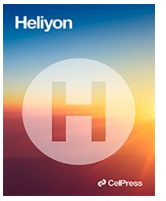
Heliyon 9 (2023) e12830

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

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

Wubshet Mamo [a,b,*], Muluken Enyew [b], Tilahun Mekonnen [b], Kassahun Tesfaye [b,c], Tileye Feyissa [b]

[a] *Ethiopian Biodiversity Institute, Addis Ababa, Ethiopia*
[b] *Institute of Biotechnology, Addis Ababa University, Addis Ababa, Ethiopia*
[c] *Ethiopian Bio & Emerging Technology Institute, Affiliated with Institute of Biotechnology, Addis Ababa University, Addis Ababa, Ethiopia*

ARTICLE INFO

Keywords:
Clustering
Gene flow
Genetic diversity
Microsatellite markers
Sorghum

ABSTRACT

In the tropical and semi-arid regions of Africa, sorghum [*Sorghum bicolor* (L.) Moench] is mainly grown as a major food security crop. Understanding the extent and pattern of genetic variability is a prerequisite criterion for sorghum improvement and conservation. The genetic diversity and population structure of 100 genotypes of sorghum were profiled using 15 microsatellite loci. A total of 108 alleles, with an overall mean of 7.2 alleles per locus, were produced by all of the microsatellite loci used due to their high polymorphism. Polymorphic information content values ranging from 0.68 to 0.89 indicated that all of the loci are effective genetic tools for analysing the genetic structure of sorghum. Different diversity metrics were used to evaluate genetic diversity among populations, and Nei's gene diversity index ranged from 0.74 to 0.81 with an overall mean of 0.78. Poor genetic differentiation (FST: 0.02; $p < 0.0001$) was found, where 98% of entire variability was accounted by the within populations genetic variability, leaving only 2.32% among populations. The highest genetic differentiation and Nis's genetic distance were observed between the sorghum populations of the Southern Nation and Nationalities Peoples and Dire Dawa regions. Due to increased gene flow ($Nm = 10.53$), the clustering, principal coordinate analysis and STRUCTURE analysis failed to categorize the populations into genetically different groups corresponding to their geographic sampling areas. In general, it was found that the microsatellite loci were highly informative and therefore valuable genetic tools to unfold the genetic diversity and population structure of Ethiopian sorghum genotypes. Among the five populations studied, sorghum populations from Amhara and Oromia had the highest genetic variation, indicating that the regions could be perhaps hotspots for useful alleles for the development of better-performing genotypes, and also for designing appropriate germplasm management strategies.

1. Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is a highly significant grain due to its adaptability to a variety of ecological and climatic

* Corresponding author. Ethiopian Biodiversity Institute, Addis Ababa, Ethiopia.
E-mail addresses: wubma53@gmail.com, wubshet.mamo@ebi.gov.et (W. Mamo).

https://doi.org/10.1016/j.heliyon.2023.e12830
Received 15 June 2022; Received in revised form 30 December 2022; Accepted 3 January 2023
Available online 7 January 2023

circumstances in the semi-arid tropics of the world [1]. It is a diploid species (2n = 2x = 20) with a genome size of 735 Mbp [2]. Thirty-one domesticated species and seventeen associated wild species belong to the genus sorghum [3], creating an excellent opportunity for further taxonomic studies of sorghum. Next to maize, wheat, rice, and barley, sorghum is the fifth most produced crop globally [4]. Sorghum is widely grown for multiple purposes such as feed, fodder, food, and fuel, and has several industrial uses, including those in pharmaceutical diagnosis, medicine, the synthesis of organic compounds, and utility products [5].

According to Ref. [6]; Ethiopia is sorghum's center of diversity and origin ranking top among nations that have made significant contributions to the sorghum germplasm collections in the world, hence, the country could be an abundant source of sorghum landraces. Landraces are key sources of beneficial genes needed to increase and maintain the production of contemporary varieties [7]. In order to maintain sustainable crop productivity, breeders can choose desired parents for breeding strategy and the introduction of genes that are not closely associated germplasms and this can be made possible by having a clear understanding of the genetic variation of the crop. Genetic diversity presents a chance for crop breeders to create new and improved cultivars with desired traits, such as large seeds, high yield potential, resistance to biotic stresses and tolerance to abiotic stresses [8].

Molecular markers reflect genetic similarities and differences without being influenced by the environment. Furthermore, no prior ancestry details are needed for the molecular marker-based characterization [9]. They have been essential for the exploitation and preservation of the plant genetic resources in traits improvement programs for various crops [10–12]. Different molecular markers such as Amplified Fragment Length Polymorphism (AFLP), Random Amplified Polymorphic DNA (RAPD), Restriction Fragment Length Polymorphism (RFLP), Diversity Arrays Technology, (DArT), Microsatellites or Simple Sequence Repeats (SSR), and Single Nucleotide Polymorphism (SNP) came to be utilized in different sorghum researches including its genetic structure analysis [13–15]. Microsatellite markers have shown great potential to detect the genetic diversity and relationships of organisms due to their allele-specific and co-dominant nature [16-18]. Their higher polymorphism, informativeness, and ease of identification from genomic sequences have made microsatellites a marker of choice in molecular studies.

Although there are many genetic diversities in the species of sorghum, numerous accessions of the crop are categorized, assessed and characterized; there exist paramount research gaps for multi-genic traits [19]. In a given period of time crop improvement programs are often limited to portions of the germplasm collection which bear important characteristics because of various reasons [20]. Thus, the sorghum-breeding strategy in aiding the selection of parental lines with desirable traits can be facilitated through a profound consideration of sorghum's genetic diversity using molecular markers. Therefore, the present study was designed to profile the genetic diversity and population structure of sorghum landraces brought together from different regions of Ethiopia revealed by microsatellite markers.

