

Genetic diversity and population structure of sorghum [*Sorghum bicolor* (L.) Moench] in Ethiopia as revealed by microsatellite markers

Bethlehem Nemera, Mulugeta Kebede, Muluken Enyew & Tileye Feyissa

To cite this article: Bethlehem Nemera, Mulugeta Kebede, Muluken Enyew & Tileye Feyissa (2022) Genetic diversity and population structure of sorghum [*Sorghum bicolor* (L.) Moench] in Ethiopia as revealed by microsatellite markers, Acta Agriculturae Scandinavica, Section B — Soil & Plant Science, 72:1, 873-884, DOI: [10.1080/09064710.2022.2117078](https://doi.org/10.1080/09064710.2022.2117078)

To link to this article: <https://doi.org/10.1080/09064710.2022.2117078>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 11 Sep 2022.



[Submit your article to this journal](#)



Article views: 356



[View related articles](#)



[View Crossmark data](#)

Genetic diversity and population structure of sorghum [*Sorghum bicolor* (L.) Moench] in Ethiopia as revealed by microsatellite markers

Bethlehem Nemera^a, Mulugeta Kebede ^b, Muluken Enyew^c and Tileye Feyissa^c

^aApplied Biology Department, School of Applied Natural Sciences, Adama Science and Technology University, Adama, Ethiopia; ^bDepartment of Plant Biology and Biodiversity Management, College of Natural and Computational Sciences, Addis Ababa University, Addis Ababa, Ethiopia; ^cInstitute of Biotechnology, Addis Ababa University, Addis Ababa, Ethiopia

ABSTRACT

Ethiopia is the center of origin of sorghum (*Sorghum bicolor* L.). Understanding of the genetic diversity of the species is fundamental to design appropriate conservation and management strategies. The present study addressed the genetic diversity of sorghum accessions collected from major growing regions of Ethiopia. Eighty sorghum accessions representing five populations namely Amhara, Dire Dawa, Oromia, Southern Nations, Nationalities and Peoples (SNNP) and Tigray were analyzed with 11 simple sequence repeat (SSR) markers. Analysis of molecular variance (AMOVA) was conducted to evaluate genetic variation within and among populations. The unweighted neighbour-joining-based cluster analysis, principal components of analysis (PCoA) and structure analysis were done to elucidate clustering of populations. The polymorphic information contents (PIC) ranges from 0.50 to 0.86. A high within-populations genetic diversity was confirmed with gene diversity values ranging from 0.51 to 0.77. AMOVA revealed 93.26% of the total genetic variation within populations and 6.74% among populations. Cluster analyses did not show clear grouping of accessions according to their geographical origins, confirming gene flow ($N_m = 6.65$) among populations. In conclusion, the SSR markers used were polymorphic and highly informative Oromia and Amhara populations displayed genetic diversity greater than mean value of 0.67 suggesting possible target populations for breeding and conservation.

ARTICLE HISTORY

Received 10 April 2022
Accepted 22 August 2022

KEYWORDS

AMOVA; Fst; gene flow; popgene; population admixture

Introduction

Sorghum (*Sorghum bicolor* (L.) Moench) belongs to family Poaceae (De Wet 1978). Sorghum originated in north-eastern Africa and then spread to different regions all around the world (De Wet and Harlan 1971). As reported by Vavilov (1951), Ethiopia is the center of origin of sorghum. *S. bicolor* includes all cultivated sorghums as well as semi-wild and wild plants (Mutegi et al. 2011). Both the cultivated and wild varieties are diploids ($2n = 2x = 20$) with a genome size of ~730 Mbp (Paterson et al. 2009). The cultivated sorghum has been classified into five races: Durra, Bicolor, Caudatum, Kaffir and Guinea (Harlan and De Wet 1972). Except Kaffir all are found in Ethiopia (Teshome et al. 1997; Girma et al. 2020).

Sorghum is the fifth in the order of cereal crops global production (Cuevas et al. 2014). According to FAOSTAT (2017), United States was the world's largest producer of sorghum. Africa is the world's regional leader in

total production of sorghum. Ethiopia is the third largest producer of sorghum in Africa behind Nigeria and Sudan (Wani et al. 2011). In Ethiopian agriculture, sorghum holds the third largest share of total cereal production after teff and maize. It is a very important crop for food and feed (Tariq et al. 2014) as well as widely used as biofuel and fibre (Murray et al. 2008; Elangovan et al. 2012; Disasa et al. 2017).

Several production constraints including drought, striga, insect pest, diseases and low-yielding local cultivars affect the productivity of sorghum (Amelework et al. 2016). To address these problems, it is important to have the knowledge of genetic variability of a crop for efficient selection process (Yaqoob et al. 2015). The application of molecular markers is more effective to assess genetic diversity as well as for selection of traits of interest in breeding programmes (Smith and Smith 1992). Molecular markers, such as simple sequence repeats (Tautz 1989) have great potential in

CONTACT Mulugeta Kebede  mulugeta.kebedet@aau.edu.et; kmulugetak@yahoo.com  Department of Plant Biology and Biodiversity Management, College of Natural and Computational Sciences, Addis Ababa University, P.O. Box 3434, Addis Ababa, Ethiopia

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/09064710.2022.2117078>.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

detecting genetic diversity and relationships of crops. Microsatellite is preferred for many sorghum genomics and molecular breeding applications due to its technical simplicity, high throughput level and potential for automation (Missihoun et al. 2015).

In sorghum, limited studies using morphological-based genetic diversity analysis have been carried out using sweet sorghum and wild sorghum germplasms (Adugna et al. 2013). Protein (allozyme) studies have been applied to the geographic and altitudinal variation among landraces collected from Ethiopia and Eritrea (Ayana et al. 2001). Some genetic diversity studies have been done on selected sorghum genotypes at the molecular level using DNA markers such as RFLP (Menz et al. 2004), RAPD (Ayana et al. 2000), AFLP (Geleta et al. 2006), ISSR (Tadesse and Feyissa 2013), SSR (Tirfessa et al. 2020), expressed sequence tags (EST) (Ramu et al. 2009) and single nucleotide polymorphism (SNP) markers (Khangura 2019; Nida et al. 2019; Mengistu et al. 2020; Enyew et al. 2022; Lee et al. 2022). Some of the genomic diversity studies of sorghum were on a global scale (Hu et al. 2019; Lasky et al. 2015; Morris et al. 2013). Others were at regional scale such as Ethiopian sorghum landraces (Girma et al. 2019; Menamo et al. 2021; Wondimu et al. 2021) and West African sorghum panel (Faye et al. 2021). In addition, sorghum nested association mapping studies were executed using an Ethiopian landrace as recurrent parent (Birhan et al. 2022; Dong et al. 2022). Recent studies reported that Ethiopian sorghum germplasm is genetically diverse with high levels of admixture (Girma et al. 2020; Menamo et al. 2021).

