Australian Journal of Crop Science AJCS
AJCS 10(4):520-527 (2016) ISSN:1835-2707
DOI: To be issued soon

Molecular evaluation of Ethiopian sweet sorghum germplasm and their contribution to regional breeding programs

Tesfaye Disasa[1,2], Tileye Feyissa[2], Belayneh Admassu[1,3], Rajneesh Paliwal[4], Santie de Villiers[5], Damaris Achieng Odeny[4*]

[1]Holetta Agricultural Research Center, P.O Box 2003, Addis Ababa, Ethiopia
[2]Faculty of Life Sciences, Addis Ababa University, P.O Box 1176, Addis Ababa, Ethiopia
[3]Current Address: USDA-ARS, 1691 S2700W, ID 83210, USA
[4]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), P.O Box 39063 – 00623 Nairobi, Kenya
[5]Pwani University, P.O Box 95-80108, Kilifi, Kenya

***Corresponding Author: d.odeny@cgiar.org**

Abstract

Sweet sorghum is an excellent feedstock for ethanol production and is also used for food and livestock feed. Germplasm collection and characterization in sweet sorghum is a crucial step towards breeding and development of superior genotypes for various end-uses. In the present study, 13 Simple Sequence Repeat (SSR) markers were used for genotyping 175 Ethiopian sweet sorghum genotypes alongside 27 improved accessions from eastern and southern Africa. All the tested markers detected 159 alleles and a high degree of polymorphism information content (PIC) averaging 0.69. A comparison between Ethiopian and improved accessions revealed higher allele numbers (124) in Ethiopian than improved accessions (92 alleles). More than half (65 out of 124) of the alleles observed in the Ethiopian accessions were rare (<5%) and 64 were private (only present within Ethiopian accessions) while in the improved accessions, 41% and 38% of the alleles detected were rare and private respectively. Both weighted Neighbor Joining-based clustering and hierarchical clustering grouped the 202 accessions into three major clusters based on geographical origin. Ethiopian accessions from the north (north Wello and south Tigray) not only clustered separately from accessions from the west central and eastern Ethiopia, but were also distinct from most of the improved genotypes. Our results reveal an unexploited highly diverse sweet sorghum genetic resource from Ethiopia that can be included in the regional breeding programs in order to efficiently optimize productivity.

Keywords: AMOVA; Genetic diversity; Population structure; *Sorghum bicolor*; SSR markers.
Abbreviations: A_A _Abundant allele; C_A _Common alleles; N_A _Total number of alleles; P_A- _Private alleles; PIC_Polymorphism information content; R_A_Rare alleles.

Introduction

multipurpose crop of great importance in the semi-arid tropics. Sorghum variants are grouped into five agronomic types depending on the end-use, namely: fiber, broomcorn, forage/fodder, grain and sweet sorghum. Sweet sorghum is taller, reaching up to 6m in height and also accumulates edible sugars in the stems (Vermerris, 2011; Huligol et al., 2004). As a C4 crop with the ability to produce high biomass in harsh conditions (Gnansounou et al., 2005), sweet sorghum is considered an attractive alternative for sustainable biofuel production (Murray et al., 2009; Kimber et al., 2013). Sweet sorghum is also used as animal feed and forage (Houx et al., 2013). Believed to have been first domesticated in Ethiopia and the surrounding countries commencing around 4000 – 3000 BC (Doggett, 1998; Dillon et al., 2007), sorghum has high genetic diversity in the eastern African region, especially in Sudan and Ethiopia (Gebrekidan, 1982). Vast collections (about 168, 500 accessions) of sorghum germplasm are maintained in various parts of the world (Billot et al., 2013) of which sweet sorghum makes a very small proportion. An effort in the 1980s to identify sweet sorghum breeding lines at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), which is a major repository for the world sorghum germplasm collection with a total of 36,774 accessions from 90 countries (Reddy et al., 2008), only ended with 70 sweet sorghum accessions (Reddy et al., 2005). The United States of America (USA) probably has the highest recorded collection of sweet sorghum (2,180 accessions) (Wang et al., 2009) originally brought from China and Africa in the 1850s (Maunder, 2000) and also the longest sweet sorghum breeding program (Murray et al., 2009). Given the increasing interest in sweet sorghum, the low numbers of accessions maintained in gene banks should be a major breeding concern, especially given the likely narrow genetic base of the collections (Murray et al., 2009). Besides, efforts to provide pedigree information that helps to distinguish sweet from grain type sorghum accessions has not received much attention (Lekgari and Dweikat, 2014) in the past. An exhaustive assessment of genetic variability among sweet sorghum accessions is an important step towards the conservation and full understanding of the contribution of sweet sorghum germplasm to global breeding programs. A recent whole-

