Journal of Crop Science and Biotechnology (2022) 25:359–371
https://doi.org/10.1007/s12892-021-00137-x

ORIGINAL RESEARCH

Genetic diversity of Sesame (*Sesamum indicum L*) using high throughput diversity array technology

Tewodros Tesfaye[1,2] · Kassahun Tesfaye[1] · Gemechu Keneni[3] · Cathrine Ziyomo[4] · Tesfahun Alemu[3]

Accepted: 17 November 2021 / Published online: 3 February 2022

Abstract

Sesame is an important oil crop widely cultivated in Africa and Asia. Characterization of genetic diversity and population structure of sesame genotypes in these continents can be used to designing breeding methods. In the present study, 300 genotypes comprising 209 Ethiopian landraces, and 75 exotic collections from different African and Asia countries, and 16 varieties were used. The panel was genotyped using two high-throughput diversity array technology markers. A total of 6115 silicoDArT and 6474 SNP markers were reported, of which 5065 silicoDArT and 5821 SNP markers were aligned with the reference sesame genome. For further analysis, it was filtered with an allele frequency for each SNP site and left 2997 high-quality SNPs. All genotypes used in this study were descended from eight geographical origins. The average diversity of the panel was 0.14. Considering the genotypes based on their geographical origin, Africa collections (0.21) without Ethiopian collection was more diverse, when further portioned Africa, North Africa (0.23) collection was more diverse than others, but at the continent level, Asia (0.17) was more diverse than Africa (0.14). The genetic distance among the populations was ranged from 0.015 to 0.394. The populations were clustered into four groups. The structure analysis was divided into four hypothetical ancestral populations and 21 genotypes were an admixture. This indicate genotypes from the same origin didn't classify on the country of origin. The genetic diversity and population structure guide future research work to design association studies and the systematic utilization of genetic resource.

Keywords DArTseq · Genetic diversity · Population structure · silicoDArT · SNP

Introduction

Sesame (*Sesamum indicum L.*, $2n = 26$), is a member of the Pedaliaceae family, and one of the ancient oil crops is grown widely in tropical and subtropical areas since the time long past (Bedigian and Harlan 1986). Sesame is produced in different parts of the world for various purposes but more than 96% of the world's sesame seed production is covered by Africa and Asia (FAOSTAT 2019). Sesame seeds are good sources of fat, protein, carbohydrates, fiber, and essential minerals.

In Ethiopia, sesame is among the foremost important oil crops both in terms of area coverage and total national annual production (CSA 2019). However, the farm level productivity of sesame in Ethiopia is very low (0.68t ha^{-1}) (CSA 2018) compared to the genetic potential of the crop yield of 2 t ha^{-1} (Mkamilo and Bedigian 2007) and other countries like Egypt (1.29 t ha^{-1}), Nigeria (1.1 t ha^{-1}) Tanzania (1 t ha^{-1}), and china (1.4 t ha^{-1}) (Sharaby and Butovchenko 2019). Improved varieties released in Ethiopia are reported to yields ranging 0.3 to 1.3 t ha^{-1} under rainfed and 1 to 2.4 t ha^{-1} under irrigation on research stations (Gebremichael 2017).

Ethiopia is considered one of the centers of genetic diversity of sesame crop and has an immense wealth of genetic diversity in the germplasm collections that can be exploited

✉ Tewodros Tesfaye
tesfaye.tewodros@yahoo.com

Tesfahun Alemu
tesfaye.tewodros@yahoo.com

1 College of Natural Sciences, Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, P. O. Box 1176, Addis Ababa, Ethiopia

2 Amhara Agricultural Research Institute, Gondar Research Center, P. O. Box 1337, Gondar, Ethiopia

3 Ethiopian Institute of Agricultural Research, Holeta Research Center, P. O. Box 2003, Holeta, Ethiopia

4 International Livestock Research Institute, Box 30709, Nairobi 00100, Kenya

through genetic improvement in the breeding program (Institute of Biodiversity Conservation (IBC) 2012). This fact also describes from the time of N.I. Vavilov (Vavilov 1951) based on morphological level studies (Gidey and Kebede 2012; Teklu and Kebede 2014; Abate et al. 2015). The effective utilization of these collections requires a systematic genetic characterization, assessment of genetic diversity, and identification of potential putative genes that can be used in the breeding program. The genetic diversity can be assessed using morphological, biochemical, and molecular markers. Several studies employ molecular markers to assess genetic diversity using various types of markers (Gebremichael and Heiko 2011; Abate and Mekbib 2015; Teshome et al. 2015). Among these markers amplified fragment length polymorphism (Laurentin and Karlovsky 2006), sequence-related amplified polymorphisms (Zhang et al. 2010, 2012), inter-simple sequence repeat (Kumar et al. 2012), simple sequence repeats (Park et al. 2011), expressed sequence tag (Farshadfar and Farshadfar 2008; Zhang et al. 2012), and insertions and deletions (Wu et al. 2014) have been used elsewhere for the analysis of germplasm genetic diversity.

More recently high-thorough put marker systems particularly single-nucleotide polymorphisms and Diversity Arrays Technology (DArT) markers have become the genetic markers of choice for genetic analyses including characterization of germplasm because of the efficiency, low cost, speed, and abundant in the genome (Gupta et al. 2008; Wei et al. 2015; Cui et al. 2017; Basak et al. 2019). SilicoDArT markers are dominant microarray markers and scored for the presence or absence of a single allele, whereas DArTseq based SNPs are co-dominant markers, both of them being successfully applied in genetic diversity (Wenzl et al. 2004; Yang et al. 2006; Bolibok-Bragoszewska et al. 2009; Sánchez-Sevilla et al. 2015; Tang et al. 2015) and population structure study of several crop species (Matthies et al. 2012; Laidò et al. 2013).

Even with all these technologies available to study genetic diversity and the importance of sesame crop as nutritional and economic importance in several parts of the world little research work was done at the national and international levels (Bedigian and Harlan 1986). However, in any breeding program assessing the genetic diversity from the diverse origins is a crucial step since the germplasm obtained from different regions can provide variants created due to the geographical adaptation that can be exploited in the breeding program. So to capture this information and incorporate it in the future breeding program, the breeder should study the genetic diversity of the available genetic pool including genotypes from diverse geographical locations. But in Ethiopian conditions, there is no comprehensive report on the genetic diversity of sesame genotypes obtained from different sources and used by the breeding program in the country using new biotechnology tools (such as SNPs). So, this showed the importance of developing research activities to assess the genetic diversity of the sesame germplasm collections available in the Ethiopian Biodiversity institute and farmers' hands to exploit it in the future breeding program to solve the sesame production constraint in the country.

Therefore, this research work was done with the objective of:-

- Assessing the genetic diversity of the existing sesame populations in Ethiopia,
- Understand the population structure of the sesame population and
- Define direction how the available diversity can be exploited in the sesame breeding program in the country.

Materials and methods

Plant materials

A total of 300 genotypes comprised of 209 Ethiopian landraces, 75 exotic collections, and 16 released varieties were used in this study. The Ethiopian landraces were collected from areas between 1931 and 2008 a.m.s.l and distributed in Amhara (56), Benshangul-Gumz (BG) (38), Oromia (52), SNNP (3), and Tigray (60) regions. The introduced germplasms were obtained from North Africa (Egypt) (27), South Africa (Zambia and Zimbabwe) (18), West Africa (Burkina Faso) (17), and East Africa (Kenya, Somalia, and Sudan) (6) without including the Ethiopian collection and from four different Asian countries (China, Japan, Israel and Philippines) (7). The germplasms were kindly provided by the Ethiopian Biodiversity Institute (EBI) and regional and federal research centers and the geographical location of the collections was presented in (Fig. 1 and Supplementary file Table S1). All genotypes showed wide ranges of variation for most of the quantitative and qualitative morphological traits studied; except plant growth type, leaf glands, anther filament colour, anther connective tip gland and anthocyanin colouration of capsule. This output is published in African crop science journal (Tewodros et al. 2021).

