Theor Appl Genet (2005) 111: 399–409
DOI 10.1007/s00122-005-1949-0

ORIGINAL PAPER

Rolf T. Folkertsma · H. Frederick W. Rattunde
Subhash Chandra · G. Soma Raju · C. Tom Hash

The pattern of genetic diversity of Guinea-race *Sorghum bicolor* (L.) Moench landraces as revealed with SSR markers

Received: 29 July 2004 / Accepted: 15 November 2004 / Published online: 18 June 2005

Abstract The Guinea-race of sorghum [*Sorghum bicolor* (L.) Moench] is a predominantly inbreeding, diploid cereal crop. It originated from West Africa and appears to have spread throughout Africa and South Asia, where it is now the dominant sorghum race, via ancient trade routes. To elucidate the genetic diversity and differentiation among Guinea-race sorghum landraces, we selected 100 accessions from the ICRISAT sorghum Guinea-race Core Collection and genotyped these using 21 simple sequence repeat (SSR) markers. The 21 SSR markers revealed a total of 123 alleles with an average Dice similarity coefficient of 0.37 across 4,950 pairs of accessions, with nearly 50% of the alleles being rare among the accessions analysed. Stratification of the accessions into 11 countries and five eco-regional groups confirmed earlier reports on the spread of Guinea-race sorghum across Africa and South Asia: most of the variation was found among the accessions from semi-arid and Sahelian Africa and the least among accessions from South Asia. In addition, accessions from South Asia most closely resembled those from southern and eastern Africa, supporting earlier suggestions that sorghum germplasm might have reached South Asia via ancient trade routes along the Arabian Sea coasts of eastern Africa, Arabia and South Asia. Stratification of the accessions according to their Snowden classification indicated clear genetic variation between margeritiferum, conspicuum and Roxburghii accessions, whereas the gambicum and guineënse accessions were genetically similar. The implications of these findings for sorghum Guinea-race plant breeding activities are discussed.

Communicated by H.C. Becker

R. T. Folkertsma (✉) · S. Chandra · G. S. Raju · C. T. Hash
International Crops Research Institute for the Semi Arid Tropics (ICRISAT)—Patancheru, Patancheru,
502 324, Andhra Pradesh, India
E-mail: r.folkertsma@cgiar.org
Tel.: +254-0-20524555
Fax: +254-0-20524001

H. F. W. Rattunde
ICRISAT—Bamako, BP 320, Bamako, Mali

Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] together with pearl millet [*Pennisetum americanum* (L.)] and finger millet [*Eleusine coracana* (L.) Gaertn] represent Africa's main contribution to the world food supply (de Vries and Toenniessen 2001). With an annual average production of 61 million tonnes over the past decade (FAOSTAT;http://appsfaoorg/defaulthtm), sorghum is the fifth most important grain crop worldwide. The bulk of African sorghum production is centred in the savannah zone of West and Central Africa where grain of this crop is a major component of the daily menu for millions of people, either as porridge or as traditional beer (Purseglove 1985). In addition, in many developing countries, sorghum stover is often used to feed cattle (de Vries and Toenniessen 2001).

Cultivated sorghum has been classified into five major and ten intermediate races on the basis of grain and glume morphology (Harlan and de Wet 1972). The Guinea-race is distinguished from the other races by grain that is more or less flat, ovate or elliptical and that twists relative to its widely opening glumes at maturity (Harlan and de Wet 1972; de Wet et al. 1972). This race is based primarily on the *Guineensia* subseries of sorghum as described by Snowden (1936). The Guinea-race has the widest geographic distribution (de Wet et al. 1972) and shows more genetic diversity than other races based on previous nuclear DNA marker studies (Deu et al. 1994).

The Guinea-race of sorghum probably originated in the West African savannah, where it might have been selected out of the original Bicolor race (de Wet et al. 1972). It subsequently spread to eastern and southern Africa and by trade routes to Asia (Doggett 1988). This

race currently accounts for more than 70% of sorghum cultivated in West and Central Africa and may account for more than 50% of all sorghum produced in Africa.

Studies characterizing the pattern of diversity within cultivated sorghum on the basis of nuclear DNA marker data have shown that Guinea-race accessions can be clearly differentiated from other races (Deu et al. 1994; Grenier et al. 2001a, 2001b). Furthermore, distinct groups of Guinea-race accessions based on region of origin (western and southern Africa) and grain/glume morphology [small grained "Guinea margaritiferums" as classified by Snowden (1936)] were identified (Deu et al. 1994; Grenier et al. 2001a, 2001b), indicating greater diversity within the Guinea-race than in other races. However, there is no known study that specifically focused on the structure of diversity within the Guinea-race.

Variation in allele frequencies at many unlinked loci is the preferred way to assess genetic diversity and differentiation and to estimate the strengths of the various forces shaping them (Fregene et al. 2003). Simple sequence repeat (SSR) markers are particularly attractive for studying genetic differentiation because they are abundant in plant and animal genomes (Condit and Hubbell 1991; Röder et al. 1995), have high levels of polymorphism (Schug et al. 1998) and are adaptable to automation (Kresovich et al. 1995; Mitchell et al. 1997). In sorghum, numerous SSR markers have been developed and mapped (see, for example, Brown et al. 1996; Taramino et al. 1997; Kong et al. 2000; Bhattramakki et al. 2000; Schloss et al. 2002). A few of these public domain markers have been employed to analyse the genetic diversity in subsets constituted from the ICRISAT sorghum collection (Grenier et al. 2000) and the USDA sorghum collection (Dean et al. 1999) and from collections originating from single countries (Djè et al. 1999; Ghebru et al. 2002).

Sorghum breeding research in West Africa emphasizes the use of genetic diversity within the Guinea-race so as to maintain the required grain quality and array of adaptive characteristics possessed by this race. The research presented here is intended to provide a detailed assessment of the pattern of diversity within the Guinea-race of sorghum in order to provide insights for tapping this diversity in current and future breeding efforts. The specific objectives of characterizing the relative extent of diversity within and between eco-geographic groups will be pursued through an analysis of germplasm from a Guinea-race core collection, established to represent the agronomic and morphological diversity within this race, using a set of 21 SSR markers.

