

## Full Length Research Paper

# DNA markers reveal genetic structure and localized diversity of Ethiopian sorghum landraces

Haile Desmae<sup>1,2\*</sup>, David R. Jordan<sup>3</sup> and Ian D. Godwin<sup>1,4</sup>

<sup>1</sup>The University of Queensland, School of Agriculture and Food Sciences, QLD, Australia.

<sup>2</sup>ICRISAT-WCA, BP320, Bamako, Mali.

<sup>3</sup>The University of Queensland, Queensland Alliance for Agriculture and Food Innovation, Warwick, QLD, Australia.

<sup>4</sup>The University of Queensland, Queensland Alliance for Agriculture and Food Innovation (QAAFI), St Lucia, QLD 4072, Australia.

Received 14 April, 2016; Accepted 26 September, 2016

North Eastern Ethiopia is a major sorghum-growing region. A total of 415 sorghum landraces were sampled to represent the range of agro-ecologies (three altitude ranges) as well as spatial heterogeneity, that is, 4 zones: North Welo, South Welo, Oromiya and North Shewa with each zone containing 2 to 5 districts. The landraces were genotyped with simple sequence repeats (SSR) and inter simple sequence repeats (ISSR) markers. High genetic diversity was observed among the landraces for both marker systems. STRUCTURE analysis revealed 4 clusters of genetically differentiated groups of landraces. Cluster analysis revealed a close relationship between landraces along geographic proximity with genetic distance between landraces increasing with an increase in geographic distance. The grouping of landraces based on districts was influenced by clinal trend and geographic proximity. The  $F_{ST}$  statistics showed significant geographic differentiation among landraces at various levels of predefined geographic origin but a large portion of the variation was among landraces within rather than between predefined populations. The landraces from North Shewa were predominantly in one cluster, and landraces from this area also exhibited the greatest allelic diversity and the highest number of private alleles. There was low variation among the highland *Zengada* landraces, but these landraces were quite strongly differentiated and fell into one population cluster. The low to moderate genetic differentiation between landraces from various geographic origins could be attributed to gene flow across the region as a consequence of seed exchange among farmers.

**Key words:** Differentiation, diversity, landraces, simple sequence repeats (SSR), sorghum.

## INTRODUCTION

Sorghum is believed to have been first domesticated in the Ethiopia/Sudan region of Eastern Africa (Vavilov,

1951; Harlan and de Wet, 1972; Stemler et al., 1977). Earlier reports indicated that four of the main five races of

\*Corresponding author. E-mail: hailshd@yahoo.com.au.

sorghum (except kafir) and the corresponding subraces are grown in Ethiopia (Doggett, 1988; Teshome et al., 1997; Ayana and Bekele, 1998). The races are found distributed over geographic locations of the country, with some areas of local concentration (Stemler et al., 1977; Prasada Rao and Mengesha, 1987; Doggett, 1988). Guinea and caudatum types are mainly grown in south western and western parts of the country. The bicolor race occurs mainly in western high rainfall highland parts of Ethiopia, and on a minor scale almost everywhere in other sorghum growing areas of the country. Large seeded and compact panicle durra and the corresponding sub-races, particularly durra-caudatum, are typical of northern and eastern Ethiopia. The country is one of the main contributors of sorghum germplasm to the world collections at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) (Reddy et al., 2002). Ethiopian sorghums have been used both in national and international sorghum improvement programs (Hawkes and Worede, 1991). For example, the important stay-green trait is from B35 (Borrell et al., 2003) and ergot resistance from IS8525 (Dahlberg et al., 2001).

North Eastern (NE) Ethiopian subsistence agriculture is dependent on sorghum, ranking first in production and second to tef in area sown (CSA, 2003). The majority of sorghum production in the country in general and NE Ethiopia in particular depends on landraces (Gebrekidan, 1973; Worede, 1992; Teshome et al., 1997, 1999; Seboka and van Hintum, 2006; Shewayrga et al., 2008). The sorghum grain is mainly used for *Injera*, a fermented flat bread and staple Ethiopian dish. Sorghum is the preferred crop after tef for making *injera* (Zegeye, 1997). The grain is also used for making local beverages (*tela* and *areke*) as well as porridge, boiled grains and roasted immature grains. The stalk is fed to animals, and is also used as fuel wood and for construction.

The sorghum breeding program for NE Ethiopia and other moisture stress areas has been focused on developing early maturing improved varieties. However, the adoption of early maturing improved varieties had limited success since the varieties do not meet farmers' needs and selection criteria (Seboka and van Hintum, 2006). Utilization of local landraces in the breeding program is important to combine important features of such landraces to meet farmers' needs. Characterizing the extent and structure of crop genetic diversity is important to understand the genetic variability available and its potential use in breeding programs. It is also important to devise appropriate sampling procedures for germplasm collection and conservation and the establishment of core collections (Brown, 1995; Hayward and Sackville-Hamilton, 1997; Ramanatha Rao and Hodgkin, 2002). Different marker systems including morphological, biochemical and molecular markers have been used to measure the extent and structure of genetic diversity in crop plants. Polymerase chain reaction (PCR) based assays for plant DNA fingerprinting such as AFLP

(Vos et al., 1995), SSR (Tautz and Renz, 1984) and ISSR (Zietkiewicz et al., 1994) have become widely used. These markers have been widely used for reliable and robust characterization of crop genetic resources (Godwin et al., 1997; Pejic et al., 1998; Smith et al., 2000; Wen et al., 2002; Uptmoor et al., 2003; Medraoui et al., 2007; Weerasooriya et al., 2016). With the development of high-throughput sequencing (or next-generation sequencing-NGS) technologies in recent years, marker techniques such as single nucleotide polymorphism (SNPs) and genotyping by sequencing (GBS) have become very useful for exploring the diversity within plant species, constructing haplotype maps and performing genome-wide association studies as well as genomics-assisted breeding (He et al., 2014). The exploitation of molecular markers in characterizing the variability of Ethiopian sorghums is very limited. The RAPD study of Ayana et al. (2000) based on sorghum samples from different parts of Ethiopia, SSR studies of Weerasooriya et al. (2016) from two areas of Western Ethiopia, and SSR and AFLP studies of sorghums from Eastern Ethiopia by Geleta et al. (2006) are investigations on small samples of accessions. Similar studies covering those of farmers' landraces within specific domains of the diverse sorghum growing conditions and agro-ecologies of the country such as NE Ethiopia would be much more informative. Adugna (2014) evaluated diversity of samples from 4 populations obtained from NE Ethiopia using SSR markers. But the samples, being obtained from four lowland sites, are hardly representative of the agro-ecological as well as landraces diversity of the area. In the present study, we used SSR and ISSR markers to assess the extent and structure of genetic diversity in large samples of sorghum landraces covering a wide area from NE Ethiopia.