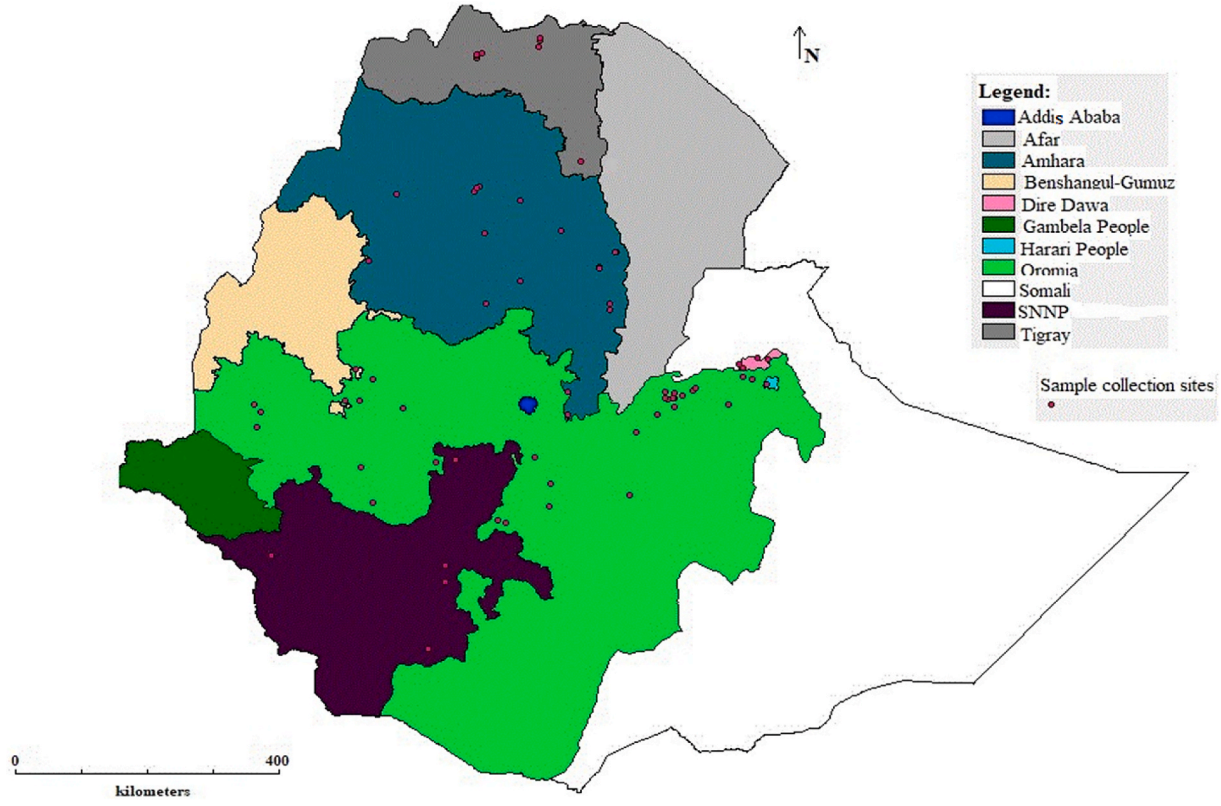


Fig. 1. Map of Ethiopia showing sorghum collection sites and regions.

2. Material and methods

2.1. Plant materials

Hundred sorghum accessions were originally collected from Amhara, Dire Dawa, Oromia, Southern Nation and Nationalities People (SNNP) and Tigray regional states of Ethiopia (Fig. 1; Supplementary Table S1) by Ethiopian Biodiversity Institute (EBI) were the subjects of the present study. These plant materials are included in one species of sorghum (*Sorghum biocolor*). Global positioning system (GPS) based points (elevation data and geographical coordinates) for each sampling site were used to make the map of the study area.

2.2. DNA extraction

For each accession, fifteen seeds were planted in the greenhouse of Melkasa Agricultural Research Center (MARC) from which five young leaves of two-week-old seedlings were picked at random and used for the genomic extraction of DNA according to the Diversity Array Technology (DArT) protocol [21]. Six milligrams of leaf sample from each individual were used to form a bulk sample. The quality of the extracted DNA was checked by loading 5 μL (μL) of DNA and 2 μL of loading dye (6X) with gel red staining on 1% agarose gel at 100 V for 45 min (min). The concentration of the DNA was checked by a NanoDrop Spectrophotometer (ND-2000, Thermo Scientific USA). Until genotyping, high molecular weight DNA was kept at 20 degree celsius (°C) below zero.

2.3. Genotyping

For genotyping, the genetic diversity of 100 genotypes of sorghum was assessed using 15 microsatellite markers (Table 1 [12,22, 23]). Microsatellite loci were chosen by taking into consideration their consistent distribution throughout the sorghum genome. Gradient temperatures were used to optimize the annealing temperatures of the primers. The first four primer pairs involved 95 °C (initial denaturation) for 5 min. Forty cycles of denaturation were used at 95 °C for 1 min. The primer annealing temperature at this stage was optimized for 2 min (Table 2), and 72 °C for primer extension within 2 min in the *polymerase chain reaction* (PCR) program. The PCR program for the next seven primer pairs (5-11) in Table 2 involved 94 °C (initial denaturation) for 5 min. At this stage, 35

Table 1
Primers sequence, repeat motif, chromosome number, annealing temperatures and their range of molecular weight detected at 15 microsatellite loci in five populations of Sorghum in Ethiopia.

Marker	Forward and Reverse primer (5′–3′)	Repeat Motif	Chromosome Number	[a]Fragment Size (base pair)		References
				Min.	Max.	
SbAGB02	F: CTCTGATATGTCGTTGTGCT R: ATAGAGAGGATAGCTTATAG-CTCA	(AG)35	7	95	126	[12]
Xcup74	F: GTCGCCATTGTGATGAAGAG R: CAGTAGTCCAGCAAAACGGC	(TG)9	2	174	185	[22]
Xtxp136	F: GCGAATAGCATCTTACAACA R: ACTGATCATTGGCAGGAC	(GCA)5	5	238	242	[12]
Xcup50	F: TGATTGATTGAGGCAGGCAC R: TTCCGGTCTCTGTCCATTTC	(ACAGG)5	10	150	159	[22]
Xcup67	F: GGTCAGTGCTTACACAGATTCC R: GGGGATTGCAGGTGTCATAG	(TA)6	10	264	294	[22]
Xcup14	F: TACATCACAGCAGGGACAGG R: CTGGAAAGCCGAGCAGTATG	(AG)10	3	209	237	[12]
Xgap084	F: CGCTCTCGGGATGAATGA R: TAACGGACCACTAACAAATGATT	(AG)14	2	176	209	[12]
Xtxp010	F: ATACTATCAAGAGGGGAGC R: AGTACTAGCCACACGTCAC	(CT)14	9	139	148	[22]
Xtxp012	F: AGATCTGGCGGCAACG R: AGTCACCCATCGATCATC	(CT)22	4	175	200	[22]
Xtxp021	F: GAGCTGCCATAGATTTGGTCG R: ACCTCGTCCCACCTTTGTTG	(AG)18	4	175	190	[23]
Xtxp141	F: TGTATGGCCTAGCTTATCT R: CAACAAGCCAACCTAAA	(GA)23	10	141	182	[23]
Xtxp145	F: GTTCCTCCTGCCATTACT R: CTTCCGCACATCCAC	(AG)22	6	120	143	[12]
Xtxp265	F: GTCTACAGGCGTGCAAATAAAA R: TTACCATGCTACCCCTAAAAGTGG	(GAA)19	6	183	220	[22]
Xtxp273	F: GTACCCATTTAAATTGTTTGCAGTAG R: CAGAGGAGGAGGAAGAGAAGG	(TTG)20	8	198	231	[22]
Xtxp320	F: TAAACTAGACCATATACTGCCATGATAA R: GTGCAAATAAGGGCTAGAGTGTT	(AAG)20	1	260	288	[12]

Min. = minimum, Max. = maximum.
[a] = resulted in this study.

cycles of denaturation were used at 95 °C for 1 min. The optimized temperatures for primer annealing were done in 1 min and primer extension at 72 °C for 2 min. The PCR program for the last four primer pairs (12–15) in Table 2 involved 94 °C (an initial denaturation) for 5 min. The denaturation was done with 44 cycles at 95 °C for 1 min. The optimized annealing temperatures were done for 2 min and primer extension at 72 °C for 2 min. For the final extension, 72 °C was used for 10 min with a holding step of 4 °C for all primers. For the PCR products, 1 × TBE buffer at 100 V for 3 h was used at 3% of the electrophoresis fractionation (agarose gel). The BIO-RAD Gel Doc TM EZ Imaging System was used to photograph via visualization under ultraviolet (UV) light stained by 2 μL of gel red. For the estimation of the amplification size of DNA products, a 100-base pair DNA ladder was used.

2.4. Scoring of molecular data and analysis

In relation to the size of the fragments, the estimation of all amplified PCR regions were carried out. PyElph 1.4 software package was used to estimate the fragment size of the PCR products [24]. Each pair of primers was assumed to amplify a single genetic locus with different molecular weight bands representing different alleles of a particular locus. Analysis of population structure and genetic diversity was performed using data from scored molecular markers. Different statistical software packages were used to calculate the standard indices of genetic diversity. To compute locus-based diversity indices across the entire population, such as polymorphic information content (PIC), major allele frequency (MAF), heterozygosity (Ho), and gene diversity (expected heterozygosity (He)) for each marker and their averages across markers (Table 3), PowerMarker software (version 3.25) was used [25].

