academicJournals

Vol. 15(23), pp. 1147-1158, 8 June, 2016
DOI: 5897/AJB2015.14604
Article Number: 557D40C58798
ISSN 1684-5315

http://www.academicjournals.org/AJB

African Journal of Biotechnology

Full Length Research Paper

Genetic diversity and population structure among sorghum (*Sorghum bicolor*, L.) germplasm collections from Western Ethiopia

Dilooshi K. Weerasooriya[1], Frank R. Maulana[1], Ananda Y. Bandara[2], Alemu Tirfessa[3], Amsalu Ayana[4], Girma Mengistu[4], Kinde Nouh[3] and Tesfaye T. Tesso[1]*

[1]Department of Agronomy, Kansas State University, Manhattan KS 66506, USA.
[2]Department of Plant Pathology, Kansas State University, Manhattan KS 66506, USA.
[3]Melkassa Research Center, Ethiopian Institute of Agricultural Research, P. O. Box 436, Nazareth Ethiopia.
[4]Oromia Agricultural Research Institute, Addis Ababa, Ethiopia.

Received 29 March, 2015; Accepted 13 May, 2016

The Western Ethiopian region harbors a unique set of sorghum germplasm adapted to conditions not conventional to sorghums grown in other parts of the world. Accessions from the region possess unique resistance to multiple leaf and grain diseases. This study is aimed at exploring the extent of genetic variation and population structure among accessions of this region. A total of 123 accessions comprising 111 from Western Ethiopia (62 from Asosa and 49 from Pawe) and 12 U.S. adapted lines were genotyped using 30 sorghum simple sequence repeat markers (SSR). Genetic diversity and population structure were analyzed using PowerMarker and STRUCTURE software, respectively, based on 23 polymorphic SSR markers. Principal component analysis (PCA) was performed to view the variability in multi-dimensional space. Population structure analysis revealed considerable admixtures between Pawe and Asosa accessions, while the PowerMarker analysis grouped the accessions into three distinct clusters largely based on collection regions. The PCA did not clearly differentiate Asosa and Pawe accessions, but U.S. adapted lines were clearly separated from the rest. The study indicated the presence of marked genetic variability among accessions from Western Ethiopia and also provided clues on shared genetic events among accessions adapted to the two areas in Western Ethiopia.

Key words: Sorghum, genetic diversity, population structure, SSR, Ethiopia.

INTRODUCTION

As a result of disruptive selection, isolation and recombination following the spread from its center of origin, sorghum amassed tremendous genetic variability for numerous traits (Doggett, 1970; Harlan and Stemer, 1976; Frederiksen, 2000; Little et al., 2012). In Ethiopia, part of the vast Northeastern Africa region where the crop

*Corresponding author. E-mail: ttesso@ksu.edu. Tel: +1 785 532 7238.

is believed to have originated (de Wet and Harlan, 1971; de Wet, 1977; Jennings and Cock, 1977), sorghum grows in diverse agroecologies ranging from the hot and dry regions in the eastern and northern lowlands to the mild mid- and high-altitude regions in the central part and to the hot and humid lowlands in western parts of the country. Through time, sorghum may have developed adaptive traits that confer fitness to specific challenges faced in different regions of the world.

The Western Ethiopia region offers a unique environment to sorghum. High rainfall, soils of various degrees of weathering and nutrient content, warm temperature and high humidity are major characteristics of the region (Nageri, 1984; Abebe, 2007). The region extends from Wanbara-Matakkal zone (Pawe) far north of the Blue Nile River to an expansive area south of the Blue Nile which includes Asosa and further south to the Gambella region. All of these areas experience warm temperature, high rainfall and a near 100% humidity during crop seasons, a unique environment to sorghums grown worldwide. Though the region is conducive for growing variety of crops, sorghum is the primary choice for communities living in these areas and is widely grown by diverse groups of people. Over the years, variants of sorghum specifically adapted to these conditions have emerged and become unique sources of germplasm of global interest, such as the Zera-Zera sorghums (Prasada Rao and Mengesha, 1981; Gebeyehu, 1993; Rai et al., 1999).

Efforts to adapt improved sorghum varieties grown in other parts of the country to these areas repeatedly failed due to extreme disease pressure, particularly grain mold and various leaf diseases. However, local sorghums from the region consistently endure the pressure and produce reasonable yields. These diseases are critical in that they seriously undermine grain yield either through reduced grain filling caused by limited photosynthesis in crops damaged by leaf diseases or due to physical damage caused to the grains (grain mold) which together result in massive loss in grain yield as well as grain and stover quality (Thakur et al., 2006; Ibrahim et al., 1985; Somani and Indira, 1999), storability (Hodges et al. 1999) and germinability (Maiti et al., 1985). Both grain mold and leaf diseases become severe when post-flowering temperature exceeds 25°C and relative humidity surpasses 85% (Garud et al., 2000; Navi et al., 2005) which are typical condition in Western Ethiopia.

The Ethiopian national sorghum research program assembled series of accessions from Pawe and Asosa regions of Western Ethiopia to initiate a new breeding program for the region. Despite similar weather characteristics in the two regions, preliminary results revealed remarkable variability among the accessions for various plant characteristics including resistance to grain mold and leaf diseases, grain yield and a range of agronomic, and morphological features. A number of previous studies on sorghums of Eastern Africa region (Ghebru et al., 2002; Mutegi et al., 2011; Ng'uni et al., 2011) including, a recent study on *in situ* diversity and population genetic structure of several Ethiopian accessions (Adugna, 2014) revealed significant genetic variation. However, there was no information whether sorghums evolved under more or less similar agroecology express the level of diversity that is of significant interest. Thus, the objectives of this study were to estimate the extent of genotypic variability among representative accessions from Pawe and Asosa regions of western Ethiopia, and assess the impact, if any, of geographic isolation on the pattern of genetic diversity and population structure among the accessions.

MATERIALS AND METHODS

Genetic

A total of 111 accessions, 49 accessions from Pawe and 62 accessions from Asosa regions in Western Ethiopia (Figure 1) were used in this study. The accessions were collected during the 2004 collection mission. In addition, 12 genotypes from the U.S. sorghum breeding programs were also included. There were no specific criteria for selecting these lines except that seven of them are public parent line (B and R) releases that are among the most widely used tester parents in U.S. sorghum improvement programs. The remaining five represent a range of sorghum types developed by different breeding programs in the U.S. Complete list of the accessions is presented in supplementary Table 1.

The Pawe and Asosa regions are part of a geographic stretch in Western Ethiopia that extends from 9°N latitude in the south to 12°N latitude in the north. The region is subdivided into Pawe in the north and Asosa in the south by the Blue Nile River and the surrounding valley is as wide as 40 to 60 km.

