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

M. N. Kitavi1*, D. K. Kiambi1, B. Haussman2, K. Semagn1, G. Muluvi3, M. Kairichi1 and J. Machuka3

1International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), P. O. Box 30709-00100 Nairobi, Kenya.
2ICRISAT-Niamey (Regional hub WCA), BP 12404, Niamey, Niger (Via Paris).
3Department 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 FST 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://apps.fao.org/defaughtm). 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 inbreds 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/ul). 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 MgCl2 (10 mM/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
Table 1. Sorghum accessions selected for the study.

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

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.

<table>
<thead>
<tr>
<th>Marker name</th>
<th>Dye label</th>
<th>Multiple x set</th>
<th>Chromosome</th>
<th>Repeat motif</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>gpsb123</td>
<td>VIC</td>
<td>8</td>
<td>(CA)7+(GA)5</td>
<td>Unpublished*</td>
<td></td>
</tr>
<tr>
<td>mSbCIR223</td>
<td>HEX</td>
<td>Set 2</td>
<td>2</td>
<td>(AC)6</td>
<td>Unpublished</td>
</tr>
<tr>
<td>mSbCIR240</td>
<td>HEX</td>
<td>Set 7</td>
<td>8</td>
<td>(TG)9</td>
<td>Unpublished</td>
</tr>
<tr>
<td>mSbCIR276</td>
<td>NED</td>
<td>Set 7</td>
<td>3</td>
<td>(AC)9</td>
<td>Unpublished</td>
</tr>
<tr>
<td>mSbCIR283</td>
<td>NED</td>
<td>Set 5</td>
<td>10</td>
<td>(CT)8 (GT)8.5</td>
<td>Unpublished</td>
</tr>
<tr>
<td>mSbCIR286</td>
<td>6-FAM</td>
<td>Set 6</td>
<td>1</td>
<td>(AC)9</td>
<td>Unpublished</td>
</tr>
<tr>
<td>SbAGB02</td>
<td>VIC</td>
<td>Set 5</td>
<td>7</td>
<td>(AG)15</td>
<td>Taramino et al. (1997)</td>
</tr>
<tr>
<td>Xcup11</td>
<td>NED</td>
<td>Set 2</td>
<td>3</td>
<td>(GCTA)4</td>
<td>Schloss et al. (2002)</td>
</tr>
<tr>
<td>Xgap072 = Sb4-72</td>
<td>VIC</td>
<td>Set 3</td>
<td>6</td>
<td>(AG)16</td>
<td>Brown et al. (1996)</td>
</tr>
<tr>
<td>Xgap084=Sb6-84</td>
<td>VIC</td>
<td>Set 2</td>
<td>2</td>
<td>(AG)14</td>
<td>Brown et al. (1996)</td>
</tr>
<tr>
<td>Xisep0310</td>
<td>HEX</td>
<td>Set 1</td>
<td>2</td>
<td>(CCAAT)4</td>
<td>unpublished, ICRISAT</td>
</tr>
<tr>
<td>Xtxp012</td>
<td>PET</td>
<td>Set 6</td>
<td>4</td>
<td>(CT)22</td>
<td>Kong et al. (2000)</td>
</tr>
<tr>
<td>Xtxp021</td>
<td>6-FAM</td>
<td>Set 4</td>
<td>4</td>
<td>(AG)18</td>
<td>Kong et al. (2000)</td>
</tr>
<tr>
<td>Xtxp114</td>
<td>6-FAM</td>
<td>Set 3</td>
<td>3</td>
<td>(AGG)8</td>
<td>Bhattiramakki et al. (2000)</td>
</tr>
<tr>
<td>Xtxp136 (Kaf3)</td>
<td>VIC</td>
<td>Set 6</td>
<td>5</td>
<td>(GCA)6</td>
<td>Bhattiramakki et al. (2000)</td>
</tr>
<tr>
<td>Xtxp141</td>
<td>NED</td>
<td>Set 3</td>
<td>10</td>
<td>(GA)23</td>
<td>Bhattiramakki et al. (2000)</td>
</tr>
<tr>
<td>Xtxp265</td>
<td>VIC</td>
<td>Set 4</td>
<td>6</td>
<td>(GAA)19</td>
<td>Bhattiramakki et al. (2000)</td>
</tr>
<tr>
<td>Xtxp273 (Pbbf)</td>
<td>6-FAM</td>
<td>Set 5</td>
<td>8</td>
<td>(TTG)20</td>
<td>Bhattiramakki et al. (2000)</td>
</tr>
<tr>
<td>Xtxp278</td>
<td>NED</td>
<td>Set 1</td>
<td>7</td>
<td>(TTG)12</td>
<td>Bhattiramakki et al. (2000)</td>
</tr>
<tr>
<td>Xtxp286</td>
<td>6-FAM</td>
<td>Set 7</td>
<td>2.00</td>
<td>(GCA)5ACA(GCA)5A(CAA)5 + (AAC)9</td>
<td>Kong et al. (2000)</td>
</tr>
<tr>
<td>Xtxp320 (PhyB)</td>
<td>NED</td>
<td>Set 4</td>
<td>1</td>
<td>(AAG)20</td>
<td>Bhattiramakki et al. (2000)</td>
</tr>
<tr>
<td>Xtxp321</td>
<td>PET</td>
<td>Set 1</td>
<td>8</td>
<td>(GT)4 + (AT)8 + (CT)21</td>
<td>Bhattiramakki et al. (2000)</td>
</tr>
</tbody>
</table>