Materials and methods

Plant material

The 100 accessions were identified from a core collection of 293 sorghum [*Sorghum bicolor* (L.) Moench] Guinea-race accessions. The Guinea-race core collection was formed by using principal components analysis (PCA) with nine quantitative traits (flowering dates, plant height, peduncle exertion, panicle width and length, grain size and seed weight) measured in India on the 3,907 Guinea accessions in the ICRISAT sorghum collection. This core collection represents 7.5% of the aforementioned Guinea accessions and 22% of the total variation for agronomic traits among the Guinea accessions (Grenier, unpublished). Approximately one-half of the accessions ($n = 46$) were chosen on the basis of desirable maturity and agronomic performance for the savannah zone of West Africa, as observed in Mali (F.W. Rattunde, unpublished), and the remainder were chosen to include countries not yet represented and to cover the range of maturity and the range of grain weights within each country of origin. These accessions represent landraces from semi-arid and Sahelian Africa, humid West Africa, East Africa and the Great Lakes, and southern Africa (corresponding to the four African eco-regions defined by de Vries and Toenniessen 2001), and South Asia (see Table 1).

Re-evaluation of the race classification according to the system of Harlan and de Wet (1972) based on two replicate samples grown in Bamako, Mali, indicated that all accessions belonged to the Guinea-race with the exception of four intermediate-Guinea-races accessions (Table 1). Ninety two of the 100 Guinea-race accessions were also classified according to the system of Snowden (1936) into margaritiferums, Roxburghii (both with small grains and glumes longer than the grain, and deciduous or persistent pedicelled spikelets, respectively), guineënse, conspicuum (grain and glumes of similar length, with deciduous or pedicelled spikelets, respectively), and exsertum and gambicum (grains longer than the glumes, with deciduous or persistent pedicelled spikelets, respectively).

The seed used for this study originated from the short-term storage collection of the ICRISAT-Patancheru gene bank. The accessions have been maintained by repeated selfing. The phenotypic expression of these accessions was observed to be quite uniform in grow-outs at ICRISAT-Mali. Grow-outs of other accessions from ICRISAT's short-term storage collection side by side with the original, non-regenerated, accession have shown that phenotypic uniformity has often increased through the maintenance process.

Ten seeds of each of the 100 Guinea-race sorghum accessions and two non-Guinea sorghum breeding line controls (BTx623 and R16) were sown in small pots in Patancheru, India. The pots were watered for 14 days, until the length of the seedlings was approximately 10 cm. A total length of approximately 6 cm of leaf-tissue from two to three seedlings of each accession was pooled and the DNA isolated from this leaf-tissue sample using the mini-prep 3% CTAB method (Mace et al. 2003) in a 96-well format. The DNA quantity for each sample was assessed using Picogreen (Juro Supply, Switzerland) and fluorescence (Spectrafluor Plus; Tecan, Switzerland) and DNA concentrations were normalized

Table 1 Guinea-race sorghum landrace accessions included in this study, their ICRISAT accession number, local name, collection site, longitude, latitude, country of origin and geographic region

IS number	Country of origin	Local name	Race[a,b]	Location	Latitude	Longitude	Altitude
Sem-arid and Sahelian Africa							
27494	Burkina Faso	Chiofida	G ga	Bomborokui	13.44	0.49	
27526	Burkina Faso	Oueni kiene	G ma	Ouakara	11.86	−3.63	
27530	Burkina Faso	Chio ye doutuon	G ga	Yaramoko	11.80	−3.28	
27564	Burkina Faso		G ga	Nionion	10.28	−3.26	
6730	Burkina Faso		G ga	Western Volta			
6731	Burkina Faso		G gu	Eastern Volta			
6749	Burkina Faso	Zilet	G gu	Western Volta			
6781	Burkina Faso		G gu	Western Volta			
10816	Chad	Oua kass	G ro				
23645	Gambia	Hafijega	G ma	Daru Salam (McCarthy Island)	13.66	−15.4	
25991	Mali	Kende die	G ma	Kebila	11.17	−7.01	
3817	Mali	Bank oumano ziamri fing	G ga	Bamako	12.39	−7.59	
3818	Mali	Bangon mana signetane	G ga	Bamako	12.39	−7.59	
3849	Mali	Tiryo kalagnigue	G ga	Bamako	12.39	−7.59	
3862	Mali		G ga	Bamako	12.39	−7.59	
3869	Mali		G ga	Bamako	12.39	−7.59	
18287	Niger	Fukimustani	G co	Maradi	13.29	7.05	
9597	Niger	Kerma	G ma	Tarna	13.26	7.06	
19952	Senegal	Saban	G gu	Namari	15.03	−13.78	
19970	Senegal	Bassi kaba	G gu	Keur Moussadrame	13.70	−15.78	
20023	Senegal	Nienikel	G gu	Youpe Amadi	14.21	−12.24	
20064	Senegal	Kinto	G ma	Kossanto	13.25	−12.17	
20081	Senegal		G gu	Sangola	12.40	−12.26	
20114	Senegal	Diouma	G ga	Badion	13.05	−14.21	
19340	Sudan		G ro	Gezira	11.19	24.15	
19368	Sudan		G ro	Gezira	11.19	24.15	
27013	Sudan		GC				
3534	Sudan	Mugbash white	G co	Tozi	10.50	30.44	
6923	Sudan	Dinderawi V	G gu				
Humid West Africa							
26570	Benin	Tchahonga	G gu				
15629	Cameroon		G gu				
16210	Cameroon		G gu				
16292	Cameroon		G co				
16340	Cameroon		G co				
16369	Cameroon		G gu				
16488	Cameroon		G gu				
16515	Cameroon		G				
16658	Cameroon		G gu				
30804	Cameroon		G co		10.86	13.88	1000
30843	Cameroon		G co	Mayotiel	10.87	13.79	550
17612	Ghana		G co				
17659	Ghana		G gu				
2430	Nigeria	Yar Baura	G ma	Sokoto	13.06	5.25	
7250	Nigeria	Query early sorghum	G gu				
7254	Nigeria		G				
7405	Nigeria	Akuki	G gu	Agyaragu; Lafia	8.40	8.55	
7509	Nigeria	Yar yavdo	G gu	Birniwa	12.47	10.14	
7567	Nigeria	Fara fara	G gu	Wamba	8.94	8.59	
7731	Nigeria	Tagyan farafara	G ga	Kubau	10.47	8.10	
7751	Nigeria	Janari	G gu	Tudai	13.06	5.13	
7772	Nigeria	Doron zabo	G	Munlusji			
7835	Nigeria	Farin dawa	G gu	Bassa	8.85	11.38	
7841	Nigeria	Etswa guntotoagi	G ga	Lafiagi	8.50	5.25	
7972	Nigeria	Farin maloka macen bantako	G gu	Illo	6.53	2.80	
27683	Sierra Leone		G ma				
27705	Sierra Leone	Kagboyoh	G ma	Mathoi	9.22	−12.27	150
27717	Sierra Leone	Keiti	G ma	Baiima	7.93	−11.55	100
27731	Sierra Leone	Kete	G ma	Ngaharun			220
27761	Sierra Leone	Tangi	G ma	Kagberi	9.20	−12.13	160

Table 1 (Contd.)