DNA extraction

The DNA of the sesame genotypes is extracted from the seed obtained in the previous harvest. Then 800 µl Lysis buffer was added to the sample of each genotype powder for the tan bead DNA extraction process. Lysis buffer is a solution for the purpose of breaking open cells for use in molecular biology experiments that analyze the labile macromolecules of the cells. The nucleic acid of the samples was extracted with an Automated Nucleic acid Extractor (Maelstrom series).

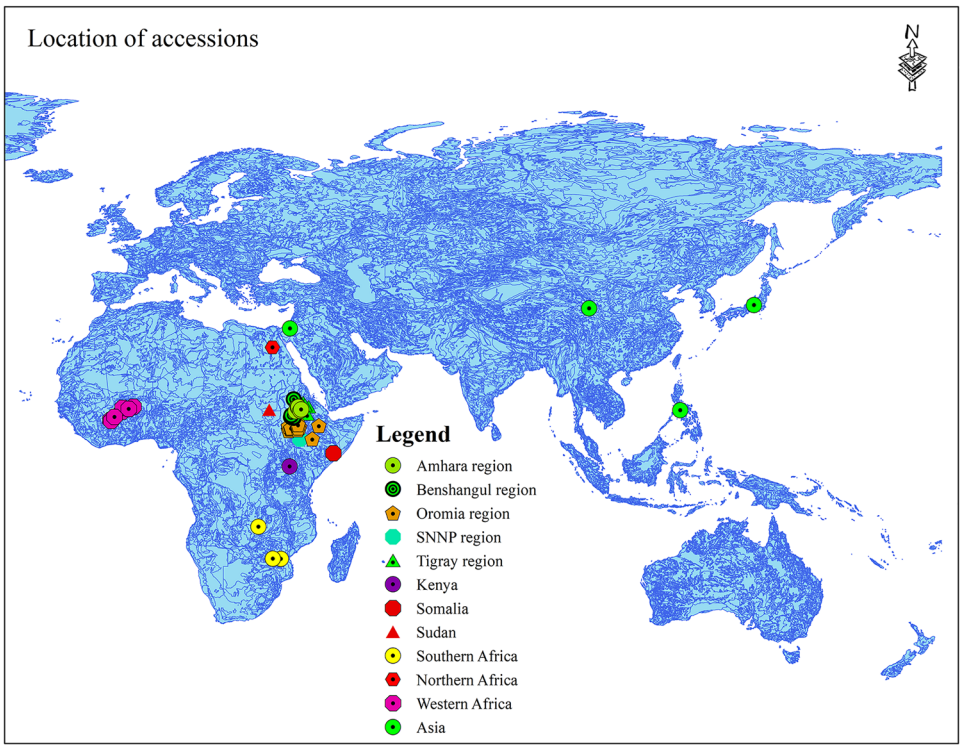

Fig. 1 Map of collection areas of sesame genotypes

At the end of the program, DNA quality was evaluated on 0.8% agarose gels and it was adjusted to 50 ng/µl for GBS analysis.

GBS library preparation and sequencing

DArTseq combines genome complexity reduction methods and next-generation sequencing platforms (Courtois et al. 2013; Cruz et al. 2013; Kilian et al. 2016; Raman et al. 2014). Therefore, DArTseq represents a new implementation of the sequencing of complexity-reduced representations (Huang et al. 2014) and more recent applications of this concept on the next-generation sequencing platforms (Bastien et al. 2014; Sonah et al. 2013). DArTseq libraries (96-plex) were prepared for the 300 accessions using 50 ng of DNA each. Libraries were constructed according to (Kilian et al. 2012) DArTseq complexity reduction method which involved digestion of DNA samples with a rare cutting enzyme *Pst*I, paired with a set of secondary frequently cutting restriction enzyme *Mse*I, ligation of barcoded adapters followed by PCR amplification of adapter ligated fragments.

Next-generation sequencing technology was implemented using the sequencer HiSeq2500 (Illumina, USA) to detect SNPs and silicoDArT markers. SilicoDArTs represent DArT markers that represents the presence/absence of restriction fragment and represents dominant markers and are scored in a binary format "1" = Presence and "0" = Absence of restriction fragment with the marker sequence in the genomic representation of the sample. "-" represents calls with non-zero counts but too low to score confidently as "1" (often representing heterozygotes). Single Nucleotide Polymorphism (SNPs) can be defined as a variation in the base composition of a single nucleotide position within a specific locus of a single chromosome of the haploid set. In standard format, SNPs markers were presented for reference and SNP alleles for each marker and genotype. This format of SNPs can be converted to other formats if required. The report was prepared as a binary or read counts file, or both depending on the order specifications. Two technical replicates of the DNA samples of each of 21 accessions were genotyped to calculate the reproducibility of the marker data. Thereafter, the SNPs and SilicoDArTs obtained were run against the sesame reference genome database (https://www.ncbi.nlm.nih.gov/genome/?term=sesame) to understand on which chromosomes of sesame the SNPs and SilicoDArTs were located. In addition, the updated genome assembly was consider for the linkage group (Wang et al. 2016). Genotype by sequencing (GBS) of the materials was carried out at the Biosciences eastern and central Africa- International Livestock Research Institute (BecA- ILRI) in Kenya.

Quality analysis of marker data

The markers were tested for reproducibility (%), call rate (%), polymorphism information content (PIC), one ratio, and minor allele frequency (MAF). Scoring of reproducibility involved the proportion of technical replicate assay pairs for which the marker score exhibited consistency. The call

rate determined the success of reading the marker sequence across the samples and was estimated from the percentage of samples for which the score was either '0' or '1'. PIC is the degree of diversity of the marker in the population and showed the usefulness of the marker for linkage analysis. One ratio constitutes the proportion of the samples for which genotype scores equaled '1'.

Data analysis

DArTseq markers were mapped using the consensus map version 4.0 (www.diversityarrays.com) developed by DArT Pty. Ltd., Australia, and the updated genome assembly and annotation issued from the Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, available online at https://www.ncbi.nlm.nih.gov/genome/?term=sesame and (Wang et al. 2016).

Data cleaning and genetic diversity analysis

DArTseq raw data were filtered according to markers criterion; minor allele frequency (MAF) > 0.01 and missing data ≤ 25%. The summary statistics of the filtered DArTseq markers such as the expected heterozygosity (He) or genetic diversity (GD), minor allele frequency, and the polymorphic information content (PIC), were calculated using Power Marker v 3.25 (Lui 2005). PIC was estimated based on the probability of finding polymorphisms between any two random samples while gene diversity is defined as the probability of two randomly chosen alleles from the population is different.

Genetic diversity analysis

The genetic distance between the tested 300 sesame genotypes and between the populations was estimated using the Euclidean and Nei genetic distance (Nei 1972) using the R software KD compute plugin system.

Clustering analysis was done based on the Euclidean genetic distance and the ward D2 agglomeration (Ward 1963) method and similarly, the dendrogram was developed. In line with it, the principal component analysis was done to assess the distribution of the genotypes according to the KD compute plugin system.

