



ASSESSMENT OF GENETIC DIVERSITY OF SORGHUM [*SORGHUM BICOLOR* (L.) MOENCH] GERMPLASM IN EAST AND CENTRAL AFRICA

^aShadia A. Salih*, ^bLiezel Herselman, ^bMarkey T. Labuschagne and ^cAbdalla H. Mohammed

^aAgricultural Research Corporation (ARC), Sorghum breeding program, Gezira Research Station, Wad Medani, Sudan,

^bUniversity of the Free State, Department of Plant Sciences, South Africa,

^cInternational Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Addis Ababa, Ethiopia.

*Corresponding email address: salihshadia@yahoo.com

ABSTRACT

The study of genetic diversity in crops has a strong impact on plant breeding and maintenance of genetic resources. Comprehensive knowledge of the genetic biodiversity of cultivated and wild sorghum germplasm is an important prerequisite for sustainability of sorghum production. Recurrent droughts resulting from climate change scenarios' in many East and Central Africa countries, where sorghum is a significant arable crop, can potentially lead to genetic erosion and loss of valuable genetic resources. This study aimed at assessing the extent and pattern of genetic diversity and population genetic structure among sorghum accessions from selected countries in East and Central Africa (Sudan, Kenya, Uganda, Ethiopia, Eritrea, Rwanda and Burundi) using 39 microsatellites markers. The studied loci were polymorphic and revealed a total of 941 alleles in 1108 sorghum genotypes. High levels of diversity were revealed with Sudan (68.5) having the highest level of genetic diversity followed by Ethiopia (65.3), whereas Burundi (0.45) and Rwanda (0.33) had the lowest level of genetic diversity. Analysis of molecular variance indicated, all variance components to be highly significant ($p < 0.001$). The bulk of the variation was partitioned within countries (68.1%) compared to among countries (31.9%). Genetic differentiation between countries based on F_{ST} values was high and highly significant ($F_{ST} = 0.32$). Neighbour-joining (NJ) analysis formed two distinct clusters according to geographic regions, namely the central region (Kenya, Burundi, Uganda and Rwanda) and the eastern region (Sudan, Ethiopia, and Eritrea). Population structure analysis revealed six distinct populations corresponding to NJ analysis and geographical origin of accessions. Countries clustered independently with small integration, which indicated the role of farmers' practices in the maintenance of landrace identity and genetic diversity. The observed high level of genetic diversity indicated that germplasm from East Africa should be preserved from genetic erosion, especially in countries with the highest diversity.

Key word: genetic structure, genetic erosion and simple sequence repeat.

INTRODUCTION

Sorghum [*Sorghum bicolor* (L.) Moench, $2n=20$] is fifth in importance among the world's cereals (Arriola, 2005) and a major crop in warm, low-rainfall areas of the world. Sorghum is Africa's second most important cereal based on both areas harvested and annual production. Africa contributes over 60% to the total land area dedicated to the cultivation of sorghum (Burch *et al.*, 2007). Sorghum thus plays an important role as a dietary staple for millions of people, especially in arid and semi-arid countries of Africa and Asia (Bantilan *et al.*, 2001).

Sorghum shows extreme genetic diversity (Sanchez *et al.*, 2002) and is predominantly self-pollinating, with varying levels of outcrossing. The highest level of variability is found in the northeast quadrant of Africa, which includes Ethiopia, Eritrea and Sudan with evidence that it is the principal area of sorghum domestication (Vavilov, 1951; Thomas *et al.*, 1996; Arriola, 2005). Based on morphology, sorghum can be divided into five races (*bicolor*, *caudatum*, *guinea*, *durra* and *kafir*), along with the ten intermediate races resulting from all possible inter-race crosses (Harlan and De Wet, 1972). The high level of genetic diversity and characterization of

accessions integrated into world collections is essential in order to classify, manage exotic germplasm, collect and ultimately utilize the different genetic improvement of the crop. Morphological characterization was the first method used by researchers to select superior sorghum genotypes (Beta and Corke, 2001). However, several studies demonstrated that morphological markers were insufficient due to low level of abundance, low heritability and the influence of the environment (Smith and Smith, 1992; Redfearn *et al.*, 1999; Cadée, 2000). Genetic variation in sorghum has been evaluated in a number of studies using morphological markers (Rao *et al.*, 1996; Djè *et al.*, 1998; Smith and Frederiksen, 2000; Geleta and Labuschagne, 2005; Barnaud *et al.*, 2007). Although this approach has been effective, it is time consuming and based on a few traits only. These limitations lead to the use of molecular markers.

Molecular marker technologies contribute towards studying genetic diversity. It has numerous advantages, including the availability of large numbers of markers that cover the entire genome and their expression are not affected by the environment (Gepts, 1993). Genetic diversity in sorghum has been estimated using several types of molecular markers such as, RFLP (Aldrich and Doebley, 1992; Deu *et al.*, 2006), RAPD (Vierling *et al.*, 1994;

Uptmoor *et al.*, 2003), AFLP (Uptmoor *et al.*, 2003; Menz *et al.*, 2004) and SSR (Brown *et al.*, 1996; Menz *et al.*, 2004; Casa *et al.*, 2005). In each of these studies, the authors studied a precise subset of sorghum germplasm.

Microsatellites or SSR markers are particularly attractive for studying genetic differentiation because they are co-dominant (Akkaya *et al.*, 1992) and abundant in the genome (Lagercrantz *et al.*, 1993). However, there are some limitations, previous genetic information is needed, huge upfront work required, problems associated with PCR (Kubik *et al.*, 2001).