genome sequencing study revealed untapped potential in Africa's indigenous sorghum (Mace et al., 2013), which global sweet sorghum collections could greatly benefit from. Being the center of diversity for sorghum, Ethiopia is one of the top contributors of sorghum genetic resource to the world (Reddy et al., 2008) and a potential source of novel genes (Haussmann et al., 2002; Reddy et al., 2009). The Ethiopian Institute of Biodiversity Conservation (IBC) maintains a huge collection of sweet sorghum, which is currently difficult to distinguish from the grain sorghum, and therefore difficult to utilize. Several past genetic diversity studies in sweet sorghum either concentrated on improved cultivars and germplasm maintained in gene banks elsewhere (Lekgari and Dweikat, 2014), or released varieties from individual countries (Ali et al., 2008). No known past diversity studies have exhaustively characterized sweet sorghum collections from Ethiopia. Molecular markers have played a significant role in the conservation and use of sorghum genetic resources (Aldrich and Doebley, 1992; Whitkus et al., 1992; Rami et al., 1998; Deu et al., 2006; Wang et al., 2006; Morris et al., 2013) and also in many aspects of sorghum improvement programs ranging from identification of diverse lines, to mapping of genomic regions controlling desirable traits and their use in marker-assisted breeding. Simple sequence repeat (SSR) markers are the preferred marker system for many sorghum genomics and molecular breeding applications (Caniato et al., 2007; Ali et al., 2008; Deu et al., 2008; Muraya et al., 2011), especially in developing countries (Sharma et al., 2010) where SNP genotyping applications are extremely rare. The present study assessed the genetic diversity and population structure of sweet sorghum accessions collected from a range of eco-geographical zones of Ethiopia alongside 27 globally known breeding lines.

Results

Polymorphism levels of SSR markers

Three SSR markers, *Xtxp289*, *SbKAK01* and *Xcup53*, were excluded from the final analysis because of poor quality. Thirteen markers produced a total of 159 alleles, across all 202 sweet sorghum accessions with an average of 12.23 alleles per marker. Most of the SSR loci were highly polymorphic. The number of alleles per marker ranged from 5 to 17 (Table 1). The PIC values for the SSR loci ranged from 0.37 to 0.85 with an average of 0.69 (Table 1). The highest PIC value of 0.85 was obtained for marker *Xtxp211*, whereas the lowest PIC value of 0.37 was recorded for marker *Xisep1029*. The overall observed heterozygosity (*Ho*) over loci varied from 0.04 (*Xgap342*) to 0.16 (*Xisep0108*) with an average of 0.09. The gene diversity index (expected heterozygosity, *He*) ranged from 0.42 (*Xisep1029*) to 0.85 (*Xtxp141*) with a mean of 0.72. Marker *Xisep1029* presented the lowest gene diversity as well as PIC value. A comparison between Ethiopian accessions and the improved sweet sorghum revealed higher average number of alleles (9.5) in the former than in the latter (7.1) (Table 1). Ethiopian accessions were also significantly richer in rare and private alleles even though the numbers of common alleles were comparable between the two groups (Table 1).

Genetic relationships

Phylogenetic analysis revealed clustering of accessions based on their geographical origin with a few discrepancies (Fig 2). Three major clusters were observed; one for accessions from the northern Ethiopia (Wello and Tigray); a second one for accessions from the eastern (east Hararge and west Hararge) and west central (Shewa, Wollega and Gojam) parts of Ethiopia; and a third cluster of improved cultivars (Fig 2). The dendrogram showed that accessions collected from different zones of northern Ethiopia were strongly differentiated from the rest of the collections. Similarly, collections from ICRISAT showed close similarity with accessions from Hararge suggesting that most of the breeding lines probably shared ancestry with accessions originally selected from Hararge. One particular improved line *ICSV93046* (ICR_19) clustered with Ethiopian accessions from Tigray and also showed potential duplication with improved cultivar *Sorcoll163/07* (PI2) from Ethiopia and high similarity with *Gambella* (PI3) (Fig 2). ICRISAT line *ICSV93046* and Ethiopian improved cultivar *Sorcoll163/07* could have been selected from the same background. There was also another potential duplication of two ICRISAT lines that clustered close to accessions from Gojam (Fig 2). The highest population differentiation was observed among improved genotypes and the rest of the Ethiopian accessions from different regions (Table 2). Expectedly, the lowest population differentiation was observed between south Tigray and north Wello (F_{ST} = 0.03), which are located adjacent to each other (Fig 1), followed by south Tigray and south Wello (F_{ST} = 0.08). The highest population differentiation was observed between the three improved genotypes (*Sorcoll163/07*, *Gambella*, *AS27*) and accessions from east Wollega (F_{ST} = 0.55), east Shewa (F_{ST} = 0.52) and Gojam (F_{ST} = 0.52) (Table 2). Results of AMOVA showed significant genetic difference in the sampled populations with 76.9% genetic variation within populations and 23.1 % among populations (Table S2).

Population structure

Analysis of population structure was done with 13 polymorphic SSR markers using STRUCTURE. A sharp peak was detected in ΔK at K=3 (Fig S1) suggesting the presence of 3 major populations (Evanno et al., 2005). The 3 detected sub-populations (Fig 3) followed a similar clustering pattern as observed in the dendrogram (Fig 2) with two major Ethiopian sub-populations and one sub-population of improved genotypes. Population I consisted mainly of accessions from the eastern and west central part of Ethiopia (Hararge, Shewa, Gojam, Wollega), while population II was composed mainly of improved accessions. Population III contained the highest numbers of accessions, which were mainly from the northern part of Ethiopia (Tigray and Wello) (Fig 3).

