Journal of Food Composition and Analysis 110 (2022) 104545

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca

Analyses of genetic diversity and population structure of sesame (*Sesamum indicum* L.) germplasm collections through seed oil and fatty acid compositions and SSR markers

Desawi Hdru Teklu [a,b,*], Hussein Shimelis [a], Abush Tesfaye [c], Admire Isaac Tichafa Shayanowako [a]

[a] African Centre for Crop Improvement, University of KwaZulu-Natal, Private Bag X01, Scottsville, Pietermaritzburg 3209, South Africa
[b] Ethiopian Agricultural Transformation Institute, P.O. Box 708, Addis Ababa, Ethiopia
[c] International Institute of Tropical Agriculture, Oyo Road, P.M.B. 5320, Ibadan, Nigeria

ARTICLE INFO

Keywords:
Fatty acids, genetic diversity
SSR markers
Seed oil content, *Sesamum indicum*

ABSTRACT

Knowledge of the genetic profiles of diverse germplasm collections of sesame using seed oil, fatty acid contents, and molecular markers is a prerequisite to develop market-preferred cultivars with quantity and quality oil. The objective of this study was to determine the genetic diversity and relationships among Ethiopia's sesame germplasm collections using seed oil content and fatty acid compositions and diagnostic simple sequence repeat (SSR) markers to select genetically complementary and promising parental lines for breeding. The contents of the seed oil and fatty acids of 100 lines grown under field conditions were determined using the near-infrared reflectance spectrometry. Twenty-seven polymorphic SSR markers were used to assess the genetic profile of the test lines and complement the seed oil and fatty acid data. The SSR markers revealed that the mean gene diversity and polymorphic information content were 0.30 and 0.25, respectively. Population structure analysis identified four major heterotic groups. Based on higher oil content and desirable fatty acid compositions and SSR markers the following superior and complementary lines such as: Hirhir Kebabo Hairless Sel-6 (from sub-cluster I-b), Hirhir Humera Sel-8 and NN0058–2 (sub-cluster II-a) and Bawnji Fiyel Kolet (sub-cluster II-b) are identified for sesame breeding programs or production globally.

1. Introduction

Sesame (*Sesamum indicum* L.; 2 n = 26), belonging to the family Pedaliaceae is one of the oldest and highly valuable oilseed crops globally (Ashri, 2010). Its domestication dates back some 5500 years ago in the Harappa Valley of India, mainly as an oilseed crop (Bedigian and Harlan, 1986). Sesame is regarded as the queen of oilseed crops due to its high quantity and quality oil and commercial value (Dossa et al., 2018). The seed oil content of sesame varies from 40.80% to 60.30%, with a mean of 53.00%, the highest value compared with other oilseed crops (Dossa et al., 2018). Sesame oil has about 85% unsaturated fatty acids (e.g., oleic acid, linoleic acid and linolenic acid), which are beneficial to human health. Consumption of sesame oil is believed to be minimising the risks of cardiovascular diseases, cancer, brain and liver damages (Yen, 1990; Yol et al., 2015). The major fatty acids present in the sesame oil include oleic acid (35.90–47.00%), linoleic acid (35.60–47.60), palmitic acid (8.70–13.80%), and stearic acid (2.10–6.40%) (Weiss, 1983; Uzun et al., 2002; Elleuch et al., 2007). Also, trace amounts of linolenic acid (1.38–2.19%) and arachidic acid (0.10–0.70%) are present in the seed oil.

Sesame is an economically important crop widely traded in local, regional and international markets (Myint et al., 2020). Global sesame consumption is steadily increasing due to high demands related to its unique nutritional values such as higher contents of vitamins (e.g., A and E), minerals, fibre, healthy fatty acids [e.g., oleic acid, linoleic acid], carbohydrate (~13.5%), and protein (~24%)] (Myint et al., 2020). Furthermore, population pressure, urbanisation and the changing life-style have increased the global demand for sesame products (Myint et al., 2020).

About 70% of the world's sesame seed is processed to produce food

* Corresponding author at: African Centre for Crop Improvement, University of KwaZulu-Natal, Private Bag X01, Scottsville, Pietermaritzburg 3209, South Africa.
E-mail addresses: desawi.Hdru@ata.gov.et (D.H. Teklu), Shimelish@ukzn.ac.za (H. Shimelis), At.Abebe@cgiar.org (A. Tesfaye), shayanowako@googlemail.com (A.I.T. Shayanowako).

https://doi.org/10.1016/j.jfca.2022.104545
Received 26 January 2022; Received in revised form 24 March 2022; Accepted 28 March 2022
Available online 31 March 2022

oil, while the seedcake is used to prepare livestock meals (Myint et al., 2020). The total annual human consumption of sesame is about 65% and 35% in the form of food oil and grain (e.g., as a garnish, snack, and flavouring agent in some foods), respectively (Morris, 2020). In 2019 the total world sesame oil production was 1286,741.00 tons. China with a total annual production of 563,637.00 tons, Myanmar (154,600.00 tons), India (96,800.00 tons), Japan (53,257.00 tons), Nigeria (45, 690.00 tons) and Turkey (38,300.00 tons) are the major global sesame oil producers in 2019 (FAOSTAT, 2019). Globally, a total of 76,140.00 tons of sesame oil was traded with a monetary value of 331.2 billion USD in 2019 (FAOSTAT, 2019). The top sesame oil-exporting countries were China (with 16,829.00 tons), Japan (9244.00 tons), India (8593.00 tons), Lebanon (6214.00 tons), and Mexico (5997.00 tons) (FAOSTAT, 2019).

In sub-Sharann Africa, Sudan, the United Republic of Tanzania, Nigeria, Burkina Faso, Ethiopia, and Mozambique are sesame grain's major producers and exporters (FAOSTAT, 2019). In Ethiopia, sesame is the leading oil crop that occupies some 45.7% of the total production area, followed by niger seed (*Guizotia abyssinica* [L.f.] Cass.) and groundnut (*Arachis hypogea* L.) (Central Statistic Authority CSA, 2019). Sesame is Ethiopia's second major export cash crop that contributes significantly to gross domestic product and foreign currency earnings (Gebremedhn et al., 2019). Globally, Ethiopia is the 9th largest sesame producer with a total annual grain production of 262,654 tons after Sudan (1210,000 tons), China (936,104 tons), Myanmar (744,498 tons), India (689,310 tons), Tanzania (680,000 tons), and Nigeria (480,000 tons) tons) (FAOSTAT, 2019).

The average grain yield of sesame in Ethiopia is 0.6 ton ha^{-1}, which is far below the attainable yield of the crop, reaching up to 4.00 t/ha. The mean yield of sesame is variable across the major producing countries such as in Lebanon (3.52 tons ha^{-1}), Saudi Arabia (2.53 tons ha^{-1}), Afghanistan (2.16 tons ha^{-1}), Tajikistan (2.12 tons ha^{-1}), Israel (2.05 tons ha^{-1}), Uzbekistan (1.77 tons ha^{-1}) and China (1.62 tons ha^{-1}) (FAOSTAT, 2019). Sesame production and productivity in East Africa, including Ethiopia have stagnated because of a lack of high-yielding and well-adapted varieties (Were et al., 2006). In the region, sesame production relies on unimproved landrace varieties selected by farmers. These varieties have intrinsic oil quality characteristics, such as unique aroma and taste by consumers. Hence, there is a need to breed lines with high oil quantity and quality for local and export markets, diverse production environments, and breeding programmes aiming for seed oil content and quality.

Ethiopia is the centre of origin and diversity for the cultivated sesame (Mehra, 1967; Mahajan et al., 2007; Bedigian, 1981; Seegeler, 1983). More than 5000 genetically diverse sesame accessions are maintained by the Ethiopian Biodiversity Institute (EBI) (Woldesenbet et al., 2015). The accessions, including landraces and introductions, are ideal sources of genetic variation to initiate a competitive breeding program focused on new variety design and deployment based on seed oil quantity and quality. These genetic resources need to be systematically evaluated for seed oil, fatty acid contents, and genetic composition to develop improved cultivars that meet the demands of sesame clients and value chains. The product profile of the new generation sesame varieties should include high oil content and well balanced fatty acid compositions such as palmitic, stearic, oleic, linoleic, and linoleic acids. The local sesame germplasm provides the genetic foundation for breeding programs to select agronomically suitable and locally adapted varieties with a suite of nutritional values for human health. There are limited studies that reported on oil quantity and quality of sesame using Ethiopia's sesame germplasm. Wang et al. (2012) used high-performance liquid chromatography (HPLC) to determine the sesamin and sesamolin content of 215 sesame lines from China core collection. The auhors found a broad genetic variation for sesamin (0.05–11.05 mg/g) and sesamolin (0–10 mg/g) contents. Dossa et al. (2018) reported the presence of great variation for seed oil content, oleic acid content, linolenic acid content and protein content amongst 139 African and Asian sesame germplasm collections using the near-infrared reflectance spectrophotometry (NIRS) scanning. Further, Biglar et al. (2012) analysed the seed oil content and fatty acid profiles of 5 Iranian sesame cultivars using a gas chromatography and reported a considerable variation. Broad variation in oleic acid, linoleic acid, linolenic acids, palmitic acid, myristic acid, capric acid, and lauric acids were reported by Wacal et al. (2019) using a total fat determination unit in sesame collections of Japan.

Molecular markers are highly reliable genetic tools that can complement phenotypic characterization for breeding (Jones et al. 2009). The following markers are widely used in genetic diversity analysis of crop species: amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), inter simple sequence repeat (ISSR), simple sequence repeat (SSR), and single nucleotide polymorphisms (SNPs). These marker systems have provided useful and complementary data revealing genetic diversity, relationships and population structure of sesame germplasm collections (Laurentin and Karlovsky, 2006; Abdellatef et al., 2008; Gebremichael and Parzies, 2011; Wei et al., 2016; Dossa et al., 2016; Asekova et al., 2018; Araújo et al., 2019, Basak et al., 2019; Teklu et al., 2021). The SSR markers have been widely used in analysing the sesame genetic diversity, population structure and heterotic groups. The markers are highly preferred for detecting higher degrees of polymorphism, reproducibility, allelic variation, and their abundance in the genomes (Wei et al., 2016; Dossa et al., 2016; Asekova et al., 2018; Araújo et al., 2019). Wei et al. (2016), Dossa et al. (2016), Asekova et al. (2018) and Araújo et al. (2019) reported considerable genetic diversities of sesame populations using the SSRs.