The entire 123 accessions were re-coded following the original order of accession number by using a two letter prefix denoting their geographic regions followed by numeric order from 1 to 123. Accordingly, Pawe accessions were coded as PW1 to PW49, Asosa accessions as AS50 to AS111 in the order they were recorded on collection book. Materials from the U.S. were arbitrarily coded US112 to US123. The accessions from Pawe and Asosa represent most of the major races of sorghum including durra, caudatum, guinea and bicolor. The U.S. public lines are results of complex crosses and may combine pedigrees from two or more of these races.

DNA extraction and SSR genotyping

The accessions were planted in the greenhouse at Melkassa Research Center of the Ethiopian Institute of Agricultural Research, Ethiopia. Ten to fifteen seedlings were raised for each accession using small nursery pots of 15 cm diameter and 10 cm height. At 20 days after emergence, tissue samples from 10 seedlings were pooled into a plastic container and the samples were immediately desiccated using silica gel and then vacuum sealed. The desiccated tissues were then shipped to Kansas State University (KSU). At KSU, the tissues were further lyopholized using a freeze dryer (Thermo Savant®, ModulyoD-115) for 3 days. For the U.S. materials, seeds of the 12 genotypes were planted in a 10 cm wide plastic pots filled with pot mix soil. At about 10 days after planting, the tissue were harvested and immediately lypholysed. About 3 g tissue samples from each of the 111 accession and 12 U.S.

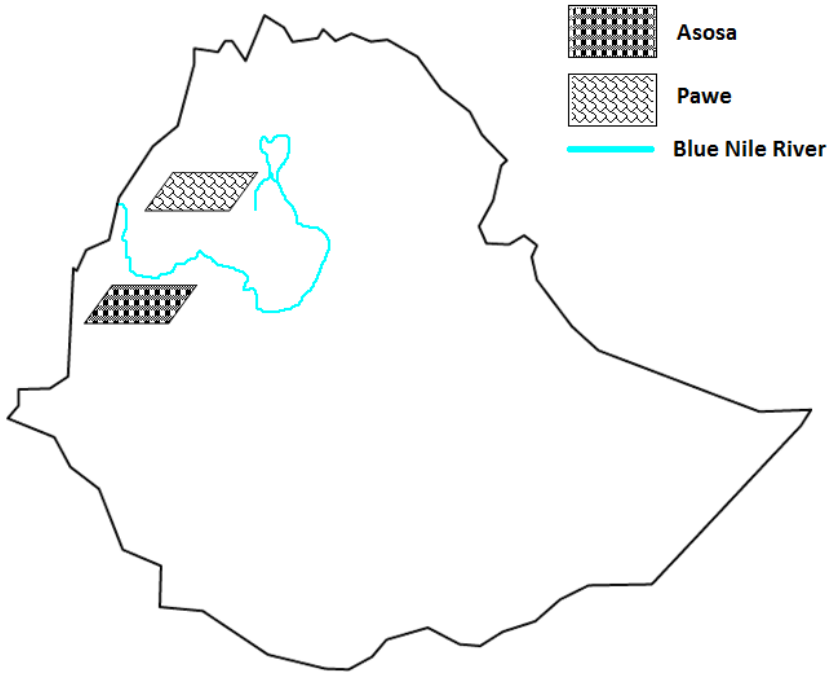

Figure 1. Geographic locations of origin of accessions in Western Ethiopia.

genotypes were pooled into plastic tubes, and the samples ground using tissue grinder, Restch Mixer mill (Retsch® MM 400) for 6 min. The DNA extraction was performed using a modified CTAB protocol (Saghai-Maroof et al., 1984). The DNA pellets were dissolved in 1x TE (10 mM Tris, 1 mM EDTA, pH 8.0) and the final DNA concentration was quantified using a spectrophotometer (NanoDrop®, ND -1000). The working DNA concentration was adjusted to 5 ng/µl for PCR analysis through further dilution with TE buffer.

Thirty simple sequence repeat (SSR) markers were used to genotype the accessions. The markers were selected from published literature based on amplicon sizes, linkage groups, and map positions (Tao et al., 1998; Bhattramakki et al., 2000; Kong et al., 2000). The markers were evenly distributed across all sorghum chromosomes except chromosome 9 (Table 1) and have varying amplicon sizes with the smallest being 145 bp based on Tx623. PCR reactions were performed in MJ Research, PTC-200 PCR thermal cycler with 4 × 96-well plate format. Twenty five nanograms of genomic DNA were used in a 20 µl total volume with a final concentration of 1.5×NH_4 buffer, 0.4 µM each primer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$ and 1 U/µl Taq DNA polymerase (Bioline, Taunton, Massachusetts). PCR amplification was performed using a touchdown protocol that consisted of one denaturation cycle at 94°C for 4 min, 8 subsequent cycles at 94°C for 1 min with annealing temperatures of 63, 61, 59, and 57°C (2 cycles each) and primer extension at 72°C for 1 min, 30 cycles at 94°C for 1 min, 55°C for 1 min and 72°C for 1 min and then a final 7 min extension at 72°C. PCR products were resolved on 3% Metaphore agarose gel using a horizontal electrophoresis apparatus in a TBE buffer loaded with 3% ethidium bromide run at 5 V/cm for 6 h along with a standard DNA ladder (Bioline, Taunton, Massachusetts). A CCD Camera was used to capture the real time gel images under UV illumination projected using Bio-Rad Gel Doc™ XR imaging system. The images were optimized and allele calling was performed using Quantity One software version 4.2.1 (Bio-Rad laboratories, Hercules, CA) using a standard 100 bp molecular weight ladder as a reference. All band sizes were compared and examined for accurate allele calling and non-specific bands were removed. This helped minimizing the ambiguity due to relatively low resolution of agarose gels.

Analyses of population structure and genetic diversity

The data from 23 polymorphic SSR markers were analyzed using specific statistical software to determine population structure and genetic diversity. SRUCTURE software version 2.2.3 (Pritchard et al., 2000a) was used to analyze population structure and group the accessions into different sub-populations. The program performs a model-based quantitative cluster analysis procedure using Bayesian approach which clusters genotypes based on their posterior membership coefficients generated via a maximum likelihood function. Thus, it allows computation of the proportion of the genome of an individual originating from each inferred population (Pritchard et al., 2000b). The number of sub-populations were assumed to be between 2 and 10 (k=2-10), and STRUCTURE was run with 10 iterations for each sub-population (k). The admixture model considering correlated allele frequencies was used with 10,000 replicates for burn-in and 10,000 replicates during analysis. The number of actual sub-populations was decided on the basis of stability of grouping patterns across 10 runs. The optimum k value was chosen to be the k at which Ln (P)D showed the least variation with respect to the number of runs reaching a stable state. Thus, based on the information generated on the simulation summary of the analysis, the optimum sub-population number was determined to be k=3. Out of 10 runs performed for k=3, the results generated for the run with the highest likelihood value was selected to determine the population structure. A composite plot was generated to show the grouping pattern of the whole sample and an expanded bar plot was generated to show the posterior membership coefficients. Individuals were assigned to subpopulations based on sub-population membership coefficient where accessions with

Table 1. Summary of 23 SSR markers used in the study.

