academicJournals

Vol. 13(14), pp. 1503-1514, 2 April, 2014
DOI: 10.5897/AJB08.1030
Article Number: AAC0F8C47113
ISSN 1684-5315

http://www.academicjournals.org/AJB

African Journal of Biotechnology

Full Length Research Paper

Assessment of the genetic diversity and pattern of relationship of West African sorghum accessions using microsatellite markers

M. N. Kitavi[1]*, D. K. Kiambi[1], B. Haussman[2], K. Semagn[1], G. Muluvi[3], M. Kairichi[1] and J. Machuka[3]

[1]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), P. O. Box 30709-00100 Nairobi, Kenya.
[2]ICRISAT-Niamey (Regional hub WCA), BP 12404, Niamey, Niger (Via Paris).
[3]Department of Biochemistry and Biotechnology, Kenyatta University; P. O. Box 43844, Nairobi, Kenya.

Received 12 November, 2008; Accepted 30 March, 2009

An understanding of the extent, distribution and patterns of genetic variation is useful for estimation of any possible loss of genetic diversity and assessment of genetic variability and its potential use in breeding programs, including establishment of heterotic groups. This study assessed patterns of genetic diversity and relationships among 30 West African sorghum accessions using 22 microsatellite markers. Population structure and within population genetic diversity was also assessed using the same markers. Genotypic data was generated using the ABI Prism 3730 and alleles called and sized using genemapper software version 3.7. Molecular data was analyzed using DARwin 4.0, powermarker 3.0 and Arlequin version 3.11. The average marker quality index was 0.27 while a mean PIC of 0.54 was observed across the 22 SSR markers. Among the 30 accessions, the markers detect a total of 146 alleles with an average of 6.6 alleles per marker. Results from the various statistical analyses performed revealed a wide range of polymorphism from 22.7 to 86.4%. The mean heterozygosity was relatively low at 0.28 while the average Nei's genetic diversity among the 30 populations was 0.57. The within population Nei's genetic diversity assessed from 49 individuals in 10 populations was lower at 0.54 and the average heterozygosity was also lower at 0.21. Cluster and principal coordinate analysis of the 30 populations revealed two distinct groups independent of their geographic origins The examination of the hierarchical partitioning of genetic variation by AMOVA demonstrated that genetic differentiation was significant at P < 0.00. Of the total diversity, 8.9% was attributed to country differences, 54.11% was attributed to population differences within the countries while 36.99% was attributed to differences within populations. The F_{ST} value (0.63) indicated a very high genetic differentiation as expected for selfing species. This study demonstrates the utility of SSR markers in detecting polymorphism, estimating genetic diversity and establishment of genetic clusters for heterotic studies.

Key words: Genetic diversity, heterozygosity, SSR microsatellites, heterotic grouping, alleles, polymorphism.

INTRODUCTION

Sorghum (*Sorghum bicolor* (L.) Moench) is the fifth most important cereal crop worldwide after wheat (*Triticum species*), rice (*Oryza species),* maize (*Zea mays*) and barley (*Hordeum vulgare*) with an annual average production of 61 million tones over the past decade (FAO, 1995; FAOSTAT, 2004; http://appsfaoorg/defaulthtm). Sorghum, together with pearl millet (*Pennisetum americanum* (L.) and finger millet (*Eleusine coracana* (L.), represent Africa's main contribution to the world food supply (de Vries and Toenniessen, 2001). The crop can be grown under a wide range of soil and climatic conditions and thus plays a major role in semi-arid regions of the world where drought, heat and poor soil conditions make production of other cereals difficult (Dogget, 1988; House, 1995). The crop occupies 25% or more of arable land in Mauritania, Gambia, Mali, Burkina Faso, Ghana, Niger, Somalia and Yemen, and > 10% of this area in Nigeria, Chad, Sudan, Tanzania and Mozambique.

Approximately 120 million people living in semi-arid West and Central Africa (WCA) depend on pearl millet and sorghum as the major food crops and sources of income. Because of its global socio-economic importance, there has been substantial interest in characterizing the levels of genetic diversity present within sorghum using both phenotypic and molecular markers (Dean et al., 1999; Dje et al., 2000; Ghebru et al., 2002; Grenier et al., 2000a, b; Menz et al., 2004; Casa et al., 2005). These analyses have provided a foundation of genetic data for making informed decisions regarding the management and utilization of genetic resources (Casa et al., 2005). Understanding the structure of genetic diversity provides the foundation for effective and sustained population breeding and hybrid development in sorghum. The genetic diversity in the germplasm of a breeding program affects the potential genetic gain through selection. Information about genetic diversity also permitted the classification of germplasm into heterotic groups, which are particularly important to hybrid breeding. Even though the genetic mechanisms that explain heterosis are not fully understood, it is well documented that crosses between unrelated and consequently genetically distant parents, show greater hybrid vigor due to an enhanced degree of heterozygosity than crosses between closely related parents (Stuber, 1994; Hallauer, 1999; Menz et al., 2004). Estimates of molecular-marker based genetic distance have been proven to be a useful way to describe existing heterotic groups to identify new heterotic groups and to assign inbreeds of unknown genetic origin to established heterotic groups (Dubreuil et al., 1996; Saghai et al., 1997; Hongtrakul et al., 1997; Pejic et al., 1998; Casa et al., 2002). The research presented here was to provide an initial assessment of the pattern of sorghum diversity of West Africa in order to provide insights for tapping this diversity in current and future breeding efforts. We used SSR loci to determine diversity and evaluate genetic relationships and population structure within and among 30 sorghum accessions.

MATERIALS AND METHODS

Plant materials

Seed samples of 30 sorghum accessions (ICRISAT B lines and parents from heterosis study) were collected from different countries and diverse geographic origins in West Africa and their passport data is provide in Table 1.

DNA extraction, quantification and normalization

Individual and bulking sampling methods were used. Approximately 6 cm of 4 seedlings leaf-tissue of each accession was pooled for all the 30 accessions. 10 accessions were selected randomly and 5 seedlings per accession sampled individually. Isolation of DNA was done using a modified CTAB protocol (Mace et al., 2004). Determination of the quality, quantity and purity of the isolated DNA was essential for PCR optimization and were done using agarose (0.7%) gel electrophoresis stained with ethidium bromide (10 mg/ml) and Nanodrop spectrophotometer. All DNA samples were diluted to the required concentration (5 ng/μl). This was necessary to ensure uniformity in results.

PCR optimization and amplification

All the 22 primers were run with the BTX623 standard DNA and optimized by varying the primer (2 pmol/ul) and $MgCl_2$ (10 nM/ul) between 0.6 - 2.0 ul. The dNTPs (2 mM/ul and enzyme (5 U/ul) were varied between 0.4 - 0.75 and 0.04 -0.08 respectively. The annealing temperatures of the primers were optimized using the touchdown PCR amplification procedure as described by Folksterman et al. (2005). PCR reactions conditions were set in 10 μl volumes in 384-well PCR plates (ABGene, Rochester, N.Y.). Temperature cycling was carried out using the GeneAmp PCR systems 9600 (PE-Applied Biosystems). A final extension of 20 min at 72°C was included to minimize the +A overhang (Smith et al., 1997). The PCR products was run on 2.5% agarose gel (Figure 1) and amplification checked with the DNA analyzer.