Nutrient profiles, including oil content and fatty acid compositions, show wide variability in sesame germplasm collections, providing opportunities for breeding for enhanced oil quantity and quality (Dossa et al., 2018). Previous studies have reported the presence of considerable variation for seed oil content in sesame genetic resources from Ethiopia (Gidey et al., 2012; Teklu et al. 2014; Hika et al., 2014, 2015). Limitedd studies assessed the genetic diversity of sesame accessions in Ethiopia using SSRs (Gebremichael and Parzies, 2011) and inter simple sequence repeat (ISSR) (Woldesenbet et al., 2015) markers. However, the previous reports did not fully represent Ethiopia's sesame core collection, and the markers used were relatively few. Therefore, there is a need for a comprehensive assessment of the genetic diversity of oil content and fatty acid compositions present in the Ethiopian sesame germplasm pool using a greater number of accessions representing the diverse growing areas through oil content and fatty acid compositions and polymorphic SSR markers. Hence, the objective of this study was to determine the genetic diversity and relationships among Ethiopia's sesame germplasm collections using seed oil content and fatty acid compositions and diagnostic simple sequence repeat markers to select genetically complementary and promising parental lines for breeding.

2. Materials and methods

2.1. Plant materials

This study evaluated 100 genetically diverse sesame genotypes, including 95 accessions, one landrace, and four improved varieties. The landrace variety (locally referred to as 'Hirhir') is widely cultivated by smallholder farmers in Ethiopia. The four improved varieties were selected from local populations released by the Humera Agricultural Research center (HuARC). The accessions were collected from five sesame-growing regions in Ethiopia by the sesame and groundnut breeding division of Werer Agricultural Research Centre of the Ethiopian Institute of Agricultural Research (EIAR). The details of the germplasm collections are given in Supplementary Table 1.

2.2. Study site and experimental design

The 100 sesame lines were cultivated in northwestern Ethiopia at the

Humera site (14°150 N, 36°370 E) to sample seeds for seed oil quality and quantity analysis. The site is the agricultural research station of the Humera Agricultural Research Centre of EIAR. Humera is situated at an altitude of 609 m above sea level and receives a total rainfall of 576.4 mm. The mean minimum and maximum temperatures at the site range from 20.3 to 36.5 °C. The site has clay soil (HuARC, 2010) predominantly. The experiment at the site was laid out using a 10 × 10 simple lattice design with two replications. In each replicate, each genotype was sown in four rows in plots of 4 m length, with a 40 cm inter-row and 10 cm intra-row spacing.

2.3. Determination of oil content and fatty acid profiles

Oil content and fatty acid compositions were determined at Wuhan city in China using the Near-Infrared Spectroscopy (NIR) (FOSS, model DS2500, Hillerod, Denmark). All mature and well-rounded seeds were separated, and true-to-type seeds were used for accurate analysis. For each genotype 3 g of seed was sampled and labelled to determine the contents of oil and the profiles of palmitic acid (16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3), all values expressed in percentage. The NIR analyzer was programmed with different calibration models for dark (black, brown and black) and light (white, light white, brown, and light brown) seeds (Dossa et al. 2018). All the sample seeds were irradiated with near-infrared monochromatic light, and the reflectance spectrum (log10 1/R) was recorded from 1100 to 2500 nm at a wavelength interval of 2 nm, according to Dossa et al. (2018). During the NIR analysis, the room temperature was maintained at 25 °C. Each sample was run in duplicate analyses, and if the difference between the two values was > 2, the analysis was repeated to increase the accuracy of the analysis (Dossa et al. 2018).

2.4. Genotyping using SSR markers

The seeds harvested from the Humera site were prepared for genotypic analysis at the Oil Crops Research Institute (OCRI)-the Chinese Academy of Agricultural Sciences (OCRI-CAAS), China.

2.4.1. DNA extraction, primer selection, polymerase chain reaction, and electrophoresis analysis

Prior to DNA extraction, seeds of the 100 sesame genotypes (Supplementary Table 1) were planted in a plastic tray in a controlled environment growth room. Fresh young leaves from two-weeks old seedlings of 10 plants were sampled for DNA extraction. Bulked leaves were grounded in liquid nitrogen and DNA extracted using the Cetyltetramethyl ammonium bromide (CTAB) method (Zhang et al. 2012a). The DNA quantity extracted from each genotype was checked with Nanodrop Lite (Thermo Scientific, USA). The concentrations of DNA were quantified by the Quantus TM Fluorometer (Promega Corporation, Madison, USA).

The microsatellite analysis was done at the Oil Crops Research Institute (OCRI), the Chinese Academy of Agricultural Sciences (OCRI-CAAS). Twenty-seven primers were selected based on their polymorphism in discriminating sesame genotypes (Supplementary Table 2). The criteria used for selecting the SSR markers were their suitability in discriminating sesame genotypes. The primers used in the present study were initially selected amongst 160 candidate primers based on their higher polymorphic information content. The selected primers provided clear and informative amplicon profiles in sesame genetic analysis.

The polymerase chain reaction (PCR) amplification of all SSR markers was performed with a total volume of 20 μL solution containing 25 ng of DNA, 4 μmol of forward primers, 4 μmol of reverse primers, 1 × buffer, 0.25 mmol of dNTPs, and 0.80 U Taq polymerase. The PCR profile was an initial de-naturation at 94 °C for 1 min, followed by primer annealing at 45.2–53 °C for 1 min, and elongation at 72 °C for 1 min. After 34 cycles, the reaction was terminated with a 10 min final extension time at 72 °C. Except for the annealing temperatures, the PCR reaction conditions were the same for all the primers. The amplified products were size separated on 6% denaturing polyacrylamide gels. The electrophoresis parameters and silver staining of gels were based on the protocols of Zhang et al. (2012a). The PCR products were separated by capillary electrophoresis on an ABI 3730 automatic sequencer. The marker data was presented as fragment sizes in an excel spreadsheet.

2.5. Data analysis

2.5.1. Oil content and fatty acid compositions

Data collected between the two analyses did not show significant differences in oil content and the major fatty acid profiles. Hence, the two data sets were pooled to conduct a one-sample t-test analysis to discern the significant difference ($P < 0.05$). The t-test analysis was conducted using the Statistical Package for the Social Sciences (SPSS) software version 20 (SPSS, 2020).

The seed oil content and fatty acid profiles data were subjected to correlation analysis to determine the magnitude of associations using R software version 4.0 (R Core Team, 2020). The oil content and fatty acid compositions were subjected to principal component analysis to determine the magnitude of variation attributable to the various components and identify the most discriminative parameter for selection. Principal component analysis was done using R software version 4.0 (R Core Team, 2020). The test genotypes were subjected to cluster analysis for oil content, and fatty acid contents based on Euclidean distances using SAS procedure CLUSTER (SAS Institute, 2018). This allowed the determination of the genetic relatedness and classification of the assessed genotypes into respective genetic groups. Bi-plot analysis was computed to infer genotype association regarding the seed oil content and fatty acid profiles using R software version 4.0 (R Core Team, 2020).

2.5.2. SSR markers data analysis

Genetic parameters, such as major allele frequency (MAF), observed heterozygosity (Ho), expected heterozygosity (He), and the polymorphic information content (PIC) were estimated using Power Marker v3.2. Cluster analysis based on the 27 SSR markers was carried out using a neighbor-joining (NJ) algorithm using the unweighted pair group method (UWPGM) in R software version 4.0 (Core Team 2000). The Bayesian clustering method in STRUCTURE version 2.3.4 was used to construct the population structure of the 100 sesame accessions (Pritchard et al. 2000). The length of the burn-in period and Markov Chain Monte Carlo (MCMC) were set at 20,000 iterations (Evanno et al. 2005). Ten runs were performed for each K-value (assumed number of subpopulations), ranging from 1 to 10 to estimate the number of populations accurately. Evanno et al. (2005) method using CLUMPK was employed to calculate Delta K values and determine the appropriate K value. The principal coordinate analysis using Darwin version 6 software was used to construct the genotypes' genetic structure.

3. Results

3.1. Seed oil content and fatty acid compositions of sesame genotypes

The one-sample t-test analysis revealed significant ($P \leq 0.05$) differences among the test genotypes for seed oil content and all the assessed fatty acid profiles (Table 1). This suggested differential responses of the assessed sesame genetic resources for production or selection programs with desirable oil and fatty acid compositions.

The mean oil content and fatty acid compositions of sesame genotypes are summarized in Table 1. The grand mean of the oil content across the studied genotypes was 49.84%, varying from 44.30% to 55.60%. The mean values for palmitic acid and stearic acid contents were 9.10% and 5.1%, respectively. The highest oil content was recorded for entries Bawnji Fiyel Kolet at 55.6%, NN0056 (55.2%), Hirhir Humera Sel-8 (54.7%), NN-0068–1 (54.6%), ACC-NS-010 (54.1%), and

Table 1
Mean values of the assessed sesame genotypes for the contents of seed oil and fatty acids (%).