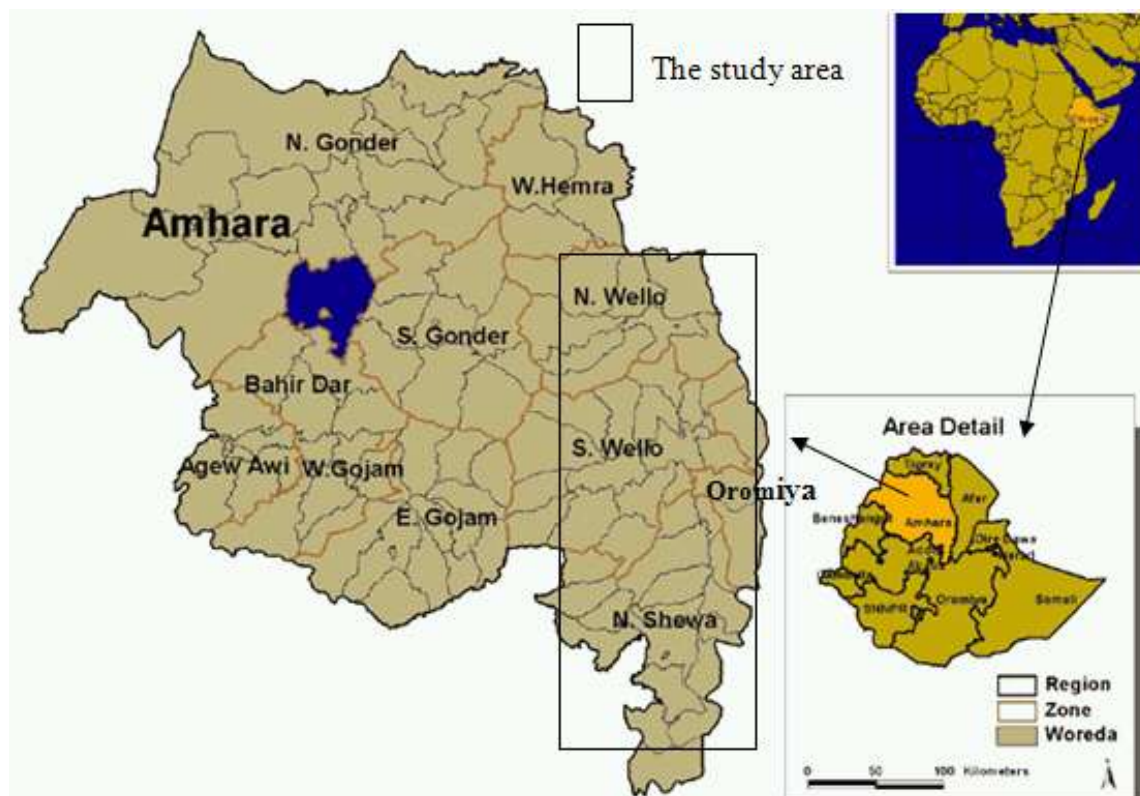
## MATERIALS AND METHODS

### Plant materials and the study area

In total, 415 sorghum landraces were sampled to represent the range of sorghum growing zones and districts of NE Ethiopia. This study area refers to sorghum growing areas of North Welo, South Welo, Oromiya and North Shewa administrative zones in Amhara Regional State (Figure 1). The landraces for the study included 103 from North Welo, 110 from South Welo, 60 from Oromiya and 142 from North Shewa. A total of 13 districts (*woredas*), small administrative units within a zone, were covered with 2 to 5 districts represented from each zone (Table 1). The landraces were collected from farmers' fields and the classification of districts is based on the passport data available for the landraces from the Institute of Biodiversity Conservation (IBC), Ethiopia. Thirty seeds were sampled from each landrace.

### DNA extraction

The sampled seeds were imported to Australia and the study was conducted at the University of Queensland, Brisbane, Australia.



**Figure 1.** Map of Ethiopia and the study area (source: <http://www.usaidethiopia.org/picture.asp>).

**Table 1.** Description of districts from the four administrative zones.

Zones	Districts	Altitude category*
North Welo	Kobo	Lowland
	Gubalafto	Lowland to highland
	Habru	Lowland to intermediate
South Welo	Ambassel	Lowland to intermediate
	Tehuledere	Intermediate to highland
	Dessie zuria	Highland
	Kalu	Lowland to intermediate
Oromiya	Debresina	Highland
	Bati	Lowland
	Artuma Jille	Lowland
North Shewa	Shewa Robit	Lowland
	Tegulet	Intermediate to highland
	Merabete	Lowland to highland

\*Lowland (<1650 m); intermediate (1650-2000 m); and highland (>2000 m) above sea level.

The seeds were planted in a quarantine glasshouse at the. Young leaves from 8 to 10 seedlings (3 to 4 weeks old) pooled together from each landrace were used for DNA extraction following a modified Cetyl-Trimethyl Ammonium Bromide (CTAB) method (Saghai-Marouf et al., 1984) as described in Ritter et al. (2007). DNA concentration and purity was determined both fluorometrically

and agarose gel with appropriate standards.

### SSR analysis

Seven sorghum SSR markers (Brown et al., 1996; Taramino et al., 1997; Dean et al., 1999; Kong et al., 2000) were selected based on their polymorphism, amplified product size range and genome coverage. The M13 primer tailing system (Schuelke, 2000) was used to fluorescently label PCR products. Primers that amplified target sequences with non-overlapping fragment sizes were tailed at the 5'-end with the same M13 sequence. A corresponding M13 primer was labeled with fluorescent dye, either 6-FAM, NED, or VIC (Applied Biosystems, Foster, CA). The M13 primer labeled with 6-FAM was used with three of the primers; NED and VIC each with two primers (Table 2). The sequence specific SSR primers (forward and reverse) were purchased from Prologo, Australia.

PCR amplification was performed for all landraces in a 10 µl reaction volume containing 20 ng of genomic DNA, 1x PCR buffer (Promega, Madison WI), 5 mM MgCl<sub>2</sub>, 0.125 mM dNTPs, 0.25 µM reverse SSR primer, 0.025 to 0.125 µM forward SSR primer, and 0.125 to 0.25 µM M13 primer, 1 unit of Taq DNA polymerase (Promega, Madison WI). Reactions were performed on an MJ Research (PTC 200) Thermal Cycler. After initial denaturation at 94°C for 5 min, the reaction mixture was subjected to 30 cycles of 94°C (30 s), 61°C/57°C/53°C (45 s), 72°C (45 s) followed by 8 cycles of 94°C (30 s), 53°C (45 s), 72°C (45 s) and a final extension of 72°C for 10 min. PCR products for each sorghum line from the seven primers were multiplexed into a single lane on an ABI 3130 Capillary electrophoresis system, with Liz-350 (ABI) as the size standard.