To assess gene flow (Nm) and population differentiation (Gst) and across the populations (Table 3), POPGENE software (version 1.32) [26] was used. Genetic and population diversity indices, namely number of observed alleles (Na), number of private alleles (NPA), allelic frequency, effective number of alleles (Ne), pairwise population genetic distances, Nei's gene diversity (h), Shannon's information index (I) and genetic differentiation test (Gst and p values) over 999 bootstrap replications for the entire population (Table 3), GenAlEx software (version 6.501) [27,28], was used.

The estimation of the variance components and the analysis of molecular variance (Table 7) were computed using Arlequin (version 3.5.2.2) [29]. Unweighted pair group method with arithmetic mean trees (UPGMA) that rely on Nei's standard genetic distance [30] were constructed by using PowerMarker (version 3.25) [25] based on 1000 bootstrap replications [31] and finally viewed using MEGA software [32] built in PowerMarker (version 3.25). MS Excel (Microsoft Inc., USA) was used to calculate abundant, common, rare (>50%, 5–50%, <5% respectively) and private (present only in one group but not in the other) alleles.

STRUCTURE software (version 2.3.4) along with the algorithm of Bayesian model-based clustering [33] was used to decide the admixture patterns and population structure. One hundred thousand burn-in period in each run and Markov Chain Monte Carlo (MCMC) (data were collected over 250 000 replications for K = 1 to K = 10 using 20 iterations for each K) was used to estimate the true number of population clusters (K). A web-based STRUCTURE HARVESTER (version 0.6.92) [34] was used to estimate the optimal K value coming after the [35] method of simulation and the Clumpak (beta version) [36] was used for the estimation of the bar plot of the best K.

3. Results

3.1. Microsatellite markers level of polymorphism

Fifteen microsatellite loci yielded 108 putative alleles that are dissimilar in fragment sizes, possessing a mean of 7.2 alleles per locus in this study (Table 3). No abundant (above 0.5), common (between 0.05 and 0.5 = 100 (92.59%)) and rare (alleles below 0.05 = 8

Table 2
Amplification parameters for the primers used.

Marker	Initial denaturation		Denaturation		[a]Annealing temperature		Extension		Final extension		Cycles
	T° (°C)	Time (Min)	T° (°C)	Time (Min)	T° (°C)	Time (Min)	T° (°C)	Time (Min)	T° (°C)	Time (Min)	
SbAGB02					44						
Xcup74					56						
Xtxp136	95	5	95	1	55	2	72	2	72	10	40
Xcup50					58						
Xcup67					56						
Xcup14					60						
Xgap084					50						
Xtxp010	94	5	95	1	60	1	72	2	72	10	35
Xtxp012					56						
Xtxp021					52						
Xtxp141					50						
Xtxp145					54						
Xtxp265	94	5	95	1	56	2	72	2	72	10	44
Xtxp273					56						
Xtxp320					58						

T° = temperature, °C = degree centigrade, Min = minutes.
[a] = results in this study.

Table 3
Informativeness and other genetic diversity summary statistics for all 15-microsatellite loci across five populations of Sorghum.

Locus	MAF	Na	Ne	I	H	H_o	Fst	Nm	PIC	P value	P_{HWE}
SBAGBO02	0.22	7.00	6.43	1.90	0.84	0.00	0.06	4.11	0.83	<0.001	<0.001
XCUP74	0.21	7.00	6.54	1.91	0.85	0.00	0.09	2.70	0.83	<0.001	<0.001
XTXP136	0.34	5.00	4.33	1.54	0.77	0.00	0.08	2.95	0.73	<0.01	<0.001
XCUP50	0.14	10.00	6.54	1.28	0.90	0.00	0.06	4.18	0.89	<0.001	<0.001
XCUP67	0.33	5.00	4.29	1.53	0.77	0.00	0.05	5.02	0.73	<0.001	<0.001
XCUP14	0.25	5.00	4.69	1.58	0.79	0.00	0.06	4.07	0.75	<0.01	<0.001
XGAP084	0.18	10.00	8.14	2.17	0.88	0.00	0.05	5.26	0.87	<0.01	<0.001
XTXP010	0.23	6.00	5.77	1.78	0.83	0.00	0.04	5.73	0.80	<0.01	<0.001
XTXP012	0.19	9.00	7.99	2.14	0.88	0.00	0.06	3.90	0.86	<0.001	<0.001
XTXP021	0.33	4.00	3.74	1.35	0.73	0.00	0.02	10.67	0.68	<0.01	<0.001
XTXP141	0.29	6.00	5.29	1.73	0.81	0.00	0.05	4.63	0.79	<0.01	<0.001
XTXP145	0.32	8.00	5.32	1.85	0.81	0.00	0.05	4.63	0.79	<0.001	<0.001
XTXP265	0.18	12.00	8.87	2.33	0.89	0.96	0.02	12.56	0.88	<0.01	<0.01
XTXP273	0.25	6.00	5.48	1.75	0.82	0.00	0.03	8.77	0.79	<0.01	<0.001
XTXP320	0.23	8.00	6.38	1.93	0.84	0.00	0.04	6.20	0.82	<0.05	<0.001
Mean	0.25	7.2	6.19	1.85	0.83	0.06	0.05	4.81	0.80		

MAF = major allele frequency, Na = Observed number of alleles, Ne = effective number of alleles, I = shannon's information statistic, h = Nei's gene diversity, H_o = observed heterozygosity, h/H_e = Expected heterozygosity, Gst = genetic differentiation statistics, Nm = gene flow = [(1/Fst) - 1]/4, where Fst = inbreeding coefficient, PIC = Polymorphic information content, *P*-value and P_{HWE} for divergence from the Hardy Weinberg Equilibrium; significant at different levels.

(7.41%) alleles were recorded from the total of 108 alleles. Per locus, the number of alleles varied from 4 to 12 for XTXP021 and XTXP265 in their order. MAF ranged from 0.14 for locus XCUP50 to 0.34 for XTXP136 a mean of 0.25 in the entire populations. The highest Na, Ne, I, Nm and PIC were recorded for XTXP265. Microsatellite loci XCUP50 (h = 0.90) and XTXP021 (h = 0.73) resulted the largest and the lowest gene diversity across the entire populations, respectively (Table 3).

The least genetic differentiation (Gst = 0.02) and the highest gene flow (12.56) among the populations was resulted in the locus XCUP265. The same locus also showed the highest observed and expected heterozygosity (H_e = 0.87), while the XTXP021 locus showed the least gene diversity (expected heterozygosity, H_e = 0.73) (Table 3). The informativeness of individual loci as measured by PIC varied from 0.68 (XTXP021) to 0.89 (XCUP50) (average = 0.80). The informativeness of the PIC values was assumed as less when they account below 0.25, medium from 0.25 to 0.5 and high for those greater than 0.5. The PIC values for all microsatellite loci recorded were greater than 0.5 indicating that their informativeness was high. All microsatellite loci were significant ($p < 0.01$) for the Hardy-Weinberg equilibrium test.

3.2. *Within and among genetic variability of populations*

Various estimates of genetic diversity are summarized in Table 4 over all the loci across populations. The number of unique alleles in a single population (private alleles) of 0.03 (0.00–0.13), shannon's information index of 1.66 (1.47–1.80), gene diversity value of 0.78 (0.74–0.81) and number of effective alleles of 5.01 (4.25–5.71) were the estimates of the average genetic diversity. Sorghum populations from Oromia showed the highest Na (7.13) and Ne (5.71) (Table 4).