However, these studies differ in their power of genetic resolution, quality of information content and extents of polymorphism. In this study, we used SSR makers, which are polymorphic loci, present in DNA that consist of repeating units of one to six base pairs. SSR makers are known for their hyper-variability, reproducibility, co-dominant nature, locus-specificity and random genome-wide distribution (Jonah et al. 2011; Jiang 2013). Therefore, the aim of this study was to assess the genetic diversity of sorghum collected from major growing areas of Ethiopia using SSR markers.

Materials and methods

Plant materials

Eighty accessions of sorghum genotypes (additional file 1) initially collected from major Sorghum growing regions of Ethiopia were obtained from Ethiopian Biodiversity Institute (EBI). Based on their source of origin/

region, they were categorised into five populations: Amhara, Dire Dawa, Oromia, Southern Nations, Nationalities and Peoples (SNNP) and Tigray. Seeds representing each accession were planted at Melkassa Agricultural Research Center (MARC). From each accession, healthy young leaves were collected from two-week-old plants and put in plastic bag containing silica gel to dry and make it ready for genomic DNA extraction.

DNA extraction

Genomic DNA was extracted from young leaves using plant DNA extraction protocol described in Diversity Array Technology (DArT, www.diversityarrays.com) with some modifications. DNA quality was checked by loading 3 µl of DNA mixed with 2 µl of 6x loading dye with gel red on a 1% agarose gel and separated at 100 V for 40 min. The extracted genomic DNA was quantified using a Nano-Drop spectrophotometer (ND-8000, 8-sample Spectrophotometer) and stored at -20°C until used for analyses.

PCR and gel electrophoresis

Fifteen SSR primer-pairs, developed by Billot et al. (2012), were used for initial screening for amplification, polymorphism and specificity to target loci. According to the manufacturer's instruction, the primers were dissolved using nuclease-free water to a final concentration of 100 mol/µl. Out of the 15 screened primer pairs, 11 of them amplified the genomic DNA and showed polymorphism across tested sorghum accessions.

DNA amplification was performed in 12.5 µl reaction volume containing 6.25 µl One Taq[®] 2X Master Mix with Standard Buffer, 2 µl 50 ng/µl template DNA, 2 µl forward and reverse primers, 0.25 µl DMSO (Dimethyl sulfoxide) and 2 µl nuclease-free water. Touch-down PCR amplification was carried out using BIO-RAD T100 Thermal cycler. The PCR programme for the primer pairs (Table 1) consisted of 15 min at 94°C for the initial denaturation followed by 12 cycles of 94°C for 30 s, 60°C for 45 s (ramp of 1°C per cycle) and 72°C for 1 min, then by 31 cycles of 94°C for 30 s, 38°C for 45 s and 72°C for 1 min, a final extension of 20 min at 72°C and then put on hold at 4°C. The PCR amplified products were separated in 3% agarose gel electrophoresis by loading 5 µl of each of the PCR product mixed with 2 µl of gel red using 1 × TAE buffer at 100 V for three hours. The product was visualised under gel documentation system (BioDoc-It[™] imaging system) and subsequently photographed. A 100 bp size molecular marker was used to estimate the size of the amplified products.

Table 1. Characteristics of the 11 SSR primers used in the study.

No	Primer name	Chromosome	DNA Motif	Primer sequence	Fragment size range (bp)
1	Xtxp021	4	(AG) ¹⁸	F: GAGCTGCCATAGATTTGGTGC R: ACCTCGTCCCACCTTTGTTG	157–214
2	Msbcir246	5	(CA) ⁷ +(GA) ⁵	F: TTTTGTGCACTTTTGAGC R: GATGATAGCGACCACAAATC	88–108
3	Xtxp141	7	(GA) ²³	F: TGATGGCCTAGCTTATCT R: CAACAAGCCAACCTAAA	139–189
4	Gpsb067	8	(GT) ¹⁰	F: TAGTCCATACACCTTTCA R: TCTCTCACACACATTCTC	163–180
5	Xtxp057	9	(GT) ²¹	F: GGAACTTTTGACGGGTAGTGC R: CGATCGTGATGTCCAATC	209–237
6	Xtxp096	2	(GA) ²⁴	F: GCTGATGTCATGTTCCCTCAC R: CATTCTGGACTCTGTCGG	161–227
7	Xtxp217	10	(GA) ²³	F: GGCTCGACTACGGAGTT R: TCGGCATATTGATTGGTTT	144–219
8	Xcup53	1	(TTTA) ⁵	F: GCAGGAGTATAGGCAGAGGC R: CGACATGACAAGCTCAAACG	187–196
9	Xcup02	6	(GCA) ⁶	F: GACGCAGCTTTGCTCCTATC R: GTCCAACCAACCCACGTATC	188–197
10	Xtxp321	8	(GT) ⁴ +(AT) ⁶ +(CT) ²¹	F: TAACCAAGCTGAGCATAAGA R: CCCATTACACATGAGACGAG	180–200
11	Msbcir276	3	(AC) ⁹	F: CCCCAATCTAACTATTGGT R: GAGGCTGAGATGCTCTGT	221–238

Data scoring and analysis

The PCR products/bands were scored using PyElph 1.4 software (Pavel and Vasile 2012). Genetic diversity and population structure analyses were carried out on the basis of the scored bands. Locus based diversity indices including major allele frequency (MAF), the number of allele (N_a), gene diversity (GD) and Polymorphic information contents (PIC) were analyzed using Power marker version 3.25 software (Liu and Muse 2005). Effective number of alleles (N_e), Shannon's Information index (I), Population differentiation (F_{st}) and Gene flow (N_m) were determined using POPGENE version 1.32 (Yeh et al. 1999).

Population's genetic pattern; number of alleles (N_a), number of private alleles (NPA), expected heterozygosity (H_e), Percentage of polymorphic loci (PPL) and estimate of the deviation from Hardy–Weinberg equilibrium (HWE) were computed using GenALEx version 6.5 software (Peakall and Smouse 2012). The same software was used to compute pairwise population genetic distances and gene flow, and to perform the genetic differentiation test (F_{st}) over 999 bootstrap replications.