**Table 2.** Description of the three ISSR, and M13 tail sequence and fluorescent dye used with the seven SSR primers.

ID	LG	Reported size range	SSR sequence (5'-3')	M13 sequence (5'-3')***	Dye	T°
<b>SSR markers</b>						
Sb1-1	08	241–277	TCCTGTTTGACAAGCGCTTATA* AAACATCATACGAGCTCATCAATG**	TTTCCCAGTCACGACGTTG	6-FAM	61
Sb4-32	07	172–216	CTCGGCGGTTAGCACAGTCAC* GCCCATAGACAGACAGCAAAGCC**	TTTCCCAGTCACGACGTTG	6-FAM	61
Txp-15	05	215	CACAAACACTAGTGCCTTATC* CATAGACACCTAGGCCATC**	TAAAACGACGGCCAGTGC	NED	57
SbKAFGK-I	07	142-166	AGCATCTTACAACAACCAAT* CTAGTGCACCTGAGTGATGAC**	TAAAACGACGGCCAGTGC	NED	53
sb5-236	10	162–222	GCCAAGAGAAACACAAACA* AGCAATGTATTTAGGCAACACA**	GCGGATAACAATTTACACAGG	VIC	57
sb6-57	03	283–313	ACAGGGCTTTAGGGAATCG* CCATCACCGTCGGCATCT**	GCGGATAACAATTTACACAGG	VIC	57
sbAGA-01	08	88-106	CGAACCATGATAAATGACTG* ATCCGTTTCACAAAAAAGT**	TTTCCCAGTCACGACGTTG	6-FAM	57
<b>ISSR markers</b>						
ISSR 2		310-902	(GA)9C	-	VIC	58
ISSR 6		255-872	(GA)9Y	-	NED	58
ISSR 7		315-912	(AG)9C	-	6-FAM	58

LG = linkage group, \*Forward primer, \*\*Reverse primer, \*\*\*sequence was tailed to 5'-end of the sequence specific forward primer. M13 primer labeled with the respective dyes was purchased from ABI. T° = Annealing temperature.

### ISSR analysis

PCR amplification of inter-simple sequence repeats was performed using three fluorescently labeled primers (Table 2). The PCR mixture was made up by: 20 ng of template DNA, 0.5 µM of primer, 0.25 mM of each dNTP, 1x colourless GoTaq® Flex Buffer (Promega, Madison WI), and 1.25 unit of Taq DNA polymerase (Promega, Madison WI) in a total volume of 20 µl. The thermocycler PTC-200 (MJ Research) was programmed as follows: 95°C for 4 min; 10 cycles of 94°C (45 s), 58 to 1°C per cycle (45 s) and 72°C (90 s); and then 26 cycles of 94°C (45 s), 52°C (45 s) and 72°C (90 s). A final extension at 72°C for 10 min was used. A separate PCR was run for each of the 3 ISSR primers and the PCR products for each sorghum line were multiplexed for a single lane capillary electrophoresis after dilution of 1:4. The genotyping was performed on an ABI 3130 Capillary electrophoresis system, with Liz-1200 (ABI) as the size standard. Each fluorescent peak from the capillary run was treated as an allele (band) and the size scored using GeneMapper software (ver.3.7) (Applied Biosystems, Foster City, CA).

### Statistical analyses

A Bayesian model-based clustering method implemented in the software STRUCTURE 2.2 (Pritchard et al., 2000) was applied to infer genetically differentiated populations or clusters (K) to reflect structuring of sorghum landraces variability in NE Ethiopia other than the predefined populations based on geographic origins. The number of clusters (K) was varied from one to ten with a burn-in of 10000 cycles and 50000 Markov Chain Monte Carlo (MCMC) iterations and 10 replicate runs performed for each K value to determine the appropriate K. The statistic delta K ( $\Delta K$ ) criterion

(Evanno et al., 2005), which is based on the log likelihood of the data [ $\ln Pr(X/K)$ ], was used to estimate K. Admixture ancestry model and correlated population allele frequencies with and without population prior information were used for the parameter set of the analyses. The final run consisted of a burn-in of 100000 cycles and 500000 Markov Chain Monte Carlo (MCMC) iterations and repeated 5 times for appropriate value of K. We also used the prior information ancestry model to identify migrants and hybrids.

Cluster (heatmap) analysis was performed based on Jaccard genetic similarity coefficients (Jaccard, 1908) using Ward's clustering method (Ward, 1963) in R-project statistical package (R Development Core Team, 2010). PowerMaker ver.3.25 (Liu and Muse, 2005) was used to estimate Nei's (1972) genetic distance between zones and between districts for each marker system and MEGA 3.1 (Kumar et al., 2004) was used to draw the dendrograms. Analysis of molecular variance (AMOVA) was performed according to Excoffier et al. (1992) to study the differentiation of landraces among predefined geographic domains such as zones and districts and among populations identified using STRUCTURE. Wright's (1951) F-statistics ( $F_{ST}$ ) was used to measure population structure and it was calculated by the methods of Weir and Cockerham (1984). Arlequin 3.5 population statistical software (Excoffier and Lischer, 2010) was used to estimate all the genetic diversity and differentiation parameters for both SSR and ISSR markers.

Genetic diversity of the landraces was measured at various levels of population structure: Model based clusters, zones and districts. For the SSRs, the mean number of alleles, and expected heterozygosity (gene diversity) at Hardy-Weinberg proportions ( $H_e$ ) (Nei, 1987) were calculated. Allelic richness was also determined based on the rarefaction method using HP-Rare package (Kalinowski, 2005). Use of rarefaction compensates for the effect of sample size on estimates of the number of alleles and number of private alleles in samples. In the case of ISSRs, the total number of

**Table 3.** Genetic diversity within groups of sorghum landraces for SSR & ISSR markers.

Geographic domains	SSR				ISSR			
	A	Ap	Ar	Arp	H <sub>o</sub>	H <sub>e</sub>	b	PIC
<b>NEE</b>	15.1	-	-	-	0.24	0.72	161	0.129
<b>Zones</b>								
North Welo	10.4	2	10.48	1.18	0.25	0.69	135	0.127
South Welo	10.6	2	10.83	0.75	0.19	0.70	144	0.132
Oromiya	8.7	0	9.39	0.35	0.18	0.73	117	0.119
North Shewa	13.3	15	11.87	1.99	0.26	0.69	136	0.127
<b>Model based populations</b>								
Cluster 1	9.6	11	6.82	0.86	0.23	0.62	129	0.118
Cluster 2	11.1	8	6.83	0.70	0.20	0.64	140	0.109
Cluster 3	9.1	5	9.88	1.06	0.25	0.77	118	0.117
Cluster 4	9.6	6	10.8	2.10	0.26	0.66	106	0.140
<b>Districts</b>								
Kobo	8.0	1	7.23	0.16	0.27	0.71	109	0.126
Gubalafto	5.6	1	6.14	0.08	0.29	0.69	71	0.123
Habru	8.3	0	7.16	0.15	0.24	0.65	111	0.125
Ambassel	7.3	2	7.14	0.23	0.26	0.64	90	0.127
Tehuledere	3.7	0	3.71	0.13	0.12	0.63	52	0.134
Dessie zuria	2.1	0	2.14	0.00	0.17	0.52	22	0.091
Kalu	8.1	0	7.15	0.01	0.17	0.66	125	0.125
Bati	3.0	0	3.00	0.00	0.03	0.47	41	0.098
Debresina	2.7	0	4.14	0.05	0.19	0.71	113	0.120
Artuma Jille	8.4	0	7.63	0.25	0.20	0.75	116	0.113
Shewa Robit	11.3	7	7.64	0.63	0.26	0.68	54	0.135
Tegulet	7.7	6	7.57	0.86	0.29	0.73	90	0.156
Merabete	7.3	2	7.14	0.39	0.24	0.64	99	0.137