Capillary electrophoresis

Genotyping was carried out by capillary electrophoresis using the ABI PRISM 3730 (Applied Biosystems), a fluorescent based

*Corresponding author. E-mail: m.kitavi@cgiar.org.

Table 1. Sorghum accessions selected for the study.

Entry	Identification	Origin	Race	Breeding origin
1	IS 3534B	Sudan	Guinea	Identified maintainer line
2	FambéB	Mali	Guinea	Maintainer, product of mutation breeding
3	IPS001B	Mali	Guinea	Maintainer, product of mutation breeding
4	CSM219B	Mali	Guinea	Maintainer, product of mutation breeding
16	(GPN99 271-20-2)	West African	Guinea	Maintainer, derived from broad-based Guinea population
17	ICSB 38	ICRISAT-Nigeria	Caudatum	Pedigree breeding
19	IS 6731bf	Burkina Faso	Guinea	Landrace
21	IS 27564bf	BurkinaFaso	Guinea	Landrace
22	Seguetana CZ	Mali	Guinea	Landrace
34	GPN01 267-9-1	ICRISAT-Mali	Guinea	Derivative of Guinea Population
40	IS 15629ca	Cameroon	Guinea	Landrace
43	Fara Fara-17	Nigeria	Guinea/Caudatum	Landrace
48	Kaura-Katsina	Nigeria	Durra/Caudatum	Landrace
49	Guinea-Katsima	Nigeria	Guinea	Landrace
50	CSM388	Mali	Guinea	Landrace
105	90SN3	Niger	Caudatum	Pedigree breeding
107	90SN5	Niger	Caudatum	Pedigree breeding
110	A223	USA	Caudatum	Pedigree breeding
111	AT623	USA	Caudatum	Pedigree breeding
139	98-BE-F5P-82	IER-Mali	Guinea/Caudatum	Pedigree breeding derivative
140	97-SB-F5DT-150 B	IER-Mali	Guinea/Caudatum	Pedigree breeding derivative
141	02-SB-F5DT-12 B	IER-Mali	Guinea/Caudatum	Pedigree breeding derivative
142	02PR-3009K B	IER-Mali	Guinea/Caudatum	Pedigree breeding derivative
144	00-KO-F5DT-19	IER-Mali	Guinea/Caudatum	Pedigree breeding derivative
145	03-F4T-38	IER-Mali	Guinea/Caudatum	Pedigree breeding derivative
146	02-SB-F4FT-298	IER-Mali	Guinea/Caudatum	Pedigree breeding derivative
147	02-SB-F4FT-189	IER-Mali	Guinea/Caudatum	Pedigree breeding derivative
153	SK5912	Nigeria	Durra/Caudatum	Derived from landrace
154	Buhu Banza	Nigeria	Durra/Caudatum	Landrace
155	Fara Fara	Nigeria	Guinea/Caudatum	Landrace

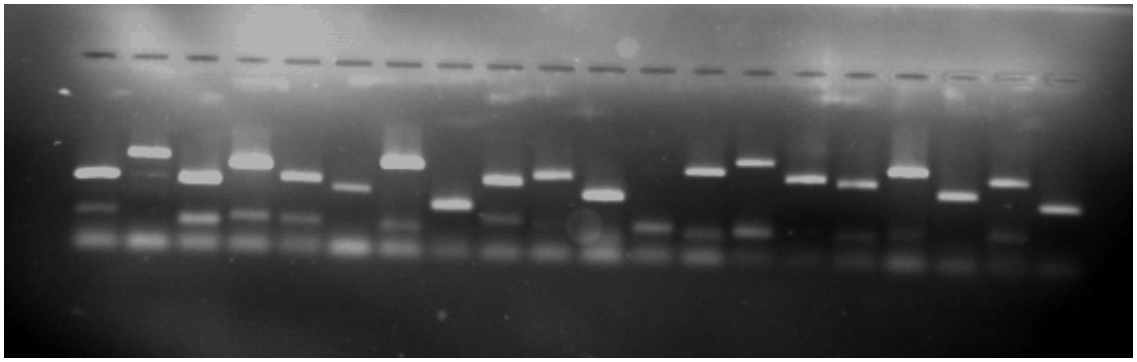

Figure 1. Primer amplification products obtained with the optimized PCR protocol.

capillary detection system that uses polymer as the separation matrix. This facilitated the accurate sizing of the microsatellite alleles to within ± 0.3 base pairs (Buhariwalla and Crouch, 2004). PCR products were co-loaded post-PCR based on dye label, fragment size and fluorescence to reduce the unit cost of high throughput genotyping (Table 2). 0.5 - 1 µl of labelled PCR

Table 2. Summary of the 22 SSR markers, primer multiplex and co-loading sets used in the study.

Marker name	Dye label	Multiple x set	Chromosome	Repeat motif	Reference
gpsb123	VIC		8	(CA)7+(GA)5	Unpublished*
mSbCIR223	HEX	Set 2	2	(AC)6	Unpublished
mSbCIR240	HEX	Set 7	8	(TG)9	Unpublished
mSbCIR276	NED	Set 7	3	(AC)9	Unpublished
mSbCIR283	NED	Set 5	10	(CT)8 (GT)8.5	Unpublished
mSbCIR286	6-FAM	Set 6	1	(AC)9	Unpublished
SbAGB02	VIC	Set 5	7	$(AG)_{35}$	Taramino et al. (1997)
Xcup11	NED	Set 2	3	$(GCTA)_4$	Schloss et al. (2002)
Xgap072 = Sb4-72	VIC	Set 3	6	(AG)16	Brown et al. (1996)
Xgap084=Sb6-84	VIC	Set 2	2	(AG)14	Brown et al. (1996)
Xisep0310	HEX	Set 1	2	(CCAAT)4	unpublished, ICRISAT
Xtxp012	PET	Set 6	4	$(CT)_{22}$	Kong et al. (2000)
Xtxp021	6-FAM	Set 4	4	$(AG)_{18}$	Kong et al. (2000)
Xtxp114	6-FAM	Set 3	3	$(AGG)_8$	Bhattramakki et al. (2000)
Xtxp136 (Kaf3)	VIC	Set 6	5	$(GCA)_5$	Bhattramakki et al. (2000)
Xtxp141	NED	Set 3	10	$(GA)_{23}$	Bhattramakki et al. (2000)
Xtxp265	VIC	Set 4	6	$(GAA)_{19}$	Bhattramakki et al. (2000)
Xtxp273 (Pbbf)	6-FAM	Set 5	8	$(TTG)_{20}$	Bhattramakki et al. (2000)
Xtxp278	NED	Set 1	7	$(TTG)_{12}$	Bhattramakki et al. (2000)
Xtxp286	6-FAM	Set 7	2.00	$(GCA)_4ACA(GCA)_5A$ $(CAA)_5 + (AAC)_9$	Kong et al. (2000)
Xtxp320 (PhyB)	NED	Set 4	1	$(AAG)_{20}$	Bhattramakki et al. (2000)
Xtxp321	PET	Set 1	8	$(GT)_4 + (AT)_6 + (CT)_{21}$	Bhattramakki et al. (2000)