Locus	LG[1]	Repeat motifs	Primer sequence (5'to 3')	References[3]	No. of alleles	Major allele freq.	PIC[2]	Allele size (bp)
Xtxp12	4	(CT)22	F: AGA TCT GGC GGC AAC G R:AGT CAC CCA TCG ATC ATC	a	6	0.34	0.72	159-234
Xtxp18	8	(AG)21	F: ACT GTC TAG AAC AAG CTG CG R: TTG CTC TAG CTA GGC ATT TC	a	7	0.39	0.72	233-324
Xtxp32	1	(AG)16	F: AGA AAT TCA CCA TGC TGC AG R: ACC TCA CAG GCC ATG TCG	a	5	0.45	0.55	97-161
Xtxp60	4	(GT)4GC (GT)5	F: GCT AGC TGA CGC ACG TCT CTG R: TGC AAC CGA GCG GTG ACT A	b	4	0.53	0.42	222-252
Xtxp75	1	(TG)10	F: CGA TGC CTC GAA AAA AAA ACG R: CCG ATC AGA GCG TGG CAG G	b	5	0.58	0.55	153-213
Xtxp88	1	(AG)31	F: CGT GAA TCA GCG AGT GTT GG R: TGC GTA ATG TTC CTG CTC	b	4	0.51	0.55	126-170
Xtxp96	2	(GA)24	F: GCT GAT GTC ATG TTC CCT CAC R: CAT TCG TGG ACT CTG TCG G	b	7	0.29	0.71	155-251
Xtxp159	7	(CT)21	F: ACC CAA AGC CCA AAT CAG R: GGG GGA GAA ACG GTG AG	b	5	0.52	0.51	138-208
Xtxp176	6	(AG)4AAC (GA)4	F: TGG CGG ACA TCC TAT T R: GGA GAG CCC GTC ACT T	b	5	0.58	0.45	133-188
Xtxp201	2	(GA)36	F: GCG TTT ATG GAA GCA AAA T R: CTC ATA AGG CAG GAC CAA C	b	7	0.41	0.73	191-284
Xtxp217	10	(GA)23	F: GGC CTC GAC TAC GGA GTT R: TCG GCA TAT TGA TTT GGT TT	b	7	0.63	0.47	136-233
Xtxp218	3	(CA)10	F: CCG GAA AAC CTG CTA CTG R: ACG CCG GAA GGA GAA G	b	5	0.46	0.53	203-254

Table 1. Contd.

Xtxp248	1	(AG)5(GA)28	F: GGG TGT CCA ATG TTG TCT GC R: GGC CGT TAC TGT CCC TTA CTC A	b	6	0.54	0.58	214-295
Xtxp297	2	(AAG)24	F: GAC CCA TAT GTG GTT TAG TCG CAA AG R: GCA CAA TCT TCG CCT AAA TCA ACA AT	b	9	0.46	0.54	183-286
Xtxp303	5	(GT)13	F: AAT GAG GAA AAT ATG AAA CAA GTA CCA A R: AAT AAC AAG CGC AAC TAT ATG AAC AAT AAA	b	4	0.47	0.57	139-192
Xtxp304	2	(TCT)42	F: ACA TAA AAG CCC CTC TTC R: CTT TCA CAC CCT TTA TTC A	b	11	0.6	0.59	141-306
Xtxp312	7	(CAA)26	F: CAG GAA AAT ACG ATC CGT GCC AAG T R: GTG AAC TAT TCG GAA GAA GTT TGG AGG AAA	b	7	0.39	0.62	137-230
Xtxp319	1	(TC)17	F: TAG ACA TCT GAA TTA AGG AGC R: CAT GCC CCT GAA AGA GA	b	6	0.47	0.54	145-239
Xtxp327	4	(TAG)3+ (GA)22	F: ACC ACT GCT CAC GCT CAC R: GCG GTG TAC AGC TTC GTC	b	8	0.74	0.42	132-216
Xtxp340	1	(TAC)15	F: AGA ACT GTG CAT GTA TTC GTC A R: AGA AAC TCC AAT TAT CAT CCA TCA	b	4	0.58	0.46	191-231
Xtxp354	8	(GA)21+ (AAG)3	F: TGG GCA GGG TAT CTA ACT GA R: GCC TTT TTC TGA GCC TTG A	b	5	0.68	0.46	148-196
Xgap1	10	(AG)16	F: TCC TGT TTG ACA AGC GCT TAT A R: AAA CAT CAT ACG AGC TCA TCA ATG	c	6	0.3	0.75	217-300
Xgap42	1	(AG)26	F: TTT TCC TCT TTC AGA TAA CCG TA R: CCC ACC AAG GGC ATC	b	5	0.48	0.57	138-215
Average					6	0.5	0.67	-

LG[1], Linkage group; PIC[2], polymorphic information content. References[3] a= Kong et al. (2000), b= Bhattramakki et al. (2000), c= Tao et al. (1998).

membership coefficient less than 0.8 for all k were considered to result from admixture, hence classified as "admixed".

Further genetic analyses was performed using Power-Marker software version 3.25 (Liu and Muse, 2005). Gene diversity, genetic distance, allelic variability, number of alleles per locus, major allele frequency and polymorphic information (PIC) of markers were determined. The gene diversity analysis provides an unbiased estimation of genetic variation at any given locus. Genetic distance matrix was generated from the marker data set using the method described by Nei (1972) in PowerMarker. Relationship between accessions was then displayed by constructing a Neighbor Joining (NJ) tree using Interactive Tree Of Life v2 (iTOL) (Letunic and Bork, 2011) through

Table 2. Summary statistics for the three major clusters (C1 through C3) detected through neighbor joining tree analysis based on 23 SSR markers.

Population/Marker parameters	C1 (Pawe)	C2 (Asosa)	C3 (USA)	Total
Sub-population size	49	62	12	123
Allele number	130	127	97	144
Number of alleles per locus	5.65	5.52	4.22	6.00
Major allele frequency	0.45	0.44	0.50	0.50
Gene diversity	0.67	0.66	0.63	0.71
Polymorphic information content	0.62	0.61	0.58	0.67
Genetic distance between clusters		C1 and C2		0.69
		C1 and C3		0.72
		C2 and C3		0.70

employing a model-free hierarchical clustering procedure. The results generated through STRUCTURE were used to color-code the neighbor-joining tree. Based on the clustering pattern displayed by the NJ tree, separate genetic analyses within and between clusters was performed using PopGen32 software (Yeh et al., 1997). Principal Component Analysis (PCA) was conducted using XLSTAT (Addinsoft™ version 2012.1, 2009) and the three dimensional plot showing association between accessions was generated.