A few studies have been devoted to assessing patterns of sorghum genetic variation at country level for example in Kenya (Nkongolo *et al.*, 2008), Sudan (Assar *et al.*, 2005), Ethiopia (Mekbib *et al.*, 2009), Malawi (Nkongolo *et al.*, 2008) and at regional levels in Western Africa (Ktavii *et al.*, 2014). The strength of the current study lies in the fact that no previous regional studies have been conducted in Eastern and Central Africa, even though it is suggested to be the center of origin of the crop. This information is important in understanding the domestication and evolution of sorghum. Therefore, this study examines the extent of genetic diversity among sorghum collections from Eastern and Central Africa is using SSR markers and to study the genetic population structure diversity distribution within and between countries.

MATERIALS AND METHODS

Study area and plant material: A set of 1108 sorghum collections from Sudan, Ethiopia, Eritrea, Kenya, Uganda, Rwanda and Burundi was selected as a core sample in Table 1.

Country	Number of accessions
Sudan	208
Ethiopia	189
Eritrea	140
Kenya	189
Uganda	118
Rwanda	99
Burundi	165
Total	1108

Table 1. Number of accessions genotyped per country.

Twenty-five seeds were sampled from each selected panicle and planted in plastic trays in the laboratory at room temperature. Fifteen leaves each from each plant then pooled for the DNA extraction.

DNA extraction: DNA was isolated from freshly harvested leaves of two-week old seedlings using the procedure described by Mace *et al.* (2004), after tissue homogenization with a GenoGrinder (Geno/Grinder 2000, SPEX SamplePrep, USA).

SSR amplification: A subset of 39 labeled SSR markers previously described by Menz *et al.* (2002) were used for this study because of their high polymorphic content and broad coverage of the sorghum genome, and genotyped at ICRISAT (Table 2). PCR reactions were performed using a Master cycler (Eppendorf) in a total volume of 10µl, containing 5µl of template DNA, 0.2 U of Amplitaq Gold *Taq* DNA polymerase, 1x PCR buffer (10 mMTris-HCl pH 8.3, 50 mMKCl, 1.5 mM MgSO₄, 1.5 mM MgCl₂), 0.16 mM of each 2'-deoxynucleotide 5'-triphosphate (dNTP), 2µM reverse

primer, 0.04 µM forward primer labeled with either 6 FAM, VIC, PET or NED.

Marker Name	Major allele frequency	No of alleles	Gene diversity	PIC
Gpsb067	0.4551	36	0.7535	0.7362
Gpsb123	0.4031	12	0.7154	0.6718
MSbCIR246	0.4121	11	0.6660	0.6077
MSbCIR262	0.5305	11	0.6578	0.6214
MSbCIR300	0.3805	13	0.7878	0.7648
MSbCIR329	0.5232	8	0.6584	0.6188
XGap206	0.1768	56	0.9423	0.9401
XGap84	0.2295	41	0.9094	0.9042
SbAGB02	0.2606	51	0.8889	0.8816
Xcup02	0.3360	17	0.7990	0.7744
Xcup14	0.3153	12	0.7754	0.7417
Xcup53	0.5363	13	0.6681	0.6413
Xcup61	0.8119	8	0.3084	0.2655
Xcup63	0.6450	9	0.5363	0.4956
XtXp010	0.2039	21	0.8878	0.8781
XtXp015	0.2808	19	0.8376	0.8197
XtXp040	0.4323	16	0.7390	0.7077
XtXp057	0.2034	36	0.9017	0.8945
XtXp145	0.5094	55	0.7303	0.7248
Xisep0310	0.6996	9	0.4790	0.4477
MSbCIR223	0.3212	16	0.8011	0.7754
MSbCIR238	0.1324	45	0.9353	0.9318
MSbCIR240	0.3433	28	0.7204	0.6688
MSbCIR248	0.5453	13	0.6664	0.6439
MSbCIR276	0.3526	8	0.7894	0.7624
MSbCIR283	0.2287	39	0.8632	0.8500
MSbCIR286	0.6075	11	0.5906	0.5603
MSbCIR306	0.3933	9	0.7687	0.7409
XGap72	0.2368	25	0.8592	0.8443
XtXp012	0.1435	45	0.9312	0.9272
XtXp021	0.3526	29	0.8310	0.8184
XtXp114	0.3300	18	0.7458	0.7007
XtXp136	0.5204	15	0.6387	0.5877
XtXp141	0.2293	36	0.9067	0.9010
XtXp265	0.2218	46	0.9145	0.9098
XtXp273	0.4601	23	0.7552	0.7403
XtXp278	0.6975	13	0.4940	0.4748
XtXp320	0.1628	31	0.9086	0.9018
XtXp321	0.1584	37	0.9299	0.9260
Mean	0.3790	24.1282	0.7613	0.7385

Table 2. Parameters of microsatellites used in the study.

PCR conditions described by Folkertsma *et al.* (2005) were used: Denaturation at 94°C for 15 min followed by 10 cycles of 94°C for 15 s, annealing for 20 min using a touchdown strategy where temperatures declined from 61°C to 50°C, and extension at 72°C for 30 s. This was followed by 35 cycles of 94°C for 10 s, 54°C for

20 s, 72°C for 30 s and a final extension step of 20 min at 72°C. The consistency of amplification was established in 3-4 individual primer pairs set by subjecting 3 µl of the PCR products to 2% (w/v) agarose gel electrophoresis at 100 V for 45 min. Then 2 µl of the PCR products was denatured at 94°C for 5 min in 8 µl of Hi-Di formamide with 0.13 µl of GeneScan 500 LIZ internal size standard (Applied Biosystems). The PCR product was subjected to capillary electrophoresis for allele detection using the ABI PRISM 3730 (Applied Biosciences). Allelic data was scored using the software Genemapper 3.7 (Applied Biosystems).