Discussion

Our study used 13 polymorphic markers evenly distributed across the sorghum genome to determine the extent of diversity among 202 sweet sorghum accessions collected from Ethiopia including improved lines either released or lined up for release in the eastern and southern Africa. The PIC values ranged from 0.37-0.85, which was comparable to a similar recent study of Ethiopian sorghum collections (Adugna, 2014) in which a PIC range of 0.39-0.85 was observed using 12 SSRs across 160 accessions. We report better results than those observed using global sweet sorghum germplasm (142 accessions) obtained from various gene banks using 29 polymorphic SSR markers (Lekgari and Dweikat, 2014) in which the PIC ranged from 0.221 to 0.75. Overall, the average PIC value of the current study (0.69) was higher than most of the previously reported values using both grain and sweet sorghum (Agrama and Tuinstra, 2003; Caniato et al., 2007; Ali et al., 2008; Wang et al., 2009;

Table 1. Comparison of genetic diversity of Ethiopian and improved sweet sorghum genotypes

	Combined analysis		Ethiopian accessions					Improved accessions				
Marker	N_A	PIC	N_A	R_A	C_A	A_A	P_A	N_A	R_A	C_A	A_A	P_A
Xcup33	8	0.51	9	6	3	-	6	3	1	2	-	0
Xgap342	5	0.45	10	5	5	-	7	10	4	6	-	7
Xisep0108	12	0.75	6	3	2	1	1	7	3	4	-	2
Xisep0617	6	0.37	3	1	1	1	0	5	1	4	-	2
Xisep0938	13	0.83	12	6	6	-	9	3	1	1	1	0
Xisep1029	10	0.75	4	1	2	1	1	5	3	1	1	2
Xtxp012	16	0.85	16	10	6	-	10	7	3	3	1	1
Xtxp141	17	0.83	12	6	6	-	6	7	3	3	1	1
Xtxp205	17	0.75	5	1	4	-	2	8	3	5	-	5
Xtxp211	12	0.64	13	6	7	-	7	9	5	4	-	3
Xtxp284	17	0.81	12	6	6	-	4	13	6	7	-	5
Xtxp312	9	0.64	12	7	5	-	8	9	4	5	-	5
Xtxp279	17	0.77	10	7	3	-	3	6	1	5	-	2
Total	159	N/A	124	65	56	3	64	92	38	50	4	35
Mean	12.23	0.69	9.5	5	4.3	1.0	4.9	7.1	2.9	3.8	1	2.7

N_A - Total number of Alleles; R_A - Rare alleles that were present in <5% of the accessions; C_A - Common alleles that were present in 5-50% of the accessions; A_A- Abundant alleles that were present in more than 50% of the accessions; P_A - Private alleles that were only found in the respective population set

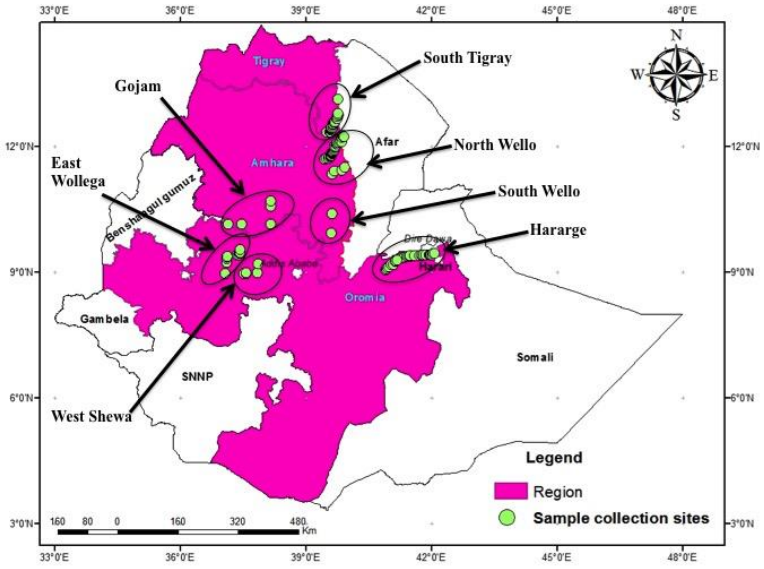

Fig 1. Map of Ethiopia showing sweet sorghum collection sites and regions.

Ramu et al., 2013) although lower in comparison to other studies (Muraya et al., 2011; Adugna et al., 2013) that included wild sorghum. Wild sorghum has been reported to be more genetically diverse resulting in higher numbers of alleles per locus (Mutegi et al., 2011; Ramu et al., 2013; Fernandez et al., 2014) than cultivated sorghum. The higher numbers of rare and private alleles observed among Ethiopian collections in comparison with improved sweet sorghum revealed that the Ethiopian accessions have preserved the most rare and private alleles, and therefore are very important resources for future sweet sorghum improvement. Our results further confirmed an earlier report by Mace et al. (2013) who observed an untapped genetic potential in indigenous sorghum (both wild and cultivated) for crop improvement. Although the significance of the rare alleles cannot be speculated from the current results, their abundance in the Ethiopian sweet sorghum accessions suggests that this diverse set of germplasm has not been extensively exploited to improve sweet sorghum material in eastern and southern Africa.

We observed that the 27 improved genotypes also contained 35 private alleles indicating their uniqueness to the Ethiopian sweet sorghum gene pool. This is not surprising, especially for ICRISAT improved lines, which have made use of global collections available at the gene bank in India. Future genome wide association studies (GWAS) will be required to determine the exact contributions of these private and rare alleles to traits of agronomic importance in sorghum. Such studies will also guide the implementation of various conservation decisions that will be necessary to ensure these important alleles are not genetically eroded. The high numbers of rare and private alleles may also be indicative of gene flow from wild to cultivated Ethiopian accessions. Ethiopia being a center of diversity for sorghum harbors abundant wild accessions that grows side by side with cultivated accessions resulting in regular hybridization between the different gene pools. It would therefore be expected that unimproved sorghum collections from the region would have a high proportion of rare and private alleles when compared with improved ones.