RESULTS

Genetic variability among accessions

Of the total of 30 SSR markers used to genotype the accessions, 23 markers showed polymorphism and were used in all analyses performed. The remaining 7 markers were not polymorphic and hence were excluded. The 23 markers generated a total of 144 distinct alleles across 123 accessions with an average of 6 alleles per locus. The average major allele frequency, PIC and mean genetic diversity were 0.50, 0.67, and 0.71, respectively (Table 2). The NJ tree analysis grouped the accessions into three broad sub-clusters primarily based on geographic patterns. Majority of the Pawe accessions were separated and pooled into one large cluster (C1), while accessions from Asosa grouped into a different cluster (C2). Two accessions each from Pawe (PW48 and PW49) and Asosa (AS73 and AS97) showed swapping of group members with PW48 and PW49 included in Asosa cluster, and AS73 and AS97 included in Pawe cluster (Figure 2). All of the seven U.S. breeding lines and five adapted genotypes were distinctly sorted and grouped into a separate cluster (C3). However, both Pawe (C1) and Asosa (C2) clusters were further split into smaller sub-clusters. The Asosa cluster was further divided into C2-1, C2-2, and C2-3 sub-clusters consisting of 26, 16, and 19 individuals, respectively, with one of the Pawe accessions PW48 sitting in the middle of C2-1 and C2-2. Similarly, Pawe accessions were apparently split into two sub-clusters C1-1 and C1-2 consisting of 26 and 23 accessions, respectively. Although no geographical coordinates were available for the collections, the pattern of numbering of the accessions which indicates the relative proximity of the accessions within a given geography also provides a clue. Accordingly, the tip branches of C2-1 sub-cluster primarily consisted of accessions AS50 to AS60 along with AS92 and 93 and the next main branch carried AS61-AS67 with few other accessions (AS85&88, AS92&93, PW48&49, AS54, 68, and 69) occurring as smaller branches or scattered in the sub-cluster. The next sub-cluster C2-2 consisted primarily of accessions AS74 to AS87 and C2-3 consisted of accessions AS94 to AS111 with few other accessions outside these number orders occasionally grouping with these sub-clusters. Likewise, the C1 cluster consisted of two sub-clusters. All of the 26 accessions grouped in this sub-cluster (C1-1) are from Pawe. Consecutive accessions PW1 to PW24 were grouped in this sub-cluster along with PW46&47 that are outside this code range. Sub-cluster C1-2 has two main branches with the ordering of accessions on these sub-branches again providing evidence on the importance of geography. Accessions PW25-34 occupied the first branch and the remaining PW35- 45 converged on the second branch except the first branch also consisted of two Asosa accessions (AS73&AS97) that were pulled off C2 (Figure 2).

Within region diversity was relatively high with number of alleles per locus for the Pawe, Asosa, and U.S. clusters being 5.65, 5.52, and 4.22, respectively, but were lower than 6, the record obtained among the entire group of accessions (Table 2). Diversity within region was higher for Pawe and Asosa accessions (0.67 and 0.66) as compared to 0.63 among the U.S. materials while major allele frequency was higher for the U.S. cluster (0.5) and lower for Pawe (0.45) and Asosa (0.44) (Table 2). Genetic distance among clusters was the highest between the U.S. cluster and Pawe (0.72) and the smallest between Asosa and Pawe clusters (0.69). The

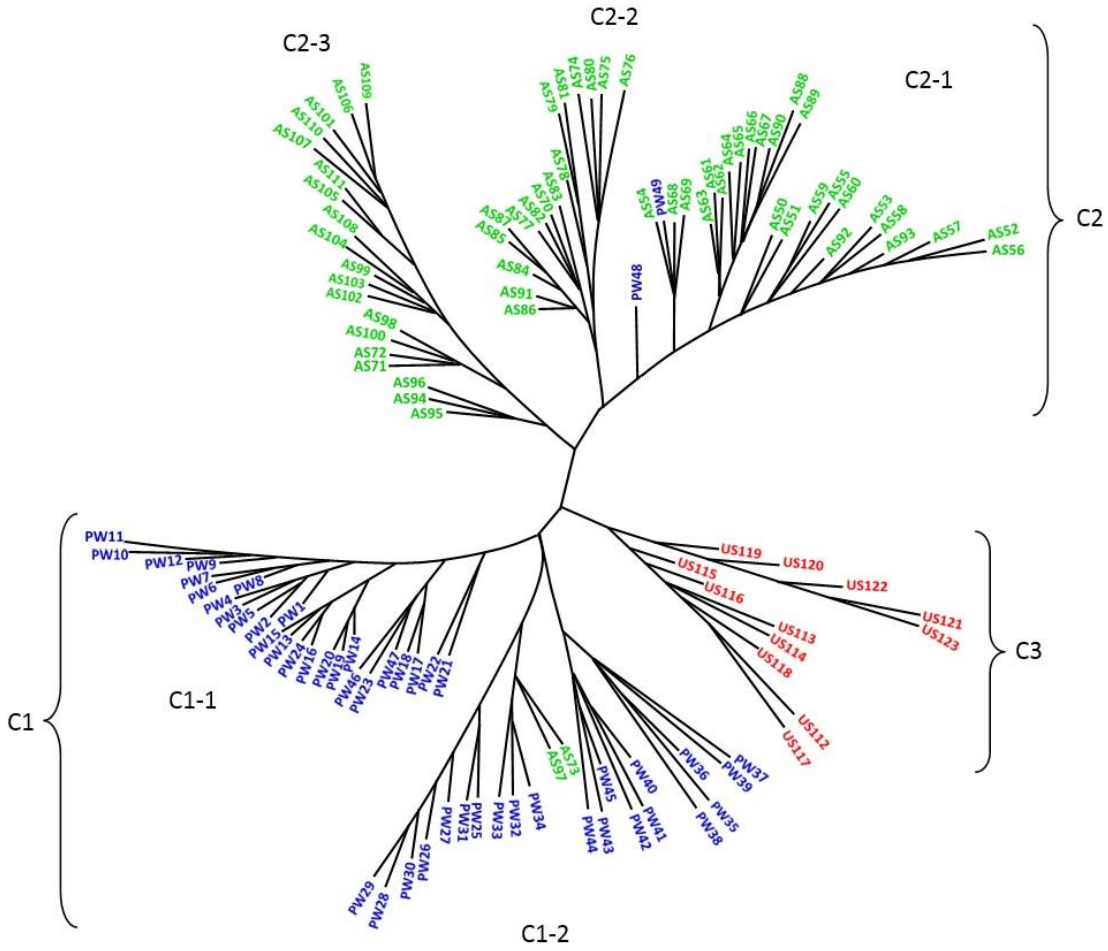

Figure 2. Neighbor-joining tree showing clustering pattern of accessions into three geographies. The accessions are color coded based on collection region, Pawe (Blue), Asosa (Green) and United States (Red).

distance between the U.S. cluster and Asosa was 0.70 (Table 2). The average PIC values for C1, C2, and C3 (0.62, 0.61, and 0.58, respectively) were in agreement with the within cluster diversity.