IS number	Country of origin	Local name	Race[a,b]	Location	Latitude	Longitude	Altitude
27768	Sierra Leone	Kagbayoh	G ma	Kagberi	9.20	−12.13	170
26230	Togo	Nadouni	G ga	Nadoti			
35180	Togo		G				
East Africa and the Great Lakes							
11099	Ethiopia		GD	Robi	10	39.89	
14732	Ethiopia		G ro	Fasha	5.29	37.35	
14541	Kenya		G ga				
7173	Tanzania		G				
23146	Tanzania		G ro				
24138	Tanzania		G gu				
24218	Tanzania		G co				
24223	Tanzania		G co				
9220	Uganda	Ekab	G ma				
9221	Uganda	Chitwa 398	G co				
Southern Africa							
19455	Botswana	Sekanakana	G ro				
14351	Malawi		G co				
14353	Malawi		G co				
14414	Malawi		G co				
14417	Malawi		G co				
21529	Malawi		G co	Choweatan			
21530	Malawi		G	Choweatan			
23767	Malawi		G co				
23698	Mozambique	Namapa	G co				200
14331	South-Africa		G co				
14317	Swaziland		G ro				
23228	Zambia		G co	Kanyelele			
23234	Zambia		G co	Kanyelele			
19670	Zimbabwe	Chikota-kota	G ro				
27113	Zimbabwe		G co	Zaka	−20.20	31.27	
27130	Zimbabwe		G co	Binga	−17.62	27.33	
27245	Zimbabwe		G co	Madziva	−16.54	31.31	
27252	Zimbabwe		G co				
30142	Zimbabwe		GC	Nyazura	−18.70	32.15	1500
South Asia							
18241	Bangladesh	Local variety	G ro				
1181	India	Madhucholam	G ro	Palladam	10.99	77.28	
17690	India		G ro	Hingoli	19.72	77.14	
19738	India	Motitura	G ro	Kasalvani	22.06	78.94	
22188	India	Pelala jonna	G ro	Kandaram	17.91	77.52	610
3967	India		G ro				
4597	India	Lalman kolumbi	GD ex	Nanded	19.15	77.33	
5505	India	Muttin guti jola	G ro	Tumkur	13.34	77.10	
3957	Nepal		G ro				

[a] Harlan and de Wet (1972) race designations: G, Guinea-race; GD, intermediate Guinea-durra; GC, intermediate Guinea-caudatum
[b] Snowden (1936) race classifications: ma, margeritiferum; ro, Roxburghii; gu, guineense; co, conspicuum; ga, gambicum; ex, exsertum

at 2.5 ng/µl. The DNA quality of each sample was evaluated by running 1 µl of DNA on a 1% agarose gel.

PCR and capillary electrophoresis

A set of 21 sorghum SSR primers was used for genotyping (see Table 2). Primer pairs for the 15 *Xcup* markers have previously been described by Schloss et al. (2002), those for *Xtxp15* and *Xtxp40* by Kong et al. (2000), for *Xtxp114* and *Xtxp320* by Bhattramakki et al. (2000), for *XSbKAF1* by Taramino et al. (1997) and for *Xgap84* by Brown et al. (1996). SSR markers were chosen based on three criteria: genome position [nine of ten sorghum linkage groups (LGs) were represented, the exception being LG H of Peng et al. (1999), which corresponds to LG E of Chittenden et al. (1994)], repeat size (ranging from di-nucleotide to hexa-nucleotide repeats) and the number of previously reported alleles (ranging from two to six). Forward primers were labelled with FAM, HEX or NED (PE-Applied Biosystems, Foster City, Calif.), allowing post-PCR pooling of the 21 primer products into seven groups of three primer products each, with each primer product in a given group being labelled with a different dye.

PCR conditions for each of the 21 SSR markers were optimized, and PCR reactions were set up in 5-µl volumes in 384-well PCR plates (ABGene, Rochester, N.Y.) using a robotic liquid handling system (Tecan, Switzerland). Each PCR reaction contained 2–4 pmol of primer, 1–4 m *M* $MgCl_2$, 0.1–0.2 m *M* dNTP, 0.1–0.125 U Amplitaq Gold Polymerase (PE-Applied Bio-

Table 2 Characteristics of the 21 SSR markers

Set	SSR locus	Repeat motif	Linkage group[a]	Total no. of alleles per locus	Common alleles (≥5%)[b]	Observed size range (bp)	Allele size difference (Δbp)	PIC[c]
1	*Xcup62*	$(GAA_)6$	A	3	2	188–192	1,3	0.22
	Xcup32	$(AAAAT)_4$	C	4	2	144–155	1,5	0.43
	Xcup07	$(CAA)_8$	G	9	3	251–273	1,2,3	0.40
2	*Xtxp40*	$(GGA)_7$	E	6	2	124–141	2,3	0.28
	Xcup28	$(TGAG)_5$	D	4	3	152–168	4	0.57
	Xcup61	$(CAG)_7$	C	2	2	196–199	3	0.34
3	*Xcup52*	$(AATT)_5$	E	3	3	235–259	2,4	0.39
	Xcup14	$(AG)_{10}$	C	12	2	206–250	2	0.37
	Xcup02	$(GCA)_6$	F	6	4	186–207	3	0.59
4	*KAF1*	$(ACA)_9$	J	9	5	129–165	1,2,3	0.66
	Xcup06	$(CTGC)_4$	A	3	1	198–206	4	0.06
	Xcup60	$(CGGT)_4$	A	3	2	151–163	1,3,4	0.26
5	*Xcup63*	$(GGATGC)_4$	B	3	3	132–144	6	0.37
	Xcup53	$(TTTA)_5$	A	7	3	181–203	1,2,3,4,5	0.44
	Xtxp114	$(AGG)_8$	C	3	2	227–239	3	0.41
6	*Xtxp15*	$(TC)_{16}$	J	11	5	196–228	2	0.71
	Xcup11	$(GCTA)_4$	C	2	2	163–172	1,4	0.37
	Xcup37	$(AG)_9$	I	5	2	202–212	2	0.42
7	*Xgap84*	$(AG)_{14}$	B	12	5	174–220	2	0.77
	Xcup69	$(ATGCG)_4$	B	5	2	237–254	1,2	0.42
	Xtxp320	$(AAG)_{20}$	A	11	6	255–297	1,2,3	0.81

[a]Linkage group (LG) designation according to Bhattramakki et al. (2000)
[b]Number of common alleles per locus with an allele frequency ≥5%
[c]PIC, Polymorphic Information Content

systems) and 1× PCR buffer (PE-Applied Biosystems). Temperature cycling was carried out using the GeneAmp PCR System 9600 (PE-Applied Biosystems) and touch-down PCR amplification: one 15-min denaturation cycle, followed first by ten cycles of 94°C for 10 s, 61°C for 20 s (ramp of 1°C per cycle) and 72°C for 30 s, then by 31 cycles of 94°C for 10 s, 54°C for 20 s and 72°C for 30 s. After completion of the 31 cycles, a final extension of 20 min at 72°C was included to try and minimize the +A overhang (Smith et al. 1995).