* Unpublished, Agropolis-CIRAD-Genoplante.

products were loaded and mixed with formamide (PE-Applied Biosystems) and ROX-labelled GS500LIZ-3730 size standard (PE-Applied Biosystems). DNA fragments were denatured and size-fractioned using capillary electrophoresis on an ABI 3730 automatic DNA sequencer (PE-Applied Biosystems). The peaks were sized and the alleles called using genemapper software and the internal ROX GS500LIZ-3730 size standard (Figure 2). This system has the advantages of automated filling of capillaries, automated sample loading and rapid electrophoresis (Buhariwalla and Crouch, 2004). To verify the repeatability of each PCR and each capillary eletrophoresis run, a control sample (accession BTx623) was included during the PCR of each SSR marker and during each capillary electrophoresis run. Only alleles with a relative fluorescent unit of > 500 were scored.

Data analysis

All SSR markers showed high reproducibility with high consistency in the amplified product between the PCR and ABI runs of the control BTx623. Allele sizing and calling was done using genemapper ver 3.7 software. Adjustment for allele size inconsistencies and checking quality of markers was done by Allelobin. Genetic diversity parameters that is polymorphic information content (PIC), heterozygosity and number of alleles for each marker, % of polymorphic loci estimates, genetic diversity within and among the accessions and countries, genetic distances within and among accessions and genetic distances within and among the countries were done using powermarker version 3.25 and tools for population genetics analysis (TFPGA) softwares. TFPGA provided 3 estimates of heterozygosity, direct count, expected heterozygosities under Hardy-Weinburg equilibrium (H-W) and Nei`s (1978) unbiased heterozygosity. PIC values give the information that each marker impacts into the study, which is the measure of the usefulness of each marker in distinguishing one individual from another. Darwin version 5.0 software was used to calculate the principle component analysis (PCA) and clustering within and among the accessions and countries. To determine the genetic relationships and differentiation, the accessions were clustered based on the matrix of genetic similarities using the un-weighted pair group method using arithmetic averages (UPGMA) clustering algorithm. Dissimilarity was calculated from alleleic data, where dissimilarity index was calculated by simple matching coefficient. Genetic distances were computed for microsatellite data and trees constructed using the neighbour-joining method in Darwin version 5.0 software.

The principle component scores were not standardized and thus had variance equal to the corresponding eigen values. The UPGMA results were used to generate dendrograms. The robustness of the phylogenies was evaluated by bootstrapping (1000 permutations) replicates over loci. An exact test was used to determine possible deviations from Hardy-Weinberg equilibrium and the existence of non-random associations of genotypes across polymorphic co-dominant loci (Guo and Thompson, 1992; Weir, 1990). Exact tests were performed using TFPGA software. Analysis of molecular

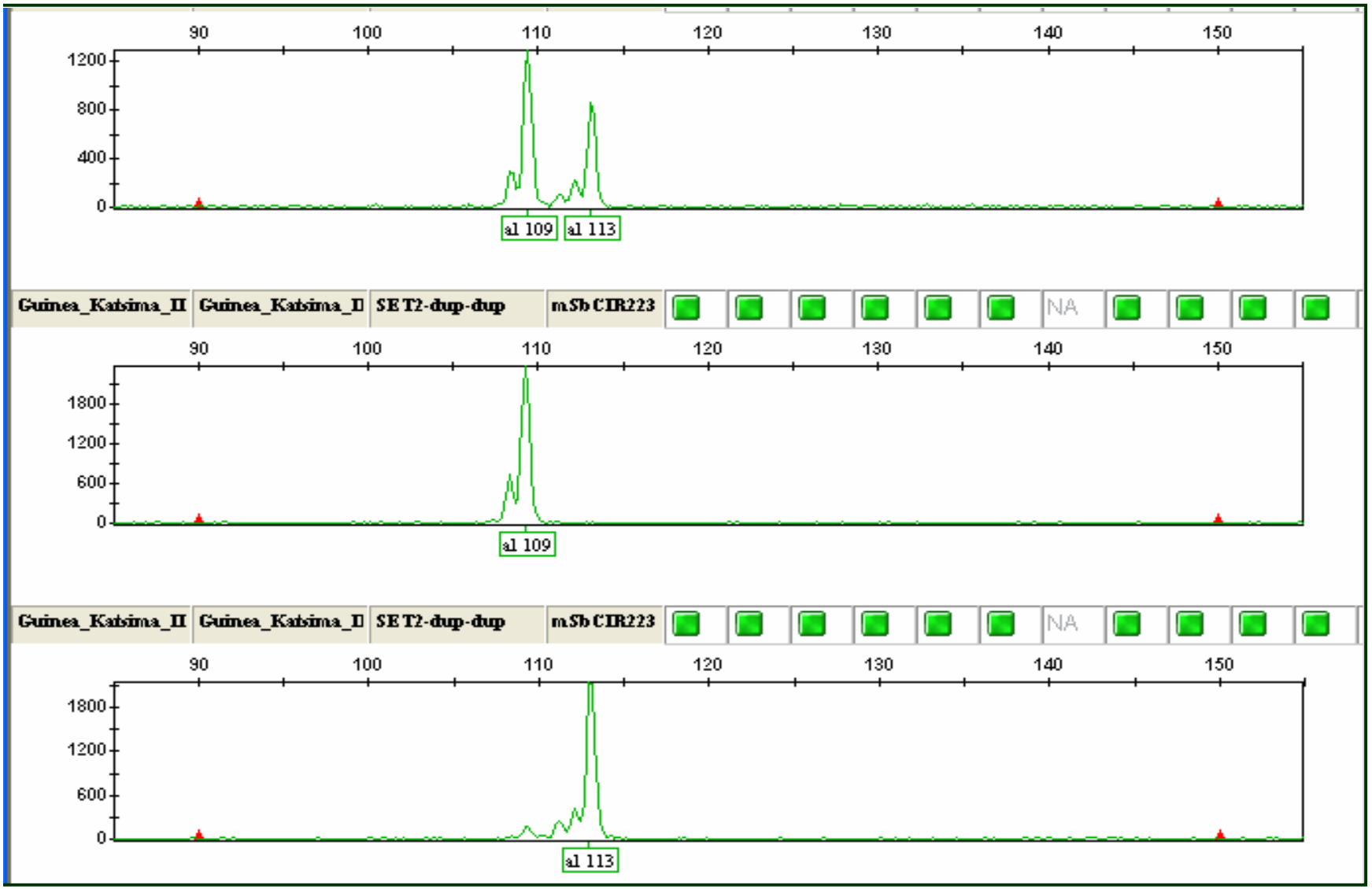

Figure 2. Electrophenograms from ABI 3730 capillary sequencer showing heterozygous individual (top) and homozygote individuals (middle and bottom). The numbers at the bottom of each peak indicate allele size in base pairs (bp).