Population structure and PCA

The population structure analysis sorted the U.S. materials from the Ethiopian accessions (Figure 3). Although the accessions were grouped into three different sub-populations (SP1 through SP3) represented in red, green, and blue, respectively (Figure 3), over 70% of the accessions from Pawe and Asosa were admixed. This differs from the results from NJ tree analysis, except for sub-population 3 (SP3) from the STRUCTURE matched C3 from NJ analysis which included all U.S. materials and one accession from Pawe that was grouped with SP3.

The PCA transformed the genotypic information to reduce the markers to a set of principal components (axis) with the first three principal components (PC1, PC2, and PC3) jointly accounting for 47.78% of the total variation (Figure 4). In agreement with both the STRUCTURE and PowerMarker analysis, all of the twelve breeding lines from United States were adequately separated from the rest of the accessions in the population when viewed in the three dimensional space. The remaining population was aggregated as one large cluster with genotypes from the two regions (Pawe and Asosa) showing an overall pattern of dispersal towards opposite directions, but with large overlaps due to considerable number of admixed individuals. This is consistent with the results of STRUCTURE analysis where accessions from Pawe and Asosa showed marked admixture based on the membership coefficients on the y-axis of Figure 3 which presents the percentage representation of accession in each sub-population.

DISCUSSION

The advent of molecular marker technology has introduced a powerful tool for analysis of genetic variability.

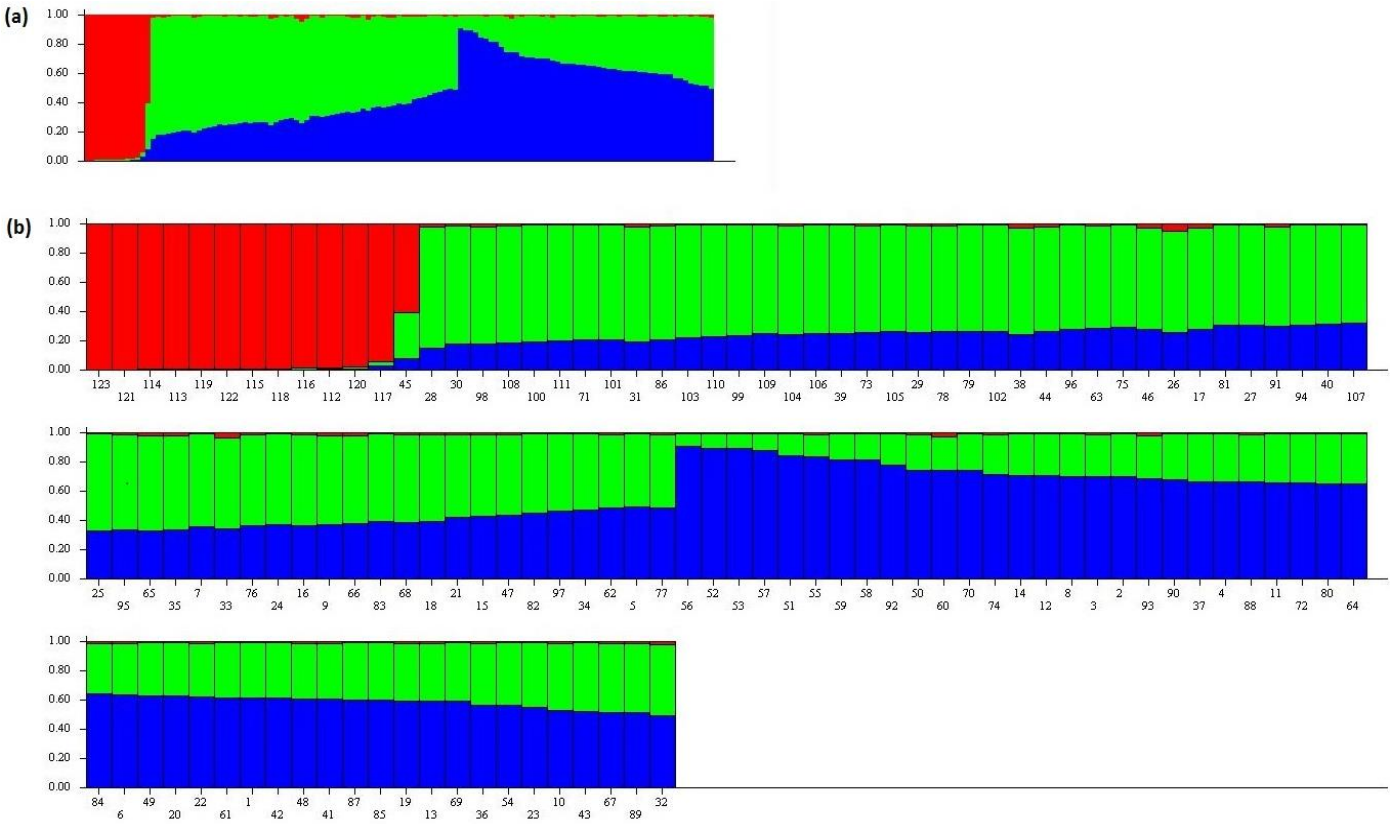

Figure 3. Graphic outputs of the STRUCTURE analysis depicting three sub-populations: (a) Composite display of sub-populations with probability of accessions assigned to a given group shown by the membership coefficients on the Y-axis; (b) Extended display of grouping of accessions into three sub-populations with the X-axis representing the accessions and the Y-axis showing membership coefficients. Red bars represent sub-population group exclusively made up of U.S. materials (SP1); Green (SP2) and blue (SP3) bars are mixture of accessions from Pawe and Asosa showing significant admixture.

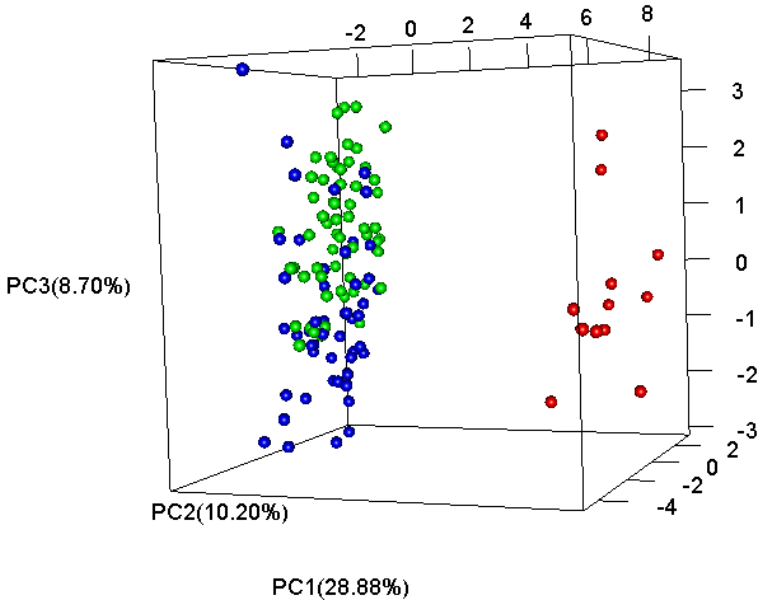

Figure 4. The principal component analysis scatter plot showing aggregation of accessions as color coded by geographical origins, Pawe (Blue), Asosa (Green) and United States (Red).