PCR products were pooled post-PCR, where 1 µl of the FAM-labelled product, 1 µl of the HEX-labelled product and 1.5 µl of the NED-labelled product were mixed with 7 µl of formamide (PE-Applied Biosystems), 0.3 µl of the ROX-labelled 400 HD size standard (PE-Applied Biosystems) and 4.2 µl of distilled water. DNA fragments were denatured and size-fractioned using capillary electrophoresis on an ABI 3700 automatic DNA sequencer (PE-Applied Biosystems). The GENESCAN 3.1 software (PE-Applied Biosystems) was applied to size peak patterns, using the internal ROX 400 HD size standard and GENOTYPER 3.1 (PE-Applied Biosystems) for allele calling. To verify the repeatability of each PCR and each capillary electrophoresis run, we included two control samples (accessions BTx623 and R16) during the PCR of each SSR marker and during each capillary electrophoresis run.

Data analysis

All SSR markers showed high reproducibility, with high consistency in the amplified product between the PCR and ABI runs of the two controls, BTx623 and R16. Therefore, all 21 markers were included in the analysis. The total number of alleles detected, the number of common alleles with allelic frequencies of at least 5%, the observed size range (in basepairs; bp), the allele size differences (in Δbp) and the polymorphism information content (PIC) values (Botstein et al. 1980; Smith et al. 2000) were determined for each SSR marker. To evaluate whether the number of SSR markers employed in this study will provide sufficient information on allele diversity in this set of 100 Guinea-race sorghum accessions, we conducted a test following a modification of Fregene et al. (2003). PIC values were calculated for combinations of 2, 3, 4 and up to 20 randomly selected SSR markers for the whole dataset of 100 accessions, with a maximum of 2,000 combinations. The PIC values were then plotted against the number of SSR markers in a boxer plot.

Pair-wise genetic similarities between individual accessions were calculated using NTSYSPC ver. 2.10t (Rohlf 2001), with binary data as input, based on Dice's genetic similarity coefficient (Dice 1945). The resulting similarity matrix was subjected to non-metric multidimensional scaling (MDS) (Kruskal 1964a, b). A scatter plot of the first and second MDS axes was drawn to visualize the inter-relationship among the accessions.

The genetic diversity within and among countries and within and among eco-regions was estimated using the program "Tools for Population Genetic Analyses" (TFPGA ver. 1.30; Miller 1997) based on estimated allele frequencies, with the following statistics: average number of common alleles per locus (allelic frequencies of at

least 5%), number of polymorphic loci, percentage of polymorphic loci, observed heterozygosity (H_{obs}) and average expected gene diversity corrected for small sample sizes (H_{unb}, Nei 1978).

Allele frequency based pair-wise genetic distances between countries and between eco-regions were calculated using POWERMARKER V3.0 (Liu and Muse 2004) based on Rogers' distance (Rogers 1972). The resulting distance matrix was subjected to sequential agglomerative hierarchical nested (SAHN) clustering using unweighted pair-group method analysis (UPGMA) (Sneath and Sokal 1973) as implemented in POWERMARKER V3.0. Bootstrapping over loci with 5,000 replications was carried out to assess the strength of evidence for the branching patterns in the resulting UPGMA dendrograms. Consensus trees were constructed using the CONSENSE procedure in PHYLIP (Felstenstein 1989).

Results

SSR marker details

A total of 21 SSR markers were used to genotype 100 Guinea-race sorghum accessions and two control accessions (BTx623 and R16). Due to low peak heights in the electropherograms, which were the result of PCR failures, we lost about 1% of the data. The 21 SSR markers revealed a total of 123 alleles among the 100 Guinea-race sorghum accessions (Table 2). The average number of alleles per SSR marker was 5.6, ranging from two alleles for markers *Xcup11* and *Xcup61* to 12 alleles for markers *Xcup14* and *Xgap84* (Table 2). When we considered only alleles with frequencies of at least 5% and defined alleles with frequencies of less than 5% as rare, the average number of common alleles per SSR marker was reduced to three, ranging from one allele for marker *Xcup06* to six alleles for marker *Xtxp320*. These results indicate the presence of a relatively large proportion of rare alleles among the Guinea-race sorghum accessions studied.

The PIC value over the 21 SSR markers averaged 0.44, ranging from 0.06 for marker *Xcup06* to 0.81 for marker *Xtxp320*. A plot of PIC against number of SSR markers yielded an asymptotic curve (Fig. 1). The plot revealed that little or no additional increase in PIC value is obtainable with more than ten SSR markers.

Pattern of genetic diversity among 100 Guinea-race sorghum accessions

The average Dice similarity among the 100 accessions was 0.37 (±0.13), ranging from 0.07 (for the accession pairs IS19340-IS27705 and IS27705-IS27013) to 1.00 [for the accessions pairs IS27761-IS27768 (both margaritiferum from Sierra Leone) and IS3957-IS3967 (both Roxburghii from India/Nepal)]. This indicated that there is a fair amount of variation among the accessions.

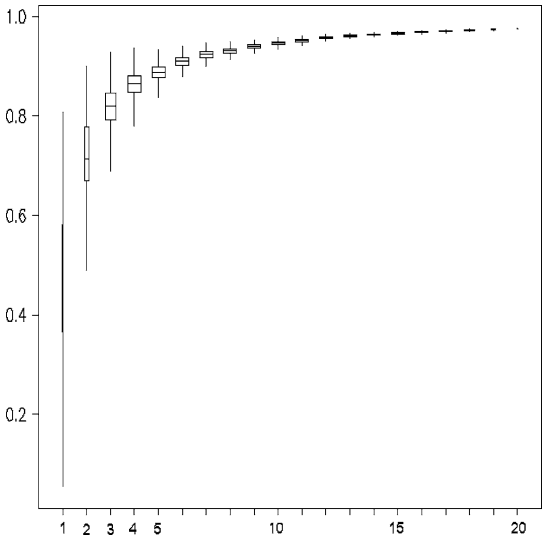

Fig. 1 Boxer plot of the PIC value against the number of loci

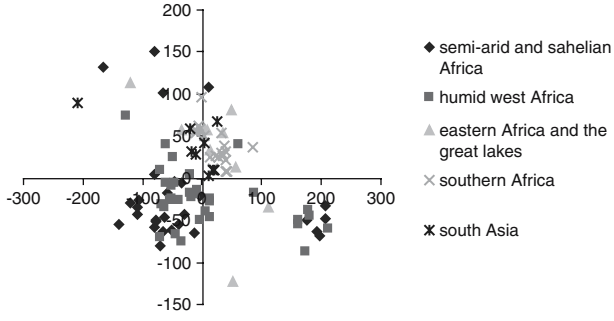

Fig. 2 MDS plot indicating the genetic relationships between 100 Guinea-race sorghum accessions. Accessions are labelled according to their eco-geographical origin