Entry number	Genotype name or designation		Fatty acid compositions (%)†				
		OC (%)	C16:0	C18:0	C18:1	C18:2	C18:3
1	Hirhir Kebabo Hairless Sel-2	50.5	8.76	4.95	46.64	38.21	0.29
2	GXT= 85(28–2)	51.00	9.46	5.05	40.26	43.95	0.28
3	Hirhir Kebabo Hairless-9	50.9	8.90	4.92	45.35	39.53	0.28
4	NN-0068–1	**54.6**	9.08	**4.84**	41.5	43.26	0.24
5	NN-0108–2	47.3	9.06	5.14	41.38	42.35	0.35
6	NN-034	47.7	9.21	5.12	40.91	44.04	0.30
7	BCS-0041	46.9	9.21	5.07	43.82	41.02	0.34
8	ACC-031–5–14	47.90	8.85	5.03	44.79	40.27	0.31
9	NN-0129–2	44.60	9.82	5.12	39.72	44.63	0.36
10	ACC-203–020	46.80	8.66	5.11	**47.34**	37.93	0.36
11	NN-0038–2	50.80	9.39	4.97	43.34	41.15	0.28
12	Bawnji Fiyel Kolet	**55.60**	9.27	**4.81**	40.04	44.19	0.28
13	Gojam Azene (Aleka)	**44.30**	9.05	**5.36**	41.77	40.99	**0.39**
14	Hirhir Humera Sel-6	53.90	9.64	4.86	37.79	**46.00**	**0.23**
15	Bawnji Gobate	48.90	8.88	5.04	45.86	38.89	0.30
16	Shwarobit (83)	50.60	9.00	5.02	40.9	43.27	0.33
17	Humera-1	52.50	9.34	4.91	40.82	43.6	0.28
18	ACC-202–950	48.20	9.01	5.01	40.68	43.32	0.35
19	NN-0026	49.10	9.04	5.07	41.52	43.5	0.28
20	ACC-NO-041	51.60	9.00	**4.73**	38.74	**45.67**	0.33
21	ACC-203–612	48.40	8.78	5.04	46.44	38.55	0.34
22	ACC-200–064–1	48.20	8.75	5.07	43.95	41.29	0.34
23	Setit-1	53.80	9.60	4.88	38.51	45.35	0.25
24	Tejareb Kokit Sel-3	49.90	8.98	5.01	**48.21**	**36.56**	0.31
25	Orofalc ACC-2	52.50	9.14	4.90	45.29	38.73	0.26
26	NN-0022	46.8	9.01	5.06	42.89	41.93	0.32
27	Setit-3	49.10	**10.23**	5.06	38.85	44.98	0.33
28	ABX= 2–01–2	48.90	8.86	5.10	42.42	42.41	0.30
29	NN-0020	49.30	8.89	5.03	43.40	41.69	0.30
30	ACC-NS-010	**54.10**	**8.51**	4.91	40.11	44.97	0.31
31	Hirhir Sel-2	51.80	9.38	4.97	39.89	44.54	0.27
32	Abxt-85-Sel-2–1	45.40	8.74	5.18	45.6	39.81	0.35
33	Setit-2	46.10	8.84	5.12	**47.54**	37.49	**0.39**
34	Hirhir Kebabo Hairless-Sel-7	53.00	9.42	4.90	44.52	39.88	0.24
35	NN-0143	48.30	9.03	5.08	45.13	39.72	0.28
36	ACC NS-031	44.90	9.04	5.17	40.67	42.53	**0.38**
37	NN-0029(2)	50.70	8.89	4.92	45.95	38.92	0.28
38	NN-0054	48.80	9.16	5.12	44.58	40.08	0.26
39	Morgo-Sel-P = 13	53.20	9.21	4.87	42.45	41.88	0.27
40	Hirhir Humera Sel-8	**54.70**	9.85	**4.83**	**36.76**	**47.14**	0.28
41	Tejareb Girar	53.80	9.19	4.94	40.64	43.75	0.27
42	NN0027	47.90	8.67	5.06	44.85	40.47	0.32
43	NN0009	46.50	9.54	5.13	44.46	39.98	0.37
44	ACC-203–610	51.80	8.86	4.94	44.37	40.4	0.30
45	NN-0146	47.40	**8.63**	5.00	45.61	39.53	0.34
46	NN-0044–2	50.60	9.17	4.99	45.41	39.36	0.29
47	NN-0018–2	49.90	9.15	5.04	41.48	42.81	0.33
48	Hirhir Nigara 1st Sel-1	50.40	8.88	5.00	44.3	40.76	0.30
49	NN00136–1	47.50	9.01	5.19	41.49	41.95	0.35
50	NN-0088–2	47.30	**8.62**	5.08	**46.71**	38.39	0.35
51	Hirhir Baeker-Sel-3	53.40	**8.63**	4.92	40.49	44.66	0.33
52	NN0068–3	51.20	8.98	4.96	43.33	41.72	0.26
53	NN0074–3	47.40	8.91	5.16	41.76	43.39	0.31
54	NN0036–1	49.40	8.71	5.02	42.26	42.61	0.32
55	Hirhir Kebabo Hairless Sel-4	53.50	9.20	4.92	38.9	**45.56**	0.24
56	NN0001–2	50.90	9.18	4.94	43.35	41.31	0.29
57	Bawnji Sel-2	46.50	9.18	4.94	43.35	41.31	0.29
58	NN0058–2	48.20	8.71	5.07	**48.80**	36.77	**0.39**
59	G-02	48.70	9.47	4.84	43.99	40.68	0.34
60	ACC-NS-007(2)	52.40	9.81	5.05	42.71	41.14	0.32
61	GA-002(3)	49.30	8.94	4.92	44.54	39.99	0.28
62	Endelemi Kirem Sel-2	52.50	9.37	5.11	43.73	40.35	0.31
63	ACC-205–299	46.30	8.76	5.01	39.27	**45.71**	0.29
64	Hirhir Kebabo Early Sel-1	53.70	9.03	5.17	40.17	43.37	0.36
65	NN0016–1	49.20	9.26	4.85	40.24	44.17	0.28
66	Hirhir Adgeshu Sel -8	49.20	9.79	4.95	43.93	40.31	0.34
67	NN0015	**54.00**	9.32	5.04	42.88	41.48	0.30
68	NN01–13	50.60	9.06	4.91	43.01	41.39	0.29
69	Bering Bawany	48.80	9.48	5.03	43.38	40.70	0.29
70	Hirhir Nigara 1st Sel-2	47.90	9.13	5.05	42.89	41.86	0.32
71	Gojam Azene (Yohans Sel-1)	53.10	8.94	5.11	42.50	42.43	0.33
72	NN0038–1	48.60	9.45	4.91	41.02	43.18	0.29
73	NN0104	51.80	9.28	5.13	44.85	39.73	0.33

(*continued on next page*)

Table 1 (*continued*)

Entry number	Genotype name or designation		Fatty acid compositions (%)†				
		OC (%)	C16:0	C18:0	C18:1	C18:2	C18:3
74	Hirhir Kebabo Hairless Sel-6	50.30	**8.31**	4.92	41.27	43.91	0.29
75	ACC 205–180	53.10	8.87	4.99	44.04	41.1	0.29
76	Hirhir	45.80	9.34	4.93	41.44	42.58	0.25
77	ACC 203–616	51.70	9.03	5.00	40.32	43.39	**0.38**
78	NN0025	46.60	8.73	4.91	45.93	38.73	0.28
79	NN-0183–3	45.80	9.22	5.08	45.35	39.53	0.36
80	Hirhir Humera	49.50	8.86	5.13	**46.71**	38.50	**0.38**
81	NN0031	50.00	8.91	5.08	44.20	40.82	0.29
82	NN0061	50.10	8.83	4.99	44.97	39.9	0.33
83	ABXC-50402	48.70	9.33	4.94	43.12	41.44	0.32
84	NN0021	51.50	9.18	5.06	44.36	40.15	0.32
85	NN0079–1	52.40	9.05	5.02	42.76	42.12	0.30
86	ACC 202–333	40.00	9.04	5.03	40.30	39.02	0.33
87	NN-0052	46.80	9.02	5.20	40.70	42.98	0.37
88	NN-0029–1	47.40	9.02	5.09	40.50	43.52	0.35
89	Teiahir Sanja Sel-6	50.90	9.19	5.02	44.35	40.27	0.29
90	ACC-202–358	48.50	9.23	5.00	44.5	40.28	0.33
91	NN0032	52.60	8.93	4.96	45.3	39.42	0.3
92	NN0071	51.20	9.15	4.96	41.97	42.29	0.31
93	NN0064–1	51.10	9.11	5.03	41.36	43.57	0.29
94	NN0056	**55.20**	9.08	**4.82**	40.76	44.06	0.27
95	NN-01–03	50.00	9.43	5.03	43.67	40.69	0.31
96	NN0032–2	49.50	9.32	5.03	42.96	41.51	0.31
97	Bawnji Maksegnt	48.30	9.71	5.09	45.67	37.82	0.32
98	Bawnji Flwha Sel-2	47.60	9.03	4.98	39.23	44.32	0.36
99	NN0068–2	49.80	9.02	4.93	39.6	44.47	0.35
100	Hirhir Filwha Large Seeded	48.10	8.90	5.11	44.13	40.8	0.34
	t-statistics, (df=98)	189.55	284.50	486.76	172.53	184.37	83.67
	Standard deviation (SD)	2.59	0.32	0.10	2.45	2.23	0.04
	Significant value (p = 0.05)	* *	* *	* *	* *	* *	* *
	Mean	49.84	9.10	5.01	42.85	41.69	0.31
	Minimum	44.30	8.31	4.73	36.76	36.56	0.23
	Maximum	55.60	10.23	5.36	48.80	47.14	0.39

* * denote significance at the 1% probability level

† OC, Oil content; C16:0, Palmitic acid; C18:0, Stearic acid; C18:1, Oleic acid; C18:2, Linoleic acid; and C18:3, Linolenic acid;
Bold-face values denote selected genotypes, while entries highlighted in red expressed the lowest values.