Variations that are obscured from human eyes can now be resolved, and genotypes can be differentiated based on variations present at genome level. The SSR markers are among the first group of marker systems invented that were based on PCR amplification. Though SNP markers are becoming more popular due to lower cost per data point and larger genome coverage, SSR markers are still widely utilized because of their abundance, co-dominant nature and ease of use as they require less sophisticated equipment and computation methods (Röder et al., 1995; Gupta and Varshney, 2000; Menz et al., 2002; Ellis and Burke, 2007; Varshney et al., 2013). SSR markers also reveal more diversity than most marker systems (Van Inghelandt et al., 2010). In sorghum, numerous SSR markers have been discovered that are linked to functional genes affecting economically important traits such as drought tolerance, *Striga* resistance, and cold tolerance (Crasta et al., 1999; Xu et al., 2000; Knoll and Ejeta, 2008; Burow et al., 2010; Satish et al., 2012) and thus have immediate application in marker based or marker assisted breeding.

In the current study, a total of 30 SSR markers were used to genotype 111 sorghum accessions collected from Western Ethiopia and another 12 U.S. adapted sorghum genotypes. A total of 144 alleles were resolved by 23 polymorphic markers with an average of 6 alleles per locus. Depending on the number of markers used and diversity of the population, previous studies in sorghum have reported 3.2 to 10.4 alleles per locus (Wang et al., 2006; Ali et al., 2008; Wang et al., 2009). In view of this, the allele diversity observed in the current study is intermediate and an average genetic diversity of 0.71 is comparable to many similar studies in sorghum and other species. Given the ecological similarity between the two regions (Asosa and Pawe) where the accessions come from the level of differentiation observed was significant.

Pawe and Asosa are the major sorghum producing areas in the far western part of Ethiopia. The two areas have similar weather patterns (high rainfall, warm temperature, and high humidity) and biotic and abiotic elements. But the areas are geographically disjointed by the 40 to 60 km wide Blue Nile river valley which interrupts sorghum production continuum in the region. This and the diversity in ethno-culture of the people between the two regions, differences in food habits and limited opportunity for interaction due to geographical and infrastructural barriers, the two areas remain fairly separated perhaps limiting movement of germplasm between the two regions. Thus accessions growing in the regions shaped by both natural and man-made circumstances peculiar to the specific regions may have resulted in some degree of genetic differentiation between accessions from the two regions. The present study seems to have shed light on these differences. The neighbor joining tree developed based on allelic variation among the accessions clearly sorted the population into three major clusters (Figure 2) with the U.S. adapted materials (C3) distinctly grouped in a separate cluster and accessions from Western Ethiopia sorted into two major clusters, C1 and C2, largely according to their geographic region with C2 consisting of mainly Asosa collections and C1 harboring that of Pawe. The two sub-groups seem to have further split into smaller sub-units perhaps based on within region geographic proximity or differences in ethnography between people in the two regions.

Population structure analysis run on arbitrary sub-population groupings of k=2 to10 also identified three sub-populations (k=3) as the optimal number of groups. However, sub-populations representing Pawe and Asosa accessions show significant admixture (green and blue bars in Figure 3), while U.S. adapted materials were clearly sorted as shown by red bar in Figure 3. It appears that despite presumed isolation, environmental factors such as rainfall, temperature, and biotic stress factors common to the two regions have played significant role in shaping natural selection, particularly toward well-adapted traits as is evident from significant admixture of alleles between the two regions. At the same time, selection for different plant attributes driven by differences in culture and food habits exercised by communities in these regions and limited movement of people between the two regions may have influenced the population structure to a certain degree. This was particularly evident from the NJ tree analysis where Pawe and Asosa accessions were clearly sorted into different groups with each having their own sub-groups. Nevertheless, accessions from the two regions were closely related to each other. This was evident from the high degree of admixture between accessions from the two regions (Figure 3a and b). It is also clear from genetic distances estimated among the clusters where both C1 (Pawe) and C2 (Asosa) clusters were found to be more distant from C3 (U.S. materials) than from each other. This shows that despite the presumed disruptive selection, accessions from the two regions tend to resemble each other more than they do to the U.S. materials, perhaps due to a common natural selection pressure. This implies that accessions from Pawe and Asosa may harbor shared alleles (admixed) that are perhaps responsible for adaptation to factors common to both environments including resistance to grain mold and leaf diseases.

Past efforts by the international sorghum research community working to improve sorghum production in Africa have successfully generated elite varieties and germplasm into which numerous desirable traits have been integrated. Many of the materials were proven useful in several countries in Africa while many others either completely failed to adapt to certain environments such as Western Ethiopia or not accepted by communities, because they do not fit to local production conditions or are not suitable for local food processing. The presence of significant genetic variability among collections representing such isolated geography such as

those in the current study, however, has significant value in that it can be effectively exploited to develop new varieties that are adapted to such specific environments or possess specific attributes needed by local communities. But these materials can also be used as sources of new alleles and can be equally useful to the global sorghum improvement efforts. Though this study did not include materials from other regions of the country, repeated failure of varieties and landraces from outside this region and the clear segregation of the accessions from the U.S. adapted genotypes observed in this study provide hints about the peculiarity of these materials. Moreover, although not widely represented in the global sorghum germplasm, materials from these regions have been long recognized as useful sources of disease resistance and grain quality such as the Zera-Zera sorghums of Gambella (Prasada Rao and Mengesha, 1981; Reddy et al., 2000; Thakur et al., 2006).