Analysis of molecular variance (AMOVA) and estimate of the variance components were conducted using Arlequin version 3.5.2.2 (Excoffier and Lischer 2010). Gene flow (N_m) among populations was estimated using the formula,

$$N_m (\text{Haploid}) = [(1 / F_{st}) - 1] / 2$$

where F_{st} = the variance among populations/total genetic variations.

A genetic dissimilarity matrix was analyzed using Neighbour Joining (NJ) method and Nei's standard genetic distance (DST, corrected) (Nei 1972) based Unweighted Pair Group Method with Arithmetic Mean (UPGMA) trees were generated using DARwin version 6.0 (Perrier and Jacquemoud-Collet 2006) and Power Marker version 3.25 (Liu and Muse 2005), respectively.

Principal Coordinate Analysis (PCoA) was performed using GenALEx version 6.5 (Peakall and Smouse 2012) software package. Population structure and admixture patterns were determined using the admixture model

based on Bayesian algorithm implemented in STRUC-TURE software version 2.3.3 (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. To estimate the true number of population cluster (K), a burn-in period of 100,000 was used in each run, and data were collected over 250,000 Markov Chain Monte Carlo (MCMC) replications for $K=1$ to $K=10$ using 20 iterations for each K . The optimum K value was predicted following the simulation method of Evanno et al. (2005), using the web-based STRUCTURE HARVESTER version 0.6.92 (Earl 2012). A bar plot for the optimum K was determined using Clumpak beta version (Kopelman et al. 2015).

Results

SSR markers and allelic diversity

The 11 SSR markers generated 70 alleles, across all 80 sorghum accessions with an average of 6.36 alleles per marker. The analysis showed that the MAF ranged from 0.20 (Xtxp217) to 0.54 (Xcup53) with mean frequency of 0.31 per locus (Table 2). The lowest MAF (0.20) and the highest number of alleles (9), effective number of alleles (7.80), Shannon's information index (2.13), gene diversity (0.87) and expected heterozygosity (0.77) were obtained for marker Xtxp217 (Table 2). Conversely, the highest MAF (0.54) and the lowest number of alleles (3), effective number of alleles (2.39), Shannon's Information Index (0.96), gene diversity (0.58) and expected heterozygosity (0.56) were obtained for marker Xcup53 (Table 2). The highest genetic differentiation (0.21) and the lowest gene flow (0.91) were obtained for marker Msbcir276, whereas the lowest genetic differentiation (0.06) and the highest gene flow (3.70) were obtained for marker Xcup02 (Table 2).

The PIC value for the SSR loci ranged from 0.50 for marker xcup53 to 0.86 for marker Xtxp217 with a mean value of 0.75. With regard to test for the HWE,

Table 2. The genetic diversity indices summary of the 11 SSR loci across five populations of Sorghum in Ethiopia.

Marker	MAF	Na	Ne	I	GD	He	Fst	Nm	PIC	P _{HWE}
gps067	0.40	5.0	3.95	1.50	0.75	0.68	0.10	2.15	0.71	0.000***
msbcir246	0.21	6.0	5.77	1.77	0.83	0.69	0.15	1.36	0.80	0.000***
xcup02	0.43	4.0	2.80	1.15	0.64	0.60	0.06	3.70	0.56	0.000***
xcup53	0.54	3.0	2.39	0.96	0.58	0.56	0.08	2.87	0.50	0.000***
txxp141	0.33	7.0	4.63	1.71	0.78	0.58	0.20	0.98	0.75	0.000***
txxp021	0.24	6.0	5.34	1.72	0.81	0.69	0.14	1.55	0.78	0.000***
txxp057	0.21	9.0	7.22	2.07	0.86	0.74	0.11	2.09	0.85	0.000***
txxp096	0.23	8.0	6.41	1.96	0.84	0.71	0.14	1.53	0.82	0.000***
txxp217	0.20	9.0	7.80	2.13	0.87	0.77	0.11	2.04	0.86	0.000***
txxp321	0.24	7.0	5.50	1.80	0.81	0.74	0.10	2.18	0.80	0.000***
msbcir276	0.35	6.0	4.70	1.67	0.79	0.59	0.21	0.91	0.76	0.000***
Mean	0.31	6.36	5.14	1.68	0.78	0.67	0.13	1.66	0.75	

Note: Where MAF = Major allele frequency, Na = Number of alleles, Ne = Effective number of alleles, I = Shannon's Information Index, GD = Gene diversity, He = Expected heterozygosity, Fst = Inbreeding coefficient within subpopulations relative to total (genetic differentiation among subpopulations), Nm = gene flow estimated from $F_{st} = 0.25 (1 - F_{st}) / F_{st}$, PIC = Polymorphic information content, P_{HWE} = P-value for deviation from HWE, ns = not significant, * = $P < 0.0001$ and hence highly significant.

all the 11 SSR markers showed highly significant ($p < 0.0001$) deviation from HWE (Table 2).

Genetic diversity within and among populations

The overall genetic diversity estimates within the populations had a number of alleles ranged from 2.64 to 6.18 with a mean value of 4.47 (Table 3). Oromia population showed the highest number of alleles (6.18), effective number of alleles (4.89), Shannon's information index (1.63) and expected heterozygosity (0.77). In the contrary, the SNNP population showed the lowest number of alleles (2.64), effective number of alleles (2.32), Shannon's information index (0.84) and expected heterozygosity (0.51). Except for the population of Amhara where number of private alleles is 0.09, there was no private allele unique to a single population. The percentage of polymorphic loci per population ranged from 90.91% (SNNP) to 100% (Oromia, Amhara, Tigray and Dire Dawa) with a mean of 98.18% (Table 3).

Genetic differentiation, distance and gene flow between populations

The pair-wise genetic differentiation between the populations ranged from 0.02 to 0.16 (Table 4). The

highest population differentiation was observed among Tigray and SNNP ($F_{st} = 0.16$), followed by Dire Dawa and SNNP ($F_{st} = 0.15$). The lowest population differentiation ($F_{st} = 0.02$) was observed among Oromia and Amhara (Table 6). The Pairwise Nei's genetic distance and gene flow (Nm) of each population from the other populations ranged from 0.11 to 0.74 and 0.94 to 16.68, respectively (Table 3). The highest measure of genetic distance (0.74) with the lowest gene flow (0.94) was observed between the populations of Tigray and SNNP (Table 5). The second-highest genetic distance (0.62) was observed between the populations of Amhara and SNNP. The lowest measure of genetic distance (0.11) with highest gene flow (16.68) was observed between the populations of Oromia and Amhara (Table 5). The second lowest gene flow (1.27) was observed between the populations of Dire Dawa and SNNP.

Table 4. Estimate of population genetic differentiation measured by Fst (below the diagonal) among different geographic regions.