Table 3. Performance of all markers in the bulks (Allelobin software).

Marker name	Repeat length	Quality index	Total alleles	Minimum allele	Maximum allele	Allele (%)	Rare alleles (< = 5%)
ISEP310	5	0.2009	2	159	204	204 (93.55)	
Xtxp321	2	0.1776	14	190	222	202 (22.58)	190, 192, 200, 206, 208, 210, 212, 222
Xtxp278	3	0.2191	4	249	285	249 (90.62)	252, 270, 285
Xcup11	4	0.26	3	154	170	170 (59.52)	154
CIR223	2	0.2139	7	108	126	116 (56.82)	120,126
sb6-84	2	0.3142	8	168	196	182 (42.86)	168,196
Sb4-72	2	0.1157	5	186	196	186 (57.89)	
Xtxp114	3	0.0719	2	211	214	214 (70.59)	
Xtxp141	2	0.2042	9	147	191	151 (51.35)	153,161,165,183
Xtxp21	2	0.4299	6	176	188	176 (61.29)	188
Xtxp265	3	0.4844	15	171	228	189 (16.67)	171, 198, 201, 204, 216, 219, 222, 228
Xtxp320	3	0.2985	8	266	290	278 (29.73)	
Xtxp273	3	0.202	11	163	211	193 (33.33)	163,184,199, 202, 211
CIR283	2	0.4556	10	99	147	117 (22.86)	131,141, 143,147
sbAG02	2	0.1923	5	96	120	96 (50.00)	116, 118, 120
Xtxp12	2	0.4695	9	171	195	175 (27.27)	179, 187, 191
Xtxp136	3	0.1702	2	241	244	241 (88.24)	
CIR286	2	0.3483	5	111	149	113 (64.71)	149
Xtxp286	3	0.4271	6	110	209	113 (37.93)	110, 194
CIR240	2	0.2117	7	112	158	112 (59.52)	136, 152, 154
CIR276	2	0.46	4	229	251	229 (39.39)	251
gpSb123	2	0.2236	4	288	296	290 (58.97)	294
Total			146				

variance (AMOVA, Excoffier et al. 1992) was used to partition SSR variation among-groups and among-populations within-groups. Within-population components were computed by a non-parametric permutation procedure, using 1000 permutations. AMOVA and F_{st} indices were calculated using the ARLEQUIN version 3.11 (http://cmpg.unibe.ch/software/arlequin3).

RESULTS

SSR marker performance

A total of 22 SSR markers were optimized and used to genotype 30 sorghum bulks and 49 individuals from 10 accessions with BTx623 being used as an internal control accession. Marker quality indices (Table 3) for both the bulk accessions and individual plants samples varied from 0.07 in Xtxp114 to 0.48 in Xtxp265 and 0.11 in SbAGB02 to 0.48 in Xtxp286, respectively, and the overall average quality index for the 22 markers was 0.27. The 22 SSR markers revealed a total of 146 different alleles among the 30 accessions, an average of 6.6 alleles per marker. The number of alleles detected in each of the 22 SSR markers was highly variable and ranged from 2 in Xisep0310, Xtxp114 and Xtxp136 to 15 by markers Xtxp 321 and Xtxp 265 while markers Sb4-72, Xtxp114, Xtxp 320 and Xtxp136 did not detect any rare alleles.

For the 49 individuals from 10 populations, the 22 SSR markers revealed a total of 108 alleles, with an average of 4.9 alleles per marker. The total number of alleles detected in the individual plants ranged from 2 in 4 markers (Xcup11, Xisep0310, Xtxp136 and Xtxp278) to 9 in Xtxp265 and Xtxp321.

In the bulk populations, polymorphism information content (PIC) ranged from 0.09 in markers ISEP0310 and Xtxp278 to 0.86 in marker Xtxp 265 while the average across all the markers was 0.54. The polymorphic information content (PIC) in the individuals from the 10 populations was much lower. The values ranged from 0.02 in Xtxp136 to 0.81 in marker mSbCIR283, with an average 0.49 across all the markers.

Pattern of genetic diversity and relationships

The total number of alleles in each of the 30 bulked accessions varied from 23 to 37 with an average of 30.17. The highest number of alleles (37) was observed in accessions CSM388 and 02SBF4189 whereas the lowest (23) was observed in accession FaraFara17, an accession originating from Nigeria. The number of alleles in the accessions was varying even among accessions of the same origin. The inter-racial pedigree breeding derivatives with 8 accessions had the most varying number of alleles between 25 - 37 while the least variability of alleles was found among the Guinea population derivatives (West African Guineas) with two accessions (between 28 to 29). The inter-racial pedigree breeding derivative that had the highest number of alleles was Nigeria with 26.85% and while the lowest was Cameroon with 2.98%.

The Nei's unbiased estimate of genetic diversity (H) among the 30 accessions across the 22 loci was relatively low at a mean value of 0.58. The mean heterozygosity for all the 30 accessions across the 22 loci was 0.29 which was lower than the mean Nei's unbiased estimate of gene diversity (H). There were relatively minimal differences between the observed heterozygosity and the Nei's unbiased estimate of gene diversity for all the microsatellite markers, which is characteristic of inbreeding species. The mean Nei's unbiased estimate of gene diversity (H) within countries was variable, ranging from the highest (H = 0.4518) in Mali to the lowest (H = 0.09) in Sudan. The average within country diversity was (H = 0.2931). The average heterozygosity was quite low (below 0.5) in all the countries. The highest level of heterozygosity was 0.3409 observed in Burkina Faso accessions while the lowest was 0.1818 in Cameroon and Sudan accessions. The difference between the average gene diversity and average heterozygosity was relatively small.

There was a high genetic differentiation within the accessions, which was significant at $p < 0.05$. Polymorphism in the ten accessions ranged from 22.73 to 86.36% across the 22 loci with a mean of 45.54%. The average Nei's gene diversity (H) within the 10 accessions was 0.54 across the 22 polymorphic loci while the average heterozygosity was 0.17. The population Nei's genetic diversity (H) was higher than within population genetic diversity, which is expected of selfing species as a result of low levels of inter-population hybridization.

The Roger's modified genetic distance was calculated to determine the relationships among the 30 populations. Pairwise genetic similarities between accessions were assessed based on Dice's genetic similarity co-efficients. The most distant accessions were ICSB38 (21), which is an introduced Caudatum, and IS3534B (25) (GD = 0.82 from and Sudan. The closest accessions were 02SB51 (5) which is a Guinea derivative and 97SB51 (8) from Burkina Faso. The genetic similarities between accessions calculated using pairwise dice's (1945) was the lowest (0.160) between IS1562 and FaraFara17 accessions from Cameroon and Nigeria respectively and the highest (0.833) between FambeB from Mali and AT623 of USA. Cluster analysis performed using the similarity matrix revealed five major groups for all acce-ssions except GPN99271 from Guinea, which appeared to be distant from all other groups (Figure 3). The number of acces-sions for each of the 5 groups varied from 2 to 8 and the patterns of grouping seemed to be irrespective of their geographical origins.