Table 2. Estimate of pairwise F_{ST} among populations collected from different regions alongside improved material.

Group	SW	NW	ST	G	EH	WH	WS	EW	ICR
NW	0.08***								
ST	0.08*	0.03***							
G	0.37***	0.37***	0.37***						
EH	0.19***	0.24***	0.24***	0.32***					
WH	0.24***	0.29***	0.28***	0.31***	0.10*				
WS	0.32***	0.36***	0.36***	0.33***	0.23***	0.14*			
EW	0.35***	0.38***	0.37***	0.22***	0.25***	0.20***	0.11		
ICR	0.15***	0.19***	0.18***	0.34***	0.17***	0.24***	0.35***	0.36***	
PI	0.23*	0.24***	0.21**	0.52**	0.31***	0.40***	0.52***	0.55*	0.19***

*Pairs were significant at *P value < 0.05, **P value < 0.01, ***P value < 0.001. SW: South Wollo; NW: North Wollo; ST: South Tigray; G: Gojam; EH: East Hararge; WH: West Hararge; WS: West Shewa; EW: East Wollega; ICR: ICRISAT*

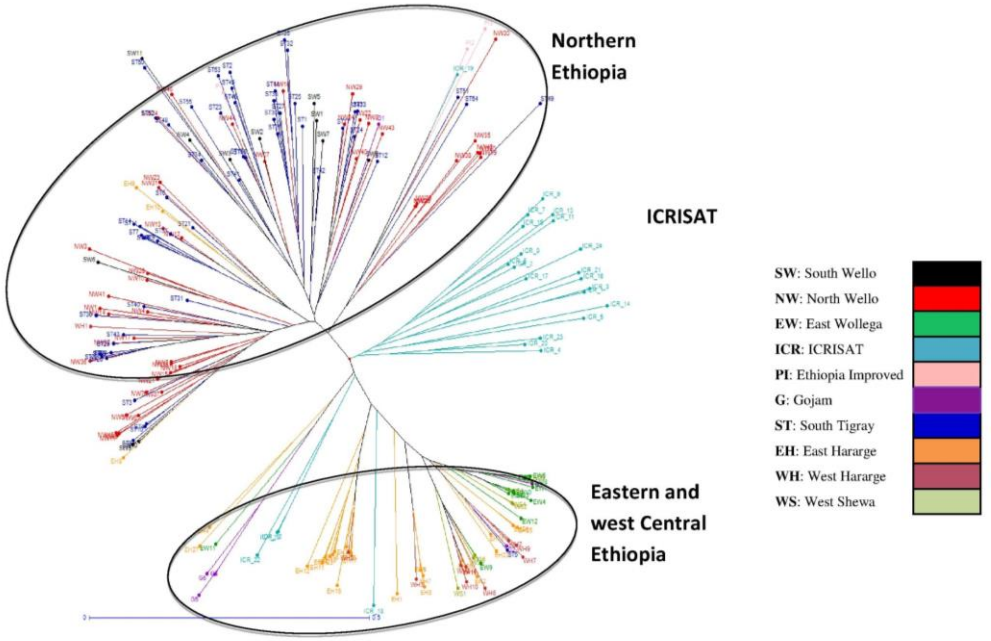

Fig 2. Unweighted Neighbor-Joining tree showing relatedness among the 202 sweet sorghum accessions. The key is shown with respective colour codes. The accession identities for the codes provided are shown in Supplementary Table S1.

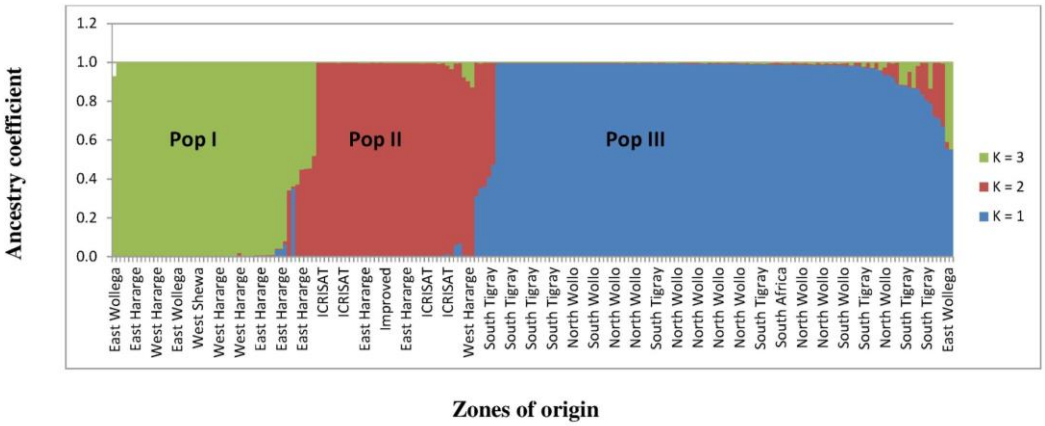

Fig 3. Bar plots of the population structure analysis of Ethiopian sweet sorghum lines alongside cultivars released in the eastern and southern Africa. Three populations were observed that were based on regions of origin.