	Amhara	Dire Dawa	Oromia	SNNP	Tigray
Amhara	0.00				
Dire Dawa	0.07	0.00			
Oromia	0.02	0.05	0.00		
SNNP	0.13	0.15	0.12	0.00	
Tigray	0.04	0.09	0.05	0.16	0.00

Table 3. Allelic patterns and diversity indices across populations averaged over the 11 SSR loci.

Population	N	Na	Ne	NPA	I	He	NPL	PPL
Amhara	28.00	6.00	4.71	0.09	1.58	0.75	11	100.00%
Dire Dawa	5.00	3.36	2.97	0.00	1.12	0.64	11	100.00%
Oromia	28.00	6.18	4.89	0.00	1.63	0.77	11	100.00%
SNNP	5.00	2.64	2.32	0.00	0.84	0.51	10	90.91%
Tigray	14.00	4.18	3.17	0.00	1.23	0.67	11	100.00%
Mean	16.00	4.47	3.61	0.02	1.28	0.67		98.18%

Note: Where N = Number of samples, Na = Number of alleles, Ne = Effective number of alleles, NPA = Number of Private Alleles, I = Shannon's Information Index, He = Expected heterozygosity, NPL = Number of polymorphic loci, PPL = Percentage of polymorphic loci.

Table 5. Pairwise Nei's genetic distance (below diagonal) and gene flow (Nm) values (above diagonal) among five sorghum populations from Ethiopia.

	Amhara	Dire Dawa	Oromia	SNNP	Tigray
Amhara	0.00	3.04	0.11	1.34	3.47
Dire Dawa	0.42	0.00	0.27	0.58	0.44
Oromia	16.68	5.86	0.00	1.46	3.34
SNNP	0.62	1.27	0.59	0.00	0.74
Tigray	0.28	2.30	0.30	0.94	0.00

Analysis of molecular variance

AMOVA showed that variability among populations and within populations accounted for 6.74% and 93.26% of the total genetic variations, respectively. The overall fixation index value, used as a measure of population differentiation was moderate ($F_{st} = 0.07$) with high gene flow (≥ 6.65) (Table 6).

Cluster analysis and population structure

The neighbour joining-cluster analysis divided the 80 accessions into four main clusters (C1, C2, C3 and C4) (Figure 1). The first cluster (C1) consisting of 22.5% from different populations (7.5% from Amhara, 3.75% from Dire Dawa, 5% from Oromia, 1.25% from SNNP and 5% from Tigray). The second cluster (C2) consisting of 13.75% from different populations (5% from Amhara, 1.25% from Dire Dawa, 6.25% from Oromia and 1.25% from SNNP). The third cluster (C3) consisting of 12.5% from different populations (2.5% from Amhara, 5% from Oromia and 5% from Tigray). The fourth cluster (C4) consisting of 51.25% from different populations (20% from Amhara, 1.25% from Dire Dawa, 18.75% from Oromia, 3.75% from SNNP and 7.5% from Tigray). UPGMA clustered the five populations into two major clusters (I and II) (Figure 2).

The genetic relatedness among sorghum genotypes was further investigated using principal coordinate analysis (PCoA) (Figure 3). PCoA revealed 29.69% of variations explained by the first three axes. The first, second and third principal coordinate axes showed 12.13%, 9.07% and 8.49%, respectively.

The population structure analysis revealed that maximum delta K becomes a sharp peak at $K = 2$ (Figure 4(a)), suggesting two clusters. Based on this value, Clumpak result (bar plot) detected a genetic

admixture and hence there was no clear geographic origin-based structuring of populations (Figure 4(b)).

Discussion

In this study, 11 SSR markers amplified 70 alleles. The number of alleles per locus ranged from three for Xcup53 to nine for Xtxp057 and Xtxp217 with a mean of 6.36 alleles per locus (Table 2). Mofokeng et al. (2014), have reported a total of 306 alleles with a mean of 6.4 alleles per locus with a range of 2–15 alleles using 30 SSR markers in 103 sorghum genotypes. However, the study by Cuevas and Prom (2013) on population structure and diversity of 137 Ethiopian sorghum germplasm conserved at USDA-ARS National Plant Germplasm System showed 14 alleles per locus. The observed differences in number of alleles could be attributed to the size of the study population, number of accession and types of SSR markers used. The mean gene diversity of the markers observed in this study (0.78) was similar to that previously reported by Thudi and Fakrudin (2011) (0.8) among rabi sorghum genotypes. Ng'uni et al. (2011) reported the mean gene diversity value of 0.53, which is lower than the value obtained in the current study.

The high level of gene diversity of SSR markers observed in this study was probably associated with the genetic diversity in the sorghum accession that represented different geographic origins. The polymorphic information content (PIC) values are usually calculated to assess the level of polymorphism of a marker. As described by Botstein et al. (1980), a marker with PIC values below 0.25 is slightly informative, a marker with PIC between 0.25 and 0.5 is considered as reasonably informative, and those greater than 0.5 are said to be highly informative. In the present study, all the 11 SSR loci used were highly informative since they had PIC values greater than 0.5. The PIC of loci ranged from 0.5 to 0.86 with a mean of 0.75. This result is fairly similar to the mean PIC value 0.78 reported in sorghum genetic diversity study by Cuevas and Prom (2013). However, the observed PIC value is higher than that of Muui et al. (2016), who reported 0.49. High PIC values indicate the high discrimination ability of the selected SSR markers for the studied sorghum genotypes.

Table 6. Analysis of molecular variance of 80 sorghum accessions genotyped using 11 SSR markers.

Source of variation	Degree of freedom	Sum of squares	Variance components	Percentage of variation	Fixation index
Among populations	4	50.41	0.30	6.74	Fst: 0.07
Within populations	155	635.84	4.10	93.26	
Total	159	686.25	4.40	100	
Nm (haploid) populations	6.65				