Allele No, Gene Diversity, Heterozygosity, PIC, and Major Allele Frequency were estimated using Power Marker genetic analysis package (version 3.25); (Liu and Muse 2005). The software Arlequin V3.5 (Excoffier and Lischer 2010) was used to calculate the genetic variation between and within geographical groups with an analysis of molecular variance (AMOVA).

Population structure

The Bayesian clustering analysis approach was used to analyze the genetic structure of the sesame genotypes using STRUCTURE v.2.3.4 software (Pritchard et al. 2000). For the analysis, five individual Markov Chain Monte Carlo (MCMC) simulations were conducted for each K-value from 1 to 11 with a burning length of 50,000, followed by 100,000 iterations. The admixture model was applied without using any prior population information and correlated allele frequencies were also employed.

The STRUCTURE results were subsequently analyzed by the STRUCTURE HARVESTER application (Earl and vonHoldt 2012) to identify a distinct peak in the change of likelihood (ΔK) at the true value of K. CLUMPAK: "a program for identifying clustering modes and packaging population structure inferences across K" (CLUMPAK server) was used. Each sesame accession was then assigned to a cluster (k) based on a membership coefficient determined by STRUCTURE V2.3.4, The cut-off probability for assignment to a cluster was > 0.50 for the clusters.

Results

Marker discovery by DArTseq

A total of 6115 polymorphic silicoDArT and 6474 SNP markers were generated of which 5065 silicoDArT and 5821 SNP were aligned with the Reference sesame genome obtained from https://www.ncbi.nlm.nih.gov/genome/?term=sesame and updated genome assembly and annotation available at (Wang et al. 2016) and the remains markers were scaffold, and unknown markers (Table 1).

The allele frequency for each SNP site was calculated: the MAF of the SNPs varied from 0 to 49.6%, with an average of 5.1%, and ~61.29% of the SNPs had a low frequency (MAF < 0.05) across the 300 accessions. After excluding the SNPs with a MAF < 0.01, there were left 2997 (~64.61%) high-quality SNPs evenly distributed across the whole genome that could be used for further analysis.

The aligned markers were distributed on all 13 chromosome of sesame with an average of 389.62 silicoDArT and 447.7 SNP markers per chromosomes. The maximum number of silicoDArT (643) and SNP (733) was found on chromosomes 6 and 3, respectively (Table 1).

Table 1 Distribution of DArTseq markers on different sesame chromosomes

Chromosome no.	Chromosome size (Mb)*	Number of silicoDArT markers	silicoDArT marker/Mbp	No. of SNP markers	SNP marker/Mbp
1	20.26	391	19.35	539	26.60
2	18.42	407	22.10	483	26.22
3	25.85	545	21.08	733	28.36
4	20.59	321	15.59	378	18.36
5	16.58	397	23.94	381	22.98
6	25.97	643	24.80	662	25.49
7	16.76	234	13.90	238	14.20
8	26.18	492	18.79	539	20.59
9	22.85	505	22.10	533	23.33
10	19.49	332	17.03	387	19.86
11	14.05	301	21.42	399	28.40
12	16.28	255	15.60	315	19.35
13	16.47	242	14.69	234	14.21
Scaffold	–	724	–	305	–
Unknown position	–	326	–	348	–
Total	259.73	6115	23.54	6474	24.93

*Indicates the chromosomal size taken from the reference genome published by Wang et al.

Analysis of genetic diversity

The number of accessions, number of alleles, genetic diversity, heterozygosity, and the polymorphism information content (PIC), and major allele frequency of the eight populations (Africa, Amhara, Asia, Benshangul-Gumz (BG), Improved, Oromia, Southern Nations, Nationalities, and People's Region (SNNP), and Tigray respectively) are shown in Table 2.

Table 2 Summary of the genetic diversity of the 300 sesame accessions based on their different geographical regions

Group	No. of accessions	Gene diversity (GD)	Heterozygosity	PIC	Miner allele frequency
Introduced from Asia	7	0.17	0.19	0.14	0.12
Introduced from different Africa countries except Ethiopia	68	0.21	0.15	0.18	0.14
Amhara	56	0.1	0.09	0.09	0.06
BG	38	0.07	0.07	0.06	0.04
Improved	16	0.12	0.12	0.1	0.07
Oromia	52	0.06	0.06	0.06	0.04
SNNP	3	0.08	0.08	0.06	0.05
Tigray	60	0.13	0.12	0.12	0.08
Continents					
Asia	7	0.17	0.19	0.14	0.12
Africa	293	0.14	0.1	0.12	0.08
By partition Africa					
E. Africa/ Ethiopia	231	0.1	0.09	0.09	0.06
N. Africa	27	0.23	0.19	0.19	0.17
S. Africa	18	0.15	0.13	0.13	0.1
W. Africa	17	0.14	0.11	0.12	0.1
Mean	300	0.14	0.11	0.12	0.08

PIC polymorphic information content

The genetic diversity study showed the sesame germplasm from Asia (0.17) is more diverse than Africa (0.14) (Table 2). The comparison of the genetic diversity among the different African regions showed the North African showed high genetic variability followed by South Africans in contrast to the low genetic variability registered in East Africa (Table 2). In Africa, the highest number of genotypes were considered from Ethiopia/East Africa but still show the lowest genetic diversity. When we compare the regions in Ethiopia, the maximum genetic diversity was observed in the Tigray region (0.13) followed by the Amhara region (0.10).

Genetic relationships among Germplasm

The Nei's genetic distance estimate among the 300 sesame germplasms evaluated in this study showed the highest value between "Najjo-68 (gabaa kamijaa)" and "17,712" landraces from the Oromia regions of Ethiopia. When we consider the eight populations evaluated, the maximum genetic distance was observed between the Asia and SNNP regions of Ethiopia (0.073) followed by Asia with the Oromia region (0.057) (Table 3).

Cluster analysis of the 300 germplasms derived from the eight different geographical origins was performed using the Ward D agglomeration based on the Euclidean genetic distance and clustered into four major cluster groups (Fig. 2).

The first cluster comprised the majority of the germplasms from different countries of Africa (28), all accessions that were introduced from Asia (7) and the different regions of Ethiopia, Amhara (8), Benshangul-Gumz (4), Oromia (10), SNNP (1), Tigray (12), and 7 improved varieties. The second cluster constitutes the highest number of accessions that were collected from the different regions of Ethiopia, Amhara (40), Benshangul-Gumz (34), Oromia (41), SNNP (2), Tigray (23), and 4 improved varieties, the remaining 13 accessions were introduced from different Africa countries. The third cluster is comprised of the majority from the Tigray region (25) and a small number from Amhara ($n=8$), Oromia (1), and 5 Improved varieties, the remaining 4 accessions were introduced from different countries of Africa. Cluster 4 comprised all the accessions that were introduced from one of the African country Egypt (23) only (Fig. 2).

An analysis of molecular variance (AMOVA)

Analysis of molecular variance (AMOVA) among the 300 sesame germplasms based on eight geographical origins indicated that 8.31% of the variance was differentiation among the populations, 15.24% of the variance was accounted by genetic differentiation among individuals within populations, while the remaining 76.44% of the variance was due to the differences within individuals (Table 4).

While in terms of continents 11.49% of the total molecular variation was observed due to differentiation between Asia and Africa, 19.45% of the variance was accounted by genetic differentiation among individuals within continents, while the remaining 69.06% of the variance was due to the differences within individuals (Table 4).