Number of private alleles (NPA = 0.13), Shannon's information index (I = 1.80) and gene diversity (H = 0.81) (Table 4) followed by the populations from Amhara. Except for the population of Oromia where NPA = 0.13, there was no private allele unique to a single population. The populations of both Amhara and Oromia showed the highest locally common alleles found in the populations and from that of Tigray the lowest Nei's gene diversity value was recorded relatively. Sixty percent of the study populations showed higher Nei's genetic diversity (H) than the overall mean value of 0.78. The percentage of polymorphic loci (PPL) per population became 100% for all populations (Oromia, SNNP, Tigray, Amhara and Dire Dawa) (Table 4).

Table 4
Fifteen microsatellite loci diversity indices[a] and allelic patterns across populations.

Population[a]	Size	Na	Ne	NPA	NLCA	I	H
Amhara	31	6.93	5.45	0.00	0.20	1.76	0.81
Dire Dawa	14	5.73	4.62	0.00	0.00	1.58	0.76
Oromia	30	7.13	5.71	0.13	0.20	1.80	0.81
SNNP	7	4.80	4.25	0.00	0.00	1.47	0.74
Tigray	18	6.47	5.01	0.00	0.13	1.69	0.79
Mean	20	6.21	5.01	0.03	0.11	1.66	0.78

Na = observed number of alleles, Ne = Number of effective alleles, NPA = number of private alleles, I = shannon's information index, H = Nei's genetic diversity, NLCA = number of locally common alleles found in populations.
[a] SNNP = Southern Nations, Nationalities, and People's Region.

3.3. Genetic relationships between the populations

Pairwise population measures of gene flow above the diagonal and Nei's genetic distance below the diagonal are presented in Table 5. Between the Oromia and Amhara populations (0.13) with the highest gene flow of 21.66, the lowest genetic distance was recorded, followed by the relationship between the Amhara and Tigray and the Oromia and Tigray populations, which scored a genetic distance of 0.16 with a gene flow of 17.48 and 17.91 respectively. The highest measure of genetic distance (0.39) with relatively low gene flow (5.27) was observed between Dire Dawa and SNNP populations. Between the SNNP and Tigray and SNNP and Oromia populations (0.32) with a gene flow of 7.92 and 8.30 the second highest genetic distance was recorded, respectively. The Tigray and SNNP populations exhibited a statistically remarkable ($p < 0.05$) highest genetic differentiation (0.91) (Table 5).

3.4. Genetic differentiation of populations

The genetic differentiation pairwise coefficient between the populations ranged from between Dire Dawa and Oromia (0.1) to between SNNP and Tigray (0.91). The lowest gene flow, highest and significant genetic differentiation was recorded between the populations of Tigray and SSNP (PhiPT = 0.91, $p < 0.05$). The second highest genetic differentiation with a gene flow rate of 0.86 was observed between the populations of Tigray and Oromia. Statistically, between the populations of Dire Dawa and SNNP a non-significant ($P = 0.05$) value of genetic differentiation was recorded (Table 6).

3.5. Analysis of molecular variance

According to major agro ecological adaptation zones (highland, intermediate and lowland) and their geographical locations (North, Central, East and South Ethiopia) the analysis of molecular variance (AMOVA) were computed without and with grouping populations. The 97.68% of genetic variation that occurred within populations was revealed by Fst values in the AMOVA analysis. Of overall genetic diversity, among geographical regions, the major agro-ecological adaptation zones and variability among populations were accounted for 1.94%, 0.72% and 2.32%, respectively. Low total genetic differentiation coefficient among the populations (Fst = 0.023, $P < 0.0001$) was recorded with high gene flow (≥10.53) (Table 7).

3.6. Clustering, principal coordination and population structure

Three major clusters (cluster 1 = C1, cluster 2 = C2 and cluster 3 = C3) and additional mini-clusters were obtained through the neighbor joining-based clustering method. Most of the individuals were categorized within C1 (72%; 72 accessions) subsequently followed by C2 which composed of 14 (14%) accessions and the least number of accessions were grouped within C3 that consisted of only 4 accessions (4%) (Fig. 2). There were no major or mini-clusters consisting of entirely accessions from a specific population (collection region), clearly indicating the presence of substantial commixture of the accessions. According to the generated unweighted pair group method with arithmetic mean (UPGMA) based dendrogram, the five populations were classified into three major clusters (cluster 1 = I, cluster 2 = II and cluster 3 = III), where I and III consisted of two populations; SNNP (southern Ethiopia) and Dire Dawa (Eastern Ethiopia) and Tigray and Amhara (Northern Ethiopia) respectively. The second cluster consisted of only a single population (Oromia; Central Ethiopia) (Fig. 3).

Principal coordinate analysis (PCoA) is a technique commonly used in multivariate statistics to visualize the pattern of genetic structure and determines the amounts of variance explained per component and cumulatively. The scatter plot obtained from PCoA revealed nearly a uniform distribution of the accessions of the entire populations in the central plane indicated a weak clustering of populations. Of the total variations explained (25.19%), each of the dimensions (1, 2 and 3) provided 4.32%, 8.52%, and 12.35%, respectively, in the first three principal coordinate axes (Fig. 4).

The value of K was used to determine the number clusters of the 100 sorghum accessions for population structure analysis on the basis of Bayesian Model. The studied sorghum accessions were clustered into three subpopulations since the K value reached a sharp peak at K = 3 in the analysis result of structure harvester website (Fig. 5a). There was no clear geographical origin-based structure result in this study, indicating the presence of a high genetic admixture among the three sub populations of the 100 Ethiopian sorghum accessions (Fig. 5b).

Table 5
Gene flow above the diagonal and pairwise Nei's genetic distance below the diagonal among populations.

Population[a]	Amhara	Dire Dawa	Oromia	SNNP	Tigray
Amhara	–	5.67	21.66	8.10	17.48
Dire Dawa	0.30	–	7.10	5.27	7.63
Oromia	0.13	0.26	–	7.92	17.91
SNNP	0.31	0.39	0.32	–	8.30
Tigray	0.16	0.26	0.16	0.32	–

– = not applicable.
[a] SNNP = Southern Nations, Nationalities, and People's Region.

Table 6
Population genetic differentiation measured by PhiPT (above the diagonal) between the populations with p-values (below the diagonal).

PhiPT/p value	Amhara	Dire Dawa	Oromia	SNNP	Tigray
Amhara	–	0.18	0.76	0.81	0.84
Dire Dawa	0.04	–	0.10	0.65	0.42
Oromia	0.01	0.03	–	0.81	0.86
SNNP	0.04	0.05	0.04	–	0.91
Tigray	0.02	0.03	0.02	0.04	–

– = not applicable, PhiPT = genetic differentiation, SNNP = Southern Nations, Nationalities, and People's Regional state.

Table 7
Among and within populations and geographical regions genetic variation divisions based on AMOVA.

Variation sources	DF	SS	MS	Estimate variance components	Percent variation	[a]Genetic differentiation	[b]P-value
Among populations	4	46.50	11.62	0.15	2.32%	Fst: 0.023	Fst
Within Populations	195	1192.54	6.12	6.12	97.68%		<0.0001
Total	199	1239.04		6.26	100%		
Nm (diploid)[c]populations	10.53						
Among Geographical Regions	3	34.16	11.39	0.12	1.94%	Fst: 0.019	Fst
Within Geographical Regions	196	1204.88	6.15	6.15	98.06%		<0.0001
Total	199	1239.04		6.27	100%		
Nm (diploid)[d] Geographical regions	12.64						
Among Altitudes	2	18.33	9.16	0.05	0.72%	Fst: 0.0072	Fst
Within Altitudes	197	1220.71	6.20	6.20	99.28%		<0.0001
Total	199	1239.04		6.24	100%		
Nm (diploid)[e]Altitudes	34.47						

[a] Fst = coefficient of genetic differentiation.
[b] significance based on 1023 permutations, Nm = gene flow (Nm).
[c] Among regions.
[d] Among geographical regions.
[e] Among altitudes (low, intermediate, highland).