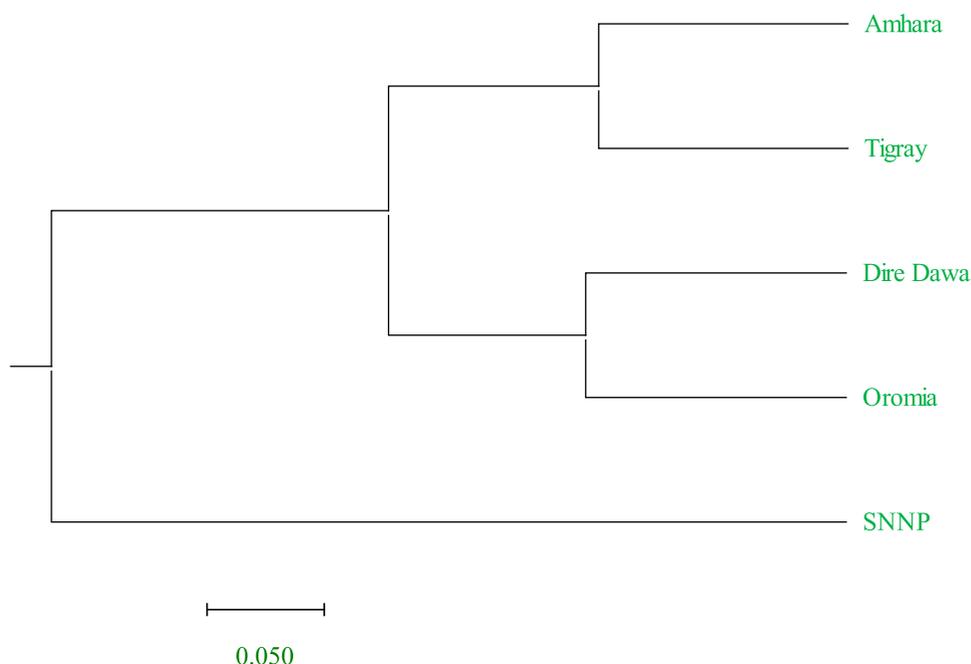


Figure 2. Unweighted pair group method with arithmetic mean dendrogram showing genetic relationships among the five populations of sorghum.

which showed the approximate number of gene flow from one population to the other. According to Slatkin (1985) and Waples (1987), N_m values grouped into three categories: 0.000–0.249 low, 0.25–0.99 intermediate and $N_m > 1.00$ high. However, seed exchange among neighbouring sorghum growing regions and high gene flow ($N_m = 6.65$) could also contribute to the observed low variation among populations. This indicates geographical locations of the populations

have no significant effect on genetic variation of sorghum accessions under study.

The unweighted neighbour-joining-based cluster (Figure 2) and PCoA analysis (Figure 3) did not reveal clear grouping of accessions according to their geographical origins possibly confirming the existence of high gene flow among populations. The clustering pattern is weak enough to support the concept of separation by distance. This is probably due to gene flow

Principal Coordinates (PCoA)

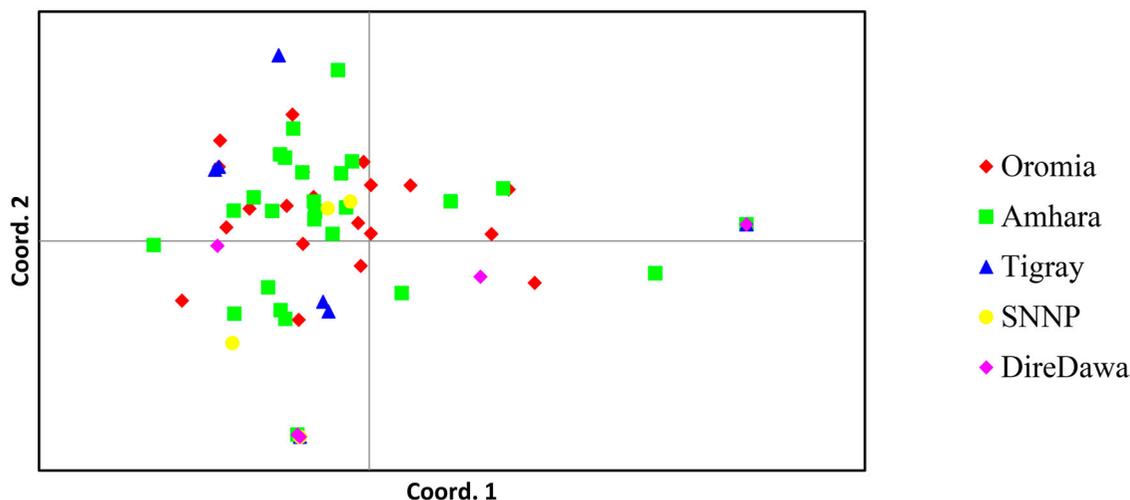


Figure 3. Principal coordinate analysis of 80 sorghum accessions using 11 SSR markers. Samples coded with the same symbol and colors belong to the same population. The percentages of variation explained 29.69% of the total variations and the first 3 axes (1, 2 and 3) are 12.13%, 9.07% and 8.49%, respectively.

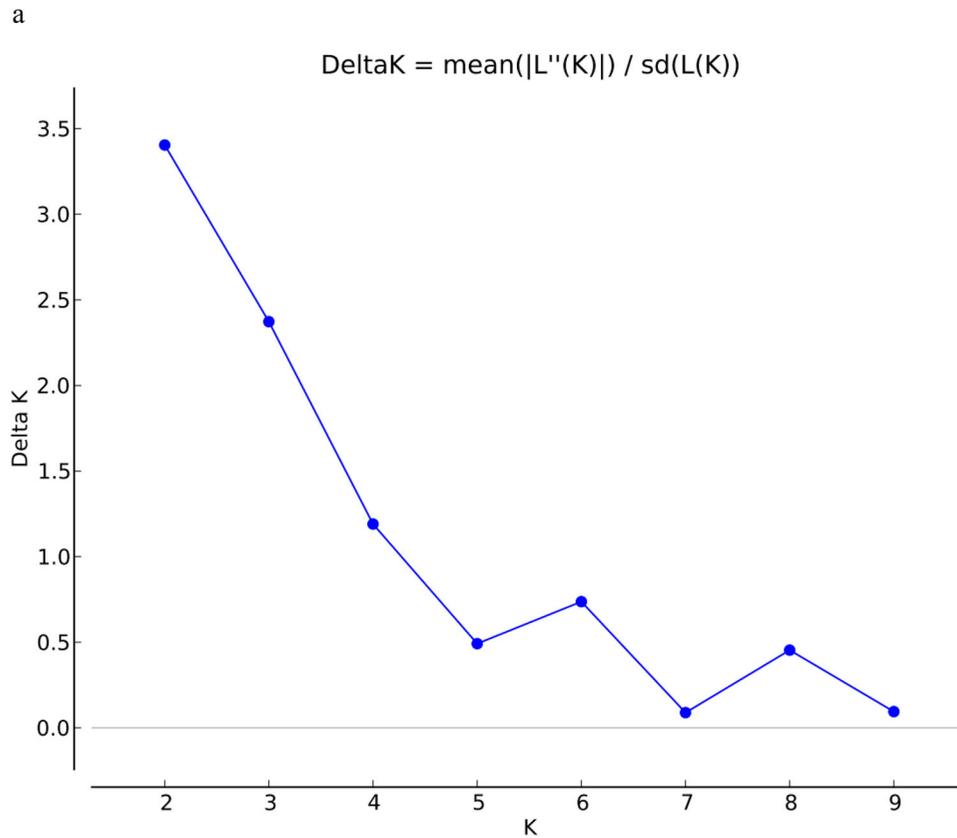


Figure 4. Population structure of the 80 sorghum accessions representing five populations in Ethiopia. (a) Best delta K value estimated using Evanno et al. (2005); method and (b) Bayesian model-based estimated population structure for $K = 2$ according to geographical locations. The different (blue and orange) colours represents genetic groups or sub-populations designated by Structure Harvester.