Data analysis

Diversity analyses: Total number of alleles (A^*), number of rare alleles (A^r , alleles with a frequency of <5% in a group), observed heterozygosity (H_o), expected heterozygosity or gene diversity (H_e) and polymorphic information content (PIC) were computed using the software PowerMarker V3.0 (Liu and Muse, 2004) to evaluate levels of genetic diversity. In view of the fact that the observed number of alleles in a sample is highly dependent on the sample size, the software HP RARE (Kalinowski, 2005) was additionally used to calculate allelic richness (A^s) and private allelic richness (A^p , alleles unique to a group), using the rarefaction technique recommended by Petit *et al.* (1998).

Analysis of population structure: In order to assess the structure of genetic diversity within and among countries, five complementary approaches were used: F -statistics, neighbour-joining (NJ) analysis, principle coordinate analysis (PCoA), Analysis of molecular variance (AMOVA), and a Bayesian model-based clustering method. Considering the seven countries for which genotypes were sampled, F_{ST} values (Weir and Cockerham, 1984) were computed within and among countries. The distance method for pairwise differentiation was used to estimate the pairwise genetic differences between countries. Permutation procedures (1000 permutations) were performed to test the significance of differences between values. Calculations were carried out using ARLEQUIN 3.0 (Excoffier *et al.*, 2005). The genetic structure of sorghum accessions was additionally investigated with AMOVA using the same software. The significance of the partitioning of genetic variance among groups was tested.

To investigate the genetic relationships among and between countries, principle coordinate analysis (PCoA) and the NJ cluster analysis algorithm were done in DARwin 5.0.155 (Perrier *et al.*, 2003). Dissimilarities between all pairs of individual genotypes were estimated based on a simple matching procedure.

To assign sorghum accessions to populations based on their genotypes, the data set was subjected to the Bayesian model-based clustering method implemented in the software STRUCTURE 2.2 (Earl, 2012) using the admixture model. The Bayesian based model assumes each individual inherited some portions of its ancestry from one of the K populations. The method of Evanno *et al.* (2005) was also used to determine the true number of K populations in the dataset. All STRUCTURE analyses were performed using the high performance computing resources of the computational

biology service unit (CBSU) from Cornell University (<http://cbsuapps.tc.cornell.edu/structure.aspx>). With the assumed number of populations (K) varying from 1 to 10, 20 replicate runs per K value, with a burn in of 50000 Markov Chain Monte Carlo (MCMC) iterations followed by 10^6 iterations of data collection. Evanno *et al.* (2005) reported that the modal value of the distribution of ΔK is located at the real K . The modal value was illustrated graphically by plotting the ΔK values against successive K values. After the identification of the 'true K ', the run showing the highest $P(X|K)$ value was considered in drawing a bar plot of the software STRUCTURE 2.2 (Earl, 2012).

RESULTS

Polymorphic level of testing microsatellites in sorghum

accessions: The 39 selected SSR loci were polymorphic and revealed a total of 941 alleles in the 1108 sorghum genotypes. The loci, number of alleles, their major allele frequency, gene diversity and PIC values are given in Table 2. The number of alleles per locus ranged from eight (MSbCIR329, and Xcup61, MSbCIR276) to 56 (XGap206) with a mean value of 24.1. PIC varied from 0.27 (Xcup61) to 0.94 (XGap206) with an average of 0.74. The highest level of genetic diversity was found in XGap206 (0.94) and the lowest in Xcup61 (0.31). The highest major allele frequency was obtained in Xcup61 (0.81) while the lowest value was found for MSbCIR238 (0.13).

Extent of genetic diversity in sorghum: Estimates of genetic diversity of the seven sorghum gene pools using various diversity parameters are shown in Table 3. Almost 59% of the detected alleles were determined to be rare (present in less than 5% of the genotypes). In terms of rare alleles, Sudan had the highest percentage (63.6%) whereas Eritrea had the lowest (52.5%). Observed heterozygosity varied from 0.114 (Uganda) to 0.371 (Eritrea) with a mean value of 0.232. Sudan had the highest genetic diversity (0.69) followed by Ethiopia (0.65), whereas Rwanda had the lowest level (0.33). The same trend could be observed for allelic richness, Sudan (8.04) and Rwanda (3.35). Private allelic richness ranged from 0.44 to 1.88 and was the highest in Ethiopia.

Genetic structure of sorghum accessions: The outcome of the partitioning of genetic diversity within and among countries using AMOVA is presented in Table 4. The analysis indicated that all variance components were highly significant and that the bulk of the variation was partitioned within countries (68%) compared to among countries (32%). Genetic differentiation between countries based on F_{ST} was high and highly significant ($F_{ST}=0.32$; $p<0.001$) which supported results obtained using AMOVA. Pairwise differentiation (Table 5) showed the highest level of divergence between Rwanda and Eritrea (0.474) and the lowest between Sudan and Kenya (0.192).

Genetic variation within and between countries: To obtain a graphical demonstration of the relationships between individual sorghum accessions, a PCoA was performed based on the dissimilarity matrix. The two main eigenvalues explained 9.8% and 5.7% of the total variance, respectively.