When we see further, In terms of population subdivision with different regions of Africa and Asia 22.17% of the total molecular variation observed was due to differentiation between different regions of Africa and Asia, 10.69% of the variance was accounted by genetic differentiation among individuals within different regions of Africa and Asia, while the remaining 67.12% of the variance was due to the differences within individuals (Table 4).

Population structure

The Bayesian model-based cluster analysis was conducted using the STRUCTURE program. The appropriate number of cluster K was determined, As *K* changed from 1 to 11 by inferring on Delta K of Evanno et al. (2005), the log-likelihood value [LnP(D)] increased continuously and inflection was evident when *K* increased numerically from 1 to 4 (Fig. 3A). Thus, the most likely numerical value of *K* was 4. The number of hypothetical ancestral populations (*K*) was further validated by the second-order statistics of ΔK. The ΔK value showed a peak at $K=4$ (Fig. 3B), which supported the classification of the panel into four major

Table 3 Pairwise population Nei's genetic distance showing the magnitude of genetic differentiation between sesame populations from different sources

	Africa	Amhara	Asia	BG	Improved	Oromia	SNNP	Tigray
Africa	0.000							
Amhara	0.021	0.000						
Asia	0.040	0.052	0.000					
BG	0.020	0.006	0.056	0.000				
Improved	0.022	0.012	0.037	0.015	0.000			
Oromia	0.021	0.010	0.057	0.004	0.014	0.000		
SNNP	0.030	0.020	0.073	0.016	0.028	0.013	0.000	
Tigray	0.024	0.006	0.047	0.015	0.009	0.017	0.027	0.000

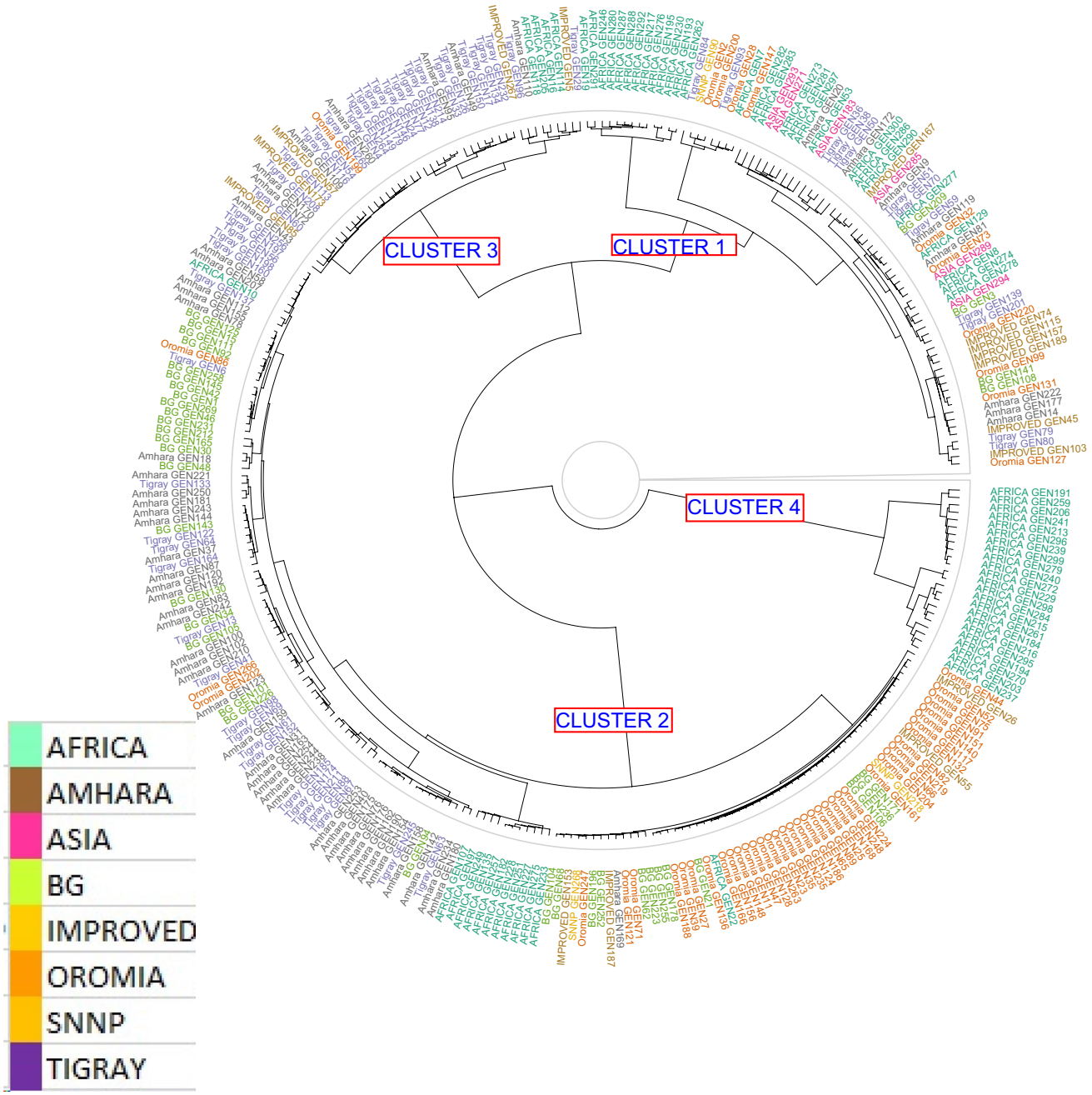

Fig. 2 UPGMA dendrogram showing the genetic relationships among 300 sesame genotypes grouped into four distinct clusters. Each geographical origin represented by a colored; green for Africa, teal for Amhara, red violet for Asia, BG: yellow green for Benshangul-Gumz; orange for Improved, red orange for Oromia, yellow orange for Southern Nations, Nationalities, and People's Region; blue violet for Tigray

hypothetical ancestral populations (Fig. 3C). The genetic diversity within each population was explained through the estimation of the expected heterozygosity, which varied from 0.06 (POP2) to 0.31 (POP4). The genetic divergence among the populations revealed by Nei's net nucleotide distance (D) indicated that a higher distance between POP3 and POP4 (0.22) and the genetic distance observed between POP1 and POP2 ($D = 0.09$) was the least among the pairs of populations. Mean fixation index of sub-populations ranged from 0.39 (POP4) to 0.77 (POP2) (Table 5).

According to the structure analysis, based on the probability of membership threshold of > 50%, 54, 159, 43,

Table 4 Analysis of molecular variance (AMOVA) among and within sesame subpopulations