NEE = North Eastern Ethiopia; A = mean number of alleles per locus; Ap = private alleles, Ar = allelic richness, Arp = private allelic richness, H<sub>o</sub>=Observed heterozygosity, H<sub>e</sub> = expected heterozygosity (genetic diversity); PIC = Polymorphic information Content; b = total number of bands. Allelic richness values for SSR markers were estimated with 150 genes for zones, and 50 genes for districts.

bands and average number of bands per primer and the Polymorphic Information Content (PIC) were calculated. The PIC was estimated according to Botstein et al. (1980) as  $PIC = \sum (1 - P_i^2) / n$ , where  $n$  is the number of band positions analyzed in the landraces,  $P_i$  is the frequency of the  $i^{th}$  pattern.

## RESULTS

### Genetic diversity

The observed values for the diversity parameters are given in Table 3. A total of 106 alleles were amplified from the 7 SSR loci across the landraces with the number of alleles per locus ranging from 8 (Sb6-57) to 26 (Sb1-1), with an average of 15.1 alleles per locus. Observed

heterozygosity among the landraces ranged from 15% (Sb5-236) to 47% (SbAGA-01) with an average of 24% per locus. The genetic diversity (expected heterozygosity) for the entire landraces ranged from 0.50 (sbKAFGK-I) to 0.85 (sb4-32) with mean 0.72. The genetic diversity for the model based populations ranged from 0.62 to 0.77 for SSR markers and from 0.11 to 0.14 for ISSR markers. Partitioning of the observed diversity for SSR markers into geographic origins showed high diversity with means of 0.70 for zones, and 0.66 for districts, while the figures for ISSR markers were 0.22 and 0.20, respectively. The mean number of alleles per locus for SSR markers at the zonal level ranged from 8.7 in Oromiya to 13.3 in North Shewa. Of the total alleles observed, 53 (50%) alleles were shared between the four zones while the remaining alleles were distributed across narrower geographic

domains. On average, 4.75 private alleles (that is, alleles found only in a single subpopulation) were observed per zone with North Shewa sorghums having the highest number of private alleles (15 alleles). There were no private alleles in Oromiya. Overall, high altitude districts such as *Tehuledere*, *Dessie zuria* and *Debresina* had lower allelic diversity compared to low altitude districts like *Kobo*, *Habru*, *Ambassel* and *Shewa Robit*. Allelic diversity was highest in *Shewa Robit* with 11.3 alleles per locus. Adjusted for sample size, districts from North Shewa showed the highest allelic richness followed by Oromiya zone. ISSR markers generated 34 to 87 bands per primer, with 161 bands in total. No significant differences were observed in banding pattern among the four zones. The numbers of bands were smaller for highland districts.

### Population genetic structure and geographic distribution

For both SSR and ISSR markers, using procedures described by Evanno et al. (2005), the maximum  $\Delta K$  occurred at  $K = 4$  for the admixture model with prior population information. Therefore, the final analysis for the assignment of landraces into four ( $K=4$ ) clusters was performed by using population information (collection zones) as prior with admixture ancestry model. The number of clusters ( $K = 4$ ) was consistent with clustering based on genetic distances (Figure 2). The landraces from North Shewa tended to be grouped together both in the STRUCTURE analysis and distance based clustering.

The ancestry coefficient ( $Q$ ) for individual landraces revealed that many of the landraces share significant ancestry from two or more populations. The landraces from North Welo and Oromiya appeared to have mixed ancestry with landraces mainly from South Welo, which is geographically located between North Welo and Oromiya zones. A limited number of landraces from the two zone shared ancestry with landraces from North Shewa. This was more evident for SSR markers (Figure 2). All the landraces were assigned to one of the four populations based on the highest proportion of ancestry even if some of them derive significant ancestry from more than one population. The number of landraces in the 4 populations varied from 36 to 181 for SSR markers, and from 39 to 211 for ISSR markers which accounted for 9 to 51% of the 415 landraces (Table 4). Cluster I included mixture of landraces like *Keteto*, *Cherekit* and *Jetere* many of which are from North Shewa. Cluster II is the largest cluster containing approximately half the landraces (211 based on ISSR markers; 181 based on SSR markers). This group included important landraces such as *Degalet* types (*Chibte*, *Abola*, *Watigela*, *Tengele*) and *Jamyo*, most of which are characterized by white, yellow and straw seed colors, and semi-compact to compact panicle types. These landraces are preferred for making *injera*, the daily

staple of farmers in the area. Cluster III included 86 landraces (ISSR markers) and 103 landraces (SSR markers) most of which are from North Welo. The cluster included landraces such as *Tikureta*, *Tinkish*, *Humera*, *Gomdade* and *Ahyo* which are mainly used for roasted grains, sweet stalk or beverages. Many of these landraces have pinkish and brown seed colors. Cluster IV, containing 36 landraces (for SSR markers) and 39 landraces (for ISSR markers), is characterized mainly by *Zengada* types. These landraces are adapted to high altitude areas like *Dessie Zuria*, *Tehuledere* and *Debre Sina* districts.