NN0015 (54.0%). Gojam Azene at 44.30% exhibited the lowest seed oil content recorded in the study. The oleic acid content of the assessed genotypes ranged from 36.7% to 48.8%, with a mean of 42.9%, while the linoleic acid content varied from 36.6% to 47.1%, with a mean of 41.7%. Genotypes that recorded the highest oleic acid content were NN0058–2 (48.80%), Tejareb Kokit Sel-3 (48.2%), Setit-2 (47.5%), ACC-203–020 (47.3%), Hirhir Humera and NN-0088–2 (46.7%). Hirhir Humera Sel-8 recorded the lowest oleic acid content of 36.76%. Some of the assessed genotypes had a linoleic acid content of > 45%, including Hirhir Humera Sel-8 (47.14%), Hirhir Humera Sel-6 (46.00%), ACC-205–299 (45.71%), ACC-NO-041 (45.67%), and Hirhir Kebabo Hairless Sel-4 (45.56%).

3.2. Correlations of oil content and fatty compositions

Phenotypic correlation coefficients for the seed oil content and fatty acid profiles of the 100 sesame genotypes are presented in Table 2. Highly significant negative correlations were observed for the contents of oil and stearic acid ($r = -0.83$; $p < 0.01$) and linolenic acid ($r = -0.81$; $p < 0.01$), whereas relatively higher and positive correlations were recorded between stearic and linolenic acid contents ($r = 0.61$; $p < 0.01$). Significant positive correlations were recorded between seed oil content and linoleic acid content ($r = 0.39$; $p < 0.01$).

Table 2
Phenotypic correlations coefficients for the assessed seed oil content and fatty acid compositions (%) of 100 sesame germplasm collections.

Traits†	OC	C16:0	C18:0	C18:1	C18:2	C18:3
OC	1.00	0.14 ns	-0.83 * *	-0.35 * *	0.39 * *	-0.81 * *
C16:0		1.00	-0.12 ns	-0.37 * *	0.28 * *	-0.19 *
C18:0			1.00	0.26 * *	-0.31 * *	0.61 * *
C18:1				1.00	-0.98 * *	0.19 *
C18:2					1.00	-0.24 *
C18:3						1.00

†OC, Oil content; C16:0, Palmitic acid; C18:0, Stearic acid; C18:1, Oleic acid; C18:2, Linoleic acid; and C18:3, Linolenic acid; * and * *, denote significant at the 5% and 1% probability levels, respectively; ns, non-significant.

3.3. Principal component analysis

Principal component analysis (PCA) was performed to ascertain each trait's contribution to the overall observed variation in seed oil and fatty acid compositions. Overall, three principal components were identified with Eigenvalues ≥ 1 explaining 81.00% of the total variation for the assessed traits (Table 3 and Fig. 1). The first two principal components (PC1) and PC2 explained the highest proportion to the total variance. PC1 explained 44.00% of the total variation, with oil content and stearic acid contributing the most significant variation to PC1. While PC2 accounted for 22.00% of the total variation and oleic and linoleic acids were the most influential traits. Fig. 1 is a bi-plot showing the sesame genotypes and traits projection based on the first two principal components. The two principal components or dimensions accounted 59.20% (Dm1 = 36.6% and Dm2 = 22.6%) to the total variation. The bi-plot delineated the test genotypes and traits across the four quadrants. For instance, entries 97 and 62 recorded the highest oleic acid and linolic acid contents, respectively and situated in quadrant II.

Table 3
Eigenvalues, explained variation, cumulative variation (%) and principal components (PCs) for oil content and fatty acid compositions amongst 100 sesame germplasm collections.

Traits†	PC1	PC2	PC3
OC	**-0.49**	-0.34	0.08
C16:0	-0.22	0.30	0.08
C18:0	**0.44**	0.35	-0.09
C18:1	0.41	**-0.53**	0.03
C18:2	-0.42	**0.48**	-0.04
C18:3	0.42	0.39	0.11
Eigenvalue	3.11	1.56	1.02
Explained variation (%)	44.00	22.00	15.00
Cumulative explained variation (%)	44.00	66.00	81.00

†OC, Oil content; C16:0, Palmitic acid; C18:0, Stearic acid; C18:1, Oleic acid; C18:2, Linoleic acid; and C18:3, Linolenic acid.
Bold-faced values in a column denote the most influential traits corresponding to the principal component.

3.4. Cluster analysis of sesame genotypes based on oil content and fatty compositions

Based on the contents of oil and fatty acids, the cluster analysis resolved the 100 sesame genotypes into two main clusters (Table 4). Each cluster was further partitioned into two sub-clusters revealing genotype differentiation for selection. Cluster I comprised 61 accessions and one landrace (Hirhir). Amongst the accessions, 45 were collected from Amhara, 12 from Tigray, 3 from Afar, and one each from Oromia and Gambela administrative regions in Ethiopia. Cluster II contained 34 accessions of which 19 were sourced from Amhara region, 8 from Tigray, 5 from Afar and 2 from Oromia regions. Furthermore, Cluster II comprises four improved varieties (Humera-1, Setit-1, Setit-2, and Setit-3) sourced from the Tigray region.

Most of the accessions allocated in Cluster I had intermediate to high oil content (50.90–55.20%) but low to intermediate oleic (36.76–44.64%), linoleic (36.77–40.80%), and linolenic acids (0.24–0.29%). Cluster I genotypes Hirhir Humera Sel-8, NN0015, and NN0056 had relatively higher oil content (≥54%). Most accessions in this cluster recorded oleic acid of < 45%, except NN-0088–2 and Hirhir Humera, which had > 46%. Furthermore, Cluster I comprised accessions such as Hirhir Humera Sel-8, Hirhir Kebabo Hairless Sel-4 and ACC-205–299, which had higher linoleic contents of > 45%. Genotypes allocated in Cluster II had low to high oil content (44.30–55.60%) but relatively higher oleic acid (39.89–48.21%), linoleic acid (38.55–46.00%), and linolenic acid (0.29–0.39%). Most test genotypes in Cluster II recorded oil contents of < 50%, except Bawnji Fiyel Kolet, NN-0068–1, Setit-1, Hirhir Kebabo Hairess-Sel-7, Hirhir Humera Sel-6, and ACC-NS-010, which had ≥ 53%. Genotypes Hirhir Kebabo Hairless Sel-2, ACC-203–020, ACC-203–612, Tejareb Kokit Sel-3, and Setit-2 expressed > 46% of oleic acid contents. Genotypes Hirhir Humera Sel-6, ACC-No-041, and Setit-1 were grouped in Cluster II and expressed linoleic compositions of higher than 45%.

3.5. SSR markers characterization

The summary of genetic parameters based on the tested SSR markers are presented in Table 5. The major alleles frequency per locus ranged from 0.52 (for markers ZMM1189) to 0.96 (ZMM2818), with a mean of 0.78 alleles per locus. The observed heterozygosity had a mean value of 0.43, varying from 0.08 to 0.96. The average expected heterozygosity value was 0.30 and ranged from 0.08 to 0.5. The polymorphic

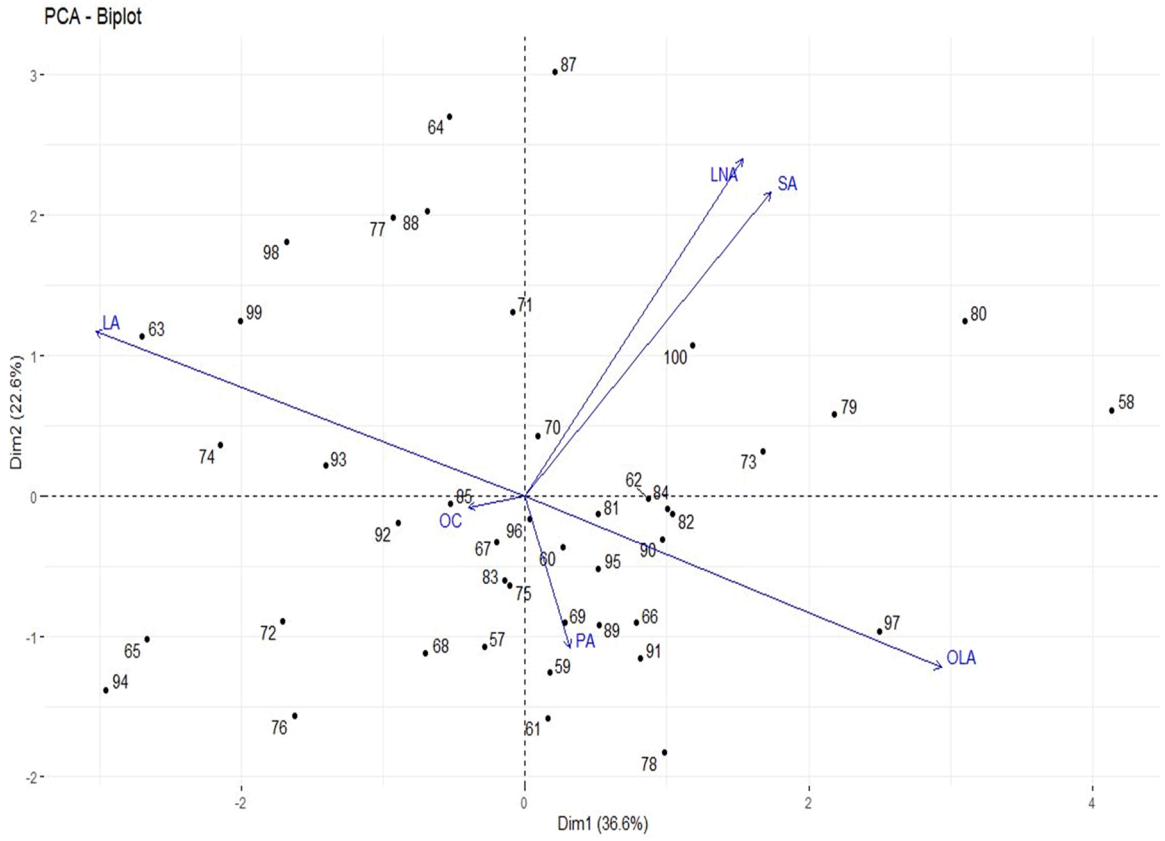

Fig. 1. A bi-plot depicting the distribution of 100 sesame genotypes using seed oil and fatty acid contents based on two principal components or dimensions (Dm1 = 36.6% and Dm2 = 22.6%). Note OC, Oil content, PA, Palmitic acid, SA, Stearic acid, OLA, Oleic acid, LA, Linoleic acid, and LNA, Linolenic acid. See Supplementary Table 1 for codes of the test genotypes. Note: the code numbers in the Figure represent the entry number of genotypes presented in Table 1.