Source of variation	df	SS	Variance components	Percentage variation
Based on different 8 geographical origins				
Among populations	7	9643.614	17.41032	8.31858
Among individuals				
Within populations	292	61,461.430	31.89832	15.24089
Within individuals	300	46,195.500	159.98566	76.44052
Total	599	117,300.544	209.29431	
Based on the Continents				
Among continents	1	941.728	26.62029	11.49050
Among individuals				
Within continents	298	70,163.316	45.06621	19.45258
Within individuals	300	46,195.500	159.98566	69.05692
Total	599	117,300.544	231.67217	
Based on different regions of Africa and Asia				
Among populations	4	12,564.718	52.85675	22.17834
Among individuals				
Within populations	295	58,540.327	25.48356	10.69273
Within individuals	300	46,195.500	159.98566	67.12892
Total	599	117,300.544	238.32598	

df degree of freedom, SS = Sum of Square

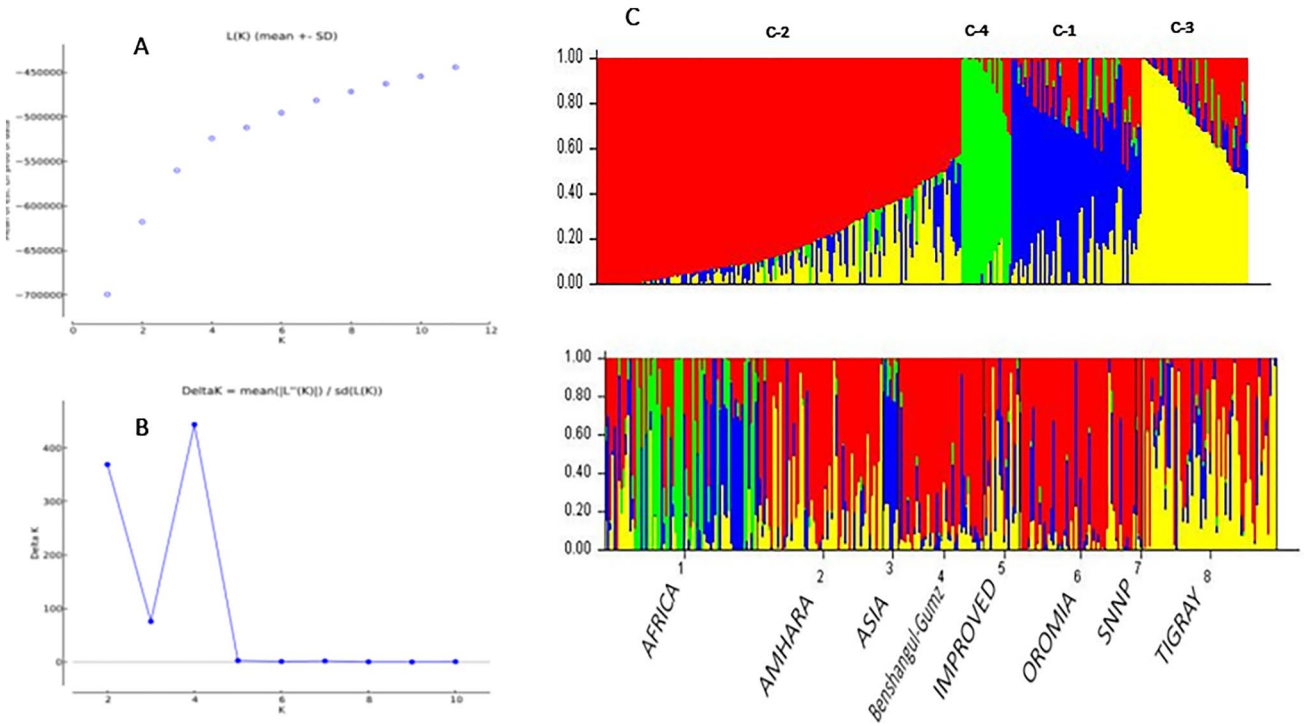

Fig. 3 Analysis of the population structure of the 300 sesame accessions using STRUCTURE. **A** Estimated LnP(D) of possible clusters (K) from 1 to 12; **B** ΔK based on the rate of change of LnP(D) between successive K; **C** population structure based on $K = 4$ with each individual is represented by a vertical bar partitioned into four colored segments, with their respective lengths representing the proportion of the individual's genome in a given group

Table 5 Genetic divergence among (net nucleotide distance) and within (expected heterozygosity) population, proportion of membership, and mean value of Fst observed from the study of the population structure of 300 sesame accessions and genotypes using DArTseq-SNP markers

Population	Net nucleotide distance			Expected heterozygosity	% of membership	Mean fixation index (Fst)
	pop2	pop3	pop4			
pop1	0.09	0.13	0.19	0.22	0.24	0.45
pop2		0.11	0.17	0.06	0.50	0.77
pop3			0.22	0.18	0.19	0.57
pop4				0.31	0.07	0.39

POP1, POP2, POP3 and POP4 represents the four sub-populations that grouped based on structure software

and 23 genotypes were respectively assigned into the four hypothetical ancestral populations, Pop 1, Pop 2, Pop 3, and Pop 4, while the remaining 21 accessions showed admixture among other subgroups.

Most accessions of Pop 1 were introduced from different countries of Africa (27), Asia countries (7), while 18 accessions from Ethiopia, Amhara ($n=4$), Benshangul-Gumz (2), Oromia (5), Tigray (7), and 2 improved varieties. The accessions and genotypes of Pop 2 constitute the largest that was mainly collected from the different regions of Ethiopia, Amhara ($n=40$), Benshangul-Gumz (35), Oromia (42), SNNP (2), Tigray (20), and 7 Improved varieties, the remaining 13 accessions were introduced from different Africa countries. The accessions of Pop 3 comprised mainly from three regions of Ethiopia, Amhara ($n=9$), Oromia (1), Tigray (26), and 4 Improved varieties, the remaining 3 accessions were introduced from different countries of Africa. Pop 4 was introduced from one of the African countries Egypt (23) only. For the Mixed group, 19 accessions were collected from different regions of Ethiopia and 2 accessions from two Africa countries.

The PCA based on DArTseq—SNP markers revealed four distinct groups of sesame genotypes and the first two principal components, accounting for 93.7% of total variation (Fig. 4A). PC1 explained 84% of the genetic variation found, while PC2 explained 7.7% of the variation, respectively. However, some intermediate lines (admixture) made the grouping less than clear-cut. When considering these intermediate lines, the panel could be neatly divided into four clusters (Fig. 4B) corresponding to the four hypothetical ancestral populations as inferred by using the STRUCTURE result.

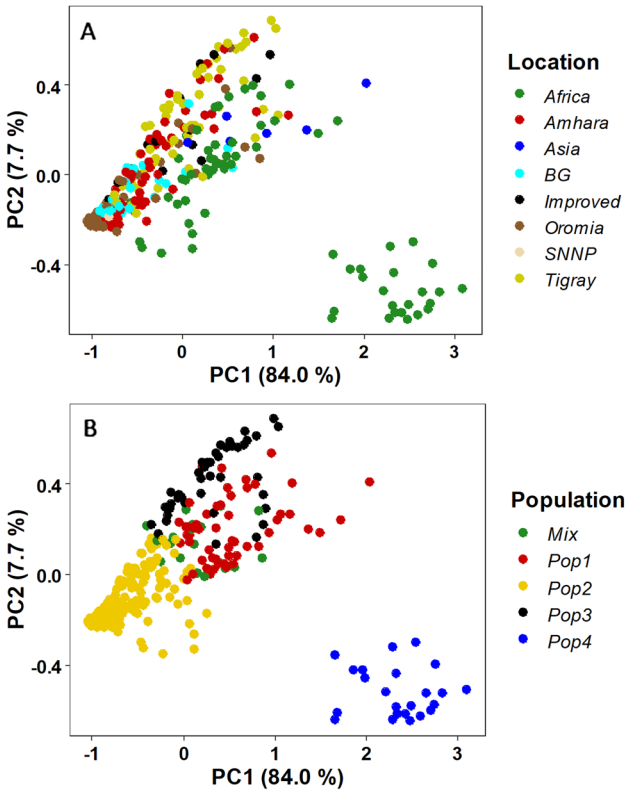

Fig. 4 Principal components analysis (PCA) of the population for 300 sesame accessions based on 2997 single-nucleotide polymorphisms (SNPs). Each individual is represented by GEN number, with its symbol color corresponding to the assigned subgroup classification. **A** PCA plots of the sesame germplasm collection based upon their geographic origins. **B** PCA plots of the same sesame germplasm collection but now based on hypothetical ancestral populations as identified by STRUCTURE

Discussion

In our study, the silicoDArT and SNP markers are distributed across the genomes and provide better genome coverage that is correlated with gene density (Dierig and Ray 2009; Kilian et al. 2012). This showed both;—silicoDArT and SNP markers used in our study are better suited for genetic diversity studies, association/linkage mapping, and/or sequence-based physical mapping in sesame. In comparison with the other existing marker technologies like microsatellite markers, DArT markers are pertinent to high-throughput work and have merits in terms of cost-effectiveness and time aspect (Kilian et al. 2003) Since it provides high information with low cost.