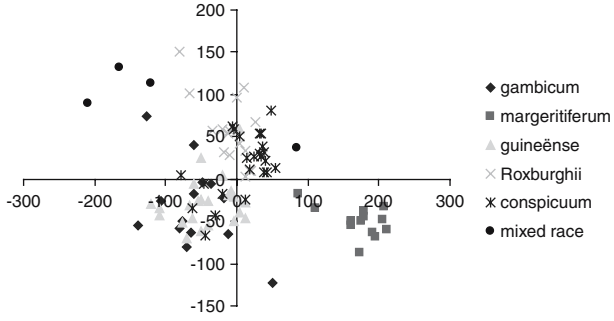

Fig. 3 MDS plot indicating the genetic relationships between 92 Guinea-race sorghum accessions. Accessions are labelled according to their Snowden Guinea-race classification

The MDS plots (Figs. 2, 3), which were constructed using the Dice similarity index, indicate that the 100 accessions can roughly be divided into two main groups—a cluster of 11 margaritiferum accessions from humid West Africa and semi-arid and Sahelian Africa in the lower-right part of the plot, and a group with the remaining 82 accessions in the left part of the plot.

Differentiation was clearly observed within the latter group of accessions: between accessions from semi-arid and Sahelian Africa (left lower part of the plot), humid West Africa (left middle part of the plot), and gradually from the *x*-axis to the top of the graph accessions from South Asia, southern Africa and accessions from East Africa and the Great Lakes.

Eco-regional pattern of genetic diversity

Genetic diversity parameters were calculated from SSR marker data at the eco-regional level (Table 3). Genetic polymorphism was found to be high to moderate within each eco-region with, on average, 16–21 of the 21 SSR loci showing polymorphism using the 5% criterion. In semi-arid and Sahelian Africa, 100% of all loci tested were polymorphic, whereas in southern Africa and South Asia, 76% of the loci tested were polymorphic. The variation in average number of alleles per locus was fairly large, with accessions from semi-arid and Sahelian Africa having an average of three alleles per locus (range: one to six), whereas those from southern Africa and South Asia had an average of two alleles per locus (range: one to five). There were a considerable number of rare alleles detected in the semi-arid and the humid West African groups, as indicated by the number of polymorphic loci dropping to an average of three alleles per SSR locus (range: one to six) based on alleles with frequencies of 5% or higher. The observed heterozygosity (H_{obs}) ranged from 0.114 (East Africa and the Great Lakes region) to 0.065 (southern Africa) and averaged 0.090, with the average expected heterozygosity corrected for small sample sizes (H_{unb}) being 0.421.

Accessions from semi-arid and Sahelian Africa and humid West Africa are clearly divergent from the accessions from southern Africa, East Africa and South Asia, as is indicated by a bootstrap value of 95 in the UPGMA dendrogram (Fig. 3). The dendrogram further illustrates the intermediate position of accessions from South Asia, which is in between the cluster with accessions of semi-arid and Sahelian African and humid West African origin and the cluster with accessions from East Africa and the Great Lakes region and Southern Africa. Both the clear divergence of accessions from semi-arid and Sahelian Africa and humid West Africa and the intermediate position of the accessions from South Asia corroborate conclusions drawn from the MDS plot (Fig. 2).

Pattern of genetic diversity based on Snowden race classification

The MDS plot coded for the Snowden-race classification of 92 Guinea-race accessions (Fig. 4) roughly illustrates the divergence of the margeritiferum accessions in the right lower portion of the plot, the conspicuum accessions in the right upper portion, the Roxburghii in the left upper portion and a mix of gambicum and guineënse

Table 3 Sample size and measures of diversity within major eco-geographical regions based on 21 SSR loci

Eco-geographical region	Sample size	Number of variable loci	Average no. allelel/locus (range)[a]	Average no. allelle/locus (≥5%) (range)[a]	Number of polymorphic loci (95%)	H_{obs}[b]	H_{unb}[b]
Semi-arid and Sahelian Africa	29	21	4 (2–9)	3 (2–6)	21	0.080	0.502
Humid West Africa	33	21	4 (2–9)	3 (1–6)	21	0.092	0.468
East Africa and the Great Lakes region	10	21	3 (1–7)	3 (1–6)	20	0.114	0.466
Southern Africa	19	21	2 (1–5)	2 (1–5)	15	0.065	0.312
South Asia	9	21	2 (1–5)	2 (1–5)	16	0.098	0.357

[a]Average and range (in parenthesis) of number of alleles per locus on a total and ≥5% allele frequency basis
[b]Heterozygosity observed (H_{obs}) and based on Nei's (1978) unbiased estimate corrected for small sample sizes (H_{unb})

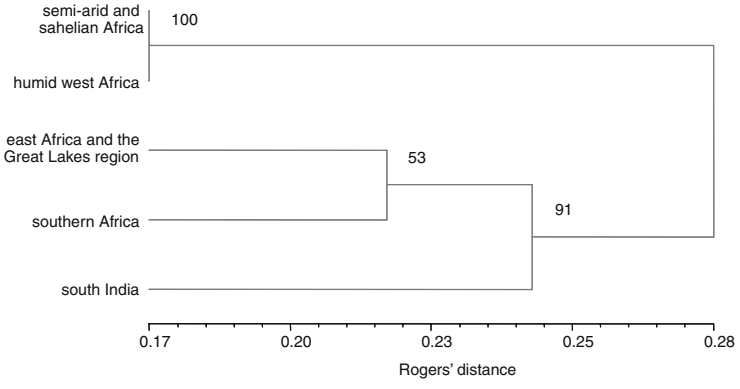

Fig. 4 UPGMA dendrogram indicating the genetic relationship among populations of sorghum accessions from four African eco-geographical regions and South Asia

accessions in the left lower portion of the plot. The accessions with a mixed-race classification can be found in the right upper and left upper portion of the plot. The margeritiferum accessions originate mainly from West humid Africa and semi-arid and Sahelian Africa, with one accession originating from Uganda. The four consipcuum accessions are humid West Africa (Cameroon, Niger and Ghana) and one from semi-arid and Sahelian Africa (Sudan). The Roxburghii accessions show the most diverse geographic origin, with most accessions originating from South Asia and a few from the four African regions. The majority of the gambicum accessions originate from the semi-arid and Sahelian Africa, whereas the majority of the guineënse accessions originate from southern Africa.