during seed exchange by the farmers and the geographical regions in Ethiopia are not completely different in terms of climate and other soil variables; rather they share similar agro-ecologies. Similarly, a study conducted by Tirfessa et al. (2020) has shown that both hierarchical clustering and PCoA analysis did not reveal clear grouping of Ethiopian sorghum accessions based on either altitude or geographical origin. In addition, Motlhaodi et al. (2014) reported that there was no observed geographical

origin or racial clustering pattern in sorghum genotypes. In contrast, Missihoun et al. (2015) reported that SSR analysis of 61 sorghum samples from Benin structured according to their botanical race and morph-physiological characteristics. In the present study (Figure 4), population genetic structure analysis also revealed the presence of weak sub-groups ($K = 2$) of the five populations of sorghum with a high potential of admixture. All populations possessed genetic background (alleles)

b

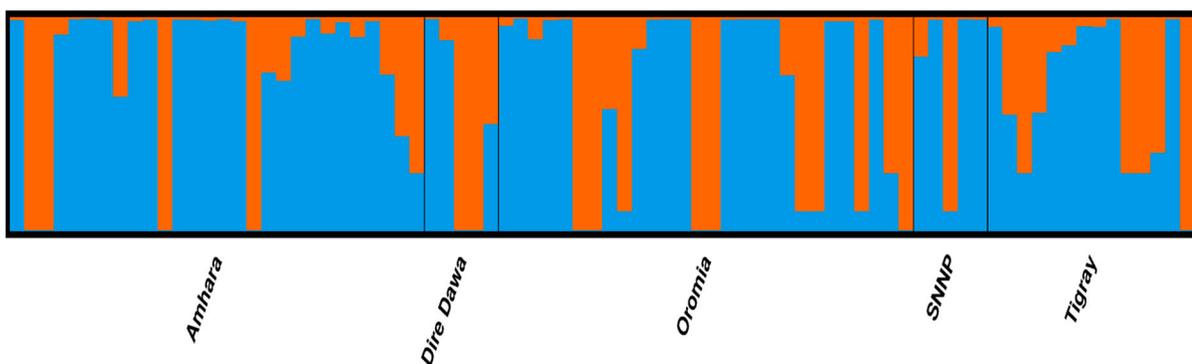


Figure 4 Continued

from both clusters confirming the presence of gene flow among the populations.

Sorghum is one of the most important multi-purpose crops grown globally. It is one of the major staple crops grown in the poorest and most food insecure regions of Ethiopia. Knowledge of the genetic diversity of sorghum is the foundation for the improvement and sustainable development of new varieties. In the present study, 80 sorghum accession collected from five sorghum growing regions of Ethiopia categorised into five populations were analyzed for allelic diversity by 11 SSR markers. All the markers were highly polymorphic and informative to describe the genetic diversity and population structure of the genotype.

The study also revealed existence of high genetic diversity within sorghum populations of Ethiopia, where 93.26% of the total genetic diversity resides within populations. Among the addressed populations, high gene diversities ranging from 0.51 to 0.77 was observed. Sorghum populations of Oromia and Amhara displayed gene diversity greater than the grand mean value of 0.67 suggesting that these locations should be targeted to generate baseline information for breeding and conservation of the crop. The unweighted neighbour-joining based cluster analysis showed the existence of weak genetic differentiation and all the sorghum population shared genetic background originated from two subpopulations confirming the presence of high genetic exchange or gene flow among populations. This information is significantly crucial for the development of pure sorghum breeding lines. The result of this study suggested that future germplasm collection and utilisation strategies of sorghum in Ethiopia should take into consideration the magnitude and pattern of genetic diversity established at genotypic level.

The present study has contributed substantial knowledge about the genetic architecture of some Ethiopian sorghum germplasm collections to the conservation, improvement and breeding programme of the crop. To come up with better resolution of the genetic diversity of sorghum in Ethiopia, we recommend an investigation which will employ more sample size, wider geographic area coverage and more molecular markers.

Acknowledgements

Swedish International Development Agency (SIDA) through 'Sorghum: screening and association mapping of drought tolerant sorghum genotypes in Ethiopia research project' and Adama Science and Technology University are acknowledged for financial support. The National Agricultural Biotechnology Research Center (NABRC) at Holeta is duly acknowledged for the access of molecular biology laboratory facility.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was financially supported partly by Swedish International Development Agency (SIDA) through 'Sorghum: screening and association mapping of drought tolerant sorghum genotypes in Ethiopia research project' and Adama Science and Technology University.

Authors' contribution

BN, MK, ME and TF contributed to the study conception and design. Material preparation, data collection and analysis were performed by BN and ME. The first draft of the manuscript was prepared by BN and MK and all the authors commented on the previous versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of interests

The authors declare that they have no competing interests.

Notes on contributors

Bethlehem Nemera is Lecturer at Applied Biology Department of Adama Science and Technology University. She specialized in plant biotechnology.

Mulugeta Kebede is Associate Professor at Department of Plant Biology and Biodiversity Management of Addis Ababa University. He is interested plant genomics and phenomics.

Muluken Eneyew is a staff as well as PhD candidate at Institute of Biotechnology, Addis Ababa University.

Tileye Feyissa is a full professor at Institute of Biotechnology, Addis Ababa University.

ORCID

Mulugeta Kebede  <http://orcid.org/0000-0002-9185-9325>

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 Evol.* 60(4):1313–1328. doi:10.1007/s10722-012-9921-8.
- Amelework BA, Shimelis HA, Tongoona P, Mengistu F, Laing MD, Ayele DG. 2016. Sorghum production systems and constraints, and coping strategies under drought-prone agroecologies of Ethiopia. *S Afr J Plant Soil.* 33(3):207–217. doi:10.1080/02571862.2016.1143043.