Then, 2997 SNP markers were filtered with a call rate of 75%, and those having > 0.01 minor allele frequency were used for the analysis, the proportion of rare SNPs (i.e., MAF < 0.05) we examined amounted to ~61.29% %, which was similar to those reported for the genomes of sesame (Cui et al. 2017), a high proportion of rare SNPs In our study may be caused due to the broad genome coverage obtained by GBS technology that is less prone to bias than low-coverage sequencing data (Wei et al. 2012) and the collection of large numbers of less abundant sesame landraces by the Ethiopian Biodiversity Institute. However, studies showed that the rare SNPs might also have control over the expression of a particular phenotype (Song et al. 2015). Providing that the number of individuals with a specific genotype will be very small, the effect of rare alleles on genome mapping could extend beyond the effect of just small population sizes. In such cases, increasing the number of individuals with rare alleles could improve the ability to check these rare alleles.

The average value of genetic diversity (0.14) was lower in the present study than in the earlier reports for the sesame collections analyzed with SNPs markers (Zhang et al. 2010; Cho et al. 2011; Cui et al. 2017; Basak et al. 2019) and SSR markers (Dossa et al. 2016; Wei et al. 2014). The broad range of variability among collections might be due to the composition of genetic resources under study, data filtering methods, sampling approaches, and the number of markers (Dossa et al. 2016). The type of marker also influences the estimated genetic diversity where high and low genetic diversity was estimated using SSR and SNPs markers. However, the accurate consideration of genetic diversity reflected the number of loci instead of the number of alleles (Cui et al. 2017).

The genetic diversity observed in Asia (0.17) sesame population is higher than Africa sesame population (0.14), in contrast, the number of germplasm used in this study from the Asia continent is only 7 as compared to the African continent (293). That indicates the potential of the Asian germplasms in the future sesame breeding program in Ethiopia. Laurentin and Karlovsky (2006) also reported high genetic diversity in sesame accessions collected from Asia that support our research finding.

The other result was obtained when we study the genetic diversity by considering different regions of Africa separately and Asian collections. The study showed the high genetic diversity obtained from the North African sesame collections (0.23) that may be because this part of the region is drier than the other part of the region and the adverse environmental situation in the North African region may result in an adaptation change that reflected in the genetic diversity. The low genetic diversity of East Africa (Ethiopian) collection as compared to other regions of Africa can be attributed to the favorable and high rainfall situation suitable for sesame production. In addition, this finding showed the importance of broadening the genetic base of the sesame population in Ethiopia through hybridization with North Africa and Asia populations and introduction to maintain its market share of the quality sesame at the international level. The clustering analysis didn't classify the sesame population according to the country of origins and showed a high level of admixture except genotype obtained from Egypt. Similar results were reported previously indifferent sesame germplasm (Ercan et al. 2004; Ali et al. 2007; Cho et al. 2011; Basak et al. 2019). The high level of admixture among the sesame genotypes from different countries and regions of the world showed the high-level gene flow from country to country through migration, trade, and seed exchange for cultivation. In line with our finding, Laurentin and Karlovsky (2006) found no association between genetic diversity and accession origin, and they proposed that ecological and geographical factors have not played a significant role in the evolution of sesame. The AMOVA analysis in our study also supported the possibility of high rates of gene flow between regions (Table 4).

The clustering analysis grouped the sesame genotypes collected from different regions of Ethiopia into Cluster 2 and 3 that showed a high level of germplasm exchange among different regions of Ethiopia consequently resulted in a high level of admixture.

The grouping of Sesame populations from Africa and Asia in Cluster I indicated a high level of genetic material exchange among the continents (Kim et al. 2002) and identical establishment stock. Evidence also showed the long history of germplasm exchange among the Asia and East African countries and the trend is also continuing today through export for industrial applications (Zohary et al. 2012).

Besides, the high level of outcross event was observed in the sesame populations when different genotypes from different locations grown in the same location due to the high level of cross-pollination observed ranged from 5 to 60% (Wei et al. 2014). This crossing could result in the similarity of accessions from the eastern part of Africa and Asia. Similar patterns have also been observed by other researchers (Cho et al. 2011; Kim et al. 2002; Laurentin and Karlovsky 2006).

Population structure analysis of the sesame population

The knowledge of the population structure in the breeding population is very important and avoids any spurious associations (Flint-Garcia et al. 2005).

In our study, the sesame population was clustered into four cluster groups with the membership coefficient value > 50%, and the clustering pattern did not follow the geographical origin of the germplasm. Among the tested

genotypes 21 accessions showed the variable percent of admixture and did not cluster in any group. This evidence supports the above genetic diversity study analysis and confirmed the high level of gene flow among the germplasm of the populations of different origins. In agreement with our finding the germplasm collected from Asia, Europe, America, and Africa clustered into three groups without considering their geographical origin (Basak et al. 2019).

The occurrence of some admixed genotypes indicated frequent hybridization and introgression events due to a high level of cross-pollination. Although the extent and significance of natural hybridization/introgression are unclear (Jarvis and Hodgkin 1998), new gene combinations between domestic cultivars and their wild or weedy relatives are important for the evolution of domesticated plants species (Jarvis and Hodgkin 1999).

The population genetic structure reflects interactions among species about their long-term evolutionary history, mutation and recombination, genetic drift, reproductive system, gene flow, and natural selection (Schaal et al. 1998; Slatkin 1987). This was also observed in our study where the genetic structure analysis showed the genetic relationship and the possible center of genetic diversity of sesame which is the Asian region. The understanding of the extent and structure of the genetic diversity of a crop could be a prerequisite for the conservation and efficient use of the germplasm available for breeding (Laidò et al. 2013). Our findings will also give information to guide our future genetic conservation strategy in sesame in Ethiopia to utilize and conserve efficiently with minimum sample high genetic variability. The various approaches (STRU CTURE, PCA, and the clustering) used in our study provide valuable information to understand the genetic diversity of the sesame population found in Ethiopia and what action will be needed to conserve and utilize the existing sesame population in the breeding program of the country.

According to the AMOVA results, low variation was explained among the population from different geographical regions of the sesame panel that showed a high level of differentiation among these populations.

In this study, most collections (225) were from Ethiopia and a specific collection was from West, South, and North Africa and seven collections were from 4 Asia countries. The Ethiopian sesame has useful characteristics and is often branded as 'Humera', 'Gondar' and 'Wellega' types, well known in the world market for their white color, sweet taste, and aroma. Even if the Ethiopian sesame populations showed unique quality characters, in the current study they clustered with germplasms of different origins. The Humera and Gondar sesame seeds are suitable for bakery and confectionary purposes and the high oil content of the Wellega sesame seed gives a major advantage for edible oil production (Wijnands et al. 2007). Collections that were introduced from different regions of Africa and Asia were showed some degree of genetic relationship and differentiation among genetic resources of Ethiopian collection. The incorporation of this germplasm in the Ethiopian breeding population (genetic pool) will broaden genetic diversity that can be used as a source of favorable alleles for agronomic traits in the breeding program. (Wang et al. 2019). The SNPs obtained from this collection could benefit future breeding and association mapping work in sesame.