Pattern of genetic diversity at the country level

The genetic diversity parameters calculated from SSR marker data at the country level for ten African countries and India shows that there is considerable diversity within each country and large differences for magnitude of diversity (Table 4). There were an average of 16 polymorphic loci across all countries (range: 9–20). Mali and Nigeria showed the highest number of polymorphic loci, with 95% of their loci polymorphic (allelic frequencies ≥ 5%). Conversely, only 43% of the loci in Sierra Leone showed polymorphism (allelic frequencies ≥ 5%). A low variation in average number of alleles per country was observed across all 11 countries, with an average of two alleles per locus (range: one to seven), whereas two countries, Cameroon and Nigeria, had on average three alleles per locus (range: one to seven). The observed heterozygosity (H_o) ranged from 0.103 (Senegal) to 0.027 (Malawi) and averaged 0.057, whereas the expected heterozygosity (H_{unb}), corrected for small sample sizes, averaged 0.357.

The UPGMA dendrogram (Fig. 5) clearly indicates the divergence of accessions from Sierra Leone from accessions originating from the other ten countries (supported by a bootstrap *P* value of 100%). Within the group of accessions originating from these ten countries, accessions from Burkina Faso, Senegal and Mali are diverged (supported by a bootstrap *P* value of 70%) from the accessions originating from the remaining seven

Table 4 Sample size and measures of diversity[a] within 11 countries from five eco-geographical regions based on 21 SSR loci

Country	Eco-geographical region[b]	Sample size	Number of variable loci	Average no. alleles/loci (≥5%) (range)	Number of. polymorphic loci (95%)	H_{obs}	H_{unb}
Sudan	SSA	5	21	2 (1–4)	18	0.067	0.450
Senegal	SSA	6	21	2 (1–5)	18	0.103	0.416
Mali	SSA	6	21	2 (1–5)	20	0.040	0.478
Burkina Faso	SSA	8	21	2 (1–3)	18	0.042	0.343
Cameroon	HWA	10	21	3 (1–7)	17	0.058	0.388
Nigeria	HWA	12	21	3 (1–6)	20	0.048	0.446
Sierra Leone	HWA	6	21	2 (1–4)	9	0.056	0.131
Tanzania	EAGL	5	21	2 (1–4)	13	0.076	0.301
Malawi	SAf	7	21	2 (1–3)	14	0.027	0.305
Zimbabwe	SAf	6	21	2 (1–4)	16	0.032	0.284
India	SAs	7	21	2 (1–4)	16	0.078	0.387

[a]See Table 3 for definitions
[b]Eco-geographical regions: SSA, Semi-arid and Sahelian Africa; HWA, humid West Africa; EAGL, East Africa and the Great Lakes region; SAf southern Africa; SAs, South Asia
[c]Average and range (in parenthesis) of number of alleles per locus on a total and ≥5% allele frequency basis
[b]Heterozygosity observed (H_{obs}) and based on Nei's (1978) unbiased estimate corrected for small sample sizes (H_{unb})

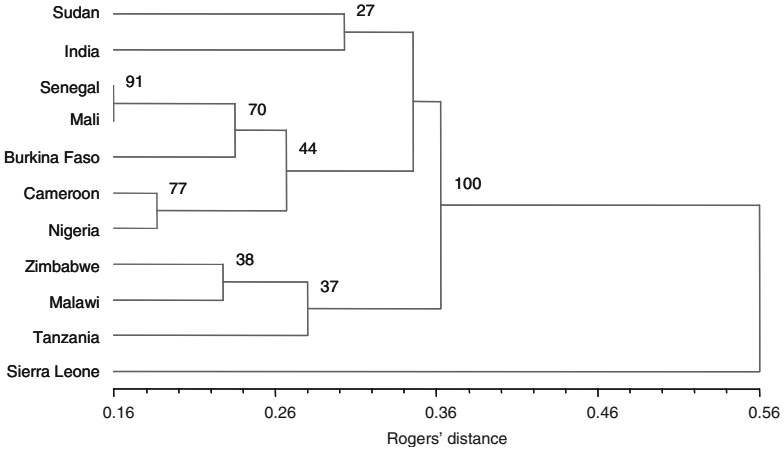

Fig. 5 UPGMA dendrogram indicating the genetic relationship among populations of sorghum accessions from ten African countries, representing four African eco-geographical regions and India (South Asia)

countries, with the clustering of accessions from Senegal and Mali supported by a bootstrap *P* value of 91%. The cluster of accessions from Nigeria and Cameroon is supported by a bootstrap *P* value of 77%. Clustering of accessions originating from southern Africa, East Africa and the Great Lakes region and India is poorly supported by a low bootstrap *P* value validation, indicating that the genetic variation among the Guinea-race sorghum accessions in these regions is considerably large and that cluster-specific alleles are very rare, if present at all, for the 21 SSR loci used in this study.

Discussion

The accessions from humid West Africa showed the largest variability for SSR allelic patterns and had the highest frequency of rare alleles. These results correspond with West Africa being the centre of domestication of the Guinea-race of sorghum (de Wet et al. 1972; Doggett 1988). The relative lower diversity exhibited by accessions from southern Africa and South Asia fits well with the historical accounts of the spread of the Guinea-race from West Africa via East Africa to southern Africa and South Asia (de Wet et al. 1972; Doggett 1988), by which the introduction of only a portion of the gene pool would result in restricted genetic variability, i.e. the founder event.

The genetic diversity within small sets of Guinea-race sorghum accessions has been analysed previously using SSR markers (Djè et al. 2000), random amplified polymorphic DNA (RAPD) markers (de Oliveira et al. 1996; Dahlberg et al. 2002) or restriction fragment length polymorphism (RFLP) markers (Deu et al. 1994). These studies focussed on the analyses of variation between the different races of sorghum. In most cases Guinea-race sorghum accessions could clearly be distinguished from the four other races and numerous Guinea-race specific markers were identified, illustrating a clear relationship between molecular divergence and racial discrimination. Employing SSR markers to analyse the genetic variation among five Guinea-race sorghum accessions, Djè et al. (2000) revealed an observed heterozygosity (H_{obv}) of 0.089, with average expected gene diversity (H_{exp}) of 0.224. These figures nicely comply with the observations among the 100 Guinea-race sorghum accessions reported here and point to the fact that Guinea-race sorghum is predominantly inbreeding, resulting in low levels of observed heterozygosity, but that the gene pool as a whole maintains a high level of allelic variation.