This is further confirmed by other studies that characterized Ethiopian sorghum accessions maintained in Ethiopia (Adugna, 2014) or elsewhere (Cuevas and Prom, 2013) as well as those from Kenya (Mutegi et al., 2011) that reported comparable proportions of rare alleles. Although an overall high level (78.7%) of rare alleles has been reported in global composite germplasm collection (GCGC) of sorghum (Billot et al., 2013), the proportion of rare alleles reported for Indian sorghum collections have been lower than in the current study (Madhusudhana et al., 2012). The Ethiopian sweet sorghum collections showed distinct clustering between accessions from the north and those from the eastern and west central Ethiopia. The results are in agreement with past studies conducted for Ethiopian (Beyene et al., 2014) and global sorghum collections (Madhusudhana et al., 2012; Ramu et al., 2013; Lekgari and Dweikat, 2014) in which clustering based on geographic origin was observed for both grain and sweet sorghum. Other studies that included different sorghum races (Wang et al., 2013) and assessed their biological status (Billot et al., 2013) further reported clustering based on geographic origin and race (Sagnard et al., 2011; Wang et al., 2013; Kitavi et al., 2014) and also according to biological status (Billot et al., 2013). There are also other studies in sorghum that did not observe geographical origin or racial clustering pattern (Motlhaodi et al., 2014) although low numbers of accessions used may have been a major contributing factor. The specific grouping of accessions obtained from ICRISAT is consistent with results of another diversity study carried out on Chinese sweet sorghum breeding lines (Zhan et al., 2012) that reported clustering of genotypes bred from the same institute. However, this unique clustering of breeding lines points to underutilization of the available genetic resources needed to maximize outputs for sweet sorghum in the region, and therefore should be a major concern to the relevant breeding institutions, especially ICRISAT. The clustering process also revealed the existence of putative duplicate genotypes, some of which were improved lines. In the case of Ethiopian accessions, the duplicates may be a result of exchanges of seeds between farmers of different regions, which subsequently got renamed and treated as different genotypes (Ng'uni et al., 2011). In the case of improved lines, the duplicate genotypes may be the same genotype released in different countries under different names, or selections from the same background that might be differing only at a few loci. The clustering results will therefore aid in future breeding programs to select parental lines that are genetically diverse if heterosis is to be achieved. Analysis of molecular variance showed high genetic diversity within population than among populations. This means that most genetic differentiation was observed when all the 202 accessions were analyzed together than when collections from different regions were analyzed independently. Our observations are similar to previous studies on grain sorghum (Adugna et al., 2013; Ramu et al., 2013; Adugna, 2014) and suggest that farmers from the same region, perhaps speaking the same local language, tend to grow the same genotypes (Labeyrie et al., 2014). The genetic similarity of materials collected from Gojam and east Wollega could be geographic as the two regions share a common boundary, which could enhance regular informal seed exchange. The 202 accessions were further classified into three distinct sub-populations (delta K = 3) using STRUCTURE. The sub-populations generated were complementary to the Neighbor-Joining (NJ) tree analysis that also grouped the 202 sweet sorghum accessions into three clusters. The similarity in the clustering of genotypes using Neighbor Joining and STRUCTURE further validates the representative nature of the markers selected for this study. Although no known genetic diversity studies have been done in Ethiopia focusing on sweet sorghum accessions, a study by Murray et al. (2009) also suggested a minimum of 3 distinct sweet sorghum populations while studying a diverse panel. Our results provide more evidence for this theory but will require more data to be conclusive. The global efforts on generation and subsequent utilization of green energy call for optimum breeding and exploitation of available genetic resources around the world. A future global detailed study on the genetic structure and diversity of sweet sorghum will be necessary for more efficient and sustainable utilization of the genetic resources for sweet sorghum breeding. The current study forms an essential foundation for future sweet sorghum breeding programs but also for ongoing programs such as ICRISAT. Both the Ethiopian sweet sorghum breeding program and that of ICRISAT will stand to gain from each other as they each represent different genetic groups. The abundance of rare and private alleles also calls for careful germplasm conservation strategies in order to ensure that these important alleles are not lost. The specific contributions of these alleles to important agronomic traits will need to be studied in more detail to ensure superior genotypes with desirable allele combinations are created.

Materials and Methods

Plant materials

A total of 178 sweet sorghum accessions were collected from all major sweet sorghum growing regions of Ethiopia (Fig 1) (Supplementary Table S1). Improved genotypes such as Gambella (sweet type) and Sorcoll 163/07 (grain type) were obtained from Melkassa Agricultural Research Center (MARC), whereas Dr. Rebeka Gebretsadik (a researcher at Amhara Agricultural Research Institute, Ethiopia) provided AS27 from South Africa (Table S1). All the improved sweet stem sorghum accessions either released or lined up for release in the eastern and southern Africa were obtained from ICRISAT, Nairobi (Table S1). Seeds of Ethiopian accessions as well as AS27 were multiplied by planting the germplasm at MARC, Ethiopia.

DNA extraction

After seed multiplication, seeds were sown in a greenhouse at Holetta Agricultural Research Center (HARC) of the Ethiopian Institute of Agricultural Research (EIAR), while global collections were grown at ICRISAT-Nairobi. Leaf tissues were collected from two to three-week-old seedlings followed by genomic DNA extraction using Promega Kit (Madison, USA) and CTAB method (Mace et al., 2003) for the Ethiopian and global collections respectively. Quantity and quality of the DNA was checked using spectrophotometry and by running on 0.8% agarose gel stained with GelRed® (Biotium, USA), respectively. The final volume of the extracted DNA was diluted to 10 ng/µl for polymerase chain reaction (PCR).