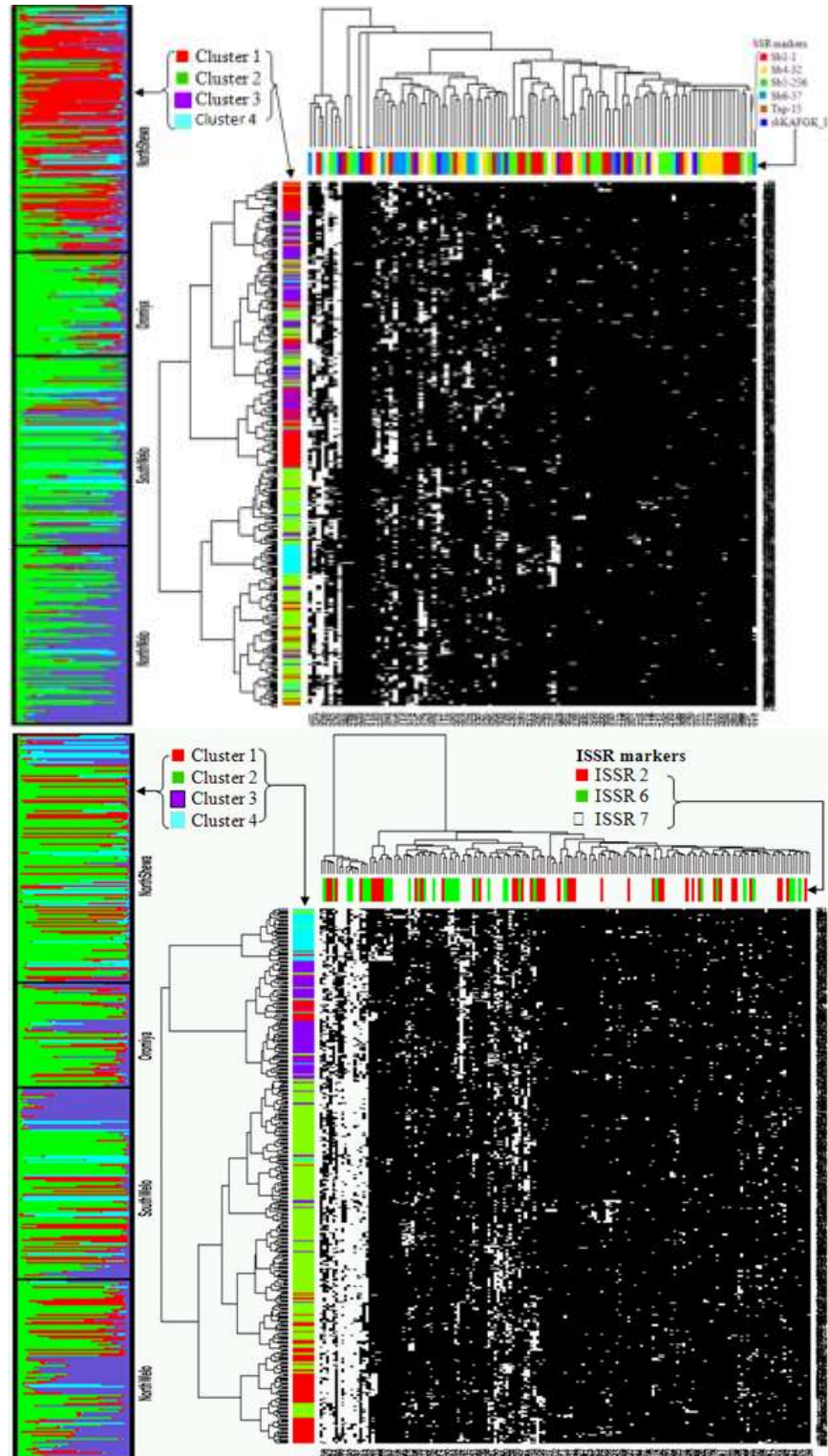
The patterns of individual landrace phylogeny revealed evidence of relationship between genetic distance and geographic distance for both SSR and ISSR markers. Many of the landraces from North Shewa were grouped together, most markedly with SSR markers. The general tendency of landraces to cluster based on geographic proximity can also be observed in the clustering of zones and districts (Figure 3). Most of the districts from low and intermediate altitude areas were clustered together (e.g. *Kobo*, *Habru*, *Ambassel*, *Kalu*) while districts from higher altitude (>2000 m) areas like *Debre Sina*, *Dessie zuria* and *Tehuledere* were grouped together. The grouping of *Shewa Robit*, *Merabete* and *Tegulet*, and that of *Tehuledere* and *Dessie Zuria* appear to reflect geographic proximity. The migrant analysis using STRUCTURE software identified fewer migrants (Supplementary Figure 1).

AMOVA analyses among the sorghum landraces, zones, districts, STRUCTURE populations and distance based clusters showed all variance components to be significant ( $P < 0.001$ ) (Table 5) for both SSR and ISSR markers. The variance at zones level for SSR markers (4%) and for ISSR markers (2.1%) appeared to be small. The values were relatively higher for districts. However, clusters obtained from STRUCTURE and distance based cluster analyses accounted about 12 to 15% of the variability.  $F$ -statistic ( $F_{ST}$ ) was significant but it revealed a small differentiation among predefined populations based on geographic origins, that is, zones and districts. The  $F_{ST}$  values ranged from 0.04 to 0.06 for SSR markers and from 0.02 to 0.05 for ISSR markers.

## DISCUSSION

### Genetic diversity

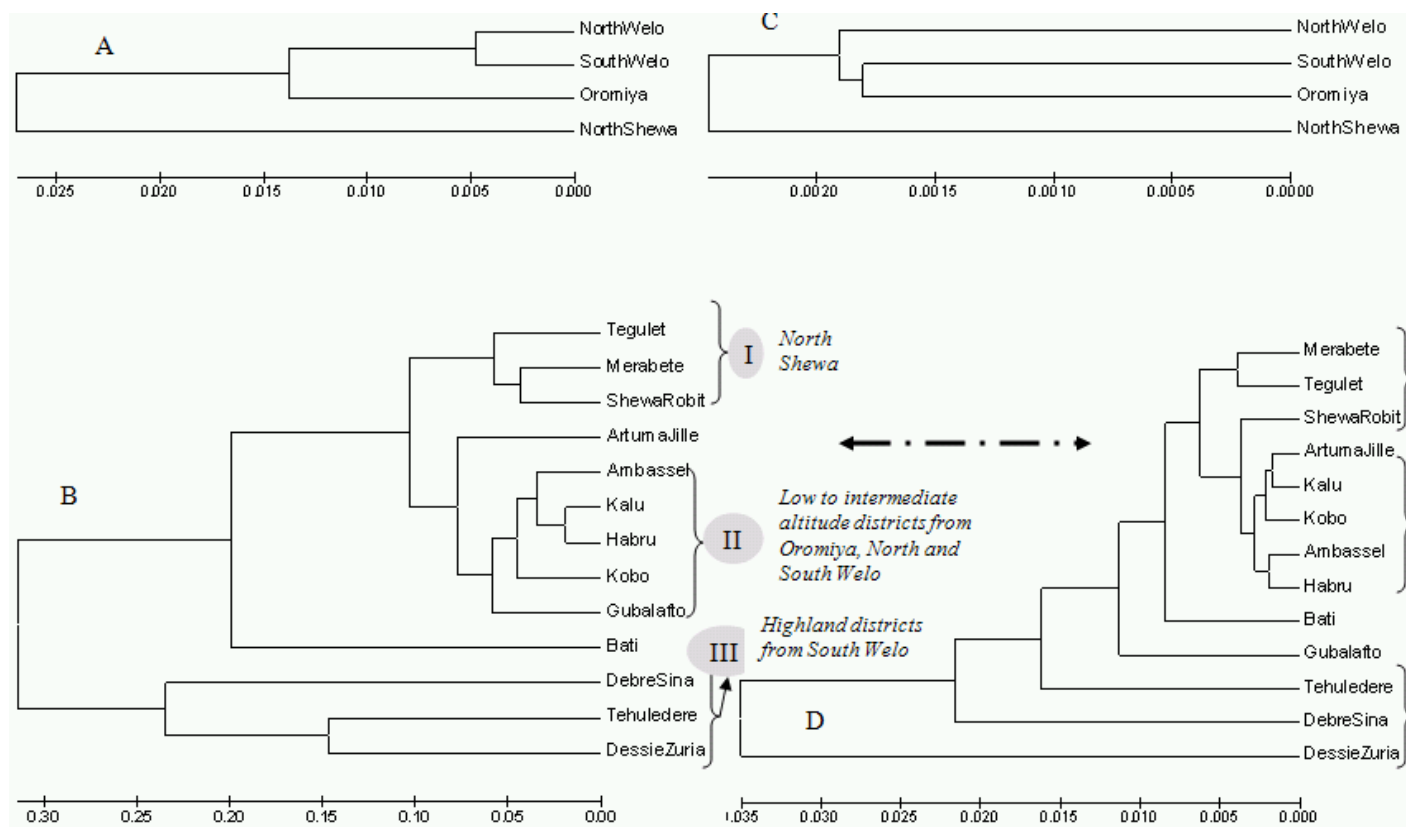
The average number of alleles, 15.1 per locus, for the seven SSR loci was higher than values reported in other studies on sorghum (Brown et al., 1996; Taramino et al., 1997; Dean et al., 1999; Kong et al., 2000; Deu et al., 2010; Weerasooriya et al., 2016). We used a higher number of landraces than previous authors, and this may account for some of these differences. However, this is also a reflection of the diversity in Ethiopia, both arising