A multidimensional scatter plot (MDS) was used to assess the genetic relationships further and the genetic patterns were confirmed by principal co-ordinate analysis (PCoA) (Figure 4). The first 5 principal components from PCoA explained 57.2% of the total variation. A scatter plot of PC1 (24.8%) and PC2 (12.2%) revealed the same patterns of grouping as the cluster analysis. The first eigen vector explained 27.94 % variation. A scatter plot of the first and second axes of non-metric multi-dimensional scaling (MDS) revealed five clusters of inter-relationships among accessions. The analysis showed that the accessions generally clustered on the basis of the geographical origins. Accessions from Nigeria were distributed in two clusters B and E which also had accession IS 2756 from Burkina Faso while Niger accessions were in their own cluster C. Accession GPN 99 271 from Guinea was placed in its own cluster A. All accessions from Mali were in cluster D and grouped with accessions from Burkina Faso. Accessions in cluster B were widely distributed. The allele frequency based pair-wise genetic distances between countries calculated using powermarker V3.25 (Liu and Muse, 2004) revealed the relatedness of accessions on a country by country basis (Table 4). The countries pair-wise comparisons revealed that accessions from Niger and Sudan were the most distant, whereas accessions from Nigeria and Cameroon were the closest genetically. The accessions from Nigeria and Mali are also quite close. Materials from Nigeria and Guinea, Nigeria and Niger, Sudan and Guinea were equal.

The matrix of pair-wise genetic distances between countries was subjected to sequential agglomerative hierarchical nested (SAHN) using un-weighted pair-group analysis (UPGMA) and resulted in 4 distinct clusters (Figure 5). Niger, Cameroon and Nigeria were closely related and grouped together in cluster.

Population structure

The examination of the hierarchical partitioning of genetic variation by AMOVA in Arlequin version 4.0 demonstrated that genetic differentiation was significant at $P < 0.00$ using the exact value for population differentiation

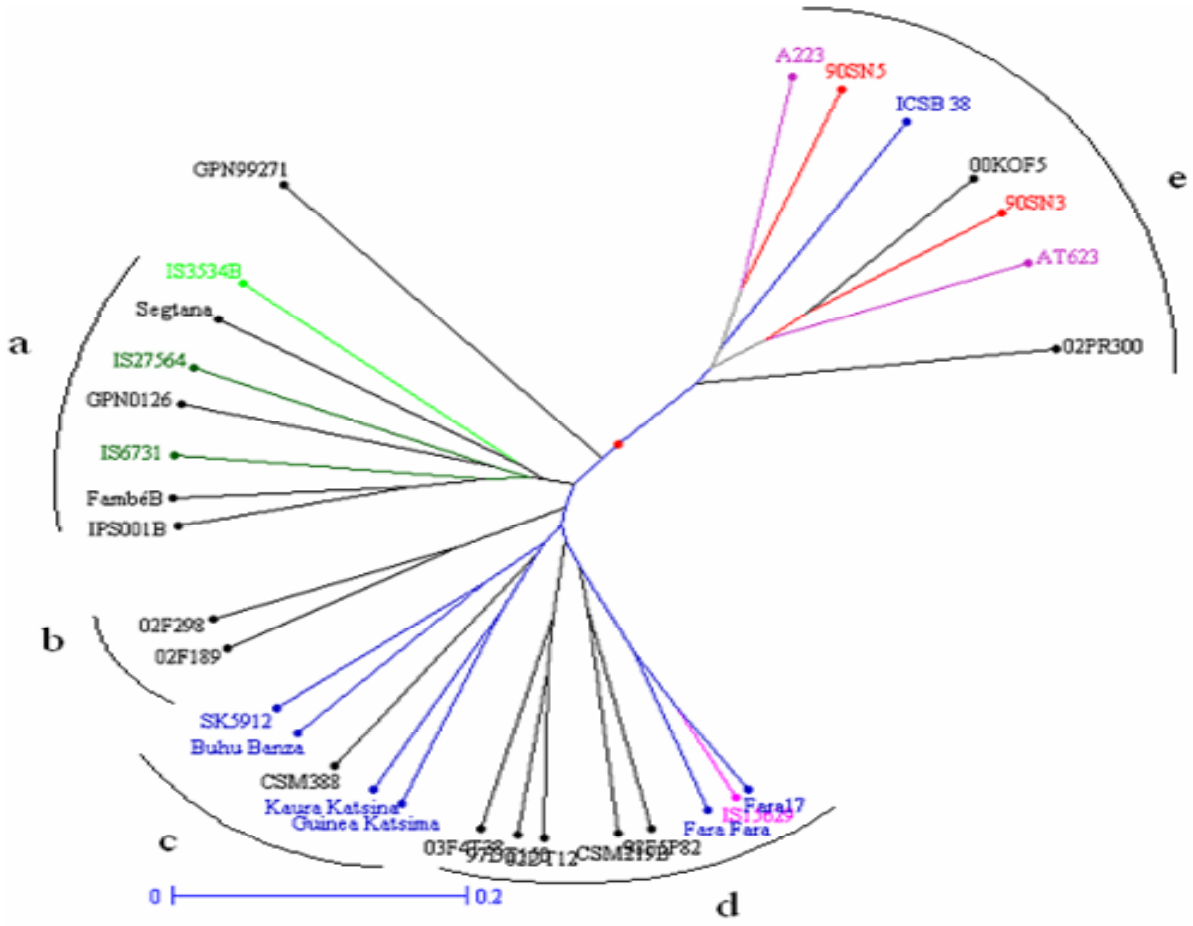

Figure 3. Neighbour joining tree showing relationships between 30 accessions from different origins using the UPGMA method of Darwin Ver. 4.0 (same colour code for all accessions that belong to the same country).

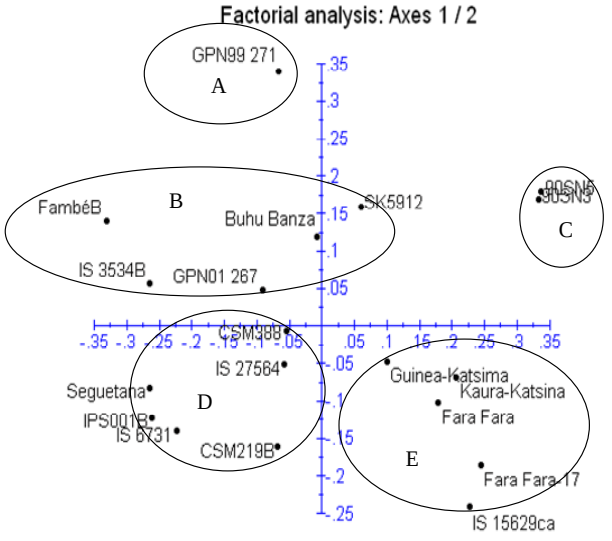

Figure 4. A Multidimensional scatter plot (MDS) showing the relationship of the accessions from West Africa.