Country	No of individuals	Total no of alleles	Allelic richness	Rare alleles %	Private allele richness	Gene diversity	Observed heterozygosity
Sudan	208	448	8.04	63.6	1.80	0.685	0.142
Ethiopia	189	362	6.81	56.6	1.88	0.653	0.321
Kenya	189	359	6.57	62.1	0.95	0.569	0.239
Eritrea	140	308	6.06	52.5	1.32	0.561	0.371
Uganda	118	308	6.07	59.4	1.77	0.537	0.114
Burundi	165	291	5.42	60.5	0.44	0.466	0.178
Rwanda	99	163	3.35	56.6	0.53	0.330	0.260
Total	1108	941					
Mean	158	320	6.05	58.8	1.24	0.543	0.232

Table 3. Genetic diversity parameters.

Source of variation	Sum of squares	Variance components	Percentage variation
Among countries	6675.372	3.524	31.95***
Within countries	16577.241	7.504	68.05***
Total	23252.612	11.028	

Table 4. Analysis of molecular variance among and within populations. All sources of variation were significant at p< 0.001.

Country	Burundi	Eritrea	Ethiopia	Kenya	Rwanda	Sudan
Eritrea	0.35087					
Ethiopia	0.37483	0.32956				
Kenya	0.23981	0.34117	0.33435			
Rwanda	0.30309	0.47377	0.45944	0.39771		
Sudan	0.24307	0.24835	0.25050	0.19234	0.34887	
Uganda	0.31395	0.39579	0.33720	0.31389	0.40197	0.28785

Table 5 Estimates of pairwise genetic differentiation between countries.

The PCoA presented in Figure 1 revealed two distinct clusters according to the origin of accessions along axis 1, namely the central region (Kenya, Burundi, Uganda, and Rwanda) to the left of the axis, and the eastern region (Sudan, Ethiopia, and Eritrea) to the right. Genotypes from most of the countries tended to cluster together and seven distinct clusters could be detected.

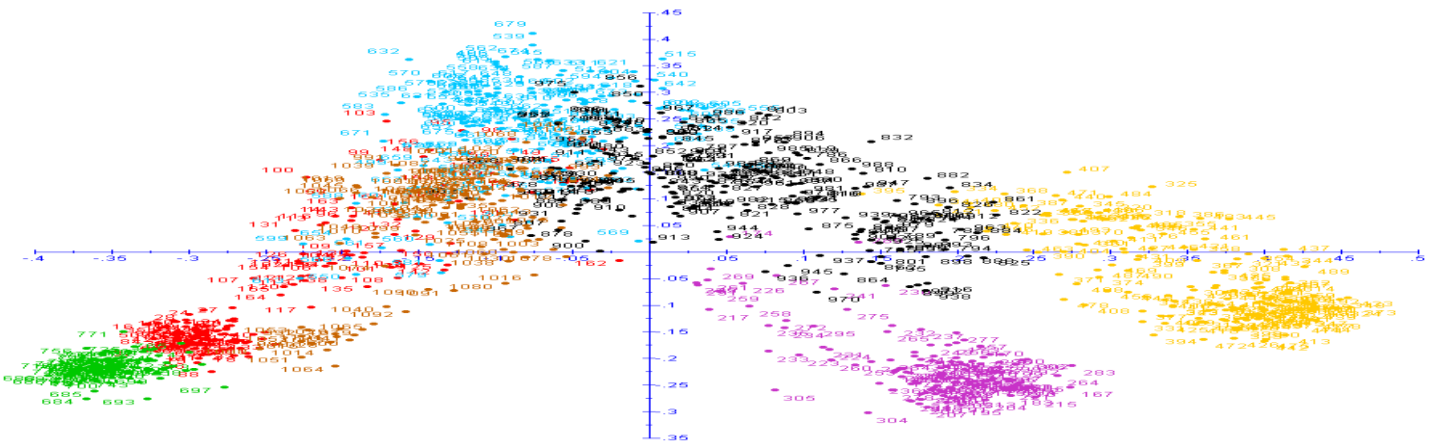


Fig 1. 1Biplot of axis 1 and 2 of the principle coordinate analysis based on the dissimilarity of 39 SSR markers among 1108 sorghum accessions. Eritrea, Ethiopia, Sudan, Kenya, Uganda, Burundi and Rwanda.

All accessions from Rwanda clustered closely together with only a few clustering with accessions from Burundi. Accessions from Eritrea and Ethiopia also formed distinct groups, although some accessions clustered somewhat away from the rest of the accessions. Although accessions from Uganda, Kenya and Sudan formed distinct clusters, some overlapping of accessions from these three countries was observed. Furthermore, although most accessions from Burundi clustered together, many accessions clustered within accessions from Uganda and to some extent accessions from Kenya. The genetic relationships within sorghum accessions determined by NJ analysis are presented in Figure 2. Sorghum

genotypes clustered into six main groups according to their geographical origin. Accessions from Kenya, Ethiopia, Eritrea, and Uganda each formed one main cluster, while two main groups were observed in Sudan and three in Burundi. Accessions from Rwanda also formed one main cluster

butgrouped within accessions from Burundi. A few accessions from Burundi clustered with Ugandan accessions while another group clustered closely together with accessions from Kenya.