- 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 Resour Crop Evol.* 47(5):471–482. doi:10.1023/A:1008751721825.
- Ayana A, Bryngelsson T, Bekele E. 2001. Geographic and altitudinal allozyme variation in sorghum (*Sorghum bicolor* (L.) Moench) landraces from Ethiopia and Eritrea. *Hereditas.* 135(1):1–12. doi:10.1111/j.1601-5223.2001.t011-00001.x.
- Billot C, Rivallan R, Sall MN, Fonckea D, Deu M, Glaszmann JC, Noyer JL, Rami JF, Risterucci AM, Wincker P, et al. 2012. A reference microsatellite kit to assess for genetic diversity of *Sorghum bicolor* (Poaceae). *Am J Bot.* 99(6):245–250. doi:10.3732/ajb.1100548.
- Birhan T, Dong H, Abajebel N, Wakjira M, Lemke C, Vadez V, Paterson AH, Bantte K. 2022. Exploiting genetic variation from unadapted germplasm – an example from improvement of sorghum in Ethiopia. *Plants People Planet.* 1–14. doi:10.1002/ppp3.10292.
- 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(3):314–331.
- 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 Evol.* 60(6):1817–1830. doi:10.1007/s10722-013-9956-5.
- Cuevas HE, Prom LK, Erpelding JE, Brotons V. 2014. Assessments of genetic diversity and anthracnose disease response among Zimbabwe Sorghum germplasm. *Plant Breed.* 133(2):234–242. doi:10.1111/pbr.12133.
- De Wet JM. 1978. Special paper: systematics and evolution of sorghum sect. Sorghum (Gramineae). *Am J Bot.* 65(4):477–484. doi:10.1002/j.1537-2197.1978.tb06096.x.
- De Wet JM, Harlan JR. 1971. The origin and domestication of *Sorghum bicolor*. *Econ Bot.* 25(2):128–135. doi:10.1007/BF02860074.
- Disasa T, Feyissa T, Admassu B. 2017. Characterization of Ethiopian sweet sorghum accessions for 0 Brix, morphological and grain yield traits. *Sugar Technol.* 19(1):72–82. doi:10.1007/s12355-016-0440-3.
- Dong H, Birhan T, Abajebel N, Wakjira M, Mitiku T, Lemke C, Vadez V, Paterson AH, Bantte K. 2022. Natural variation further increases resilience of sorghum bred for chronically drought-prone environments. *J Exp Bot.* erac217. doi:10.1093/jxb/erac217.
- Earl DA. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 4(2):359–361. doi:10.1007/s12686-011-9548-7.
- Elangovan M, Jain SK, Patel NV. 2012. Characterization of sorghum germplasm collected from Gujarat. *Indian J Plant Genet Resour.* 26(1):42–46.
- Enyew M, Feyissa T, Carlsson AS, Tesfaye K, Hammenhag C, Geleta M. 2022. Genetic diversity and population structure of sorghum [*Sorghum bicolor* (L.) Moench] accessions as revealed by single nucleotide polymorphism markers. *Front Plant Sci.* 5(12):799482. doi:10.3389/fpls.2021.799482.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 14(8):2611–2620. doi:10.1111/j.1365-294X.2005.02553.x.
- Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 10(3):564–567. doi:10.1111/j.1755-0998.2010.02847.x.
- FAOSTAT. 2017. Production – crops – area harvested/production quantity – sorghum –2014. FAO Statistics (online database), Food and Agriculture Organization; [Accessed 2017 September 22] www.fao.org/faostat/en.
- Faye JM, Maina F, Akata EA, Sine B, Diatta C, Mamadou A, Marla S, Bouchet S, Teme N, Rami JF, et al. 2021. A genomics resource for genetics, physiology, and breeding of West African sorghum. *Plant Genome.* 2021(14):e20075. doi:10.1002/tpg2.20075.
- 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. doi:10.1007/s10531-005-0313-7.
- Girma G, Nida H, Seyoum A, Mekonen M, Nega A, Lule D, Dessalegn K, Bekele A, Gebreyohannes A, Adeyanju A, et al. 2019. A large-scale genome-wide association analyses of Ethiopian sorghum landrace collection reveal loci associated with important traits. *Front Plant Sci.* 10:691. doi:10.3389/fpls.2019.00691.
- Girma G, Nida H, Tirfessa A, Lule D, Bejiga T, Seyoum A, Mekonen M, Nega A, Dessalegn K, Birhanu C, et al. 2020. A comprehensive phenotypic and genomic characterization of Ethiopian sorghum germplasm defines core collection and reveals rich genetic potential in adaptive traits. *Plant Genome.* 13:e20055. doi:10.1002/tpg2.20055.
- Harlan JR, De Wet JM. 1972. A simplified classification of cultivated sorghum. *Crop Sci.* 12(2):172–176. doi:10.2135/cropsci1972.0011183X001200020005x.
- Hu Z, Olatoye MO, Marla S, Morris GP. 2019. An integrated genotyping-by-sequencing polymorphism map for over 10,000 sorghum genotypes. *Plant Genome.* 12:180044. doi:10.3835/plantgenome2018.06.0044.
- Jiang GL. 2013. Molecular markers and marker-assisted breeding in plants. In: Anderson SB, editor. *Plant breeding from laboratories to fields*. Copenhagen: University of Copenhagen; p. 45–83.
- Jonah PM, Bello LL, Lucky O, Midau A, Moruppa SM, Moruppa SM. 2011. Review: the importance of molecular markers in plant breeding programmes. *Global J Sci Front Res.* 11(5):4–12.
- Khangura ME. 2019. A genome-wide association study of the quantitative resistance to *Striga hermonthica* and plant architecture of *Sorghum bicolor* in Northwestern Ethiopia. Purdue University Graduate School. PhD dissertation. doi:10.25394/PGS.10286864.v1.
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour.* 15(5):1179–1191. doi:10.1111/1755-0998.12387.
- Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, Juenger TE, Hyma K, Acharya C, Mitchell SE, et al. 2015. Genome-environment associations in sorghum landraces predict adaptive traits. *Sci Adv.* 1(6):e1400218. doi:10.1126/sciadv.1400218.
- Lee S, Fu F, Liao CJ, Mewa DB, Adeyanju A, Ejeta G, Lisch D, Mengiste T. 2022. Broad-spectrum fungal resistance in sorghum is conferred through the complex regulation of