(Raymond and Rousset, 1995) (Table 5). There was a clear genetic differentiation both among countries, among populations within countries and within the populations using significance tests based on 1000 permutations calculated according to Weir (1990; 1996). The genetic variation was higher among populations within countries with a variance component of 3.50 than within populations (with a variance component of 2.39). The genetic variation was lowest among countries with variance components of 0.576. Of the total diversity, 8.90% was attributed to country differences, 54.11% was attributed to population differences within the countries while 36.99% was attributed to differences within populations. The Fst index was 0.630, indicating a high level of genetic differentiation among the populations. The value of F, which is equivalent to F_{IT} and estimates the overall (total population) level of inbreeding, was 0.70 indicating a relatively high level of inbreeding while the value of theta (P) was 0.2342 showing an excess of homozygotes.

Table 4. Genetic distance matrices between countries calculated according to Nei and Li (1979) using powermarker ver. 3.25.

	Burkina Faso	Cameroon	Guinea	Mali	Niger	Nigeria	Sudan
Burkina Faso	****	****	****	****	****	****	****
Cameroon	0.52	****	****	****	****	****	****
Guinea	0.51	0.67	****	****	****	****	****
Mali	0.41	0.57	0.45	****	****	****	****
Niger	0.68	0.53	0.58	0.68	****	****	****
Nigeria	0.48	0.39	0.47	0.42	0.47	****	****
Sudan	0.55	0.66	0.47	0.51	0.75	0.61	****

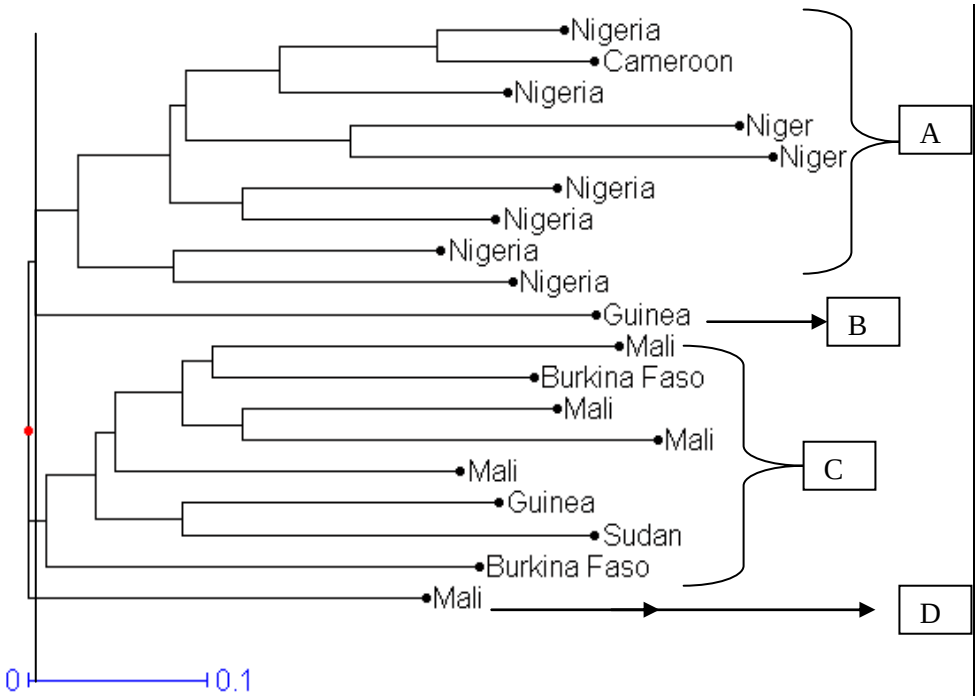

Figure 5. Dendrogram of distribution of diversity between Western Africa countries using binary data (Darwin 4.0).

The Hardy-Weinberg equilibrium assumes a random mating system, which can be assessed by computing the deviation of the observed heterozygosity from Nei`s unbiased estimate of gene diversity (H). The mean chi-square exact p-value test for deviation of the observed heterozygosities from H (Nei's unbiased estimate of gene diversity) using Markov chain method showed that the deviation from Hardy-Weinberg equilibrium was significant (p = 0.05) when the entire dataset was analyzed. The deviation from Hardy-Weinberg equilibrium was highly significant in all the 22 loci (P = 0.0000) except in loci Xtxp278 and Xtxp136. The deviation from the Hardy-Weinberg equilibrium is expected inbreeding species as the mating system systems is not expected to be random.

DISCUSSION

Most of the SSR markers that were used in this study showed high polymorphic information contents (PIC) with very good marker quality indices (Table 3). Brown et al. (1996) as well as Anas and Yoshinda (2004) while working on sorghum observed similarly high levels of polymorphisms. The mean diversity index per each SSR locus was 0.53 in the individuals and 0.58 in the bulks, which allowed us to discriminate each of the sorghum accessions. The SSRs used in this study covered all the sorghum genome (Dean et al., 1999). Thus, these markers should represent the genetic diversity among these sorghum lines. All the 22 SSR loci were poly-

Table 5. AMOVA calculated according to weir, (Weir, 1996).

Source of variation	d.f	Sum of squares	Variance component	% of variation	P-value	Fst
Among groups (country)	3	144.96	0.58	8.9	0.094	
Among accessions within group	6	218.38	3.50	54.11	< 0.001	0.63
Within accessions	88	210.58	2.39	36.99	< 0.001	

morphic in the individuals (except locus mSbCIR 286) and bulks, as seen in previous studies of sorghum germplasm with diverse geographic origins (Grenier et al., 2000b; Casa et al., 2005).

The average number of alleles per locus identified in this study was lower in individuals (4.91) and higher in bulks (6.27) than the average of 5.9 previously reported in elite sorghum lines (Smith et al., 2000), though both were lower than obtained in other studies (Menz et al., 2004; Uptmoor et al., 2003; Dje et al., 2000; Grenier et al., 2000). The SSR loci were able to uniquely identify each of our accessions. Fragment sizes obtained in this study, were usually across a wider range than previously reported in studies of the same loci in other sorghum varieties (Brown et al., 1996; Dean et al., 1999; Dje et al., 2000; Ghebru et al., 2002). The number of observed alleles for most of the loci was higher than those observed by Schlötterer (1998). This suggests that the sorghum accessions studied may be exceptionally polymorphic, providing more size variation within 30 accessions compared to other studies (Brown et al., 1996; Dean et al., 1999; Dje et al., 2000; Grenier et al., 2000; Ghebru et al., 2002). This could also be due to the]inclusion of interracial varieties in the study.