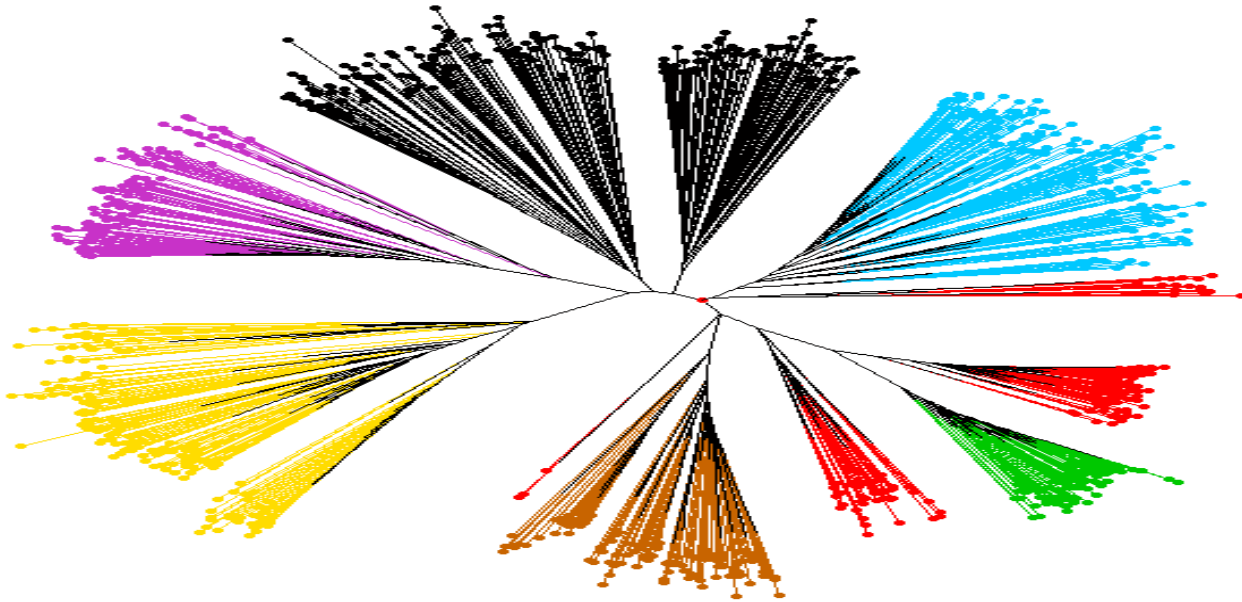


Fig 2. Neighbour-joining cluster analysis showing the genetic relationship among 1108 sorghum accessions using 39 SSR markers based on the simple matching index. Eritrea, Ethiopia, Sudan, Kenya, Uganda, Burundi and Rwanda.

Bayesian model-based cluster analysis: According to the method described by Evanno *et al.* (2005), the initial STRUCTURE analysis identified $K = 6$ to be the most probable number of populations (Figure 3).

The Bayesian model-based cluster analysis at $K=6$ was successful to identify distinct differentiation among sorghum accessions based on country of origin, which confirmed results obtained from the PCoA and NJ analyses. Sorghum genotypes grouped into six populations according to geographical origin (Figure 4). Accessions from Eritrea, Ethiopia, Kenya, Sudan and Uganda each formed part of one specific population (green, blue, pink, yellow, and navy respectively). Accessions from Burundi and Rwanda belonged to the same population. Accessions from Burundi were assigned to two different populations that corresponded to accessions of Rwanda and Kenya.

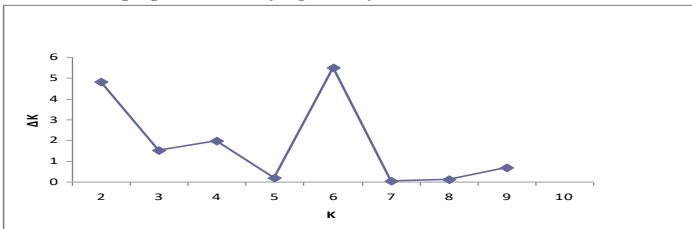


Fig 3. Evanno's ΔK statistic for $K=1$ to $K=10$. The modal value is at $K=6$

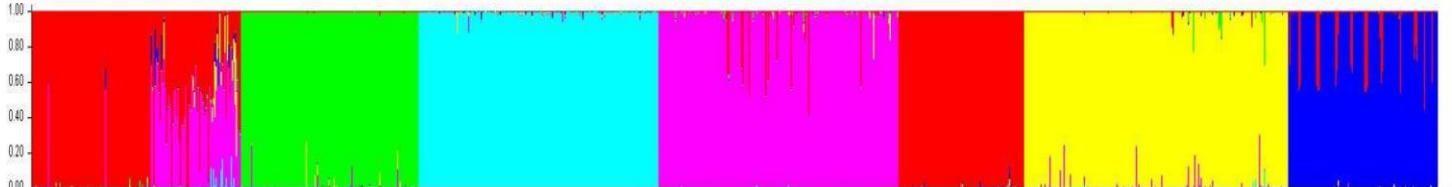


Fig 4. Bar plot of the estimated genetic structure at $K=6$ using the default STRUCTURE parameters with the individuals ordered by country of origin. Each individual is represented by a vertical line which is partitioned into colored segments that represent its proportion of genome in K (colored) clusters.

Discussion

For all SSR loci studied, a high level of genetic polymorphism was observed across country groups, confirming that sorghum microsatellites are an excellent tool to accession diversity. The highest observed frequency of individual alleles per locus was 0.81 which is higher than the value of 0.52 obtained by Ghebru *et al.* (2002). The range of the number of

alleles observed (8-56) was much higher than the range reported for sorghum by Dje *et al.* (2000) of 14-24, 3-9 by Ghebru *et al.* (2002), 4-10 by Anas and Yoshida (2004) and 2-10 by Ali *et al.* (2008). These reported high values might be due to the larger sample size and wider geographic zone. Smith and Frederiksen (2000) demonstrated that the discriminatory power of the PIC of an SSR marker provides an

estimate of that marker by taking into account not only the number of alleles that are detected but also the relative frequencies of those alleles. Mean PIC value of SSR markers used in the current study was 0.74, while Smith and Frederiksen (2000) and Ali *et al.* (2008) working on sorghum reported PICs for SSR markers on 0.645 and 0.400, respectively. The detected high PIC value of 0.74 indicates that the SSR marker system has sufficient resolution and that the selected marker set has a high discriminatory power and could be applied in similar studies on sorghum.