PCR and fragment analyses

Sixteen polymorphic SSR primers (Menz et al., 2002; Schloss et al., 2002; Ramu et al., 2009) that were evenly distributed across the whole sorghum nuclear genome were selected for use in genotyping the 202 accessions. All forward primers contained an M13-tag (5'-CACGACGTTGTAAAACGAC - 3') on the 5' end that was fluorescently labeled to allow detection of amplification products (Schuelke, 2000). PCR amplification was performed in 10 µl reaction volume comprising of 1 x PCR buffer (20 mM Tris-HCl, pH 7.6; 100 mM KCl; 0.1 mM EDTA; 1 mM DTT; 0.5% (w/v) Triton X-100; 50% (v/v) glycerol), 2 mM $MgCl_2$, 0.16 mM dNTPs, 0.16 µM fluorescent labeled M13-forward primer, 0.04 µM forward primer, 0.2 µM reverse primer, 0.2 units of Taq DNA polymerase (SibEnzyme Ltd, Russia) and 30ng of template DNA. Forward primers were labeled with FAM, PET, NED or VIC (Applied Biosystems, USA). PCR was carried out in a GeneAmp® PCR System 9700 thermal cycler (Applied Biosystems) programmed for initial denaturation at 94°C for 15 min, followed by second denaturation at 94°C for 30 sec, annealing at 50°C for 1 min, extension at 72°C for 2 min and final elongation at 72°C for 20 min.

Successful amplification was confirmed by running 2.0 µl of the PCR products on a 2% (w/v) agarose gel stained with GelRed® (Biotium) and visualized under UV. Depending on the nature of fluorescent label and strength of the amplification bands used, a volume ranging from 2.5µl to 3.5µl of four different amplification products were co-loaded along with the internal size standard, GeneScan™ –500 LIZ® (Applied Biosystems) and Hi-Di™ Formamide (Applied Biosystems). The fragments were separated by capillary electrophoresis using an ABI Prism® 3730 Genetic analyzer (Applied Biosystems). PCR fragment sizes were manually scored using GeneMapper 4.0 software (Applied Biosystems). PowerMarker v.3.25 (Liu and Muse 2005) was used to compute PIC and total number of alleles (Table 1). Rare (<5%), common (5-50%), abundant (>50%) and private alleles (alleles only present in one group and not the other) were manually computed in MS Excel (Microsoft Inc., Seattle, USA) for improved (ICRISAT, *Sorcoll163/07*, *Gambella* and *AS27*) and Ethiopian accessions. Polymorphism information content (PIC) was calculated using the method of Botstein et al. (1980).

$$PIC = 1 - \sum_{i=0}^{k} p_i^2 - \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} 2\, p_i^2 p_j^2$$

Where, p_i and p_j are the frequencies of alleles i and j, respectively

Markers with a PIC value of more than 0.5 were considered highly informative, between 0.25 and 0.5 as informative and less than 0.25 as less informative.

Phylogenetic analysis

To identify the pair-wise genetic relationships between accessions, a genetic dissimilarity matrix was analyzed using Neighbor Joining (NJ) method, as implemented in DARwin v5 (Perrier and Jacquemoud-Collet, 2006). The dendrogram was constructed using the same software. Genetic distance between groups was estimated by F_{ST} statistics. F statistics parameters were performed as follows based on the method of Wright (1951);

$F_{ST} = (F_{IT} - F_{IS}) / (1-F_{IS})$

Where,

F_{ST} is the fixation index describing the correlation of genes of different individuals in the same population;

F_{IS} is the inbreeding coefficient, describing the correlation of genes within individuals in the population;

F_{IT} is the overall inbreeding coefficient, describing the correlation of genes within individuals relative to the total population.

Analysis of Molecular Variance (AMOVA) was determined using Arlequin 3.11 software (Excoffier et al., 2005). A genetic dissimilarity matrix was calculated using pairwise F_{ST} in order to determine pair-wise genetic relationships among regions.

Population structure

Population structure refers to the non-random distribution of genotypes among individuals within population. It occurs from the unequal distribution of alleles among subpopulations of different ancestries. Two complementary approaches, Bayesian clustering and Neighbor-Joining tree, were used to assess the genetic structure without defining a-priori populations. The genetic structure of sweet sorghum populations was determined using the admixture model based on Monte Carlo Markov Chain algorithm implemented in program STRUCTURE 2.3.3 software (Pritchard et al., 2000). The admixture model with correlated allele frequencies was used, assuming that the genome of each individual resulted from the mixture of K ancestral populations. The estimated proportions of each individual's genotype originating from each of the K ancestral populations (q) was calculated for K ranging from 1 to 12 ancestral populations (or clusters), with fifteen runs for each K value. For each run, a burn-in period of 50000 and MCMC (Markov Chain Monte Carlo) replications of 10000 was used. The

optimum *K* value was calculated using structure harvester http://taylor0.biology.ucla.edu/structureHarvester/, which computed the log likelihood of the data [LnP(D)] in the STRUCTURE output and an *adhoc* statistic Δk based on the rate of change in LnP(D) between successive *k* (Evanno et al., 2005). Results from each replicate run were combined using the CLUMPP software (Jakobsson and Rosenberg, 2007). The results of both the Bayesian clustering and Neighbor-Joining methods were then compared to check for consistency of the clusters.