Within accessions, we observed substantial inbreeding (F_{IS} = 0.70), which would be expected as a consequence of self-fertilization at a rate of S = 2 (F_{IS}/ (1+FIS) = 0.71. A similar level of selfing has been reported for sorghum by Ellstrand and Foster (1983). Similar values of inbreeding coefficient (F_{IS} = 0.70) were obtained using both alloenzyme and microsatellite markers in cultivated sorghum sampled *in situ* in North-Western Morroco (Dje et al., 1999). The coefficients obtained were higher than those of Dje et al. (2000). The large F_{IS} values show the large degree of relatedness among the individuals within the sorghum accessions studied. This may be due to high levels of selection for uniformity exercised by farmers in the collection areas (Ghebru et al., 2002; Ghebru, personal interview of local farmers; Araya et al. 1997; Moa, 1999; Dje et al., 2000).

Wright (1978) cited by Kiambi et al. (2005); Semagn et al. (2001) and Hartl (1987), suggested that an Fst range of 0 - 0.05 indicates little differentiation, 0.05 - 0.15 moderate and 0.15 - 0.25 large differentiation and above 0.25 indicates very large differentiation. In this study, the level of population differentiation was Fst = 0.63, which is very large using the suggested parameters. This F_{ST} value was slightly lower than in other sorghum population genetics studies. Dje et al. (2000) reported Fst = 0.68 for landraces on the basis of only three different SSR loci. The western African sorghums show a high level of allelic fixation similar to previous reports by Ghebru et al. (2002) in sorghum landrace populations. Apparently, this indicates an increased level of inbreeding and hence high levels of homozygosity. The Fst values observed in this study were relatively high, indicating an increased degree of allele fixation. New alleles may be generated because of outcrossing and subsequent intralocus recombination including gene conversion. Because local farmers practice a lot of selection, effective population sizes are reduced, therefore increasing the opportunity for fixation of alleles.

In this study, a high inter-population differentiation was found (54.11%). Isoenzyme studies have shown that selfing species have nearly 60% of their genetic diversity distributed among populations while less than 25% of the genetic diversity of mixed mating and outcrossing species is found among populations (Semagn et al. 2001; Hamrick and Godt, 1997).

Sorghum accessions from western Africa tends to have a wide genetic background, the accessions divided into 4 groups A - E. In this study, with a few exceptions, a clear tendency of clustering was observed based on the accessions race and breeding origin as well as their country of origin, similar clustering was also observed by Geleta et al. (2006). Most of the accessions given the same name or similar identification characters by farmers were grouped together, similar results were also obtained by Ghebru et al. (2002) based on the accession names (given by farmers) and again on the basis of their country of origin. Similar results were also obtained by Anas and Yoshinda (2004). In many instances, species exhibit a spatial structure of genetic variation across their eco-geographic range. Different levels of genetic diversity among countries may be due to several factors including mating systems, rate of mutation, migration and dispersal mechanisms, biotic and abiotic selection intensities which are determined by geographic location, climate and soil (Kiambi et al., 2005).

Information about the relationship among breeding materials and the genetic diversity in the available germplasm is important in making choices of parents in

breeding programs. This applies particularly to hybrid breeding where recognition and exploitation of heterotic patterns between different sources of germplasm are important for success. The principal coordinate analysis PCoA based on genetic distance (GD) estimates determined by SSR data for the 49 individuals, provided a distinct separation of lines from different germplasm groups in West African sorghum. The ICRISAT B lines, which are also maintainer lines, were close together though they did not appear in the same cluster while the other hybrid parents from heterosis study also classified as landraces clustered in the same axis but in two groups. Similar results were also observed in maize by (Melchinger et al., 1991; Messmer et al., 1992; Dubreuil et al., 1996) and barley (Melchinger et al., 1994). Likewise, cluster analysis (CA) of 148 US maize inbreeds (Dubreuil et al., 1996) as well as B- and R- lines in sorghum (Ahnert et al., 1996) and sunflower (Hongtrakul et al., 1997) partitioned the lines in accordance with their origin from different breeding groups and pedigree information.

The cluster analysis based on genetic similarity among the 30 bulked sorghum accessions examined showed a clear demarcation of the germplasm mainly according to their races and breeding origin and not to their B- or parental classification. The sorghum races represented in this study included guinea, durra/caudatum, guinea/ caudatum and caudatum while the breeding origins included pedigree breeding lines and their derivatives, landraces and landrace derivatives, maintainer lines, maintainer products of mutation breeding, and maintainer derivative from guinea populations. Each of the clusters E, D and B were clearly differentiated from other groups and all represented the guinea race, landraces (D and E) and a maintainer line from guinea derivative (cluster B). Clusters A and C had the guinea/caudatum and durra/ caudatum races respectively while cluster F contained predominantly the guinea race. Cluster G had representatives from caudatum and guinea/caudatum while cluster H had only guinea/caudatum. All materials in both clusters were of pedigree breeding and their derivatives breeding origin. Cluster I was a mixture of guinea/ caudatum, durra/caudatun and guinea races but the materials were all landraces. Similar results were obtained in sorghum by Menz et al. 2004 where the sorghum lines were grouped according to their working groups.

This study provides a detailed analysis and quantification of genetic diversity in sorghum accessions of West African origin. The data also reaffirm the power of SSR markers to distinctly group closely related accessions. Several authors have indicated that SSR technology is highly cost effective (Smith et al., 2000) and that this technology could easily be employed in resource-poor countries. It could provide efficient and fast screening for both germplasm conservation and crop improvement. The data demonstrates that the West African sorghum accessions contain a great deal of genetic diversity as indicated by the observed number of alleles. The high genetic diversity value among the sorghum accessions (landraces, breeding entries and maintainer lines) indicates that the level of genetic diversity was not influenced by breeding activities. The consistent clustering of most breeding entries/races from Western Africa close to each other in the present study apparently substantiates that the marker system used has a high potential in quantifying the level of similarity and relationships among sorghum germplasm.

The results of this study suggest that elite sorghum germplasm should be grouped by genetic background and not by existing B- or R-line classification. Moreover, the traditional inter-group crossing approach utilized by public breeders may actually dilute potential heterotic patterns. The results reported by Ahnert et al. (1996) also support this conclusion.

REFERENCES

Ahnert D, Lee M, Austin DF, Livini C, Woodman WL (1996). Genetic Diversity among Elite Sorghum Inbred Lines Assessed with DNA Markers and Pedigree Information. Crop Sci. 36(5):1385-1392.

Anas Yoshida T (2004). Genetic diversity among Japanese cultivated sorghum assessed with simple sequence repeats markers. Plant Product. Sci. 7(2):217-223.

Araya D, Tesfamariam S, Sium T (1997). Survey of traditional breeding practices and maintenance of the landraces in Zoba Debub University of Asmara, Eritrea.

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

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.