The genetic relationship among and within the sesame populations from different origins may be a valuable source of parental material, genetically divergent germplasm like Njjoo-68 and 17,712 may be showed the possibility of identifying some divergent genotypes within Ethiopian sesame populations and that can be used in the future crossing program to develop high yielding sesame varieties to improve the production and productivity in the country.

Conclusions

The present research showed the effectiveness of DArTseq in characterizing the genetic diversity and population structure of sesame collection in Ethiopia. The gene diversity study showed the Asian population of sesame is more divergent than the sesame populations from Africa at continent level and North Africa germplasm are the most diverse population when further portioned based the sample size and geographical location level and, which indicate the importance of further collection from the Asian continent, and North Africa (Egypt) to enrich the genetic pool of the sesame in the breeding program and for further genetic diversity study. Even if Ethiopian sesame has useful characteristics, it has low genetic diversity as compared to the population obtained from other regions of the world. This tells us the breeding program should focus on obtaining more sesame accessions from Asia and North Africa to enrich the genetic pool and conducting further crossing programs to introgress the favorable genes from the imported germplasm. This study also supports the idea; ecological and geographical factors are less effective in the evolution of sesame. This finding guides the systematic utilization and conservation of the genetic resource of sesame and underlines the requirement of conducting a further collection of sesame genotypes from all sesame growing regions of the world.

Therefore, our next objective is to identify sesame genotypes with desirable traits and to conduct association mapping on the economically important traits in sesame to implement markers assisted selection (MAS) in the sesame breeding program in the country to increase the efficiency of the breeding program.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12892-021-00137-x.

Acknowledgements The authors are grateful to Ethiopian Biodiversity Institute (EBI) and Assosa, Bako, Gondar, Humera, and, Werer Agricultural Research Centers for their generous provision of collection and improved varieties used in this study. The financial support of Amhara Agricultural Research Institute of (ARARI) and Gondar Agricultural Research Center through BENEFIT-SBN project, is appreciated. The first author further obliged to Addis Ababa University and IGSS program in BecA-ILRI Hub, Nairobi-Kenya for the training provided.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

Abate M, Mekbib F (2015) Assessment of genetic diversity in Ethiopian sesame (Sesamum indicum L.) germplasm using random amplified polymorphic DNA (RAPD) markers council for innovative research. J Adv Agric 5:639–649

Abate M, Mekbib F, Ayana A, Nigussie M (2015) Genetic variability and association of traits in mid-altitude sesame (Sesamum indicum L.) Germplasm of Ethiopia. Am J Exp Agric 9:1–14

Ali GM, Yasumoto S, Seki-Katsuta M (2007) Assessment of genetic diversity in sesame (Sesamum indicum L.) detected by amplified fragment length polymorphism markers. Electron J Biotechnol 10:12–23

Basak M, Uzun B, Id EY (2019) Genetic diversity and population structure of the Mediterranean sesame core collection with the use of genome-wide SNPs developed by double digest RAD-Seq, pp 1–15

Bastien M, Sonah H, Belzile F (2014) Genome wide association mapping of Sclerotinia sclerotiorum resistance in soybean with a genotyping-by-sequencing approach. The Plant Genome 7:1–13

Bedigian D, Harlan JR (1986) Evidence for cultivation of sesame in the ancient world. The New York Botanical Garden 40:137–154

Bolibok-Bragoszewska H, Heller-Uszyńska K, Wenzl P, Uszyński G, Kilian A, Rakoczy-Trojanowska M (2009) DArT markers for the rye genome - genetic diversity and mapping. BMC Genomics 10:1–11

Cho YI, Park JH, Lee CW, Ra WH, Chung JW, Lee JR, Ma KH, Lee SY, Lee KS, Lee MCPY (2011) Evaluation of the genetic diversity and population structure of sesame (Sesamum indicum L.) using microsatellite markers. Genes Genomics 33:187–195

Courtois B, Audebert A, Dardou A, Roques S, Ghneim-Herrera T, Droc G, Frouin J, Rouan L, Gozé E, Kilian AAN (2013) Genome-wide association mapping of root traits in a japonica rice panel. PLoS ONE 8(11):e78037

Cruz VMV, Kilian A, Dierig DA (2013) Development of DArT marker platforms and genetic diversity assessment of the US collection of the new oilseed crop lesquerella and related species. PLoS ONE 8:e64062

CSA (Central Statistical Agency) (2018) Agricultural sample survey report on area and production of major crops (Private Peasant Holdings, Meher Season)

CSA (Central Statistical Agency) (2019) Agricultural sample survey 2019/2020 (2012 E.C.) (September–January 2019/2020) Volume VII report on crop and livestock product utilization (Private Peasant Holdings, Meher Season)

Cui C, Mei H, Liu Y, Zhang H, Zheng Y (2017) Genetic diversity, population structure, and linkage disequilibrium of an association-mapping panel revealed by genome-wide SNP markers in sesame. Front Plant Sci 8:1–10

Dierig D, Ray DT (2009) New crops breeding: Lesquerella. In: Vollmann J, Rajcan I (eds) Oil crops. Springer, New York, pp 507–516

Dossa K, Wei X, Zhang Y, Fonceka D, Yang W, Diouf D (2016) Analysis of genetic diversity and population structure of sesame accessions from Africa and Asia as major centers of its cultivation. Genes 7:1–13

Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Ercan AG, Taskin M, Turgut K (2004) Analysis of genetic diversity in Turkish sesame (Sesamum indicum L.) populations using RAPD markers. Genet Resour Crop Evol 51:599–607

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

Excoffier L, Lischer HEL (2010) A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

FAOSTAT. 2019. Food and agriculture organization statistical databases. Accessed from http://www.fao.org/faostat/en/#home

Farshadfar M, Farshadfar E (2008) Genetic variability and path analysis of chickpea (Cicer arientinum L.) landraces and lines. J Appl Sci 8:3951–3956

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054–1064

Gebremichael DE (2017) Sesame (Sesamum indicum L.) breeding in Ethiopia. Int J Novel Res Life Sci 4:1–11

Gebremichael DE, Heiko KP (2011) Genetic variability among landraces of sesame in Ethiopia. Afr Crop Sci J 19:1–13

Gidey YT, Kebede SAGG (2012) Extent and pattern of genetic diversity for morpho-agronomic traits in Ethiopian sesame landraces (Sesamum indicum L.). Asian J Agric Res 6:118–128

Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity 101:5–18

Huang YF, Poland JA, Wight CP, Jackson EW, Tinker NA (2014) Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. PLoS ONE 9:1–16

Institute of Biodiversity Conservation (IBC) (2012) Ethiopia: third country report on the state of plant genetic resources for food and agriculture. Addis Ababa, Ethiopia

Jarvis DI, Hodgkin T (1998) Wild relatives and crop cultivars: conserving the connection. In The Proceedings of an International Symposium on in situ conservation of plant genetic diversity. George Allen & Unwin, London, UK. pp 163–179

Jarvis DI, Hodgkin T (1999) Wild relatives and crop cultivars: detecting natural introgression and farmer selection of new genetic combinations in agroecosystems. Mol Ecol 8:S159–S173