Although, there is no racial classification data for the accessions, field observations indicate that they are of multiple racial groups. However, unlike other parts of the country that are dominated by the durra race, sorghums of this region appear to be largely made up of caudatum and caudatum-guinea hybrids and their derivatives (Prasada Rao and Mengesha, 1981). But the grouping of the populations into clusters and sub-clusters seem to be largely influenced by geographic origin, while earlier studies have implicated racial ancestry as the major factor influencing the pattern of genetic diversity (Tao et al., 1993; Cui et al., 1995; de Oliveira et al., 1996). The lack of racial differentiation in this study may be due to absence of association between genes controlling panicle morphology that are often used to classify genotypes into racial groups (Snowden, 1936) and the markers used in this study. Nevertheless, several other studies conducted on sample sets representing global sorghum germplasm also grouped genotypes according to geographic origin or breeding history rather than race (Tesso et al., 2005; Billot et al., 2013; Lekgari and Dweikat, 2014). Moreover, studies conducted in other species also agree with our results. Genetic structure among common self-pollinated field crops including wheat (Zhang et al., 2009; Vanzetti et al., 2013), barley (Naeem et al., 2011), and horticultural crops such as grape (Bacilieri et al., 2013) as well as pathogen species of economically important crops such as oil-seed brassica (Strehlow et al., 2014) have been shown to be strongly influenced by geographic adaptation.

Owing to the extreme variation in ecological habitats under which the crop is cultivated, sorghum genetic diversity in Ethiopia appears to be compartmentalized according to regional climatic variables. The Western Ethiopia region harbors sorghum types that are resistant to leaf and panicle diseases and produce high grain quality crop despite the warm weather and very high humidity. But in contrast to the common view that materials from the region were derived from narrow genetic base, at least based on the polymorphic loci examined in this study, the accessions represent broad genetic pool similar to those collected from other parts of the country. In view of the failed attempts to adapt varieties developed elsewhere to the Western Ethiopia region, the diversity in the local collection observed in this study can serve as primary resource for developing new varieties that suit the region or as sources of new alleles for global sorghum breeding programs. Nevertheless, the recent introduction of large multi-national agricultural companies into the region is risking displacement of the traditional crops including sorghum by export crops such as rice which eventually will pose threat to this unique genetic resource. Urgent action is needed to collect and protect the materials from certain loss.

Conflict of Interests

The authors have not declared any conflict of interests

ACKNOWLEDGEMENTS

The authors are thankful to the Global Crop Diversity Trust for financial support and the Ethiopian Institute of Agricultural Research, Oromia Agricultural Research Institute, and Kansas State University for providing field and laboratory facilities used for conducting the study. This work is contributed by Kansas Agricultural Experiment Station, contribution No. 2015-188-J.

REFERENCES

Abebe M (2007). Nature and Management of Acid Soils in Ethiopia. EIAR, Addis Ababa, Ethiopia.

Addinsoft (2009). XLSTAT statistical analysis software, 2012. IOP Publishing Physics Web www.xlstat.com (Accessed 26 December 2013).

Adugna A (2014). Analysis of in situ diversity and population structure in Ethiopian cultivated *Sorghum bicolor* (L.) landraces using phenotypic traits and SSR markers. Springerplus 3:212.

Ali ML, Rajewski JF, Baenziger PS, Gill KS, Eskridge KM, Dweikat I (2008). Assessment of genetic diversity and relationship among a collection of U.S. sweet sorghum germplasm. Mol. Breed. 21:497-509.

Bacilieri R, Lacombe T, Le Cunff L, Di Vecchi-Staraz M, Laucou V, Genna B, Péros JP, This P, Boursiquot JM (2013). Genetic structure in cultivated grape vines is linked to geography and human selection. BMC Plant Biol. 13(1):1.

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

Billot C, Ramu P, Bouchet S, Chantereau J, Deu M, Gardes L, Noyer JL, Rami JF, Rivallan R, Li Y, Lu P, Wang T, Folkertsma RT, Arnaud E, Upadhyaya HD, Glaszmann JC, Hash CT (2013). Massive Sorghum Collection Genotyped with SSR Markers to Enhance Use of Global Genetic Resources. Plos One 8:1-16.

Burow G, Burke JJ, Xin Z, Franks C (2010). Genetic dissection of early-season cold tolerance in sorghum (*Sorghum bicolor* (L.) Moench). Mol. Breed. 28:391-402.