Most SSR markers revealed a high discriminatory power. The allelic diversity might likely be due to the high levels of polymorphism of these markers. These high levels of allelic variability but low levels of heterozygosity observed confirm the previous study using SSR markers to test five Guinea-race accessions by Dje *et al.* (2000) and 100 Guinea-race accessions by Folkertsma *et al.* (2005) and fit the predominantly inbreeding nature of sorghum. The high levels of polymorphism detected by these SSR marker loci permitted the selection of a group of six markers, XGap206, MSbCIR238, Xtxp012, Xtxp141, XGap84 and Xtxp265, whose allele combinations offered exclusive genotyping for all 1108 accessions. Moreover, combinations of two or more of these primers in a single PCR reaction might facilitate fast fingerprinting of sorghum accessions.

The study furthermore revealed a high level of rare alleles (59%), which is in agreement with Folkertsma *et al.* (2005) who detected 50% rare alleles when they studied the pattern of genetic diversity in Guinea-race sorghum and Casa *et al.* (2005) who also detected 64% rare alleles. The high detection rate of rare alleles might indicate the relatively high mutation rate of SSR loci regions. It might also be due to a big sample size (1108 accessions) used and seven countries being sampled. Both PCoA and NJ cluster analysis indicated that accessions from different countries clustered separately, suggesting the uniqueness of genotypes from each country, which might probably explain the high level of private alleles. Relatively low levels of genetic diversity were detected [average gene diversity (GD) per country=0.543; average GD per locus=0.761] compared to values in other studies. For example, Djè *et al.* (1998) reported GD=0.83 for regional scale studies from five regions in Morocco, and Uptmoor *et al.* (2003) reported a GD=0.59 for 23 landraces. Multiple origins for the domesticated sorghum, cross-pollination between races, and outcrossing between genotypes and highly variable wild species are considered to be factors contributing to the extensive genetic diversity observed in the crop (Arriola, 2005).

High levels of genetic variability were detected within countries (68.1%) using AMOVA analysis. This could perhaps be attributed to sorghum landraces being under selection and existence of effective barriers impeding a continuous exchange of genes among accessions. Farmers' practices of selecting particular sorghum varieties depending on use in most cases result in a reduction of effective population sizes which in turn increases chances for fixation of alleles. Previous genetic diversity studies involving microsatellites on cultivated sorghum also revealed a higher genetic diversity

among than within accessions. For instance, Ghebru *et al.* (2002) observed significant genetic variation among 28 Eritrean sorghum accessions for all variance components in which differences among accessions accounted for 50.4% of the variation while within accession diversity accounted for 49.6%. This might suggest that the study underestimated the genetic diversity present within the region. The F_{ST} value observed in the current study was relatively low (0.32), compared to those observed by Dje *et al.* (2000) who reported an F_{ST} of 0.68 in a world collection (25 accessions) on the basis of only three different SSR loci and Ghebru *et al.* (2002) who reported an F_{ST} of 0.50 among 28 Eritrean sorghum. Djè *et al.* (1998) explained low F_{ST} values caused either by the occurrence of frequent gene flow among countries, or by a restriction of the intensity of genetic drift due to effective population sizes. The low F_{ST} value could be explained by the fact that the study took place *insitu* at a large scale (1108 collections), with the potential for high gene flow among accessions that are planted in close proximity in mixed fields. However, the significant differences detected in this study among countries suggested the existence of geographical barriers to gene flow.

Genetic relationships observed among sorghum accessions using distance methods, pairwise differentiation, PCoA, NJ analysis and Bayesian model-based cluster analysis confirmed the differentiation of accessions according to countries. Even though there was a low value of differentiation ($F_{ST}=0.32$) among countries, there were high levels of genetic variation among accessions to separate them (as was observed for F_{ST} pairwise distances).

The cluster analysis in this study showed that sorghum accessions from the same country were clustered together. This pattern of genetic relationships where accessions from the same country were genetically similar could be attributed to existence of variety exchange patterns of such landraces between relatives or friends in the communities. A landrace, which may constitute an accession, is the outcome of a continuous and dynamic development process involving maintenance and adaptation of germplasm to the environment and specific local needs by a community. Farmers often exchange seeds of landraces with other farmers from within or involving outlining localities through which they gain access to new landraces, which were adapted to similar environments.

Clustering and PCoA results revealed similarities of accessions from Rwanda and Burundi, Sudan and Kenya, respectively. This could be caused either by the incidence of recurrent of gene flow between the two countries, seed interactions amongst farmers, or by constraint of the concentration of genetic flow due to an effective population size.

Additionally, the PCoA analysis of the 1108 accessions produced significant groupings that support the earlier studies of East Africa (Sudan, Ethiopia and Eritrea) and Central Africa (Kenya, Uganda, Burundi and Rwanda) collections, done by, Harlan (1976), Stemler *et al.* (1977), Salih (2011), Morris *et al.* (2013), Arriola (2005) and Harlan and De Wet (1972).