Kilian A, Sanewski G, Ko L (2016) The application of DArTseq technology to pineapple. In: XXIX international horticultural congress on horticulture: sustaining lives, livelihoods and landscapes (IHC2014) IV, vol 1111. Humana Press, Totowa, pp 181–188

Kilian A, Huttner E, Wenzl P, Jaccoud D, Carling J, Caig V, Evers M, Heller-Uszynska K, Cayla C, Patarapuwadol S, Xia L (2003) The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In Proceedings of the international congress in the wake of the double helix: from the green revolution to the gene revolution. pp 443–461

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C A-KM (2012) Diversity arrays technology: a generic genome profiling technology on open

platforms. In: Data production and analysis in population genomics, pp 67–89. Humana Press, Totowa, NJ.
Kim DH, Zur G, Danin-Poleg Y, Lee SW, Shim KB, Kang CWKY (2002) Genetic relationships of sesame germplasm collection as revealed by inter-simple sequence repeats. Plant Breeding 121:259–262
Kumar H, Kaur G, Banga S (2012) Molecular characterization and assessment of genetic diversity in sesame (Sesamum indicum L.) germplasm collection using ISSR markers. J Crop Improv 26:540–557
Laidò G, Mangini G, Taranto F, Gadaleta A, Blanco A, Cattivelli L, Marone D, Mastrangelo AM, Papa RDVP (2013) Genetic diversity and population structure of tetraploid wheats (Triticum turgidum L.) estimated by SSR, DArT and PEdigree data. PLoS ONE 8:1–17
Laurentin HE, Karlovsky P (2006) Genetic relationship and diversity in a sesame (Sesamum indicum L.) germplasm collection using amplified fragment length polymorphism (AFLP). BMC Genet 7:1–10
Lui K (2005) PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics 21:2128–2129
Matthies IE, van Hintum T, Weise S, Röder MS (2012) Population structure revealed by different marker types (SSR or DArT) has an impact on the results of genome-wide association mapping in European barley cultivars. Mol Breeding 30:951–966
Merve B, Uzun Bulent YE (2019) Genetic diversity and population structure of the Mediterranean sesame core collection with use of genome-wide SNPs developed by double digest RAD-Seq. PLoS ONE 14:1–15
Mkamilo GS, Bedigian D (2007) In PROTA (Plant Resources of Tropical Africa/Ressources végétales de l'Afrique tropicale). Wageningen, Netherlands
Nei M (1972) Genetic distance between populations. Am Nat 106:283–292
Park YCJ, Ra CLW, Lee JCJ (2011) Evaluation of the genetic diversity and population structure of sesame (Sesamum indicum L.) using microsatellite markers. Genes & Genomics 33:187–195
Pritchard JK, Stephens M, Donnelly P (2000) Structure software for population genetics inference (V.2.3.4)
Raman H, Raman R, Kilian A, Detering F, Carling J, Coombes N, Diffey S, Kadkol G, Edwards D, McCully MRP (2014) Genome-wide delineation of natural variation for pod shatter resistance in Brassica napus. PLoS ONE. https://doi.org/10.1371/journal.pone.0101673
Sánchez-Sevilla JF, Horvath A, Botella MA, Gaston A, Folta K, Kilian A, Denoyes B, Amaya I (2015) Diversity arrays technology (DArT) marker platforms for diversity analysis and linkage mapping in a complex crop, the octoploid cultivated strawberry (Fragaria × ananassa). PLoS ONE 10:1–22
Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA (1998) Phylogeographic studies in plants: problems and prospects. Mol Ecol 7:465–474
Sharaby N, Butovchenko A (2019) Cultivation technology of sesame seeds and its production in the world and in Egypt. IOP Conf Series: Earth Environ Sci 403:1–8
Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236:787–792
Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, Normandeau É, Laroche J, Larose S, Jean MBF (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS ONE 8:1–9
Song XJ, Kuroha T, Ayano M, Furuta T, Nagai K, Komeda N, Segami S, Miura K, Ogawa D, Kamura TST (2015) Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. Proc Natl Acad Sci USA 112:76–81
Tang J, Daroch M, Kilian A, Jeżowski S, Pogrzeba M, Mos M (2015) DArT-based characterisation of genetic diversity in a Miscanthus collection from Poland. Planta 242:985–996
Teklu DH, DE Kebede SAG (2014) Assessment of genetic variability, genetic advance, correlation and path analysis for morphological traits in sesame genotypes. Asian J Agric Res 8:181–194
Teshome D, Kassahun T, Bekele E (2015) Genetic diversity of sesame germplasm collection (SESAMUM INDICUM L.): implication for conservation, improvement and use. Int J Biotechnol Mol Biol Res 6:7–18
Tewodros T, Kassahun T, Gemechu K, Alemu T (2021) Morphological characteristics and genetic diversity of Ethiopian sesame genotypes. Afr Crop Sci J 29:59–77
Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants, vol 72. LWW, p 482
Wang L, Xia Q, Zhang Y, Zhu X, Zhu X, Li D, Ni X, Gao Y, Xiang H, Wei XYJ (2016) Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map. BMC Genomics 17:1–13
Wang Y, Rashid MA, Li X, Yao C, Lu L, Bai J, Li Y, Xu N, Yang Q, Zhang LBG (2019) Collection and evaluation of genetic diversity and population structure of potato landraces and varieties in China. Front Plant Sci 10:1–11
Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236–244
Wei X, Liu K, Zhang Y, Feng Q, Wang L, Zhao Y, Li D, Zhao Q, Zhu X, Zhu XLW (2015) Genetic discovery for oil production and quality in sesame. Nat Commun 6:1–10
Wei X, Qiao WH, Chen YT, Wang RS, Cao LR, Zhang WX, Yuan NN, Li ZC, Zeng HLYQ (2012) Domestication and geographic origin of Oryza sativa in China: insights from multilocus analysis of nucleotide variation of O. sativa and O. rufipogon. Mol Ecol 21:5073–5087
Wei X, Wang L, Zhang Y, Qi X, Wang X, Ding X, Zhang JZX (2014) Development of simple sequence repeat (SSR) markers of sesame (Sesamum indicum) from a genome survey. Molecules 19:5150–5162
Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci USA 101:9915–9920
Wijnands JHM, Biersteker J, Hiel R (2007) Oilseeds business opportunities in Ethiopia. Ministry of Agriculture, Nature and Food Quality
Wu K, Liu H, Yang M, Tao Y, Ma H, Wu W, Zuo YZY (2014) High-density genetic map construction and QTLs analysis of grain yield-related traits in Sesame (Sesamum indicum L.) based on RAD-Seq technology. BMC Plant Biol 14:1–14
Yang S, Pang W, Ash G, Harper J, Carling J, Wenzl P, Huttner E, Zong XKA (2006) Low level of genetic diversity in cultivated Pigeonpea compared to its wild relatives is revealed by diversity arrays technology. Theor Appl Genet 113:585–595
Zhang YX, Zhang XR, Hua W, Wang LH, Che Z (2010) Analysis of genetic diversity among indigenous landraces from sesame (Sesamum indicum L.) core collection in China as revealed by SRAP and SSR markers. Genes & Genomics 32:207–215
Zhang Y, Zhang X, Che Z, Wang L, Wei W, Li D (2012) Genetic diversity assessment of sesame core collection in China by phenotype and molecular markers and extraction of a mini-core collection. BMC Genet 13:1
Zohary D, Hopf M, Weiss E (2012) Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin. Oxford University Press