Table 4
Distribution of the 100 sesame germplasm collections based on oil content and fatty acid compositions.

Germplasm	Source (regions or research centre in Ethiopia)	Cluster and collection name			
		I		II	
		I-a	I-b	II-a	II-b
Accessions	Amhara Region	NN0016–1, NN0015, NN01–13, Gojam Azene(Yohans Sel-1), NN0038–1, NN0104, ACC 203–616, NN0025, NN-0183–3, NN0031, NN0061, NN0021, NN0079–1, NN-0052, NN-0029–1, Teiahir Sanja Sel-6, ACC-202–358, NN0032, NN0071, NN0064–1, NN0056, NN-01–03, NN0032–2, Bawnji Maksegnt, Bawnji Flwha Sel-2, NN0068–2, Hirhir Filwha Large Seeded, ACC 202–333	Tejareb Girar, NN0027, NN0009, ACC-203–610, NN-0146, NN-0044–2, NN-0018–2, NN00136–1, NN-0088–2, NN0068–3, NN0074–3, NN0036–1, NN0001–2, Bawnji Sel-2, NN0058–2, G-02, Endelemi Kirem Sel-2,	ACC-203–612, Tejareb Kokit Sel-3, NN-0022, NN-0020, NN-0143, NN-0029(2), NN-0054	NN-0068–1, NN-0108–2, NN-034, NN-0129–2, ACC-203–020, NN-0038–2, Bawnji Fiyel Kolet, Gonjam Azen(Aleka), Bawnji Gobate, Shwarobit(83, ACC-202–950, NN-0026
	Tigray Region	Hirhir Adgeshu Sel −8, Hirhir Nigara 1st Sel-2, Hirhir Kebabo Hairless Sel-6, ACC 205–180, Hirhirh Humera	Hirhir Humera Sel-8, Hirhir Nigara 1st Sel-1, Hirhir Baeker-Sel-3, Hirhir Kebabo Hairless Sel-4, ACC-205–299, Hirhir Kebabo Early Sel-1	ACC-200–064–1, Hirhir Sel-2, Hirhir Kebabo Hairless-Sel-7	Hirhir kebabo Hairless sel-2, Hirhir kebabo Hairless-9, ACC 031–5–14, Hirhir Humera Sel-6, ACC-NO-041,
	Afar Region	Bering Bawany, ABXC-50402	Morgo-Sel-P = 13	Orofalc ACC-2, ABX=2–01–2, ABXT-85-SEL-2–1	GXT=85(28–2), BCS-0041
	Oromia Region	–	ACC-NS-007(2)	ACC-NS-010, ACC NS-031	
	Gambela Region		GA-002(3)	–	–
Landrace	Tigray Region	Hirhir	–		
Improved varieties	Humera Agricultural Research Center	–	–	Setit-2, Setit-3	Humera-1, Setit-1
Sub-total		36	26	17	21
Total		100			

Table 5
Genetic parameter estimates of 27 SSR markers using 100 sesame genotypes.

Locus	Primer size (bp)	Genetic parameter			
		MAF	He	Ho	PIC
ID0046	101	0.72	0.40	0.56	0.32
ZMM1043	184	0.75	0.38	0.51	0.31
ZMM3261	244	0.59	0.48	0.82	0.37
ID0041	280	0.96	0.08	0.08	0.07
ZMM5015	151	0.79	0.34	0.43	0.28
ZMM4664	184	0.60	0.48	0.80	0.36
ZMM1809	256	0.86	0.24	0.28	0.21
ZMM2321	280	0.90	0.18	0.20	0.16
ZMM5358	164	0.62	0.47	0.77	0.36
ID0068	199	0.86	0.25	0.29	0.22
ZMM3312	264	0.56	0.49	0.89	0.37
ZMM1033	179	0.76	0.37	0.49	0.30
ZMM1189	212	0.52	0.50	0.96	0.37
ZMM2202	276	0.89	0.20	0.22	0.18
ZMM1637	265	0.68	0.44	0.65	0.34
ZMM4645	179	0.81	0.31	0.39	0.26
ZMM1700	258	0.95	0.10	0.10	0.09
ID0175	271	0.96	0.08	0.08	0.07
ZMM1353	169	0.94	0.12	0.13	0.11
ID0145	196	0.77	0.36	0.47	0.29
ZMM4803	268	0.95	0.10	0.11	0.10
ZMM6141	167	0.75	0.38	0.51	0.31
ZMM3013	216	0.69	0.43	0.63	0.34
ZMM2818	279	0.96	0.08	0.08	0.07
ZMM3223	279	0.82	0.30	0.36	0.25
ZMM1691	220	0.73	0.39	0.54	0.32
ZMM1851	280	0.90	0.18	0.20	0.16
Mean	221	0.78	0.30	0.43	0.25

MAF= Major allele frequency, He = Unbiased expected heterozygosity (gene diversity), Ho = Observed heterozygosity, PIC = Polymorphic information content. bp= base pairs

information content values of the markers varied from 0.07 (markers ID0041, ID0175, and ZMM2818) to 0.37 (ZMM3261, ZMM3312 and ZMM1189), with a mean of 0.25. The magnitude of the PIC values determines the informativeness of the markers for genetic diversity analysis.

3.6. Population structure analysis

The Bayesian-based analysis of population structure using SSR markers showed that the log-likelihood at K = 4 was optimal to group the subset of 100 sesame accessions into four genetically distinct subpopulations (Fig. 2a). The population structure analysis revealed the existence of four distinct subpopulations among the 100 sesame genotypes (Fig. 2b). A total of 63 accessions were stratified into four subpopulations, while 37 accessions were admixtures (Table 7). Subpopulation I is the largest which consisted of 24 accessions collected from the Amhara (17 collections), Tigray (3), Afar (3), and Oromia (1) regions. Subpopulation II had 13 accessions collected from the following regions: Amhara (8), Afar (4), and Tigray (1). Subpopulation III comprised nine accessions sourced from the Amhara (5) and Tigray (4) regions. Subpopulations IV is the second-largest population consisting of 17 accessions sourced from Tigray (10), Amhara (6), and Afar (1) regions. The principal coordinate analysis resolved admixture groups among the genotypes (Fig. 2c).

3.7. Cluster analysis of 100 Sesame accessions

In agreement with the seed oil and fatty acid compositions (Table 4), the cluster analysis based on SSR markers resolved two main clusters and two sub-clusters (Fig. 3). Cluster I consisted of 49 accessions and one improved variety sourced from the following regions: 37 accessions (from Amhara), 5 accessions and one improved variety (Tigray), 6 accessions (Afar), and one accession (Oromia). Cluster II was the most divergent, which comprised 50 genotypes sourced from Amhara (28 accessions), Tigray (13 accessions, one landrace, and three improved

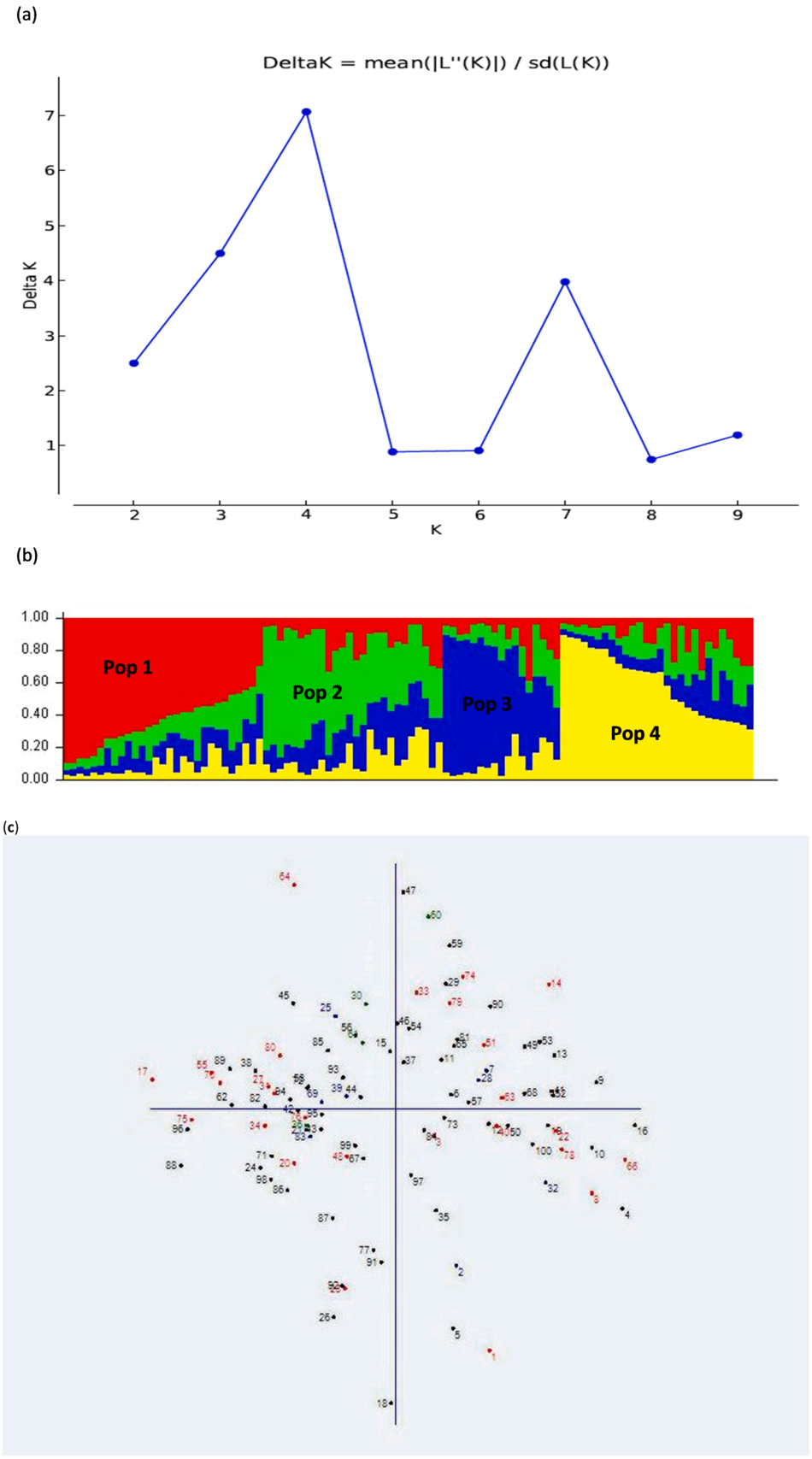

Fig. 2. Population inference among100 sesame entries based on 27 SSR markers: (a) Delta K estimation based on the Evanno procedure, (b) Sub-populations for the best delta K value of four. Pop 1, 2, 3, and 4 denote Populations 1, 2, 3, and 4, respectively, and (c) principal coordinate clustering of genotypes. Note: the code numbers in the Figure represent the entry number of genotypes presented in Table 1.