Conclusions

We successfully characterized 202 sorghum accessions using 13 polymorphic markers and identified distinct grouping between Ethiopian and ICRISAT sweet sorghum lines released in eastern and southern Africa. The distinctness of these accessions reveals the lack of exploitation of the abundant sweet sorghum genetic variability from Ethiopia and calls for more inclusion of the unique germplasm in various local and global breeding programs in the future. The abundance of rare and private alleles observed among Ethiopian collections provides more evidence for novel alleles that can be efficiently exploited through future genome-wide association studies (GWAS) for sugar related traits. Our results create an opportunity to enrich both national and global gene banks with diverse sweet sorghum landraces but also provide immediate germplasm resources that can be used directly in breeding programs.

Acknowledgements

This research is part of a PhD work by the first author. The study was conducted by financial support provided by the Swedish International Development Agency (SIDA) through Bio-Innovate project "Delivering new sorghum and millets innovations for food security and improving livelihoods in Eastern Africa"- project No. 01/2010. The authors acknowledge additional financial support of the Ethiopian Institute of Agricultural Research. The authors are also grateful to Miss. Annis Saiyiorri, Mr. Vincent Njunge and Samuel Manthi for their technical assistance.

References

Adugna A, Snow AA, Sweeney PM, Bekele E, Mutegi E (2013) Population genetic structure of *in situ* wild *Sorghum bicolor* in its Ethiopian center of origin based on SSR markers. Genet Resour Crop Ev. 60:1313-1328.

Adugna A (2014) Analysis of *in situ* diversity and population structure in Ethiopian cultivated *Sorghum bicolor* (L.) landraces using phenotypic traits and SSR markers. Springer Plus 3:1-14.

Agrama HA, Tuinstra MR (2003) Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. Afr J Biotechnol. 2:334-340.

Aldrich PR, Doebley J (1992) Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. Theor Appl Genet. 85:293-302.

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

Beyene A, Hussien S, Pangirayi T, Mark L, Fentahun M (2014) Genetic variation in lowland sorghum (*Sorghum bicolor* (L.) Moench) landraces assessed by simple sequence repeats. Plant Genet Resour. 14:1-11.

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:e59714.

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

Caniato F, Guimara˜es C, Schaffert R, Alves V, Kochian L, Borém A, Klein PE, Magalhaes JV (2007) Genetic diversity for aluminum tolerance in sorghum. Theor Appl Genet. 114:863-876.

Cuevas HE, Prom LK (2013) Assessment of molecular diversity and population structure of the Ethiopian sorghum [*Sorghum bicolor* (L.) Moench] germplasm collection maintained by the USDA–ARS national plant germplasm system using SSR markers. Genet Resour Crop Ev. 60:1817-1830.

Deu M, Rattunde H, Chantereau J (2006) A global view of genetic diversity in cultivated sorghums using a core collection. Genome. 49:168-180.

Deu M, Sagnard F, Chantereau J, Calatayud C, He´rault D, Mariac C, Pham JL, Vigouroux Y, Kapran I, Traore PS, Mamadou A, Gerard B, Ndjeunga J, Bezançon G (2008) Niger-wide assessment of in situ sorghum genetic diversity with microsatellite markers. Theor Appl Genet. 116:903-913.

Dillon SL, Shapter FM, Henry RJ, Cordeiro G, Izquierdo L, Lee S (2007) Domestication to crop improvement: genetic resources for sorghum and saccharum (Andropogoneae). Ann Bot. 100:975–989.

Doggett H (1988) Sorghum (2nd edn). London: Longman.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 14:2611-2620.

Excoffier L, Laval LG, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol Bioinform Online 1:47-50.

Fernandez MGS, Okeno JA, Mutegi E, Fessehaie A and Chalfant S (2014) Assessment of genetic diversity among sorghum landraces and their wild/weedy relatives in western Kenya using Simple Sequence Repeat markers. Conserv Genet. 15:1269-1280.

Gebrekidan B (1982) Utilization of accession in sorghum improvement. In: House LR, Mughogho LK, Peacock JM (eds) Sorghum in the eighties. Proceedings of the international symposium on sorghum, ICRISAT, Patancheru, 1982.

Gnansounou E, Dauriat A, Wyman CE (2005). Refining sweet sorghum to ethanol and sugar: economic trade-offs in the context of North China. Biores Technol. 96:985-1002.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 23:1801-1806.

Houx JH, Roberts CA and Fritschi FB (2013) Evaluation of sweet sorghum bagasse as an alternative livestock feed. Crop Sci. 53:1784-1790.

Haussmann BI, Mahalakshmi V, Reddy BV, Seetharama N, Hash CT, Geiger HH (2002) QTL mapping of stay-green in two sorghum recombinant inbred populations. Theor Appl Genet. 106:133-142.

Huligol RV et al (2004) A trial with sweet sorghum. In: Alternative uses of sorghum and pearl millet in Asia*:* Proceedings of the Expert Meeting, ICRISAT, Andhra Pradesh, India. 1–4 July 2003. CFC Technical Paper No. 34. The Netherlands (CFC) and ICRISAT Patancheru, p 364.

Kimber CT, Dahlberg JA and Kresovich S (2013) The gene pool of Sorghum bicolor and its improvement. In: Paterson AH (ed) Genomics of the Saccharinae. New York: Springer, p 23-41.

Kitavi MN, Kiambi DK, Haussman B, Semagn K, Muluvi G, Kairichi M and Machuka J (2014) Assessment of the genetic diversity and pattern of relationship of West African sorghum accessions using microsatellite markers. Afr J Biotechnol. 13:1503-1514.