The PIC of an SSR marker provides an estimate of the discriminatory power of that SSR marker by taking into account not only the number of alleles that are detected but also the relative frequencies of those alleles (Smith et al. 2000). Most of the SSR markers used in this study revealed a moderate discriminatory power. The use of PIC to evaluate the number of SSR markers needed to provide sufficient information on allele diversity in a given dataset (Fregene et al. 2003) was applied in this study. The asymptotic curve revealed by plotting PIC versus number of SSR markers (Fig. 1) shows that the 21 SSR markers used in this study were more than sufficient to assess the genetic variation in this set of Guinea-race accessions. The plot reveals that little or no increase of PIC is obtainable with more than ten markers. This number of SSR markers is considerably smaller than found necessary by Fregene et al. (2003), who indicated that a minimum of 30 unlinked SSR markers was required to assess genetic diversity of 283 cassava landrace accessions. This difference is likely due to cassava being an out-crossing, highly heterozygous, vegetatively propagated species comprising a relatively large number of alleles. Sorghum is a primarily inbreeding, genetically homogeneous species comprising a relatively small number of alleles [up to 12 alleles per SSR marker, including rare alleles (frequency less than 5%)]. The high levels of allelic variability but low levels of heterozygosity observed in this study correspond with a previous SSR marker study of five Guinea-race accessions (Djé et al. 2000) and fits with the predominantly inbreeding nature of sorghum.

Clear patterns of genetic divergence among the Guinea-race accessions were observed based on the Dice similarity index of SSR allelic patterns. The clear distinction of West African from East African, southern African and South Asia accessions confirms earlier findings of divergence between West African and southern African accessions based on RFLPs (Deu et al. 1994) and isozymes (Ollitrault et al. 1989). However, a mitochondrial DNA diversity study of 43 Guinea accessions did not reveal this divergence (Deu et al. 1995). The shared origins of East African, southern African and South Asian accessions is further suggested by their overlap in the MDS plot of Dice's similarity coefficient of SSR allelic patterns.

The major divergence of the margaritiferums from all other Guinea accessions observed in this study confirms the previous observations based on diversity for RFLP markers (Deu et al. 1994), mitochondrial DNA (Deu et al. 1995) and isozymes (Ollitrault et al. 1989). These results support the suggestion of an independent origin of the margaritiferums, possibly with introgression from wild species (Ollitrault et al. 1989; Deu et al. 1995), but do not support the possibility of southern African origins. It is not clear if the lack of allelic variation among the margaritiferums, particularly those from Sierra Leone, is due to selection for specific adaptive constraints or to a sampling effect. However, previous observations of the absence of alleles in margaritiferums relative to other Guinea sorghums (K. von Brocke, personal communication) suggests that genetic drift, due to adaptation or sampling effects, may be an important determinant of this pattern.

The apparent clear divergence among Guinea-race sorghum landraces based on the eco-geographic and Snowden morphological classification could be of great importance to applied breeding programmes to the extent to which the patterns for useful genetic diversity

correspond to the neutral SSR allelic patterns observed here. These results suggest that Guinea hybrid breeding work could develop heterotic pools on the basis of eco-geographic/morphological characteristics so as to maximize genetic distance and heterosis. For example, the guineënse and gambicum accessions from West Africa could serve as one pool, and the conspicuums from humid West-, East-, and southern Africa could serve as a separate pool. Hybrid combinations representing eco-geographic and morphologically contrasting parents are currently being evaluated (Rattunde, unpublished). Jordan et al. (2003) observed only weak associations between whole genome heterozygosity based on RFLP markers and heterosis in non-Guinea-race sorghum materials. This association still needs to be assessed for the more diverse Guinea-race gene pool.

The 21 SSR markers evaluated in this study consisted of four di-nucleotide, seven tri-nucleotide, seven tetra-nucleotide, one penta-nucleotide and two hexa-nucleotide repeat units. Mutations in SSR markers resulting in allele size differences are often caused by deletions or insertions of single or multiple repeat units due to unequal crossing over followed by concerted evolution. The stepwise mutation model (SMM; Kimura and Crow 1964) assumes that alleles mutate back and forth by a small number of repeats and that the same allelic states are created repeatedly over time. An alternative model is the infinite alleles model (IAM; Ohta and Kimura 1973), which assumes that each mutation creates a new allele in the population. Of 21 the SSR markers tested in this study, 14 seem to comply with the SMM, although the different allelic states have not been confirmed by sequencing. *Xtxp320*, for instance, a marker with a tri-nucleotide repeat unit, revealed 12 alleles, differing in size by three or six nucleotides. However, six SSR markers— *Xcup07*, *Xtxp40*, *Xcup32*, *XSbKAF1*, *Xcup53* and *Xcup69*—showed discrepancies in the expected allele size differences, and some allelic states clearly do not fit the SMM.

Similarly, in maize, Matsuoka et al. (2002) sequenced 31 alleles of six SSR loci in *Zea mays*. The DNA sequences revealed complex mutation patterns resulting from insertions/deletions (indels) in the regions flanking the repeat motifs. Although micro-satellite repeat changes were observed, the indels in the flanking regions were more frequent and thus responsible for most of the variation in allele size. Uptmoor et al. (2003) compared the performance of an SSM- and an IAM-based genetic distance when analysing the genetic variation among 46 southern African sorghum accessions using SSR markers. They concluded that the SMM-based genetic distance measure outperformed the IAM-based genetic distance in terms of higher correlations between SMM-based matrices and AFLP and RAPD distance matrices as compared to IAM-based matrices and AFLP and RAPD distance matrices. However, since 7 out of 21 SSR markers employed in this study violate the assumption of SMM and the loci of the remaining 14 SSR alleles were not sequenced to check their compliance with SMM, it was felt that the use of an SMM-based distance measure was inappropriate.

The main objective of the research reported here was to analyse the pattern of genetic diversity within Guinea-race landraces to facilitate their use in breeding programmes for West and Central Africa. The relatively large proportion of rare alleles (allele frequency $< 5\%$) and overall divergence among Guinea-race sorghum accessions indicates opportunities to select divergent parents with enough genotypic and phenotypic variation to allow mapping of genes and quantitative trait loci and for marker-assisted introgression of traits into elite breeding lines within this race itself.

Acknowledgements We dedicate this manuscript to Dr. Hartwig H. Geiger, Professor of Population Genetics at the University of Hohenheim, Stuttgart, Germany, for his lifelong contribution to cereal breeding. We wish to thank Dr. Cecile Grenier for creating the Guinea-race Core Collection and Dr. Jacques Chantereau for assistance with the Snowden classification. This research was financed using a Core contribution from ICRISAT.