* 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 electrophoresis 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 heterozygositues 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 codominant 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).

<table>
<thead>
<tr>
<th>Marker name</th>
<th>Repeat length</th>
<th>Quality index</th>
<th>Total alleles</th>
<th>Minimum allele</th>
<th>Maximum allele</th>
<th>Allele (%)</th>
<th>Rare alleles (&lt; 5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISEP310</td>
<td>5</td>
<td>0.2009</td>
<td>2</td>
<td>159</td>
<td>204</td>
<td>204 (93.55)</td>
<td></td>
</tr>
<tr>
<td>Xtxp321</td>
<td>2</td>
<td>0.1776</td>
<td>14</td>
<td>190</td>
<td>222</td>
<td>202 (22.58)</td>
<td></td>
</tr>
<tr>
<td>Xtxp278</td>
<td>3</td>
<td>0.2191</td>
<td>4</td>
<td>249</td>
<td>285</td>
<td>249 (90.62)</td>
<td></td>
</tr>
<tr>
<td>Xcup11</td>
<td>4</td>
<td>0.26</td>
<td>3</td>
<td>154</td>
<td>170</td>
<td>170 (59.52)</td>
<td></td>
</tr>
<tr>
<td>CIR223</td>
<td>2</td>
<td>0.2139</td>
<td>7</td>
<td>108</td>
<td>126</td>
<td>116 (56.82)</td>
<td></td>
</tr>
<tr>
<td>sb6-84</td>
<td>2</td>
<td>0.3142</td>
<td>8</td>
<td>168</td>
<td>196</td>
<td>182 (42.86)</td>
<td></td>
</tr>
<tr>
<td>Sb4-72</td>
<td>2</td>
<td>0.1157</td>
<td>5</td>
<td>186</td>
<td>196</td>
<td>186 (57.89)</td>
<td></td>
</tr>
<tr>
<td>Xtxp114</td>
<td>3</td>
<td>0.0719</td>
<td>2</td>
<td>211</td>
<td>214</td>
<td>214 (70.59)</td>
<td></td>
</tr>
<tr>
<td>Xtxp141</td>
<td>2</td>
<td>0.2042</td>
<td>9</td>
<td>147</td>
<td>191</td>
<td>151 (51.35)</td>
<td></td>
</tr>
<tr>
<td>Xtxp21</td>
<td>2</td>
<td>0.4299</td>
<td>6</td>
<td>176</td>
<td>188</td>
<td>176 (61.29)</td>
<td></td>
</tr>
<tr>
<td>Xtxp265</td>
<td>3</td>
<td>0.4844</td>
<td>15</td>
<td>171</td>
<td>228</td>
<td>189 (16.67)</td>
<td></td>
</tr>
<tr>
<td>Xtxp320</td>
<td>3</td>
<td>0.2985</td>
<td>8</td>
<td>266</td>
<td>290</td>
<td>278 (29.73)</td>
<td></td>
</tr>
<tr>
<td>Xtxp273</td>
<td>3</td>
<td>0.202</td>
<td>11</td>
<td>163</td>
<td>211</td>
<td>193 (33.33)</td>
<td></td>
</tr>
<tr>
<td>CIR283</td>
<td>2</td>
<td>0.4556</td>
<td>10</td>
<td>99</td>
<td>147</td>
<td>117 (22.86)</td>
<td></td>
</tr>
<tr>
<td>sbAG02</td>
<td>2</td>
<td>0.1923</td>
<td>5</td>
<td>96</td>
<td>120</td>
<td>96 (50.00)</td>
<td></td>
</tr>
<tr>
<td>Xtxp12</td>
<td>2</td>
<td>0.4695</td>
<td>9</td>
<td>171</td>
<td>195</td>
<td>175 (27.27)</td>
<td></td>
</tr>
<tr>
<td>Xtxp136</td>
<td>3</td>
<td>0.1702</td>
<td>2</td>
<td>241</td>
<td>244</td>
<td>241 (88.24)</td>
<td></td>
</tr>
<tr>
<td>CIR286</td>
<td>2</td>
<td>0.3483</td>
<td>5</td>
<td>111</td>
<td>149</td>
<td>113 (64.71)</td>
<td></td>
</tr>
<tr>
<td>Xtxp286</td>
<td>3</td>
<td>0.4271</td>
<td>6</td>
<td>110</td>
<td>209</td>
<td>113 (37.93)</td>
<td></td>
</tr>
<tr>
<td>CIR240</td>
<td>2</td>
<td>0.2117</td>
<td>7</td>
<td>112</td>
<td>158</td>
<td>112 (59.52)</td>
<td></td>
</tr>
<tr>
<td>CIR276</td>
<td>2</td>
<td>0.46</td>
<td>4</td>
<td>229</td>
<td>251</td>
<td>229 (39.39)</td>
<td></td>
</tr>
<tr>
<td>gpSb123</td>
<td>2</td>
<td>0.2236</td>
<td>4</td>
<td>288</td>
<td>296</td>
<td>290 (58.97)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>146</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 Xbp 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 between 25 - 37 was observed in the Guinea population 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 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 accessions except GPN99271 from Guinea, which appeared to be distant from all other groups (Figure 3). The number of accessions 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 eigenvector 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-relations 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.