Labeyrie V, Deu M, Barnaud A, Calatayud C, Buiron M, Wambugu P, Manel S, Glaszmann JC, Leclerc C (2014) Influence of ethnolinguistic diversity on the sorghum genetic patterns in subsistence farming systems in eastern Kenya. PLoS One. 9:e92178.

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.

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 21:2128-2129.

Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, Campbell BC, Hu W, Innes DJ, Han X, Cruickshank A, Dai C, Fre`re C, Zhang H, Hunt CH, Wang X, Shatte T, Wang M, Su Z, Li J, Lin X, Godwin ID, Jordan DR, Wang J (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat Commun. 4:2320.

Madhusudhana R, Balakrishna D, Rajendrakumar P, Seetharama N, Patil JV (2012) Molecular characterization and assessment of genetic diversity of sorghum inbred lines. Afr J Biotechnol. 11:15626-15635.

Maunder AB (2000) History of cultivar development in the United States: From memoirs of AB Maunder-sorghum breeder. In: Smith CW, Frederickson RA (ed) Sorghum: Origin, history, technology and production. John Wiley and Sons, New York.

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.

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. P Natl Acad Sci USA. 110:453-458.

Motlhaodi T, Geleta M, Bryngelsson T, Fatih M, Chite S, Ortiz R (2014) Genetic diversity in *ex-situ* conserved sorghum accessions of Botswana as estimated by microsatellite markers. Aust J Crop Sci. 8:35-43.

Muraya MM, de Villiers S, Parzies HK, Mutegi E, Sagnard F, Kanyenji BM, Kiambi D, Geiger HH (2011) Genetic structure and diversity of wild sorghum populations (*Sorghum spp*.) from different eco-geographical regions of Kenya. Theor Appl Genet. 123:571–583.

Murray SC, Rooney WL, Mitchell SE, Kresovich S (2009) Sweet sorghum genetic diversity and association mapping for Brix and height. Plant Genome. 2:48-62.

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. 122:989-1004.

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. 148:52-62.

Perrier X, Jacquemoud-Collet JP (2006) DARwin software, http://darwin.cirad.fr/darwin

Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

Rami JF, Dufour P, Trouche G, Fliedel G, Mestres C, Davrieux F, Blanchard P, Hamon P (1998) Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). Theor Appl Genet. 97:605-616.

Ramu P, Kassahun B, Senthilvel S, Kumar CA, Jayashree B, Folkertsma RT, Reddy LA, Kuruvinashetti MS, Haussmann BIG, Hash CT (2009) Exploiting rice–sorghum synteny for targeted development of EST-SSRs to enrich the sorghum genetic linkage map. Theor Appl Genet. 119:1193-1204.

Ramu P, Billot C, Rami JF, Senthilvel S, Upadhyaya HD, Reddy L A, Hash CT (2013) Assessment of genetic diversity in the sorghum reference set using EST-SSR markers. Theor Appl Genet. 126:2051-2064.

Reddy BVS, Ramesh S, Sanjan Reddy P, Ramaih B, Salimath PM, Kachapar R (2005) Sweet sorghum - a potential alternate raw material for bio-ethanol and bio-energy. Int Sorghum Millets Newslett. 46:79–86.

Reddy BVS, Kumar AA, Reddy PS, Elangovan M (2008) Sorghum germplasm: diversity and utilization. In: Reddy BVS, Ramesh S, Kumar AA, Gowda CLL (eds) Sorghum improvement in the new millennium Patancheru, Andhra Pradesh, India: International Crops Research Institute for the Semi-Arid Tropics.

Reddy BVS, Ramesh S, Reddy PS, Kumar AA (2009) Genetic enhancement for drought tolerance in sorghum. Plant Breed Rev. 31:189–222.

Sagnard F, Deu M, Dembélé D, Leblois R, Toure L, Diakité M,Calatayud C, Vaksmann M, Bouchet S, Mallé Y, Togola S, Traoré PCS (2011) Genetic diversity, structure, gene flow and evolutionary relationships within the *Sorghum bicolor* wild-weedy-crop complex in a western African region. Theor Appl Genet. 123:1231-1246.

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

Schuelke M (2000) An economic method for the fluorescent labelling of PCR fragments. Nat Biotechnol. 18:233-234.

Sharma R, Deshpande SP, Senthilvel S, Rao VP, Rajaram V, Hash CT, Thakur RP (2010) SSR allelic diversity in relation to morphological traits and resistance to grain mould in sorghum. Crop Pasture Sci. 61:230-240.

Vermerris W (2011) Survey of genomics approaches to improve bioenergy traits in maize, sorghum and sugarcane. J Integr Plant Biol. 53:105-116.

Wang ML, Dean R, Erpelding J, Pederson G (2006) Molecular genetic evaluation of sorghum accession 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 US historic sweet sorghum collection. Theor Appl Genet. 120:13-23.

Wang YH, Upadhyaya HD, Burrell AM, Sahraeian SME, Klein RR, Klein PE (2013) Genetic structure and linkage disequilibrium in a diverse, representative collection of the C4 model plant, *Sorghum bicolor*. Genes Genome Genet. 3:783-793.

Whitkus R, Doebley J, Lee M (1992) Comparative genome mapping of sorghum and maize. Genetics 132:1119-1130.

Wright S (1951) The genetical structure of populations. Ann Eugenic. 15:323-354.

Zhan Q, Zhou L, Bi N, Wu H, Li J, Lu J, Lu J, Lin P (2012) A comprehensive analysis of genetic diversity and relationships of 13 sweet sorghum varieties as energy sources. J Sustain Bioenerg Syst. 2:86-91