References

Bhattramakki D, Dong JM, Chhabra AK, Hart GE (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L) Moench. Genome 43:988–1002

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

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 sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet 93:190–198

Chittenden LM, Schertz KF, Lin Y-R, Wing RA, Paterson AH (1994) A detailed RFLP map of *Sorghum bicolor* × *S propinquum*, suitable for high-density mapping, suggests ancestral duplication of *Sorghum* chromosomes or chromosomal segments. Theor Appl Genet 87:925–933

Condit R, Hubble SP (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome 34:66–71

Dahlberg JA, Zhang X, Hart GE, Mullet JE (2002) Comparative assessment of variation among sorghum germplasm accessions using seed morphology and RAPD Measurements. Crop Sci 42:291–296

Dean RE, Dahlberg JA, Hopkins MS, Mitchell SE, Kresovich 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

Deu M, Gonzalez-de-Leon D, Glaszmann J-C, 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

Dice LR (1945) Measures of the amount of ecological associations between species. Ecology 26:297–302

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

Djè Y, Heuertz M, Lefèbvre C, Vekemans X (2000) Assessment of genetic diversity within and among germplasm accessions in

cultivated sorghum using microsatellite markers. Theor Appl Genet 100:918–925
Doggett H (1988) Sorghum, 2nd edn. Longman Scientific & Technical, London
FAOSTAT (2004) http://appsfaoorg/defaulthtm
Felsenstein J (1989) PHYLIP—Phylogeny inference package (version 32). Cladistics 5:164–166
Fregene MA, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A, Mitchel S, Gullberg U, Rosling H, Dixon AGO, Dean R, Kresovich S (2003) Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. Theor Appl Genet 107:1083–1093
Ghebru B, Schmidt RJ, Bennetzen JL (2002) Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. Theor Appl Genet 105:229–236
Grenier C, Deu M, Kresovich S, Bramel-Cox PJ, 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
Grenier C, Bramel-Cox PJ, Hamon P (2001a) Core collection of sorghum: I. Stratification based on eco-geographical data. Crop Sci 41:234–240
Grenier C, Hamon P, Bramel-Cox PJ (2001b) Core collection of sorghum: II. Comparison of three random sampling strategies. Crop Sci 41:241–246
Harlan JR, de Wet JMJ (1972) A simplified classification of cultivated sorghum. Crop Sci 12:172–176
Jordan DR, Tao Y, Godwin ID, Henzell RG, Cooper M, McIntyre CL (2003) Prediction of hybrid performance in grain sorghum using RFLP markers. Theor Appl Genet 106:559–567
Kimura M, Crow JF (1964) The numbers of alleles that can be maintained in a finite population. Genetics 49:725–738
Kong L, Dong J, Hart GE (2000) Characteristics, linkage-map positions, and allelic differentiation of *Sorghum bicolour* (L.) Moench DNA simple-sequence repeats (SSRs). Theor Appl Genet 101:438–448
Kresovich S, Szewc-McFadden AK, Bliek SM, McFerson JR (1995) Abundance and characterisation of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of Brassica napus L. (rapeseed). Theor Appl Genet 91:206–211
Kruskal JB (1964a) Multidimensional scaling by optimising goodness of fit to a non-metric hypothesis. Psychometrika 29:1–27
Kruskal JB (1964b) Nonmetric multidimensional scaling: a numerical method. Psychometrika 29:28–42
Liu K, Muse S (2004) POWERMARKER: new genetic data analysis software Version 3.0. Free program distributed by the author over the internet from http://www.powermarker.net
Mace ES, Buhariwalla HK, Crouch JH (2003) A high throughput DNA extraction protocol for molecular breeding programs. Plant Mol Biol Rep 21:459a–459h
Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebly J (2002) Microsatellites in *Zea*; variability, patterns of mutations, and use for evolutionary studies. Theor Appl Genet 104:436–450
Miller MP (1997) Tools for population genetic analyses (TFPGA) version 13: a windows program for the analysis of allozyme and population genetic data. Program distributed by the author over the internet from http://biowebusuedu/mpmbio/tfpgaasp
Mitchell SE, Kresovich S, Jester CA, Javier Hernandez C, Szewc-McFadden AK (1997) Application of multiplex PCR and fluorescence-based, semi automated allele sizing technology for genotyping plant genetic resources. Crop Sci 37:617–624
Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590
Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet Res 22:201–204
de Oliveira AC, Richter T, Bennetzen JL (1996) Regional and racial specificities in sorghum germplasm assessed with DNA markers. Genome 39:579–587
Ollitrault P, Arnaud M, Chantereau J (1989) Enzyme polymorphism in sorghums II Genetic and evolutionary constitution of cultivated sorghum. Agron Trop 44:211–222
Peng Y, Schertz KF, Cartinhour S, Hart GE (1999) Comparative genome mapping of *Sorghum bicolor* (L.) Moench using an RFLP map constructed in a population of recombinant inbred lines. Plant Breed 118:225–235
Purseglove JW (1985) Tropical crops: monocotyledons. Longman, London
Röder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Genet Gen 246:327–333
Rogers JS (1972) Measures of genetic similarity and genetic distance. Stud Genet VII Univ Tex Publ 7213:145–153
Rohlf (2001) NTSYSPC, ver. 210t. Applied Biostatistics. SSPS, Chicago, Ill.
Schloss SJ, Mitchell SE, White GM, Kukatla R, Bowers JE, Paterson AH, Kresovich S (2002) Characterization of RFLP clone sequences for gene discovery and SSR development in *Sorghum bicolor* (L). Moench. Theor Appl Genet 105:912–920
Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF (1998) The mutation rates of di-, tri- and tetra-nucleotide repeats in *Drosophila melanogaster*. Mol Biol Evol 15:1751–1760
Smith JR, Carpten JD, Brownstein MJ, Ghosh S, Magnuson VL, Gilbert DA, Trent JM, Collins FS (1995) Approach to genotyping errors caused by non-templated nucleotide addition by *Taq* DNA polymerase. Genome Res 5:312–317
Smith JSC, Kresovich S, Hopkins MS, Mitchell SE, Dean RE, Woodman WL, Lee M, Porter K (2000) Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. Crop Sci 40:226–232
Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman, New York
Snowden JD (1936) The cultivated races of sorghum. Adlard, UK
Taramino G, Tarchini R, Ferrario S, Lee M, Pé ME (1997) Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolour*. Theor Appl Genet 95:66–72
Uptmoor R, Wenzel W, Friedt W, Donaldson G, Ayisi K, Ordon F (2003) Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Southern Africa by RAPDs, AFLPs and SSRs. Theor Appl Genet 106:1316–1325
de Vries J, Toenniessen G (2001) Securing the harvest: biotechnology, breeding and seed systems for African crops. CABI Publ, Wallingford
de Wet JMJ, Harlan JR, Kurmarohita B (1972) Origin and evolution of Guinea sorghums. East Afr Agric For J 114–119