4. Discussion

4.1. Microsatellite markers polymorphism

The diversity levels of the 100 accessions were decided using 15 polymorphic microsatellite markers dispersed throughout the sorghum genome. In the current study, the fragment sizes of the microsatellite markers were within the domain of earlier research results [12,18,22,23,37]. An average of 7.2 alleles were recorded across all the 15 loci where individuals counted 4 to 12 alleles. In this study, the mean number of alleles was remarkably higher than the one reported by Ref. [12]; where the mean was 6.7 alleles per locus for 200 sorghum accessions and was lower than that reported by Ref. [38]; where the mean value was 19 alleles per locus for 3367 sorghum accessions using 41 microsatellite markers collected around the world. These variations could be due to the diverse nature of the sorghum accessions and the level of difference in the polymorphism of the SSR markers used.

In the present study, the PIC values ranged from 0.68 to 0.89 and are comparable to a similar recent study on the same crop [12]; who reported a PIC value varying from 0.06 to 0.81 for 39 microsatellite markers in 200 accessions, and by Danquah et al. (2019) with a PIC value ranging from 0.05 to 0.78. The mean PIC value was 0.80 in this study which is greater than the one reported by Refs. [12,39,40]; suggesting the microsatellite loci have greater resolution capacity to show variations among sorghum accessions [41]. This implied that the selected pair of primers are useful genetic tools to uncover the genetic structure of sorghum in this study.

The genetic diversity (h) values varied from 0.73 to 0.90 with a mean of 0.83 in this study and was considerably greater than the reported value by Ref. [38]. In this study, the increase in genetic diversity implied the presence of significant genetic diversity along sorghum accessions [39], indicating the possibility of improving sorghum agronomic traits through breeding, and its importance for germplasm conservation.

4.2. Population genetic diversity

More than 9432 sorghum accessions were collected from various regions of Ethiopia and conserved in the gene bank of Ethiopian Biodiversity Institute (EBI) (https://www.%20ebi.gov.et). Of these genetic resources, very few portions of them are studied in a similar way. This study helps breeders to uncover promising sorghum parental lines those have great importance in the sorghum breeding programs. In addition, it is important to further utilise the existing genetic assets for breeding purposes and also design and implement appropriate germplasm management strategies, as all previous studies revealed the existence of large genetic variations in Ethiopian

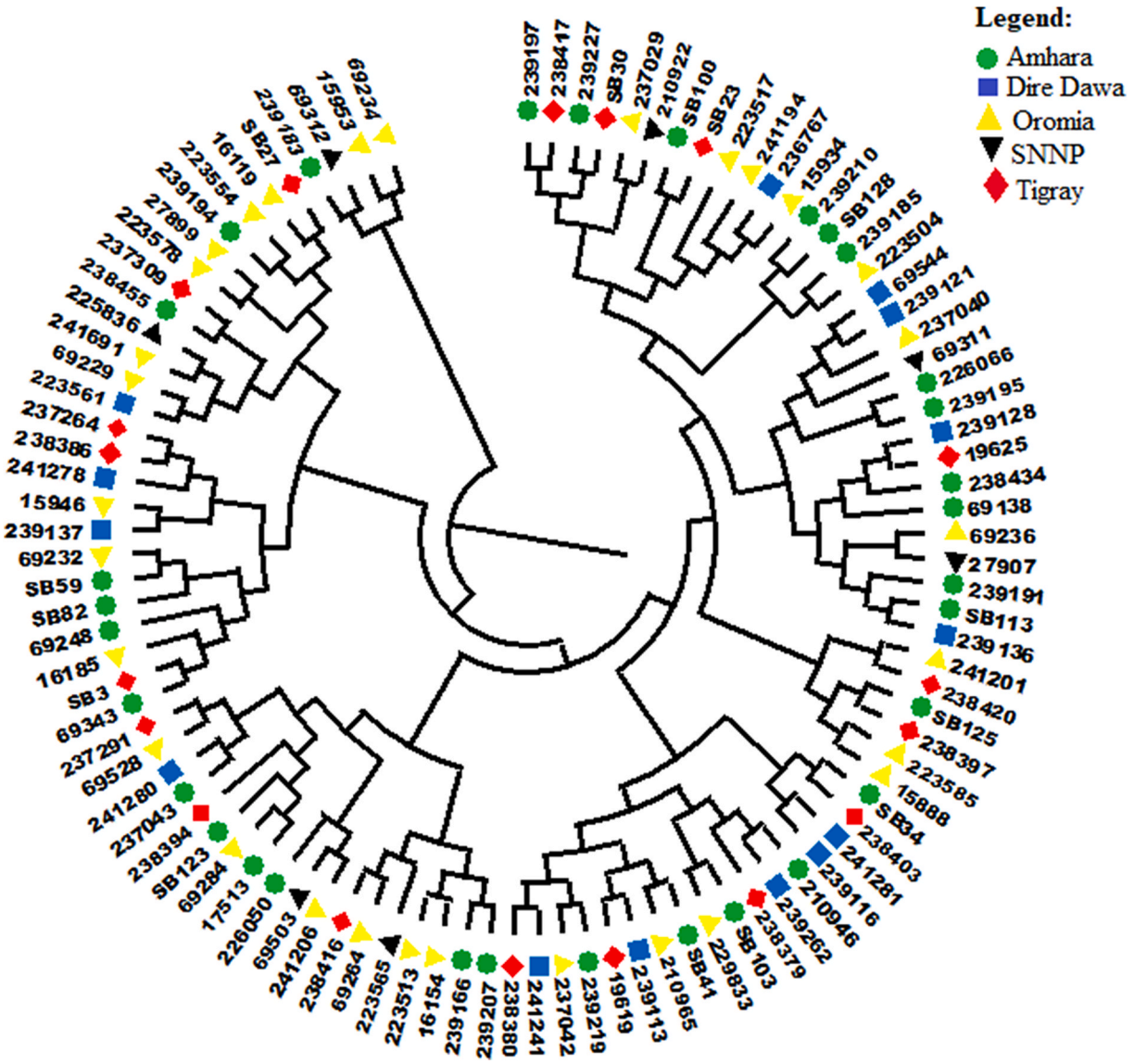


Fig. 2. The relationship of hundred sorghum accessions using unweighted neighbor-joining (UNJ) tree.

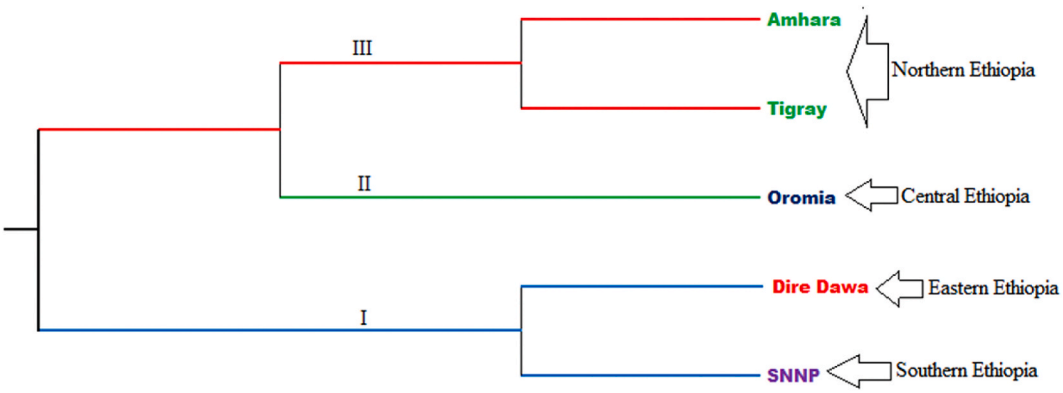


Fig. 3. Relationship among 5 populations of sorghum accessions using unweighted pair-group method with arithmetic mean (UPGMA).