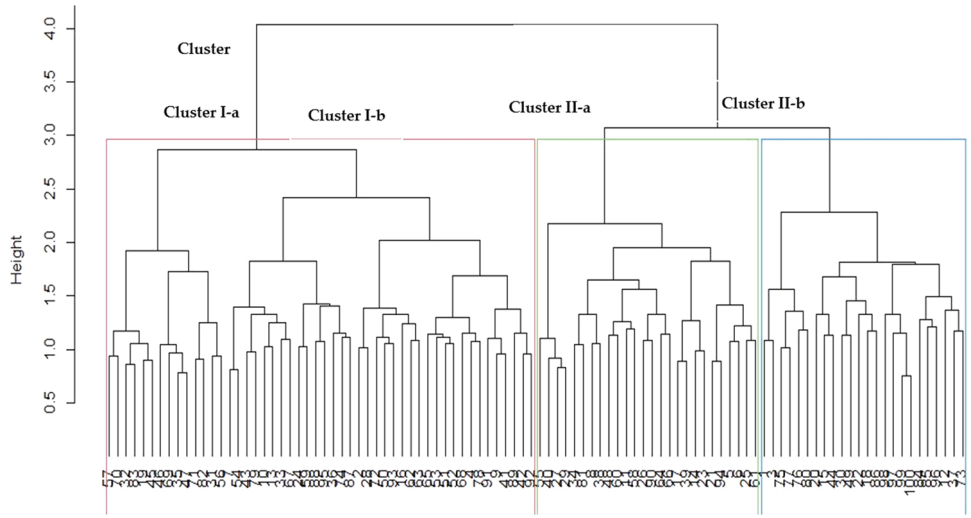

Fig. 3. Dendrogram based on Unweighted Pair Group Method with Arithmetic Mean (UPGMA) showing the genetic relationship among 100 sesame entries using 27 SSR markers. Note: see Supplementary Table 1 for codes of entries.

varieties), Afar (two accessions), and Oromia (two accessions), and Gambela (one accession). Similar to the population structure analysis, all the two main clusters and their respective sub-branches had genotypes with high oil and fatty acid contents.

Accessions allocated in Cluster I had intermediate (50.00–53.8%) to high (54.00–54.6%) oil (e.g. NN-0068–1, NN0015, Gojam Azene [Yohans Sel-1], Tejareb Girar, and Hirhir Baeker-Sel-3), high oleic acid (>45%) (Tejareb Kokit Sel-3, Setit-2, and NN-0088–2), high linoleic acid (>44%) (ACC-205–299, Hirhir Baeker-Sel-3, Hirhir Sel-2, and NN0016–1) and high linolenic acid (>0.38) (Gojam Azene and ABX=2–01–2 contents.

Cluster II accessions recorded higher oil, oleic and linoleic acid contents. Accessions Bawnji Fiyel Kolet, NN0056, Hirhir Humera Sel-8, and ACC-NS-010 had relatively higher oil content (>54%). The majority of accessions in this cluster recorded oleic acid content of < 45%, except NN0058–2, Hirhir Humera, Hirhir Kebabo Hairless Sel-2, and ACC-203–612, which had > 46%. In addition, cluster II comprises accessions such as Hirhir Humera Sel-8, Hirhir Humera Sel-6, ACC-NO-041, Hirhir Kebabo Hairless Sel-4, and Setit-1 with a relatively higher linoleic acid content (>44%) and higher linolenic acid content recorded on some accessions: NN0058–2 (0.39%), ACC 203–616 (0.38%), and Hirhir Humera (0.38%).

Crosses should be made between parental lines selected from different clusters with positive values to develop new breeding populations with desirable oil content and fatty acid compositions. Hence accessions Hirhir Kebabo Hairless Sel-6 (from sub-cluster I-b), Hirhir Humera Sel-8 and NN0058–2 (sub-cluster II-a) and Bawnji Fiyel Kolet (sub-cluster II-b), are ideal candidates with complementary fatty acid profiles for direct production and/or further breeding.

4. Discussion

4.1. Oil content and fatty acid compositions

The test genotypes showed significant ($p \leq 0.05$) variation for seed oil content and fatty acid profiles (Table 1). This suggests that the germplasm pool contains vital seed oil content and fatty acid profiles for sesame improvement through hybridization and selections. Oil content and fatty acid compositions of sesame genotypes for assessed traits are summarized in Table 1. The mean oil content across the studied genotypes was 49.84%, which was similar to 50.1% reported by Biglar et al. (2012). However, Dossa et al. (2018) reported a relatively higher mean oil content of 53.00%. In addition, Agidew et al. (2021) reported a significantly higher oil content (53.2–58.2%) in the Ethiopian sesame germplasm. The mean palmitic and stearic acid content were 9.10% and 5.1% in that order (Table 1). In line with this, Park et al. (2015) reported a mean value of 9.9% and 5.8% palmitic and stearic acid contents, respectively, in sesame accession in Korea. Contrary to the present study, Biglar et al. (2012) reported low palmitic acid (9.6%) and stearic acid (4.7%) contents in world collection of sesame. Agidew et al. (2021) reported higher palmitic and stearic acid contents of 10.2% and 6.15%, respectively, in the Ethiopian collections. The highest oil content was recorded for entries Bawnji Fiyel Kolet (55.6%), NN0056 (55.2%), Hirhir Humera Sel-8 (54.7%), NN-0068–1 (54.6%), ACC-NS-010 (54.1%) and NN0015 (54.0%). The oil content of the identified genotypes was higher than 31.00–48.00% of previously evaluated sesame germplasm in East Africa, including Kenya, Tanzania and Uganda (Were et al. 2006). Cultivation of these genotypes is vital for export market standards and to meet the quality attributes of the confectionary industry globally. The oleic acid content ranged from 36.7% to 48.8% (with a mean of 42.9%), which was higher than the mean oleic content of 38.10% and lower than the mean value of linoleic acid content of 41.70% in the present study reported by Dossa et al. (2018). Also, Biglar et al. (2012) reported a mean value of oleic acid content of 43.3% (range of 32.7–53.9%) in world sesame collection. Park et al. (2015) reported that oleic acid content ranged from42.0 to 43.0%, which is lower than the present study. Conversely, Agidew et al. (2021) reported relatively lower oleic acid (37.2–38.9%) and higher linoleic acid (42.5–44.3%) contents in the Ethiopian sesame collection. Genotypes Tejareb Kokit Sel-3 (48.2%), Setit-2 (47.5%), ACC-203–020 (47.3%), both Hirhir Humera and NN-0088–2 (46.7%) expressed the highest oleic acid content. Hirhir Humera Sel-6, ACC-NO-041, Setit-1, Hirhir Humera Sel-8, Hirhir Kebabo Hairless Sel-4 and Hirhir Kebabo Early Sel-1 with a linoleic acid content of > 45% were identified in the present study. The unsaturated fatty acids (C18:1 and C18:2), constitute the major fatty acids in sesame oil (84.0%) (Wei et al. 2015). Cultivating sesame genotypes with high oleic and linolic acid composition is beneficial to human health by minimising the risks of cardiovascular diseases, cancer, and brain and liver damage (Yen, 1990; Yol et al. 2015).

4.2. Traits Associations

In the present study, oil content was negatively correlated with oleic acid content and corroborated with the findings of Dossa et al. (2018). This suggests that breeding sesame simultaneously for high oil and oleic acid contents would be difficult. A moderately high and positive

correlation was recorded between stearic and linolenic acid content, indicating both traits could be improved simultaneously among the studied sesame genotypes. Oleic acid content was negatively correlated with linoleic acid content (Table 2) and corroborated with the findings of Park et al. (2015). Similarly, Dossa et al. (2018) reported a negative correlation between sesame oleic and linoleic acid contents. Therefore, improvement of oil content among the studied sesame population should be based on direct selection of high oil and fatty acid content genotypes and transgressive segregants in future improvement programs.

Identifying and selecting sesame genotypes with high oil content and fatty acid compositions is important in utilizing the germplasm in sesame breeding programs. Dossa et al. (2018) identified two PCs, which explained 79.43% of the total variation in 139 sesame genotypes in Africa and Asia for oil content, protein, and fatty acid compositions. They reported that linolenic acid and protein content were the largest contributors to the explained variation in PC1, whereas oil content and oleic acid content were in PC2. In the present study, the main contributors to the observed phenotypic variation were oil content, stearic acid, oleic and linolic acids with high loading coefficients in PC1 and PC2.