NJ clustering and Bayesian model-based cluster analysis grouped accessions according to their country of origin. Several factors could have contributed to the detected patterns. The predominantly autogamous breeding system can contribute towards explaining patterns of genetic diversity and structure observed. Secondly, environmental, biological, cultural, and socio-economic factors all play a role in farmer's decisions to choose or keep a particular sorghum cultivar at any given time. Farmers make decisions on how much of each accession to plant, the percentage of seed or germplasm to save and the percentage to buy or exchange from other sources. Each of these decisions affects the genetic diversity of crop cultivars and is linked to a complex set of environmental and socio-economic influences (Hobbs and Huenneke, 1992).

CONCLUSION

Sorghum is essential for the achievement of food security and for the establishment of sustainable production systems in the semi-arid regions of eastern Africa. This study assessed the pattern and extent of genetic diversity of sorghum accessions originally collected from a large scale in Eastern and Central Africa. High levels of diversity were detected; which provides farmers and plant breeders with options to develop, through selection and breeding, new and more productive varieties that are adapted to changing environments. Although some sorghum accessions could have a low pairwise genetic distance between them, this study showed that considerable genetic differentiation exists between each country used in this study. The evaluation of the genetic variation of accessions from different countries will help confirm or find new differentiation patterns which are important for the development of sorghum breeding programs. A high level of genetic polymorphism across all country groups was observed, which confirmed that the sorghum microsatellite kit is an excellent tool to access diversity of this crop.

REFERENCES

- Akkaya, M. S., A. A. Bhagwat and P. B. Cregan, 1992. Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics*, 132(4): 1131-1139.
- Aldrich, P. and J. Doebley, 1992. Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild sorghum bicolor. *Theoretical and Applied Genetics*, 85(2-3): 293-302.
- Ali, M., J. Rajewski, P. Baenziger, K. Gill, K. M. Eskridge and I. Dweikat, 2008. Assessment of genetic diversity and relationship among a collection of us sweet sorghum germplasm by ssr markers. *Molecular Breeding*, 21(4): 497-509.
- Anas and T. Yoshida, 2004. Genetic diversity among japanese cultivated sorghum assessed with simple sequence repeats markers. *Plant production science*, 7(2): 217-223.
- Arriola, P., 2005. I. Gene flow, hybridization and introgression: Definitions and explanations. *Gene flow and germplasm management*: 1.
- Assar, A., R. Uptmoor, A. Abdelmula, M. Salih, F. Ordon and W. Friedt, 2005. Genetic variation in sorghum germplasm from sudan, icrisat, and USA assessed by simple sequence repeats (ssrs). *Crop science*, 45(4): 1636-1644.
- Bantilan, M. C. S., P. P. Rao and R. Padmaja, 2001. Future of agriculture in the semi-arid tropics: Proceedings of an international symposium on future of agriculture in semi-arid tropics, international crops research institute for the semi-arid tropics, patancheru, india, 14 nov 2000. International Crops Research Institute for the Semi-Arid Tropics.
- Barnaud, A., M. Deu, E. Garine, D. McKey and H. I. Joly, 2007. Local genetic diversity of sorghum in a village in northern cameroon: Structure and dynamics of landraces. *Theoretical and Applied Genetics*, 114(2): 237-248.
- Beta, T. and H. Corke, 2001. Genetic and environmental variation in sorghum starch properties. *Journal of Cereal Science*, 34(3): 261-268.
- Brown, S., M. Hopkins, S. Mitchell, M. Senior, T. Wang, R. Duncan, F. Gonzalez-Candelas and S. Kresovich, 1996. Multiple methods for the identification of polymorphic simple sequence repeats (ssrs) in sorghum [*sorghum bicolor* (L.) moench]. *Theoretical and Applied Genetics*, 93(1-2): 190-198.
- Burch, D., G. Lawrence, G. Green, K. Ichijo, I. Nonaka, M. Pimentel, J. Bower, C. Gilbert, V. Couto Filho and L. Flavio, 2007. World development report 2008: Agriculture for development. The World Bank.
- Cadée, N., 2000. Genetic and environmental effects on morphology and fluctuating asymmetry in nestling barn swallows. *Journal of Evolutionary Biology*, 13(3): 359-370.
- Casa, A., S. Mitchell, M. Hamblin, H. Sun, J. Bowers, A. Paterson, C. Aquadro and S. Kresovich, 2005. Diversity and selection in sorghum: Simultaneous analyses using simple sequence repeats. *Theoretical and Applied Genetics*, 111(1): 23-30.
- Deu, M., F. Rattunde and J. Chantereau, 2006. A global view of genetic diversity in cultivated sorghums using a core collection. *Genome*, 49(2): 168-180.
- Djè, Y., M. Ater, C. Lefèbvre and X. Vekemans, 1998. Patterns of morphological and allozyme variation in sorghum landraces of northwestern morocco. *Genetic Resources and Crop Evolution*, 45(6): 541-548.
- Dje, Y., M. Heuertz, C. Lefebvre and X. Vekemans, 2000. Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. *Theoretical and Applied Genetics*, 100(6): 918-925.
- Earl, D. A., 2012. Structure harvester: A website and program for visualizing structure output and implementing the evanno method. *Conservation genetics resources*, 4(2): 359-361.
- Evanno, G., S. Regnaut and J. Goudet, 2005. Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular ecology*, 14(8): 2611-2620.
- Excoffier, L., G. Laval and S. Schneider, 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary bioinformatics online*, 1: 47.
- Folkertsma, R. T., H. F. W. Rattunde, S. Chandra, G. S. Raju and C. T. Hash, 2005. The pattern of genetic diversity of guinea-race sorghum bicolor (L.) moench landraces as revealed with ssr markers. *Theoretical and Applied Genetics*, 111(3): 399-409.