<table>
<thead>
<tr>
<th></th>
<th>Burkina Faso</th>
<th>Cameroon</th>
<th>Guinea</th>
<th>Mali</th>
<th>Niger</th>
<th>Nigeria</th>
<th>Sudan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burkina Faso</td>
<td>****</td>
<td>****</td>
<td>****</td>
<td>****</td>
<td>****</td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td>Cameroon</td>
<td>0.52</td>
<td>****</td>
<td>****</td>
<td>****</td>
<td>****</td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td>Guinea</td>
<td>0.51</td>
<td>0.67</td>
<td>****</td>
<td>****</td>
<td>****</td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td>Mali</td>
<td>0.41</td>
<td>0.57</td>
<td>0.45</td>
<td>****</td>
<td>****</td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td>Niger</td>
<td>0.68</td>
<td>0.53</td>
<td>0.58</td>
<td>0.68</td>
<td>****</td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td>Nigeria</td>
<td>0.48</td>
<td>0.39</td>
<td>0.47</td>
<td>0.42</td>
<td>0.47</td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td>Sudan</td>
<td>0.55</td>
<td>0.66</td>
<td>0.47</td>
<td>0.51</td>
<td>0.75</td>
<td>0.61</td>
<td>****</td>
</tr>
</tbody>
</table>

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-
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; Dje 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 Morocco (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 ecogeographic 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

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f</th>
<th>Sum of squares</th>
<th>Variance component</th>
<th>% of variation</th>
<th>P-value</th>
<th>Fst</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among groups (country)</td>
<td>3</td>
<td>144.96</td>
<td>0.58</td>
<td>8.9</td>
<td>0.094</td>
<td>0.63</td>
</tr>
<tr>
<td>Among accessions within group</td>
<td>6</td>
<td>218.38</td>
<td>3.50</td>
<td>54.11</td>
<td>&lt; 0.001</td>
<td>0.63</td>
</tr>
<tr>
<td>Within accessions</td>
<td>88</td>
<td>210.58</td>
<td>2.39</td>
<td>36.99</td>
<td>&lt; 0.001</td>
<td>0.63</td>
</tr>
</tbody>
</table>
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 (PCA) 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 inbreds (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/caudatum 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