Crasta OR, Xu WW, Rosenow DT, Mullet J, Nguyen HT (1999). Mapping of post-flowering droughtresistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. Mol. Genet Genomics 262(3):579-588.
Cui YX, Xu GW, Magill CW, Schertz KF, Hart GE (1995). RFLP-based assay of *Sorghum bicolor* (L.) Moench genetic diversity. Theor. Appl. Genet. 90:787-796.
de Oliveira AC, Richter T, Bennetzen JL (1996). Regional and racial specifications in sorghum germplasm assessed with DNA markers. Genome 39:579-587.
de Wet (1977). Domestication of African Cereals. Afr. Econ. Hist. 3:15-32.
de Wet JMJ, Harlan JR (1971). The origins and domestication of Sorghum bicolor. Econ. Bot. 25:128-135.
Doggett H (1970). Sorghum (Tropical Agriculture Series). Longmans, London.
Ellis JR, Burke JM (2007). EST-SSRs as a resource for population genetic analyses. Heredity 99:125-132.
Frederiksen RA (2000). Diseases and disease management in sorghum. In: Smith CW, Frrederiksen RA (eds.) Sorghum: Origin, History, Technology and Production. New York: John Wiley and Sons. Pp.497-533.
Garud TB, Ismail S, Shinde BM (2000). Effect of two mold-causing fungi on germination of sorghum seed. International Sorghum and Millets Newsletter, pp. 41-54.
Gebeyehu G (1993). Characterization and Evaluation of Sorghum [*Sorghum bicolor* Moench (L.)] Germplasms Collected from Gambella. Alemaya University of Agriculture, Ethiopia.
Ghebru B, Schmidt J, Bennetzen L (2002). Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. Theor. Appl. Genet. 105(2-3):229-236.
Gupta PK, Varshney RK (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113(3):163-185.
Harlan JR, Stemler ABL (1976). The races of sorghum in Africa. In. Harlan JR, de Wet JMJ and Stemer ABL (eds.) Origins of African Plant Domestication, Mouton, The Hague, The Netherlands, pp. 465-478.
Hodges RJ, Hall AJ, Jayaraj K, Jaiswal P, Potdar N, Yoganand B, Navi SS (1999). Quality changes in farm-stored sorghum grain grown in the wet or dry season in South India – A technical and social study. NRI Report 2412. Chatham Avenue, Kent ME4 4TB, UK: Natural Resources Institute. 33p.
Ibrahim OE, Nyquist WE, Axtell JD (1985). Quantitative inheritance and correlations of agronomic and grain quality traits of sorghum. Crop Sci. 25:649-654.
Jennings P, Cock J (1977). Centres of origin of crops and their productivity. Econ. Bot. 31:51-54.
Knoll J, Ejeta G (2008). Marker-assisted selection for early-season cold tolerance in sorghum: QTL validation across populations and environments. Theor. Appl. Genet.116:541-553.
Kong L, Dong J, Hart GE (2000). Characteristics, linkage-map positions, and allelic differentiation of *Sorghum bicolor* (L.) Moench DNA simple-sequence repeats (SSRs). Theor. Appl. Genet.101:438-448.
Lekgari A, Dweikat I (2014). Assessment of Genetic Variability of 142 Sweet Sorghum Germplasm of Diverse Origin with Molecular and Morphological Markers. Open J. Ecol. 4:371-393.
Letunic I, Bork P (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 39 (Web Server issue):W475-8.
Little CR., Perumal R, Tesso T, Prom LK, Odvody GN, Magill CW (2012). Sorghum Pathology and Biotechnology - A Fungal Disease Perspective: Part I. Grain Mold, Head Smut, and Ergot. Eur. J. Plant Sci. Biotechnol. 6(Special Issue 1):10-30.
Liu K, Muse SV (2005). PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129.
Maiti RK, Raju PS, Bidinger FR (1985). Studies on germinability and some aspects of preharvest physiology of sorghum grain. Seed Sci. Technol. 13:27-35.
Menz MA, Klein RR, Mullet JE, Obert JA, Unruh NC,Klein PE (2002). A high-density genetic map of *Sorghum bicolor* (L) Moench based on 2926 AFLP, RFLP and SSR markers. Plant Mol. Biol. 48:483-499.
Mutegi E, Sagnard F, Semagn K, Deu M, Muraya M, Kanyenji B, de Villiers S, Kiambi D, Herselman L, Labuschagne M (2011). Genetic structure and relationships within and between cultivated and wild sorghum (*Sorghum bicolor* (L.) Moench) in Kenya as revealed by microsatellite markers. Theor. Appl. Genet. 122(5):989-1004.
Naeem M, Khan MMA, Moinuddin, Idrees M, Aftab T (2011). Triacontanol-mediated regulation of growth and other physiological attributes, active constituents and yield of *Mentha arvensis* L. Plant Growth Regul. 65(1):195-206.
Nageri A (1984). The effect of lime and phosphorus application on the dry matter and grain yield of bread wheat (*Triticum aestivum*) grown on red soils of Ghimbi and Nedjo Woreda of Wollega Region, Western Ethiopia. M.Sc. Thesis. School of Graduate Studies, Alamaya College of Agriculture, Addis Ababa University, Addis Ababa, Ethiopia. 87 p.
Navi SS, Bandyopadhyay R, Reddy RK, Thakur RP, Yang XB (2005). Effects of wetness duration and grain development stages on sorghum grain mold infection. Plant Dis. 89:872-878.
Nei M (1972). Genetic distance between populations. Am. Nat. 106:283-292.
Ng'uni D, Geleta M, Bryngelsson T (2011). Genetic diversity in sorghum (*Sorghum bicolor* (L.) Moench) accessions of Zambia as revealed by simple sequence repeats (SSR). Hereditas 148(2):52-62.
Prasada Rao KE, Mengesha MH (1981). A pointed collection of Zera-Zera sorghum in the Gambella area of Ethiopia, Genetic Resources Progress Report no. 33. ICRISAT, Patancheru, India.
Pritchard JK, Stephens M, Donnelly P (2000a). Inference of population structure using multilocus genotype data. Genetics 155:945-959.
Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b). Association mapping in structured populations. Am. J. Hum. Genet. 67:170-181.
Rai KN, Murty DS, Andrews DJ, Bramel-Cox PJ (1999). Genetic enhancement of pearl millet and sorghum for the semi-arid tropics of Asia and Africa. Genome 42:617-628.
Reddy BVS, Bandyopadhyay R, Ramaiah B, Ortiz R (2000). Breeding grain mold resistant sorghum cultivars. In. Technical and institutional options for sorghum grain mold management: proceedings of an international consultation, 18-19 May 2000, ICRISAT, Patancheru, India (Chandrashekar A, Bandyopadhyay R, Hall AJ, eds.). Patancheru, India, ICRISAT. Pp. 195-224.
Röder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD, Ganal MW (1995). Abundance, variability and chromosomal location of microsatellites in wheat. Mol. Genet. 246:327-333.
Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW(1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. PNAS (USA). 81:8014-8018.
Satish K, Gutema Z, Grenier C, Rich PJ, Ejeta G (2012).Molecular tagging and validation of microsatellite markers linked to the low germination stimulant gene (lgs) for Striga resistance in sorghum [*Sorghum bicolor* (L.) Moench]. Theor. Appl. Genet. 124(6):989-1003.
Snowden JD (1936). The cultivated races of sorghum. Adlard and Sons, London, UK.
Somani RB, Indira I (1999). Effect of grain molds on grain weight in sorghum. J. Mycol. Plant Pathol. 29:22-24.
Strehlow B, de Mol F, Struck C (2014). History of oilseed rape cropping and geographic origin influence the genetic structure of *Plasmodiophora brassicae* populations. Phytopathology 104(5):532-538.
Tao Y, Manners JM, Ludlow M, Henzel RJ (1993). DNA polymorphisms in grain sorghum (*Sorghum bicolor* (L.) Moench). Theor. Appl. Genet. 86:679-688.
Tao YZ, Jordan DR, Henzell RG, McIntyre CL (1998). Construction of a genetic map in a sorghum RIL population using probes from different sources and its comparison with other sorghum maps. Aust. J. Agric. Res. 49:729-736.
Tesso T, Claflin LE, Tuinstra MR (2005). Analysis of stalk rot resistance and genetic diversity among drought tolerant sorghum gentoypes. Crop Sci. 45:645-652.
Thakur RP, Reddy BVS, Indira S, Rao VP, Navi SS, Yang X-B, Ramesh S (2006). Sorghum Grain Mold. Plant Pathology and Microbiology

Publications and Papers. Paper 1.
Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor. Appl. Genet. 120:1289-1299.
Vanzetti LS, Yerkovich N, Chialvo E, Lombardo L, Vaschetto L and Helguer M (2013). Genetic structure of Argentinean hexaploid wheat germplasm. Genet. Mol. Biol. 36(3):391-399.
Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat. Biotechnol. 31:240-246.
Wang ML, Dean R, Erpelding J, Pederson G (2006). Molecular genetic evaluation of sorghum germplasm differing in response to fungal diseases: Rust (*Puccinia purpurea*) and anthracnose (*Collectotrichum graminicola*). Euphytica 148:319-330.
Wang ML, Zhu C, Barkley NA, Chen Z, Erpelding JE, Murray SC, Tuinstra MR, Tesso T, Pederson GA, Yu J (2009). Genetic diversity and population structure analysis of accessions in the U.S. historic sweet sorghum collection. Theor. Appl. Genet. 120:13-23.
Xu W, Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen HT (2000). Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench). Genome 43(3):461-469.
Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX (1997). POPGENE, the user-friendly shareware for population genetic analysis. Canada: Molecular Biology and Biotechnology Centre, University of Alberta.
Zhang Z, Lin Z, Xin Z (2009). Research progress in BYDV resistance genes derived from wheat and its wild relatives. J. Genet. Genomics 36(9):567-573.