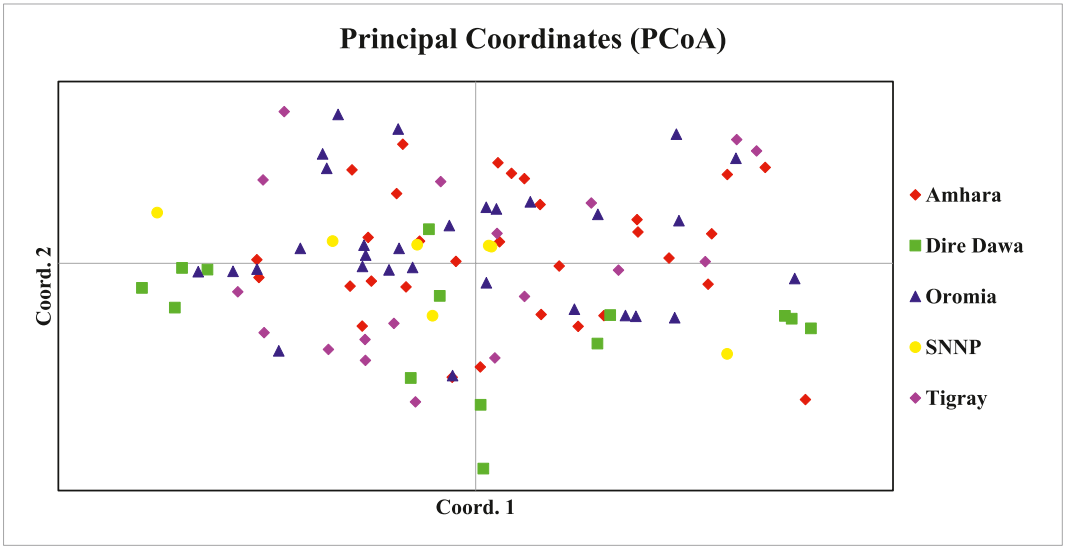


Fig. 4. The principal coordinate analysis (PCoA) of 100 sorghum accessions. SNNP = Southern Nations, Nationalities, and People's Region.

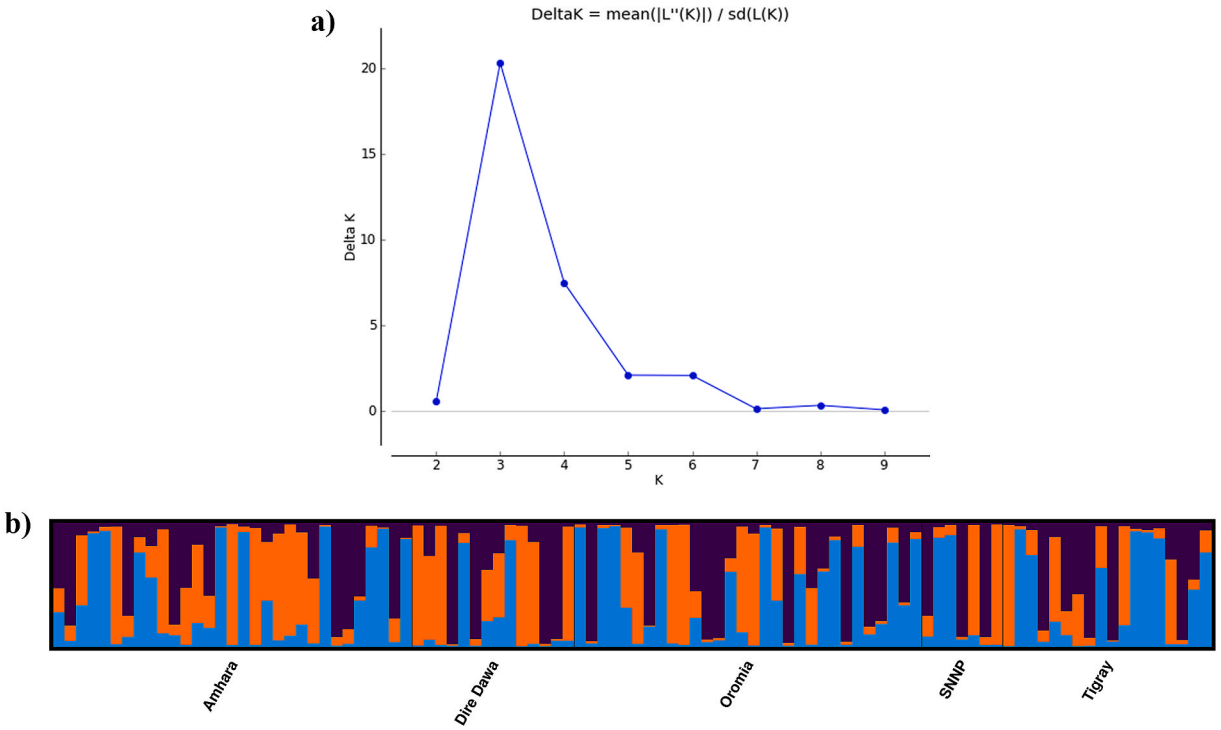


Fig. 5. Population structure of 100 sorghum accessions representing five populations in Ethiopia. (a) Outstanding delta K = 3 (peak) value by Evano et al. (2005) method; (b) Expected population structure along with geographical locations. SNNP = Southern Nations, Nationalities, and People's Region.

sorghum germplasm. Hence, the focus of this research was to unlock the population structure of sorghum accessions and their genetic diversity recovered from various regions of Ethiopia to generate useful information that will supplement previous studies to aid in development of better-performing verities, and also proper management of the available sorghum genetic resources. The values of shannon's information index (1.66), effective alleles (5.01), observed alleles (6.21) and Nei's gene diversity (0.71) showed a high average for all the five populations. According to the report of [42]; the value of the mean gene diversity (0.57) was lower than the one obtained in the current study, that was 0.78. The presence of extensive variations across accessions of sorghum representing different geographic regions of Ethiopia showed a higher gene diversity in this study.

In the current study, the private alleles were detected only in the population of Oromia (0.13) and the percentage value of polymorphic loci was high in all the five populations. This is considerably lower than a report from East and Central Africa (NPA = 1.24) and a previous study from Ethiopia (NPA = 1.88) [41]. Since Ethiopia is believed to be sorghum's center of diversity, the detected private alleles in the sorghum population of Oromia could be the result of gene flow from wild types that have ample relatives in close proximity along with the cultivated ones arising from various gene pools of normal hybridization. Hence, they are very important genetic sources for future sorghum improvement programs in Ethiopia [22] to boost the breeding activities of the crop, as the diverse nature of its germplasm has not been widely utilized. The increased value of gene diversity (0.81), Shannon's information index (1.80), percentage of polymorphic loci (0.20) and successful alleles (5.71) within sorghum populations of Oromia was an indication that the region is very appropriate for further analysis of the genetic variability of sorghum, germplasm conservation purpose and also as a source of useful alleles for breeding purposes.