- Geleta, N. and M. Labuschagne, 2005. Qualitative traits variation in sorghum (*sorghum bicolor* (L.) moench) germplasm from, eastern highlands of ethiopia. *Biodiversity & Conservation*, 14(13): 3055-3064.
- Gepts, P., 1993. The use of molecular and biochemical markers in crop evolution studies. In: *Evolutionary biology*. Springer: pp: 51-94.
- Ghebru, B., R. Schmidt and J. Bennetzen, 2002. Genetic diversity of eritrean sorghum landraces assessed with simple sequence repeat (ssr) markers. *Theoretical and Applied Genetics*, 105(2-3): 229-236.
- Harlan, J. and J. De Wet, 1972. A simplified classification of cultivated sorghum. *Crop science*, 12(2): 172-176.
- Harlan, J. R., 1976. Origins of african plant domestication. Walter de Gruyter.
- Hobbs, R. J. and L. F. Huenneke, 1992. Disturbance, diversity, and invasion: Implications for conservation. *Conservation biology*, 6(3): 324-337.
- Ktavii, M., D. Kiambi, B. Haussman, K. Semagn, G. Muluvi, M. Kairichi and J. Machuka, 2014. Assessment of the genetic diversity and pattern of relationship of west african sorghum accessions using microsatellite markers. *African Journal of Biotechnology*, 13(14).
- Kubik, C., M. Sawkins, W. A. Meyer and B. S. Gaut, 2001. Genetic diversity in seven perennial ryegrass (*L.*) cultivars based on ssr markers. *Crop Science*, 41(5): 1565-1572.
- Lagercrantz, U., H. Ellegren and L. Andersson, 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Research*, 21(5): 1111-1115.
- Mekbib, F., A. Bjoslash, L. Sperling and G. Synnevaring, 2009. Factors shaping on-farm genetic resources of sorghum (*sorghum bicolor* (L.) moench) in the centre of diversity, ethiopia. *International Journal of Biodiversity and Conservation*, 1(2): 045-059.
- Menz, M., R. Klein, J. Mullet, J. Obert, N. Unruh and P. Klein, 2002. A high-density genetic map of sorghum bicolor (L.) moench based on 2926 aflp®, rflp and ssr markers. *Plant molecular biology*, 48(5-6): 483-499.
- Menz, M. A., R. R. Klein, N. C. Unruh, W. L. Rooney, P. E. Klein and J. E. Mullet, 2004. Genetic diversity of public inbreds of sorghum determined by mapped aflp and ssr markers. *Crop Science*, 44(4): 1236-1244.
- Morris, G. P., P. Ramu, S. P. Deshpande, C. T. Hash, T. Shah, H. D. Upadhyaya, O. Riera-Lizarazu, P. J. Brown, C. B. Acharya and S. E. Mitchell, 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences*, 110(2): 453-458.
- Nkongolo, K., K. Chinthu, M. Malusi and Z. Vokhiwa, 2008. Participatory variety selection and characterization of sorghum (*sorghum bicolor* (L.) moench) elite accessions from malawian gene pool using farmer and breeder knowledge. *African Journal of Agricultural Research*, 3(4): 273-283.
- Rao, S. A., K. P. Rao, M. Mengesha and V. G. Reddy, 1996. Morphological diversity in sorghum germplasm from india. *Genetic Resources and Crop Evolution*, 43(6): 559-567.
- Redfearn, D. D., D. R. Buxton and T. E. Devine, 1999. Sorghum intercropping effects on yield, morphology, and quality of forage soybean. *Crop Science*, 39(5): 1380-1384.
- Salih, S. A., 2011. Genetic diversity analysis in sorghum germplasm collections from eastern africa as estimated by morpho-agronomical and ssr markers. University of the Free State.
- Sanchez, A., P. Subudhi, D. Rosenow and H. Nguyen, 2002. Mapping qtls associated with drought resistance in sorghum (*sorghum bicolor* L. Moench). *Plant molecular biology*, 48(5-6): 713-726.
- Smith, C. W. and R. A. Frederiksen, 2000. Sorghum: Origin, history, technology, and production. John Wiley & Sons.
- Smith, J. and O. Smith, 1992. Fingerprinting crop varieties. *Advances in agronomy*, 47: 85-140.
- Stemler, A., J. Harlan and J. De Wet, 1977. The sorghums of ethiopia. *Economic Botany*, 31(4): 446-460.
- Thomas, M. D., I. Sissoko and M. Sacko, 1996. Development of leaf anthracnose and its effect on yield and grain weight of sorghum in west africa. *Plant Disease*, 80(2): 151-153.
- Uptmoor, R., W. Wenzel, W. Friedt, G. Donaldson, K. Ayisi and F. Ordon, 2003. Comparative analysis on the genetic relatedness of sorghum bicolor accessions from southern africa by rapds, aflps and ssrs. *Theoretical and Applied Genetics*, 106(7): 1316-1325.
- Vavilov, N. I., 1951. The origin, variation, immunity and breeding of cultivated plants. *Soil Science*, 72(6): 482.
- Vierling, R., Z. Xiang, C. Joshi, M. Gilbert and H. Nguyen, 1994. Genetic diversity among elite sorghum lines revealed by restriction fragment length polymorphisms and random amplified polymorphic dnas. *Theoretical and Applied Genetics*, 87(7): 816-820.
- Weir, B. S. and C. C. Cockerham, 1984. Estimating f-statistics for the analysis of population structure. *evolution*: 1358-1370.