4.3. Cluster analysis of 100 sesame accessions

Cluster analysis identified two major clusters and four sub-clusters, revealing genetic variation among the assessed sesame genotypes (Table 4). Dossa et al. (2018) grouped 139 sesame genotypes into two clusters using oil, protein, and fatty acid compositions. In the present study, some sesame genotypes collected from different regions were grouped in the same cluster, such as Hirhir Humera Sel-8 (Tigray), Hirhir Filwha Large Seeded (Amhara), GA-002(3) (Gambela) and ABX= 2–01–2 (Afar) and ACC NS-031 (Oromia). Therefore, the geographic origin of germplasm collections is not necessarily a key indicator of genetic diversity. The exchange of genetic materials among farmers and traders in the regions contributes to genetic variation across different regions. For improved oil content and oleic, linoleic, and linolenic acids content the ideal genotypes for future crosses are situated in Clusters I and II, which possessed candidates with excellent oil content and oleic, linoleic, and linolenic acids.

4.4. Genetic diversity and population structure of sesame germplasm based on SSR markers

Simple sequence repeats markers are useful genomic resources to complement phenotypic data for effective selection. In the present study the major alleles frequency per locus recorded a mean value of 0.78 among the sesame genotypes (Table 5), which was much higher than that reported by Asekova et al. (2018) and Adu-Gyamfi et al. (2019). The genotypic differences and the number of SSR markers used in the genetic analysis are attributable to variation in allele frequency (He et al. 2011; Baraket et al. 2011; Jifar et al. 2020). The average observed heterozygosity value of 0.43 reported in the present study is higher than the findings of Gebremichael and Parzies (2011); Asekova et al. (2018), and Araújo et al. (2019), who reported values of 0.23, 0.01, and 0.12 in sesame respectively. This study's observed heterozygosity was lower than the findings of He et al. (2011), who reported a value of 0.56 in sesame. The mean expected heterozygosity recorded in the present study (He = 0.30) (Table 5), which was lower than the values of 0.72 and 0.34 reported by Asekova et al. (2018) and Araújo et al. (2019) using 23 and 10 SSR markers among 129 and 36 sesame accessions, respectively. The higher heterozygosity recorded in the present study suggested that the Ethiopian sesame populations have a high genetic variation for selection. In the present study, the higher heterozygosity recorded suggested that the Ethiopian sesame populations have a high genetic variation for selection.

The population structure revealed four populations (Fig. 2). Using 44 and 23 SSR markers, Wei et al. (2016) and Asekova et al. (2018) found two populations among 94 and 129 sesame genotypes from China and Korea, respectively. Only 63 genotypes were structured into the four populations. Nonetheless, the test accessions sourced from one region were distributed across the different populations. This indicates that geographical separation does not affect the genetic differentiation of sesame genotypes. Results showed that the four sub-populations comprised genotypes collected from different sources, although most of the released genotypes (Humera-1, setit-1 and 3) were grouped in subpopulation IV, except setit-2 was admixture.

The expected heterozygosity was 0.15, 0.22, 0.29 and 0.20 in subpopulation 1, 2, 3, and 4, respectively (Table 5). The level of genetic differentiation among the subpopulations was measured by estimating the fixation index (FST). The results showed that subpopulation 1 with a higher FST of 0.39 was more differentiated than subpopulations 2, 3 and 4, with FST values of 0.23, 0.02, and 0.20. The greater gene fixation index of 0.39 in population I, which included accessions from the Amhara, Tigray, Afar, and Oromia regions, suggests that these regions have more genetic differentiation due to substantial gene flow. In contrast, the low gene fixation index found in population III, which includes accessions from the Amhara and Tigray regions, revealed a lack of differentiation. This could be attributed to gene flow from collection to collection via germplasm exchange.

The UPGMA cluster analysis identified two major clusters and four sub-clusters (Fig. 3). Using 23 SSR markers, Asekova et al. (20018) divided 129 sesame genotypes into two clusters. The genotype clustering patterns in this study did not match the intended population structure based on the collection regions. This could be because genotypes from similar places are from the same gene pool or have similar ancestral ties (Mulualem et al. 2018). On the other hand, William et al. (2019) suggested that genetic dissimilarity between test genotypes could be generated by a variety of ancestral origins, significant gene flow driven by cross-pollination, and probable gene or chromosomal mutation. Some sesame genotypes collected from different regions were grouped in the same cluster in the current study. This result corroborated the findings of Zhang et al. (2012b) among 24 sesame genotypes in China. This indicating that geographical separation does not affect the genetic differentiation of germplasm (Ganesamurthy et al. 2010). As a result, substantial gene flow and lack of genetic differentiation result from the exchange of genetic materials among farmers and traders in the regions. Farmers' selections and management practices affect genetic diversity patterns (Barnaud et al. 2008).

Parental lines with complementary traits selected from different clusters could be utilized to develop new breeding populations possessing desirable oil content and fatty acid compositions. Accordingly, Hirhir Kebabo Hairless Sel-6 (from sub-cluster I-b), Hirhir Humera Sel-8 and NN0058–2 (sub-cluster II-a) and Bawnji Fiyel Kolet (sub-cluster II-b) are some of the accessions that might be considered desirable parents. These clusters comprised ideal candidates with high oil, oleic, linoleic, and linolenic contents for production and further breeding.

5. Conclusions

The current study determined the genetic diversity and relationships among Ethiopia's sesame germplasm collections using seed oil content and fatty acid compositions and diagnostic simple sequence repeat (SSR) markers to select genetically complementary and promising parental lines for breeding. The test genotypes showed wide variation for seed oil content and fatty acid compositions. The mean oil content of the assessed lines was 49.84% ranging from 44.30% to 55.60%. The oleic acid content ranged from 36.70% to 48.80%, with a mean of 42.90%, followed by linoleic acid (36.60–47.10%, mean 41.70%). The SSR markers revealed that the mean gene diversity and polymorphic information content were 0.30 and 0.25, respectively, indicating that the assessed sesame germplasms were diverse for selection. Population structure analysis identified four major heterotic groups useful for selection. Based on higher oil content and desirable fatty acid

compositions and SSR markers the following superior and complementary lines were selected: Hirhir Kebabo Hairless Sel-6 (from sub-cluster I-b), Hirhir Humera Sel-8 and NN0058–2 (sub-cluster II-a) and Bawnji Fiyel Kolet (sub-cluster II-b). The identified genetic resources are useful parental lines for sesame breeding programs in Ethiopia and elsewhere.

Funding

The research was financially supported by the Ethiopian Agricultural Transformation Institute (ATI) and the Bill & Melinda Gates Foundation (BMGF-CORE) through Education Grant; the Oil Crops Research Institute of the Chinese Academy of Agricultural Science (OCRI-CAAS) through the Crop Germplasm Resources Protection project (2019NWB033), and the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP 2016 OCRI).

CRediT authorship contribution statement

DHT: Carried out the phenotyping and lab studies, Data analyses, Conceptualization, Data curation, Writing – original draft. HS: Supervision, Conceptualization, designed the study, Writing – review & editing. AT: In-country co-supervision, Writing – review & editing. AS: Molecular data analysis, Writing – review & editing. The author (s) read and approved the final manuscript.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgment

The Ethiopian Agricultural Transformation Institute (ATI), and the Oil Crops Research Institute of the Chinese Academy of Agricultural Science (OCRI-CAAS) through the Crop Germplasm Resources Protection project (2019NWB033), and the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP 2016 OCRI) are thanked for research support. The Ethiopian Institute of Agricultural Research (EIAR), staff of the Humera Agricultural Research Centre (HuARC) and Melka Werer Agricultural Research Centre (WARC) are also greatly appreciated for germplasm supply and the overall field research support.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jfca.2022.104545.

References

Abdellatef, E., Sirelkhatem, R., Ahmed, M.M., Radwan, K.H., Khalafalla, M.M., 2008. Study of genetic diversity in Sudanese sesame (*Sesamum indicum* L.) germplasm using random amplified polymorphic DNA (RAPD) markers. Afr. J. Biotechnol. 7 (24), 4423–4427.

Adu-Gyamfi, R., Prempeh, R., Zakari, I., 2019. Diversity assessment of some sesame (*Sesamum indicum* L.) genotypes cultivated in northern Ghana using morphological and SSR markers. Adv. Agric. 3, 1–10. https://doi.org/10.1155/2019/6067891.

Agidew, M.G., Dubale, A.A., Atlabachew, M., et al., 2021. Fatty acid composition, total phenolic contents and antioxidant activity of white and black sesame seed varieties from different localities of Ethiopia. Chem. Biol. Technol. Agric. 8, 14. https://doi.org/10.1186/s40538-021-00215-w.

Araújo, E.D.S., Arriel, N.H.C., Santos, R.C.D., Lima, L.M.D., 2019. Assessment of genetic variability in sesame accessions using SSR markers and morpho agronomic traits. Aust. J. Crop Sci. 13 (01), 45–54. https://doi.org/10.21475/ajcs.19.13.01 p1157.

Asekova, S., Kulkarni, K.P., Oh, K.W., Lee, M.H., Oh, E., Kim, J.I., Yeo, U., Pae, U.S., Ha, T.J., Kim, S.U., 2018. Analysis of molecular variance and population structure of sesame (*Sesamum indicum* L.) genotypes using SSR markers. Plant Breed. Biotechnol. 6 (4), 321–336. https://doi.org/10.9787/PBB.2018.6.4.321.

Ashri, A., 2010. Sesame breeding. In: Janick, J. (Ed.), In Plant Breeding Reviews, Vol. 16. Oxford: John Wiley.

Baraket, G., Chatti, K., Saddoud, O., Abdelkarim, A.B., Mars, M., 2011. Comparative assessment of SSR and AFLP markers for evaluation of genetic diversity and conservation of Fig (Ficus carica L.) genetic resources in Tunisia. Plant Mol. Biol. Rep. 29, 171–184.