4.3. Population genetic structure

The observed high gene flow among populations and sexual recombination within the population was probably due to the greater within population genetic variations than among population genetic variations in the analysis of molecular variance. The genetic diversity among geographical regions and altitudes in the major agro-ecologies of Ethiopia [7], was lower than the diversity within them the analysis revealed. The sorghum accessions studied were differentiated accounting just 2.32% of the overall genetic dissimilarity according to the result of the analysis of variance. Similar lower proportions of among population genetic variations in Kenya [43], South Africa [11], Ethiopia (Tirfessa et al., 2020) and China [14] were reported for S. *bicolor* populations as 2.75%, 9.56%, 0.38% and 8%, respectively. For the same crop, variable rates of genetic differentiation were reported from different parts of the world. For *S. bicolor* populations in Ethiopia [16] and China [14], a statistically significant moderate to high among population genetic differentiation (0.03–0.08 and 0.08 to 0.082) was reported, respectively. High genetic variability within the population (97.68%) could be attributed to sexual recombination and spontaneous mutation [44] in the current study.

According to Ref. [45]; gene flow that encloses various systems of gene interchange across populations can affect their genetic variation. In this study, seed exchange practices between communities and marketing [46] as well as socio-economic conditions may have facilitated a high gene flow (Nm = 10.53) among populations from different seed collection regions leading to a reduced among-populations genetic differentiation (Fst: 0.023) [46]. A similar result was observed for geographical regions and major agro-ecologies (Nm = 12.64, Fst: 0.019 and Nm = 34.47, Fst: 0.01) with significant magnitude ($P < 0.0001$) of genetic variations respectively. The principal coordinate analysis revealed that presence of a high within populations than among populations of sorghum genetic variation level showed a mixing up along the axis [47], rather particular categories were unsuccessful to yield definite group. This could be due to the movement of genetic materials from one growing region to the other.

The result of STRUCTURE software divulged three (K = 3) mini populations were obtained from all the hundred accessions collected from the five different regions of Ethiopia. The existence of high gene flow among collection regions at bordering areas through pollen grain flow and germplasm exchange through common marketing places was evidenced by the results of principal coordinate analysis (PCoA), unweighted pair-group method with arithmetic mean (UPGMA) dendrogram and unweighted neighbor-joining tree (UNJ) in the STRUCTURE analysis. Previous genetic structure analysis in the same crop has also shown the presence of poor population clustering in sorghum accessions of India [47] and Ethiopia [22,48].

Utilizing genome sequencing technologies to characterize untapped genetic resources genome-wide opens up new possibilities for effective crop genetic materials use and sustainable breeding [49] in the future. We suggest further research on sorghum genetic diversity and population structure in Ethiopia that covers a wide geographic area and a larger sample size through single nucleotide polymorphism (SNP) markers to achieve better genome coverage and resolution since the entire genome of sorghum is already sequenced [50,51].

5. Conclusion

Microsatellite markers were used to produce valuable evidence for sorghum breeding programs, effective development of various important traits, conservation purposes, and germplasm management strategies to explore the genetic diversity and population structure of Ethiopian sorghum landraces. A high population polymorphism was detected and the genetic materials were also found to be relevant to aid the degree of genetic variability of S. *bicolor* in the whole microsatellite markers utilized. A comparatively increased genetic diversity (0.81) was obtained from Amhara and Oromia out of the five populations studied. Therefore, these regions could be central sources for desirable alleles for breeding values, and also for germplasm conservation. Finally, we suggest further studies based on high-density markers that include accessions from other parts of Ethiopia to generate a better insight from the whole nation to disclose the genetic structure and diversity of sorghum, and also to develop the best-performing varieties.

Author contribution statement

All authors have significantly contributed to the writing and development of this study.

Funding statement

This research was assisted financially by the Swedish International Development Cooperation Agency (Sida); sub-project 'Sorghum:

screening and association mapping of drought tolerant sorghum genotypes in Ethiopia' - SP1. The Ethiopian Biodiversity Institute (EBI) also supported the study.

Data availability statement

No data was used for the research described in the article.

Declaration of interest's statement

The authors declare no conflict of interest.

Acknowledgements

We thank the Ethiopian Biodiversity Institute (EBI) for providing sorghum accessions for the study and Melkassa Agricultural Research Center (MARC) for allowing the research experiment sites in different agro-ecologies of Ethiopia.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.heliyon.2023.e12830.

References

[1] M.A. Ali, K. Jabran, S.I. Awan, A. Abbas, M. Zulkiffal, T. Acet, A. Rehman, Morpho-physiological diversity and its implications for improving drought tolerance in grain sorghum at different growth stages, Aust. J. Crop. Sci. 5 (2011) 311.
[2] S.L. Dillon, P.K. Lawrence, R.J. Henry, The new use of sorghum bicolor-derived SSR markers to evaluate genetic diversity in 17 Australian sorghum species, Plant Genet. Resour. 3 (2005) 19–28.
[3] J.D. Snowden, The wild fodder sorghums of the section eu-sorghum, J. Linn Soc. London 55 (1955) 1–191.
[4] FAOSTAT, Food and Agriculture Organization of the United Nations, FAO, Rome, 2019.
[5] K. Hariprasanna, S. Rakshit, Economic importance of sorghum, in: S. Rakshit, Y.H. Wang (Eds.), Compendium of Plant Genomes, vol. 1, Sorgsd., 2016, pp. 1–25.
[6] H. Doggett, The development of the cultivated sorghums, Essays Crop Plant Evol. (1965) 50–69.
[7] A. Ayana, T. Bryngelsson, E. Bekele, Geographic and altitudinal allozyme variation in sorghum (Sorghum bicolor (L.) Moench) landraces from Ethiopia and Eritrea, Hereditas 135 (2001) 1–12.
[8] R.W. Allard, Principles of Plant Breeding, John Wiley and Sons, 1999.
[9] M. Bohn, H.F. Utz, A.E. Melchinger, Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance, Crop Sci. 39 (1999) 228–237.
[10] G.P. Morris, P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, J. Harriman, Population genomic and genome-wide association studies of agroclimatic traits in sorghum, Proc. Natl. Acad. Sci. USA 110 (2013) 453–458.
[11] T. Motlhaodi, M. Geleta, S. Chite, M. Fatih, R. Ortiz, T. Bryngelsson, Genetic diversity in sorghum [Sorghum bicolor (L.) Moench] germplasm from Southern Africa as revealed by microsatellite markers and agro-morphological traits, Genet. Resour. Crop Evol. 64 (2017) 599–610.
[12] A. Tirfessa, T. Tesso, A. Adugna, M. Hussein, K. Dan, Genetic diversity among Ethiopian sorghum [Sorghum bicolor (L.) Moench] gene bank accessions as revealed by SSR markers, Afr. J. Biotechnol. 19 (2020) 84–91.
[13] J.A. Ruiz-Chutan, J. Salava, D. Janovska, J. Ziarovska, M. Kalousova, E. Fernandez, Assessment of genetic diversity in sorghum bicolor using RAPD markers, Genetika 51 (2019) 789–803.
[14] M. Zhu, J. Chen, N. Yuyama, L. Luo, X. Xiao, Y. Lv, Y. Liu, H. Cai, Genetic diversity and population structure of broomcorn sorghum investigated with simple sequence repeat markers, Trop. Plant Biol. 13 (2020) 62–72.
[15] M.B. Enyew, T. Feyissa, A.S. Carlsson, K. Tesfaye, C. Hammenhag, M. Geleta, Genetic Diversity and population structure of sorghum [Sorghum Bicolor (L.) Moench] sccessions as revealed by single nucleotide polymorphism markers, Front. Plant Sci. 12 (2021), 799482.
[16] A. Adugna, A.A. Snow, P.M. Sweeney, E. Bekele, E. Mutegi, Population genetic structure of in situ wild sorghum bicolor in its Ethiopian center of origin based on SSR markers, Genet. Resour. Crop Evol. 60 (2013) 1313–1328.
[17] A. Adugna, Analysis of in situ diversity and population structure in Ethiopian cultivated sorghum bicolor (L.) landraces using phenotypic traits and SSR markers, SpringerPlus 3 (2014) 212.
[18] Z. Wondimu, H. Dong, A.H. Paterson, W. Worku, K. Bantte, Genetic diversity, population structure, and selection signature in Ethiopian sorghum [Sorghum bicolor L. (Moench)] germplasm, G3-Genes Genom. Genet. 11 (2021) 6–7.
[19] D.T. Rosenow, J.A. Dahlberg, Collection, conversion and utilization of sorghum, in: C. Wayne, A. Richard Smith (Eds.), Sorghum: Origin, History, Technology and Production, vol. 471, JWS, 2000, p. 3, 24237.
[20] G. Brhane, Utilization of germplasm in sorghum improvement, Proc. Int. Seaweed Symp. 2 (1982) 7.
[21] DArT, Plant DNA Extraction Protocol for DArT. Diversity Array Technology, 2019.
[22] T. Disasa, Genetic Diversity Analysis and QTL Mapping of Selected Traits in Sorghum (Sorghum Bicolor (L.) Moench), Doctoral Dissertation, Addis Ababa University, Addis Ababa, Ethiopia, 2016.
[23] G. Olani, T. Feyissa, T. Disasa, Genetic diversity analysis of sorghum [Sorghum bicolor (L.) Moench] races in Ethiopia using simple sequence repeats (SSR) markers, Afr. J. Biotechnol. 20 (2021) 431–439.
[24] A.B. Pavel, C.I. Vasile, PyElph-a software tool for gel images analysis and phylogenetics, BMC Bioinf. 13 (2012) 1–6.
[25] K. Liu, S.V. Muse, Powermarker: an integrated analysis environment for genetic marker analysis, Bioinformatics 21 (2005) 2128–2129.
[26] F.C. Yeh, R. Yang, Popegene Version 1.31: Microsoft Window-Based Freeware for Population Genetic Analysis, vol. 1, Department of renewable resources. University of, Alberta, Canada, 1999, p. 29.
[27] R. Peakall, P.E. Smouse, GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research, Mol. Ecol. Notes 6 (2006) 288–295.
[28] R. Peakall, P.E. Smouse, GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update, J. Bioinform. 28 (2012) 2537–2539.
[29] L. Excoffier, H. Lischer, An Integrated Software Package for Population Genetics Data Analysis. Computational and Molecular Population Genetics Lab (CMPG), Institute of Zoology, University of Berne, Switzerland, 2006.