**Figure 2.** A clustering profile of landraces from STRUCTURE analysis for  $K = 4$  and genetic distance based heatmap using (A) SSR and (B) ISSR markers. In both cases, geographic (zones of collection) information was used as a prior information in running the STRUCTURE analyses. Each landrace is represented by a line divided into  $K$  coloured fragments proportional to its membership in the corresponding genetic cluster. The names on the side of STRUCTURE graph denote the predefined population names, that is, administrative zones (North Welo, South Welo, Oromiya and North Shewa) from which each landrace is sampled. The landraces on distance based heatmap were colored according to STRUCTURE membership to show the correspondence of the two clustering results, that is, red color on the STRUCTURE histogram corresponds to red on the heatmap (dendrogram) leaf of distance based cluster and so on.

**Table 4.** Distribution of landraces in the four clusters (K = 4) from the 4 administrative zones based on STRUCTURE analyses using SSR and ISSR markers.

STRUCTURE clusters	SSR					ISSR					Prominent landraces
	North Welo	South Welo	Oromiya	North Shewa	Total	North Welo	South Welo	Oromiya	North Shewa	Total	
1	2	6	3	84	<b>95</b>	18	18	10	33	79	<i>Keteto, Jetere, Cherekit</i>
2	45	59	48	29	<b>181</b>	42	52	37	80	211	<i>Degalet</i> types and <i>Jamyo</i>
3	53	30	5	15	<b>103</b>	40	26	10	10	86	<i>Tikureta, Merabete, Ahyo, Tinkish, Moge</i>
4	3	15	4	14	<b>36</b>	3	14	3	19	39	<i>Zengada</i>



**Figure 3.** Dendrograms showing phylogeny of landraces from the 4 zones and 13 districts based on Nei (72) distance calculated from SSR (A, B); and ISSR markers (C, D).



**Table 5.** AMOVA showing the structure of variability in sorghum landraces of NE Ethiopia using SSR and ISSR markers.

Source of variation	Groups	Percentage of variance components		F <sub>ST</sub>
		Among groups	Among landraces	
<b>SSR markers</b>				
Zones	4	3.99	96.01	0.04
Districts	13	5.65	94.35	0.06
STRUCTURE populations	4	13.71	86.29	0.14
Distance based clusters	4	12.30	87.70	0.12
<b>ISSR markers</b>				
Zones	4	2.06	97.94	0.02
Districts	13	4.62	95.38	0.05
STRUCTURE populations	4	14.80	85.20	0.15
Distance based clusters	4	13.10	86.90	0.13

from the fact that this is one of the, if not the major, centre of origin/domestication of sorghum, and the fact that there are diverse end-uses for sorghum in this region. The result for proportions of rare alleles for the SSR markers is in general agreement with values reported for sorghum using SSR markers (Grenier et al., 2000; Dje et al., 2000; Ghebru et al., 2001). The number of bands per ISSR primer in the present study was higher than bands reported in Moroccan sorghums (Medraoui et al., 2007).

The genetic diversity values for both the ISSR and the SSR markers were high for all landraces as well as for landraces in each zone and the model based populations. Other studies on sorghum (Grenier et al., 2000; Dje et al., 2000; Ghebru et al., 2001; Medraoui et al., 2007; Weerasooriya et al., 2016) reported values close to the present study. The high diversity observed in NE Ethiopian sorghum landraces could be attributed to various factors including subsistence farming practice that rely on landraces, introgression with wild and weedy relatives, the geographic and agro-climatic variability of the area affecting adaptability of landraces. The area is characterized by a very diverse topography with valley bottoms and hills creating niche or variable micro-environments of sorghum production. It is also prone to crop failures due to moisture shortage. The diverse landraces are insurance to avert risk of crop failure as well as to meet niche environments. Consequently, farmers of the area have been maintaining invaluable diversity for generations. It has been documented that farmers make conscious decision and management efforts based on agro-ecological condition and end-use to maintain landraces diversity (Teshome et al., 1997; Seboka and van Hintum, 2006). Preferences of different landraces for various end-uses like sweet sorghum (juicy stalk), roasted grain (milky dough stage), local beverages (e.g. *tela*), and daily dishes (e.g. *injera*) affect the selection and maintenance of landraces, which in turn affect genetic diversity. Phenotypic study of a larger

sample of 974 accessions (from which the 415 accessions were subsampled) indicated a high variability for both quantitative and qualitative traits (Desmae et al., 2016)

#### Genetic structure and differentiation of landraces

The model based structure analyses using population information (LOCPRIOR) revealed the presence of 4 populations of sorghum landraces in NE Ethiopia. Many landraces share ancestry with landraces from neighboring geographic origins. The landraces in the 4 model based clusters were significantly differentiated by both ISSR and SSR markers with F<sub>ST</sub> values ranging from 0.12 to 0.15 (Table 5). The model based clusters were consistent with distance based clusters. Odong et al. (2011) reported a close similarity between clusters obtained by Ward's (Ward, 1963) and STRUCTURE (Pritchard et al., 2000).

