KS Varaprasad and Dr RDVJ Prasada Rao, OiC and Principal Scientist respectively at NBPGR Regional Station Hyderabad, Dr S Senthilvel and Dr RT Folkertsma, Molecular Breeders at ICRISAT, for their constructive comments.

References

Allard R.W. 1992. Predictive methods for germplasm identification. In: Stalker H.T. and Murphy J.P. (eds), Plant Breeding in the 1990s. CAB International, UK, pp. 119–146.

Audilakshmi S., Stenhouse J.W., Reddy T.P. and Prasad M.V.R. 1999. Grain mould resistance and associated characters of sorghum genotypes. Euphytica 107: 91–103.

Bowcock A.M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J.R. and Cavalli-Sforza L.L. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368: 455–457.

Brown S.M., Hopkins M.S., Mitchell S.E., Senior M.L., Wang T.Y., Duncan R.R., Gonzales-Candelas F. and Kresovich S. 1996. Multiple methods for identification of polymorphic simple sequence repeats (SSRs) in sorghum (*Sorghum bicolor* (L.) Moench). Theor. Appl. Genet. 93: 190–198.

Crow J.F. and Kimura M. 1970. An Introduction to Population Genetics Theory. Harper and Row, NY.

Dean R.E., Dahlberg J.A. and Hopkins M.S. 1999. Genetic redundancy and diversity among 'Orange' accessions in the US National Sorghum collection as assessed with simple sequence repeat (SSR) markers. Crop. Sci. 39: 1215–1221.

De Oliveira Antonio C., Richter Todd and Bennetzen Jefferey L. 1996. Regional and racial specificities in sorghum germplasm assessed with DNA markers. Genome 39: 579–587.

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

Duncan R.R., Bramel-Cox P.J. and Miller F.R. 1991. Contributions of Introduced Sorghum Germplasm to Hybrid Development in the USA. Crop Science Society of America – Use of Plant Introductions in Cultivar Development. Special Publ. Part 1. CSSA No. 17.

Gower J.C. 1971. A general coefficient of similarity and some of its properties. Biometrics 27: 857–871.

Grenier C., Deu M., Kresovich S., Bramel-Cox P.J. and Hamon P. 2000. Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures B: using molecular markers. Theor. Appl. Genet. 101: 197–202.

Kong L., Dong J. and Hart G.E. 2000. Characteristics, linkage-map positions, and allelic differentiation of *Sorghum bicolor* (L.) Moench DNA simple-sequence repeats (SSRs). Theor. Appl. Genet. 101: 438–448.

Marshall D.R. and Brown A.H.O. 1975. Optimum sampling strategies in genetic conservation. In: Frankel O.H. and Hawkes J.G. (eds), Crop Genetic Resources for Today and Tomorrow. Cambridge University Press, Cambridge, UK, pp. 53–80

Miller M.P. 1997. Tools for Population Genetic Analysis (TFPGA) 1.3: A Windows Programme for the Analysis of Allozyme and Molecular Population Genetic Data.

Morden W.C., Doebley J. and Schertz K.F. 1989. Allozyme variation in old world races of *Sorghum bicolor* (Poaceae). Am. J. Bot. 76: 245–255.

Nei M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Nei M. and Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. PNAS (USA) 76: 5269–5273.

Pande S., Bock C.H., Bandyopadhyay R., Narayana Y.D., Reddy B.V.S., Lenne J.M. and Jeger M.J. 1997. Downy Mildew of Sorghum. Information Bulletin No. 51 Patancheru 502 324, A.P., India.

Payne R.W. (ed.) 2002. The Guide to GenStat Release 6.1. Part 2: Statistics.

Rogers J.S. 1972. Measures of genetic similarity and genetic distance. In: Studies in Genetic VII. University of Texas Publication no. 7213, Austin, pp. 145–154.

Rohlf F.J. 1994. NTSYS-pc: Numerical Taxonomy and Multivariate System, Version 2.1.1. Exeter Software, Setauket, New York.

Schoen D.J. and Brown A.H.D. 1993. Conservation and allelic richness in wild crop relations in aided by assessment of genetic markers. PNAS 90: 10623–10627.

Sivaramakrishnan S., Seetha K., Nageshwar Rao A. and Laxman Singh 1997. RFLP analysis of cytoplasmic male-sterile lines of pigeon pea (*Cajanus cajan* (L.) Mill sp.) developed by interspecific crosses. Euphytica 93: 307–312.

Snedecor G.W. and Cochran W.G. 1967. Statistical Methods. Oxford and IBH Publishing Co.

Tautz D., Trick M. and Dover G.A. 1986. Cryptic simplicity in DNA is a major source of genetic variation. Nature 322: 652–656.

Weir B.S. and Cockerham C.C. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358–1370.