[30] M. Nei, N. Takezaki, Estimation of genetic distances and phylogenetic trees from DNA analysis, in: Proc 5th World Cong Genet Appl Livstock Prod. 21, 1983, pp. 405–412.
[31] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, Evolution 4 (1985) 783–791.
[32] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, Mega x: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547.
[33] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2000) 945–959.
[34] D.A. Earl, Structure harvester: a website and program for visualizing structure output and implementing the Evanno method, Conservation 4 (2012) 359–361.
[35] G. Evanno, R. Sebastien, G. Jerome, Detecting the number of clusters of individuals using the software structure: a simulation study, Mol. Ecol. 14 (2005) 2611–2620.
[36] N.M. Kopelman, J. Mayzel, M. Jakobsson, N.A. Rosenberg, I. Mayrose, Clumpak: a program for identifying clustering modes and packaging population structure inferences across K, Mol. Ecol. Resour. 15 (2015) 1179–1191.
[37] D. Ng'uni, M. Geleta, P. Hofvander, M. Fatih, T. Bryngelsson, Comparative genetic diversity and nutritional quality variation among some important Southern African sorghum accessions [Sorghum bicolor (L.) Moench], Aust. J. Crop. Sci. 6 (2012) 56.
[38] C. Billot, P. Ramu, S. Bouchet, J. Chantereau, M. Deu, L. Gardes, J.L. Noyer, J.F. Rami, R. Rivallan, Y. Li, P. Lu, Massive sorghum collection genotyped with SSR markers to enhance use of global genetic resources, PLoS One 8 (2013) 4.
[39] A. Mofokeng, H. Shimelis, P. Tongoona, M. Laing, A genetic diversity analysis of South African sorghum genotypes using SSR markers, Afr. J. Plant Soil. 31 (2014) 145–152.
[40] A. Danquah, I.K. Galyuon, E.P. Otwe, D.K. Asante, Genetic diversity in some Ghanaian and Malian sorghum [Sorghum bicolor (L) Moench] accessions using SSR markers, Afr. J. Biotechnol. 18 (2019) 591–602.
[41] S.A. Salih, L. Herslman, M.T. Labuschange, A.H. Mohammed, Assessment of genetic diversity of sorghum [Sorghum bicolor (L.) Moench] germplasm in East and central Africa, World J. Microbiol. Biotechnol. 1 (2016) 113–120.
[42] A. Deresa, A. Tesso, A. Adugna, Sorghum production and research experiences, Afr. J. Biotechnol. 8 (2005) 18–29.
[43] C.W. Muui, R.M. Muasya, D.T. Kirubi, S.M. Runo, A. Karugu, Genetic variability of sorghum landraces from lower Eastern Kenya based on simple sequence repeats (SSRs) markers, Afr. J. Biotechnol. 15 (2016) 264–271.
[44] S. Berraies, M.S. Gharbi, F. Belzile, A. Yahyaoui, M.R. Hajlaoui, M. Trifi, M. Jean, S. Rezgui, High genetic diversity of Mycospaherella graminicola (Zymoseptoria tritici) from a single wheat field in Tunisia as revealed by SSR markers, Afr. J. Biotechnol. 2 (2013) 12, 2013.
[45] M. Slatkin, Gene flow and the geographic structure of natural populations, Science 236 (1987) 787–792.
[46] M. Dalvand, D. Zafari, M.J. Soleimani Pari, R. Roohparvar, S.M. Tabib Ghafari, Studying genetic diversity in Zymoseptoria tritici, causal agent of Septoria Tritici Blotch, by using ISSR and SSR markers, J. Agric. Sci. Technol. 20 (2018) 1307–1316.
[47] S. Pandian, L. Satish, J. Shilpha, M. Ramesh, Genetic diversity analysis reveals strong population structure in sorghum germplasm collection. Proceedings of the National Academy of Sciences, India Section B, Biol. Sci. 90 (2019) 179–190.
[48] D.K. Weerasooriya, F.R. Maulana, A.Y. Bandara, A. Tirfessa, A. Ayana, G. Mengistu, K. Nouh, T. Tesso, Genetic diversity and population structure among sorghum (Sorghum bicolor, L.) germplasm collections from western Ethiopia, Afr. J. Biotechnol. 15 (2016) 1147–1158.
[49] S. Chakrabarty, R. Mufumbo, S. Windpassinger, D. Jordan, E. Mace, R.J. Snowdon, A. Hathorn, Genetic and genomic diversity in the sorghum gene bank collection of Uganda, BMC Plant Biol. 22 (2022) 1–11.
[50] A.H. Paterson, J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Poliakov, J. Schmutz, The Sorghum bicolor genome and the diversification of grasses, Nature 457 (2009) 551–556.
[51] R.F. McCormick, S.K. Truong, A. Sreedasyam, J. Jenkins, S. Shu, D. Sims, M. Kennedy, M. Amirebrahimi, B.D. Weers, B. McKinley, A. Mattison, The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization, Plant J. 93 (2018) 338–354.