The important evolutionary factors that affect the extent of population differentiation are gene flow, genetic drift and selection. Significant geographic differentiation was observed among landraces at various levels of predefined geographic origin but a large portion of variation was among landraces within rather than between predefined populations (zones and districts). The low to moderate geographic differentiation could be attributed to frequent gene flow among fields as a consequence of seed exchanges among farmers, and/or to restriction of the intensity of genetic drift due to a high effective population size (Dje et al., 1999). Earlier observations (Shewayyrga et al., 2008) showed that many landraces with the same local name are grown in two or more zones attributed to the existing farmers' seed system. Seeds are shared through gift, exchange in kind, and purchase. Low differentiations between landraces from different geographic regions have been reported in previous studies on sorghum (Dean et al., 1999; Dje et

al., 1999, 2000; Ayana et al., 2000; Ghebru et al., 2001). In addition, the phylogenetic analyses showed a significant trend but it did not reveal a strong association between general clustering pattern of individual landraces and geographic origin. Specifically, moderate to strong clustering in population cluster I was evident for landraces from North Shewa in particular, and this region exhibited both the highest allelic diversity and the highest number of private alleles. In addition, there was clustering together of many landraces from North Welo in population cluster III, whereas the high altitude *Zengada* types were in population cluster IV. The clustering of districts was influenced by agro-ecological similarity and spatial proximity. Agro-ecologically similar and spatially close districts were grouped together. Linhart and Grant (1996) remarked that different environments generate different selection pressures and significant barriers to gene flow, which in turn enhance genetic heterogeneity and differentiation among semi-isolated or isolated populations. The present study area stretches more than 400 km from north to south with North Welo the most northerly and North Shewa the most southerly. It is of significance that some of the highland districts, particularly Tegulet and Merabete are situated on the West (Blue Nile) escarpment while the lowland and intermediate altitude districts are situated on the East (Rift Valley) escarpment. Further, districts from lowland and intermediate altitudes have warmer climates compared to those in higher altitudes. Such distance and climatic factors may enforce physical and adaptive barriers to gene flow.

In summary, from the observed variability, it can be argued that NE Ethiopia is a region with high sorghum landraces diversity. There was indication that some areas, for example North Shewa, had a high level of diversity with more private alleles than other regions. By contrast, the highland areas had lower diversity than the lowland and intermediate areas but there were several alleles unique to the highlands. Overall, however, the diversity of sorghum landraces was widely distributed across NE Ethiopia. This diversity can be exploited for improvement of sorghum in the area through incorporating landraces in the breeding program as parents for traits of interest. Systematic screening of the landraces would be important to identify potential parents. In that case, improved varieties can be developed that meet the needs of farmers, which will ultimately be adopted.

### Conflict of Interests

The authors have not declared any conflict of interests.

### ACKNOWLEDGEMENT

The first author acknowledges Sirinka Agricultural

Research Center and Amhara Region Agricultural Research Institute, Ethiopia for the provision of research facilities, financial support and other necessary assistance throughout the field study, and IBC for provision of germplasm.

### REFERENCES

- 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.
- Ayana A, Bekele E (1998). Geographical patterns of morphological variation in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Ethiopia and Eritrea: qualitative characters. Hereditas 129:195-205.
- Ayana A, Bryngelsson T, Bekele E (2000). Genetic variation of Ethiopian and Eritrean sorghum (*Sorghum bicolor* (L.) Moench) germplasm assessed by random amplified polymorphic DNA (RAPD). Genet. Res. Crop Evol. 47:471-482.
- Borrell A, van Oosterom E, Hammer G, Jordan D, Douglas A (2003). The physiology of "stay-green" in sorghum. In: Solutions for a better environment. Edited by Murray Unkovich and Garry O'Leary. Proceedings of the 11th Australian Agronomy Conference, 2-6 February 2003, Geelong, Victoria.
- Botstein D, White RL, Skolnick MH, Davies RW (1980). Construction of a genetic map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. 32:314-331.
- Brown AHD (1995). The core collection at the crossroads. In: Core Collection of Plant Genetic Resources. Hodgkin T, Brown AHD, van Hintum TJJ and Morales EAV (eds.). IPGRI pp. 3-19.
- Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalezcandelas F, Kresovich S (1996). Multiple Methods for the Identification of Polymorphic Simple Sequence Repeats (SSRs) in Sorghum [*Sorghum bicolor* (L.) Moench]. Theor. Appl. Genet. 93(1-2):190-198.
- Dahlberg JA, Bandyopadhyay R, Rooney WL, Odvody GN, Madera-Torres P (2001). Evaluation of sorghum germplasm used in US breeding programmes for sources of sugary disease resistance. Plant Pathol. 50:681-689.
- Dean RE, Dahlberg JA, Hopkins MS, Mitchell SE, Kresovich S (1999). Genetic Redundancy and Diversity among 'Orange' Landraces in the US National Sorghum Collection as assessed with Simple Sequence Repeat (SSR) Markers. Crop Sci. 39(4):1215-1221.
- Desmae H, Jordan D, Godwin I (2016) Geographic patterns of phenotypic diversity in sorghum (*Sorghum bicolor* (L.) Moench) landraces from North Eastern Ethiopia. Afr. J. Agric. Res. 11(33):3111-3122.
- Dje Y, Heuertz M, Lefebvre C, Vekemans X (2000). Assessment of genetic diversity within and among germplasm landraces in cultivated sorghum using microsatellite markers. Theor. Appl. Genet. 100:918-925.
- Dje Y, Forcioli D, Ater M, Lefebvre C, Vekemans X (1999). Assessing population genetic structure of sorghum landraces from Northwestern Morocco using allozyme and microsatellite markers. Theor. Appl. Genet. 99:157-163.
- Doggett H (1988). Sorghum. 2<sup>nd</sup> 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, Smouse P, Quattro J (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479-491.
- Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564-567.
- Gebrekidan B (1973). The importance of the Ethiopian sorghum germplasm in the world collection. Econ. Bot. 27:442-445.
- Geleta N, Labuschagne MT, Viljoen CD (2006). Genetic diversity analysis in sorghum germplasm as estimated by AFLP, SSR and morpho-agronomical markers. Biodivers. Conserv. 15(10):3251-3265.
- Ghebru B, Schmidt RJ, Bennetzen JL (2002). Genetic diversity of

- Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. *Theor. Appl. Genet.* 105:229-236.
- Godwin ID, Aitken EAB, Smith LW (1997). Application of inter simple sequence repeat (ISSR) markers to plant genetics. *Electrophoresis* 18:1524-1528
- Grenier C, Deu M, Kresovich S, Bramel-Cox PJ, Hamon P (2000). Assessment of Genetic Diversity in Three Subsets Constituted from the ICRISAT Sorghum Collection Using Random vs Non-Random Sampling Procedures, B. Using molecular markers. *Theor. Appl. Genet.* 101:197-202.
- Harlan JR, de Wet JMJ (1972). A simplified classification of cultivated sorghum. *Crop Sci.* 12(2):172-176.
- Hawkes JG, Worede M (1991). Use of Ethiopian germplasm in national and international programmes. In: *Plant genetic resources of Ethiopia* (Engels JMM, Hawkes JG, Worede M (eds.)). Pp.360-368.
- Hayward MD, Sackville-Hamilton NR (1997). Genetic Diversity: Population Structure and Conservation. In: *Biotechnology and plant genetic resources-Conservation and uses*. Callow JA, Ford-Lloyd BV and Newbury HJ (eds.). CAB International. Pp. 49-76
- He J, Zhao X, Laroche A, Lu Z-X, Liu HK, Li Z (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Review Front. Plant Sci.* 5:484.
- Jaccard P (1908). Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.* 44:223-270.
- Kalinowski ST (2005). HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Mol. Ecol. Notes* 5:187-189.
- 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(3):438-448.
- Kumar S, Tamura K, Nei M (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* 5:150-163.
- Linhart YB, Grant MC (1996). Evolutionary significance of local genetic differentiation in plants. *Annu. Rev. Ecol. Syst.* 27:237-77.
- Liu J, Muse SV (2005). PowerMarker: Integrated analysis environment for genetic marker data. *Bioinformatics* 21(9):2128-2129.
- Medraoui L, Ater M, Benhabib O, Msikine D, Filali-Maltouf A (2007). Evaluation of genetic variability of sorghum (*Sorghum bicolor* L. Moench) in northwestern Morocco by ISSR and RAPD markers. *Plant Biol. Pathol.* 330:789-797.
- Nei M (1972). Genetic distance between populations. *Am. Nat.* 106:283-292.
- Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY, USA.
- Odong TL, van Heerwaarden J, Jansen J, van Hintum TJL, van Eeuwijk FA (2011). Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data? *Theor. Appl. Genet.* 123(2):195-205.
- Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M (1998). Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. *Theor. Appl. Genet.* 97:1248-1255
- Prasada Rao KE, Mengesha MH (1987). Morphology and distribution of Zerazera sorghums. *J. Agric. Tradit. Bot. Appl.* 34:51-55.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- R Development Core Team (2010). R: A language and environment for statistical computing. R foundation for statistical computing.
- Ramanatha Rao VR, Hodgkin T (2002). Genetic diversity and conservation and utilization of plant genetic Resources. *Plant Cell Tissue Organ Cult.* 68:1-19.
- Reddy VG, Rao NK, Reddy BS, PrasadaRao KE (2002). Geographic distribution of Basic and Intermediate races in the World Collection of Sorghum Germplasm. *International Sorghum and Millet Newsletter.* 43:15-16.
- Ritter KB, McIntyre CL, Godwin ID, Jordan DR, Chapman SC (2007). An assessment of the genetic relationship between sweet and grain sorghums, within *Sorghum bicolor* ssp. *bicolor* (L.) Moench, using AFLP markers. *Euphytica* 157(1-2):161-176.
- Saghai-Marouf MA, Soliman KM, Jorgenson RA, Allard RW (1984). Ribosomal DNA spacer length polymorphism in barley. Mendelian inheritance, chromosomal location and population dynamics. *Proc. Natl. Acad. Sci. USA* 81:8014-8018.
- Schuelke M (2000). An economic method for fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18:233-234.
- Seboka B, van Hintum T (2006). The dynamics of on-farm management of sorghum in Ethiopia: Implication for the conservation and improvement of plant genetic resources. *Genet. Res. Crop Evol.* 53(7):1385-1403.
- Shewayrga H, Jordan DR, Godwin ID (2008). Genetic erosion and changes in distribution of sorghum (*Sorghum bicolor* L. Moench) landraces in North-Eastern Ethiopia. *Plant Genet. Resour.* 6(1):1-10.
- Smith JSC, Kresovich S, Hopkins MS, Mitchell SE, Dean RE, Woodman WL, Lee M, Porter K (2000). Genetic Diversity among Elite Sorghum Inbred Lines Assessed with Simple Sequence Repeats. *Crop Sci.* 40:226-232.
- Stemler ABL, Harlan JR, de Wet JMJ (1977). The sorghums of Ethiopia. *Econ. Bot.* 31:446-460.
- Taramino G, Tarchini R, Ferrario S, Lee M, Pe' ME (1997). Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. *Theor. Appl. Genet.* 95:66-72.
- Tautz D, Renz M (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.* 12(12):4127-4138.
- Teshome A, Baum BR, Fahrig L, Torrance JK, Arnason TJ, Lambert JD (1997). Sorghum (*sorghum bicolor* (L.) Moench) landrace variation and classification in North Shewa and South Welo, Ethiopia. *Euphytica* 97:255-263.
- Teshome A, Fahrig L, Torrance JK, Lambert JD, Arnason TJ, Baum BR (1999). Maintenance of Sorghum (*Sorghum bicolor*, Poaceae) Landrace Diversity by Farmers' Selection in Ethiopia. *Econ. Bot.* 53(1):79-88.
- Uptmoor R, Wenzel W, Friedt W, Donaldson G, Ayisi K, Ordon F (2003). Comparative Analysis on the Genetic Relatedness of *Sorghum bicolor* landraces from Southern Africa by RAPDs, AFLPs and SSRs. *Theor. Appl. Genet.* 106(7):1316-1325.
- Vavilov NI (1951). The origin, variation, immunity and breeding of cultivated plants. *Chron. Bot.* 13:1-366.
- Vos P, Hogers R, Bleeker M, Reijmans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995). AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* 23:4407-4414.
- Ward JH (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58(301):236-244.
- Weerasooriya DK, Maulana FR, Bandara AY, Tifessa A, Ayana A, Mengistu G, Nouh K, Tesso TT (2016). Genetic diversity and population structure among sorghum (*Sorghum bicolor*, L.) germplasm collections from Western Ethiopia. *Afr. J. Biotechnol.* 15(23):1147-1158.
- Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358-1370.
- Wen L, Tang HV, Chen W, Chang R, Pring DR, Klein PE, Childs KL, Klein RR (2002). Development and mapping of AFLP markers linked to the sorghum fertility restorer gene *rf4*. *Theor. Appl. Genet.* 104:577-585
- Worede M (1992). Ethiopia: a genebank working with farmers. In: Cooper D, Vellve R and Hobbelink H (eds) *Growing Diversity*. London: Intermediate Technology Publications, Pp. 78-94.
- Wright S (1951). The genetical structure of populations. *Ann. Eugen.* 15:323-354
- Zegeye A (1997). Acceptability of injera with stewed chicken. *Food Qual. Prefer.* 8(4):293-295.
- Zietkiewicz E, Rafalski A, Labuda D (1994). Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20:176-183.