Buhariwalla HK, Crouch JH (2004). Optimization of marker screening protocol to assess the degree and distribution of genetic diversity in landrecas of pigeonpea. In: Bramel, PJ (2004). Assesing the risk of losses of Biodiversity in Traditinal Cropping Systems: A case study of pigeonpea in Andhara Pradesh. International Crops Institute for the semi-Arid tropics Pantacheru, 502, 324, Andhra Pradesh, India. p. 168

Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, Paterson AH, Aquadro CF, Kresovich S (2005). Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theor. Appl. Genet. 111(1):23-30.

Casa AM, Mitchell SE, Smith OS, Register III JC, Wessler SR, Kresovich S (2002). Evaluation of *Hbr* (MITE) markers for assessment of genetic relationships among maize (*Zea mays* L) Inbred lines. Theor. Appl. Genet. 104:104-110. darwinnmsuedu/~molbio/plants/sorghumhtml

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

Dje` Y, Forcioli D, Ater M, Lefe`bvre C, Vekemans X (1999). Assessing population genetic structure of sorghum landraces from Northwestern Morocco using allozyme and microsatellite markers. Theor. Appl. Genet. 99: 157-163

Dje` Y, Heuertz M, Lefe`bvre C, Vekemans X (2000). Assessment of genetic diversity within and among germplasm accession in cultivated sorghum using microsatellite markers.

De Vries J, Toenniessen G (2001) Securing the harvest: biotechnology, breeding and seed systems for African crops: CABI Publ Wallingford 59-60.

Dogget H (1988) Sorghum, 2nd edition Longman, UK , Burnt Mill, Harlow, Essex, England.

Dubreuil P, Dufour P, Krejci E, Causse M, De Vienne D, Gallais A, Charcosset A (1996). Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. Crop Sci. 36:790-799.

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.

FAO (1995) Production Yearbook Vol. 49, FAO. Roe, Italy

FAOSTAT (2004) http://appsfaoorg/defaulthtm.

Folksterman RT, Fredrick H, Rattunde W, Chandra S, Raju SG, Hash TC (2005). The pattern of genetic diversity of Guinea-race *Sorghum bicolor* (L) Moench landraces as revealed with SSR markers.

Hallauer AR (1999). Temperate maize and heterosis pp. 353-361 *In:* Coors j, Pandey S (ed). Genetics and Exploitation of Heterosis in Crops CIMMYT, Mexico City, 17-22 Aug 1997 ASA, Madison, WI.

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

Hartl DL (1987). A primer of population genetics 2nd Edition Sinauer, Associates Inc Publishers Sunderland, Massachusetts http://cmpgunibech/software/arlequin3.

Hongtrakul V, Huestis GM, Knapp SJ (1997). Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: Genetic diversity among oilseed inbred lines. Theor. Appl. Genet. 95:400-407.

House LR (1995). Sorghum: One of the world's great cereals. Afr. Crop Sci. J. 3: 135-142

Geleta N, Maryke T, Labuschagne Viljoen CD (2006). Genetic diversity analysis in sorghum germplasm as estimated by AFLP, SSR and morpho-agronomical markers. Biodivers. Conserv. 15:3251-3265.

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, Bramel-Cox PJ, Noirot M, Prasada Rao KA, Hamon P (2000a). Assessment of genetic diversity in three subsets constituted from ICRISAT sorghum collection using random vs non-random sampling procedures A using morpho-agronomical and passport data. Theor. Appl. Genet. 101:190-196.

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

Guo SW, Thompson EA (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. 48: 361-372.

Kiambi DK, Newbury HJ, Ford-Lloyd BV, Dawson I (2005). Contrasting genetic diversity among *Oryza longistaminata* (A Chev et Roehr) from different geographic origins using AFLP. Afr. J. Biotechnol. Vol. 4(4):308-317.

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

Liu K, Muse S (2004). POWERMARKER: New genetic data analysis software Version 30 Free program distributed by the author over the internet from http://www Powermarkernet.

Mace ES, Hutokshi k, Buhariwalla Crouch JH (2004). A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol. Biol. Rep. 21:459a-459h.

Melchinger AE, Messmer MM, Lee M, Woodman WL, Lamkey KR (1991). Diversity and relationships among US maize inbreds revealed by restriction fragment length polymorphisms. Crop Sci. 31: 669-678.

Melchinger AE, Andreas G, Mahendra S, Messmer MM (1994). Relationships among European Barley Germplasm: I Genetic Diversity among Winter and Spring Cultivars Revealed by RFLPs. 34(5):1199-1205.

Menz MA, Klein RR, Unhruh NC, Rooney WI, Klein PE, Mullet JE (2004). Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Sci. 44:1236-1244.

Messmer MM, Melchinger AE, Boppenmaier J, Herrmann RG (1992). Relationships among early European maize (Zea maysL) inbred lines: II Comparison of pedigree and RFLP data. Crop Sci. 33:944-950.

Ministry of Agriculture (MOA) (1999). Lowland sorghum in Eritrea: Survey of local landraces of lowland sorghums as described by farmers, MOA, Asmara, Eritrea `1.

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

Nei M, Li W (1979) Mathematical model for studying genetic variation in terms of Restriction endonucleases. Proc. Natl. Acad. Sci. 76(10):5269-5273.

Pejic IP, Ajamone-Marsan M, 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. Thoer. Appl. Genet. 97: 1248-1255.

Raymond M, Rousset F (1995). GENEPOP (ver 12): A population genetics software for exact test and ecumenicism. J. Hered. 86:248-249.

Saghai MA, Yang GP, Zhang Q, Gravois KA (1997). Correlation between molecular marker distance and hybrid performance in US Southern long grain rice. Crop Sci. 37:145-150.

Schloss SJ, Mitchell SE, White GM, Kukatla R, Bowers JE, Paterson AH, Kresovich S (2002). Charecterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* (L) moench. Theor. Appl. Genet. 105:912-920.

Schlötterer C, Ritter R, Harr B, Brem G (1998). High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. Mol. Biol. Evol. 15:1269-1274.

Semagn K, Stedje B, Bjornstad A (2001). Analysis of genetic diversity and structure in Ethiopian populations of *Phytolacca dodecandra* using RAPD. 135:51-60.

Senior ML, Murphy JP, Goodman MM, Stuber CW (1998). Utility of SSRs for determining genetic similarities and relationships in maize using agarose gel system. Crop Sci. 38:1088-1098.

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997). An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): Comparisons with data from RFLPs and pedigree. Theor. Appl. Genet. 95:163-173.

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.

Stuber CW (1994). Hereosis in plant breeding. Plant Breed Rev. 12:227-251.

Taramino G, Tarchini R, Ferrario S, Lee M, Pe` ME (1997). Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor* . 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 Sourthern Africa by RAPDs, AFLPs and SSRs. Thoer. Appl. Genet. 106: 1316-1325.

Weir BS (1990). Genetic data analysis Sinauer, Sunderland, MA.

Weir BS (1996). Genetic data analysis II. Sinauer, Sunderland.

Wright S (1978). Evolution and the genetics of populations Vol. 4 variability within and among natural populationsuniversity of Chicago Press Chicago.