- an immune receptor gene embedded in a natural antisense transcript. *Plant Cell*. 34:1641–1665. doi:10.1093/plcell/koab305.
- Liu K, Muse SV. 2005. Power marker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. 21(9):2128–2129. doi:10.1093/bioinformatics/bti282.
- Menamo T, Kassahun B, Borrell AK, Jordan DR, Tao Y, Hunt CH, Mace ES. 2021. Genetic diversity of Ethiopian sorghum reveals signatures of climatic adaptation. *Theor Appl Genet*. 134(2):731–742. doi:10.1007/s00122-020-03727-5.
- Mengistu G, Shimelis H, Laing M, Lule D, Assefa E, Mathew I. 2020. Genetic diversity assessment of sorghum (*Sorghum bicolor* (L.) Moench) landraces using SNP markers. *S Afr J Plant Soil*. 37(3):220–226. doi:10.1080/02571862.2020.1736346.
- Menz MA, Klein RR, Unruh NC, Rooney WL, Klein PE, Mullet JE. 2004. Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. *Crop Sci*. 44(4):1236–1244. doi:10.2135/cropsci2004.1236.
- Missihoun AA, Adoukonou-Sagbadja H, Sedah P, Agbangla C, Ahanhanzo C, Dagba RA. 2015. Genetic diversity of *Sorghum bicolor* (L.) Moench landraces from Northwestern Benin as revealed by microsatellite markers. *Afr J Biotechnol*. 14(16):1342–1353. doi:10.4314/acsj.v24i2.6.
- Mofokeng A, Shimelis H, Tongoona P, Laing M. 2014. A genetic diversity analysis of South African sorghum genotypes using SSR markers. *S Afr J Plant Soil*. 31(3):145–152. doi:10.1080/02571862.2014.923051.
- Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, et al. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *PNAS*. 110(2):453–458. doi:10.1073/pnas.1215985110.
- 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(1):35–43. doi:10.1017/S1479262119000212.
- Murray SC, Sharma A, Rooney WL, Klein PE, Mullet JE, Mitchell SE, Kresovich S. 2008. Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. *Crop Sci*. 48(6):2165–2179. doi:10.2135/cropsci2008.01.0016.
- 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. doi:10.1007/s00122-010-1504-5.
- Muui CW, Muasya RM, Kirubi DT, Runo SM, Karugu A. 2016. Genetic variability of sorghum landraces from lower Eastern Kenya based on simple sequence repeats (SSRs) markers. *Afr J Biotechnol*. 15(8):264–271. doi:10.5897/AJB2015.14680.
- 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. doi:10.1111/j.1601-5223.2011.02208.x.
- Nida H, Girma G, Mekonen M, Lee S, Seyoum A, Dessalegn K, Tadesse T, Ayana A, Senbetay T, Tesso T, et al. 2019. Identification of sorghum grain mold resistance loci through genome wide association mapping. *J Cereal Sci*. 85:295–304. doi:10.1016/j.jcs.2018.12.016.
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, et al. 2009. The Sorghum bicolor genome and the diversification of grasses. *Nature*. 457(7229):551–556. doi:10.1038/nature07723.
- Pavel AB, Vasile CI. 2012. PyElph – a software tool for gel images analysis and phylogenetics. *BMC Bioinf*. 13(1):1–6. doi:10.1186/1471-2105-13-9.
- Peakall R, Smouse PE. 2012. Genalex 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*. 28(19):2537–2539. doi:10.1093/bioinformatics/bts460.
- Perrier X, Jacquemoud-Collet JP. 2006. DARwin – dissimilarity analysis and representation for windows. Version. 5:158.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155(2):945–959.
- Ramu P, Kassahun B, Senthilvel S, Kumar CA, Jayashree B, Folkertsma RT, Reddy LA, Kuruvina Shetti 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(7):1193–1204. doi:10.1007/s00122-009-1120-4.
- Slatkin M. 1985. Rare alleles as indicators of gene flow. *Evolution*. 39(1):53–65. doi:10.1111/j.1558-5646.1985.tb04079.x.
- Smith JSC, Smith OS. 1992. Fingerprinting crop varieties. *Adv Agron*. 47:85–140.
- Tadesse H, Feyissa T. 2013. Analysis of genetic diversity of *Sorghum bicolor* ssp. *bicolor* (L.) Moench using ISSR markers. *Asian J Plant Sci*. 12(2):61–70. doi:10.3923/ajps.2013.61.70.
- Tariq AS, Akram Z, Shabbir G, Khan KS, Mahmood T, Iqbal MS. 2014. Heterosis and combining ability evaluation for quality traits in forage sorghum (*Sorghum bicolor* L.). *SABRAO J Breed Genet*. 46(2):174–182.
- Tautz D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res*. 17(16):6463–6471. doi:10.1093/nar/17.16.6463.
- Teshome A, Baum BR, Fahrige 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(3):255–263. doi:10.1023/A:1003074008785.
- Thudi M, Fakrudin B. 2011. Identification of unique alleles and assessment of genetic diversity of rabi sorghum accessions using simple sequence repeat markers. *J Plant Biochem Biotechnol*. 20(1):74–83. doi:10.1080/02571862.2014.923051.
- Tirfessa A, Tesso T, Adugna A, Mohammed H, Kiambi D. 2020. Genetic diversity among Ethiopian sorghum [*Sorghum bicolor* (L.) Moench] gene bank accessions as revealed by SSR markers. *Afr J Biotechnol*. 19(2):84–91. doi:10.5897/AJB2019.16981.
- Vavilov NI. 1951. The origin, variation, immunity and breeding of cultivated plants, (Translated by SK. Chestitee). *Chronica Botanica*. 13:1–364.
- Wani SP, Albrizio NR, Vaija R. 2011. Sorghum crop yield response to water. Rome: FAO.
- Waples RS. 1987. A multispecies approach to the analysis of gene flow in marine shore fishes. *Evolution*. 41(2):385–400. FAO. doi:10.1111/j.1558-5646.1987.tb05805.x.
- Wondimu Z, Dong H, Paterson A, Worku W, Bantte K. 2021. Genetic diversity, population structure and selection signature

- in Ethiopian sorghum [*Sorghum bicolor* L. (Moench)] germplasm. 3G. 11(6):jkab087. doi:[10.1093/g3journal/jkab087](https://doi.org/10.1093/g3journal/jkab087).
- Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*. 19:395–420.
- Yaqoob M, Hussain N, Rashid A. 2015. Genetic variability and heritability analysis for yield and morphological traits in sorghum (*Sorghum bicolor* (L.) Moench) genotypes. *J Agri Res*. 53(3):331–343.
- Yeh FC, Yang RC, Boyle T. 1999. POPGENE version 1.32: Microsoft Windows-based freeware for population genetic analysis, quick user guide. Center for International Forestry Research, University of Alberta, Edmonton, Alberta.