Barnaud, A., Trigueros, G., McKey, D., Joly, H.I., 2008. High outcrossing rates in fields with mixed sorghum landraces: How are landraces maintained? Heredity 101, 445–452.

Basak, M., Uzun, B., Yol, E., 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 (10), e0223757. https://doi.org/10.1371/journal.pone.0223757.

Bedigian, D. (1981). Origin, diversity, exploration and collection of sesame. In: Sesame: Status and Improvement, Proc. Expert Consultation, FAO, Rome, Italy 8–12, 164–16.

Bedigian, D., Harlan, J., 1986. Evidence for cultivation of sesame in the ancient world. Econ. Bot. 40 (2), 137–154.

Biglar, M., Moghaddam, G., Sadeghi, N., Oveisi, M.R., Jannat, B., Kaboli, Z., Hassani, S., Hajimahmoodi, M., 2012. Profiling of major fatty acids in different raw and roasted sesame seeds cultivars. Afr. J. Biotechnol. 11 (24), 6619–6623.

Central Statistic Authority (CSA). (2019). Ethiopian agricultural sample enumeration: Report on the primary results of area, production and yield of temporary crops of private peasant holdings in Meher Season. *Central Statistic Authority*: Addis Ababa, Ethiopia, Addis Ababa, Ethiopia.

Dossa, K., Wei, X., Niang, M., Liu, P., Zhang, Y., Wang, L., Liao, B., Cissé, N., Zhang, X., Diouf, D., 2018. Near-infrared reflectance spectroscopy reveals wide variation in major components of sesame seeds from Africa and Asia. Crop J. 6, 202–206.

Dossa, K., Wei, X., Zhang, Y., Fonceka, D., Wenjuan, Y., Diouf, D., Cisse, N., Liao, B., Zhang, X., 2016. Analysis of genetic diversity and population structure of sesame samples from Africa and Asia as major centers of its cultivation. Genes 7, 14.

Elleuch, M., Besbes, S., Roiseux, O., Blecker, C., Attia, H., 2007. Quality characteristics of sesame seeds and by-products. Food Chem. 103, 641–650.

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.

FAOSTAT. (2019). Food and Agriculture Organization of the United Nations; FAOSTAT Online Statistical Service: Rome, Italy, Available online: ⟨http://www.fao.org/faostat/en/#data/QC⟩ (accessed on 21 September 2021).

Ganesamurthy, K., Punitha, D., Elangovan, M., 2010. Genetic diversity among the land races of sorghum collected in Tamil Nadu. Electron. J. Plant Breed. 1, 1375–1379.

Gebremedhn, M.B., Tessema, W., Gebre, G.G., Mawcha, K.T., Assefa, M.K., 2019. Value chain analysis of sesame (*Sesamum indicum* L.) in Humera district, Tigray, Ethiopia. *Cogent Food &*. Agriculture 5 (1), 1705741.

Gebremichael, D.E., Parzies, H.K., 2011. Genetic variability among landraces of sesame in Ethiopia. Afr. Crop Sci. J. 19 (1), 1–13.

Gidey, Y.T., Kebede, S.A., Gashawbeza, G.T., 2012. Extent and pattern of the genetic diversity for morpho-agronomic traits in Ethiopian sesame landraces (*Sesamum indicum* L.). Asian J. Agric. Res. 6, 118–128.

He, Q., Li, X.W., Liang, G.L., Ji, K., Guo, Q.G., 2011. Genetic diversity and identity of Chinese loquat cultivars/accessions (Eriobotrya japonica) using apple SSR markers. Plant Mol. Biol. Rep. 29, 197–208.

Hika, G., Geleta, N., Jaleta, Z., 2014. Correlation and divergence analysis for phenotypic traits in sesame (*Sesamum indicum* L.) Genotypes. Sci. Technol. Arts Res. J. 3, 01–09.

Hika, G., Geleta, N., Jaleta, Z., 2015. Genetic variability, heritability and genetic advance for the phenotypic sesame (*Sesamum indicum* L.) Populations from Ethiopia. Sci Technol. Arts Res. J. 4, 20–26.

HuARC. (2010). Annual Research Report for the Period 2009/10; Humera Agricultural Research Centre: Tigray, Ethiopia.

Jifar, H., Tesfaye, K., Dagne, K., Assef,a, K., Tadele, Z., 2020. Genetic diversity and population structure of tef [Eragrostis tef (Zucc.) Trotter] as Revealed by SSR Markers. Adv. Crop Sci. Technol. 8 (1), 1000438.

Jones, N., Ougham, H., Thomas, H., Pasakinskiene, I., 2009. Markers and mapping revisited: finding your gene. New Phytol. 183, 935–966.

Laurentin, E.H., Karlovsky, P., 2006. Genetic relationship and diversity in a sesame (Sesamum indicum L.) germplasm collection using amplified fragment length polymorphism (AFLP). BMC Genet. 7 (10) https://doi.org/10.1186/1471-2156-7-10.

Mahajan, R.K., Bisht, I.S., Dhillon, B.S., 2007. Establishment of a core collection of world sesame (*Sesamum indicum* L.) germplasm accessions. SABRAO J. Breed. Genet. 39 (1), 53–64.

Mehra, K.L., 1967. Sesame in India. In: Weiss, E.A. (Ed.), Oilseed Crops, Tropical Agriculture Series. Longman, London, pp. 282–340.

Morris, J.B., 2020. Food, industrial, nutraceutical, and pharmaceutical uses of sesame genetic resources. In: Janick, J., Whipkey, A. (Eds.), In Trends In News Crops And New Uses. ASHS Press, Alexandria, Egypt, pp. 153–156.

Mulualem, T., Mekbib, F., Shimelis, H., Gebre, E., Amelework, B., 2018. Genetic diversity of yam (Dioscorea spp.) landrace collections from Ethiopia using simple sequence repeat markers. Aust. J. Crop Sci. 12, 1223–1230.

Myint, D., Gilani, S.A., Kawase, M., Watanabe, K.N., 2020. Sustainable sesame (*Sesamum indicum* L.) production through improved technology: an overview of production, challenges, and opportunities in Myanmar. Sustainability 12, 3515.

Park, J.H., Suresh, S., Raveendar, S., Baek, H.J., Kim, C.K., Lee, S., Cho, G.T., Ma, K.H., Lee, C.W., Chung, J.W., 2015. Development and evaluation of core collection using qualitative and quantitative trait descriptor in sesame (*Sesamum indicum* L.) germplasm. Korean J. Crop Sci. 60, 1. https://doi.org/10.7740/kjcs.2014.60.1.

Pritchard, J., Stephens, M., Rosenberg, N., Donnelly, P., 2000. Association mapping in structured populations. Am. J. Human Genet. 67, 170–180.

R Core Team. (2020). A Language and Environment for Statistical Computing; R Foundation for Computing: Vienna, Austria.

SAS Institute. (2018). Statistical Analysis Software; Version 9.4; SAS Institute Inc.: Cary, NC, USA.
Seegeler, C.J. (1983). Oil seeds in Ethiopia: their Taxonomy and Agricultural Significance. Centre for Agricultural Publication and Documentation, Wageningen, the Netherlands.
SPSS. (2020). Statistical Package for Social Sciences; SPSS: Chicago, IL, USA.
Teklu, D.H., Kebede, S.A., Gebremichael, D.E., 2014. Assessment of genetic variability, genetic advance, correlation, and path analysis for morphological traits in sesame genotypes. Asian J. Agric. Res. 7, 118–128.
Teklu, D.H., Shimelis, H., Tesfaye, A., Mashilo, J., Zhang, X., Zhang, Y., Dossa, K., Shayanowako, A.I.T., 2021. Genetic variability and population structure of ethiopian sesame (*Sesamum indicum* L.) germplasm assessed through phenotypic traits and simple sequence repeats markers. Plants 10, 1129.
Uzun, B., Ülger, S., Çağirgan, M.I., 2002. Comparison of determinate and indeterminate types of sesame for oil content and fatty acid composition. Turk. J. Agric. For. 26, 269–274.
Wacal, C., Ogata, N., Basalirwa, D., Sasagawa, D., Kato, M., Handa, T., Masunaga, T., Yamamoto, S., Nishihara, E., 2019. Fatty acid composition of sesame (*Sesamum indicum* L.) seeds in relation to yield and soil chemical properties on continuously monocropped upland fields converted from paddy fields. Agronomy 9, 801.
Wang, L., Zhang, Y., Li, P., Wang, X., Zhang, W., Wei, W., Zhang, X., 2012. HPLC analysis of seed sesamin and sesamolin variation in a sesame germplasm collection in China. J. Am. Oil Chem. Soc. 89, 1011–1020.
Weiss, E.A., 1983. Sesame. "Oilseed Crops". Longman Inc., New York, pp. 282–340.
Wei, X., Liu, K., Zhang, Feng, Y., Wang, Q.L., Zhao, Y., Li, D., Zhao, Q., Zhu, X., Zhu, X., Li, W., Fan, D., Gao, Y., Lu, Y., Zhang, X., Tang, X., Zhou, C., Zhu, C., Liu, L., Zhong, R., Tian, Q., Wen, Z., Weng, Q., Han, B., Huang, X., Zhang, X., 2015. Genetic discovery for oil production and quality in sesame. Nat. Commun. 6, 8609.
Wei, W., Zhang, Y., Wang, L., Li, D., Gao, Y., Zhang, X., 2016. Genetic diversity, population structure, and association mapping of 10 agronomic traits in sesame. Crop Sci. 56, 331–343. https://doi.org/10.2135/cropsci2015.03.0153.
Were, B.A., Onkware, A.O., Gudu, S., Welander, M., Carlsson, A.S., 2006. Seed oil content and fatty acid composition in East African sesame (*Sesamum indicum* L.) accessions evaluated over 3 years. Field Crops Res. 97 (2), 254–260.
William, T.S., Hussein, S., Mark, L., Isack, M., Admire, I.T.S., 2019. Assessment of the genetic diversity and population structure of rice genotypes using SSR markers. Acta Agric. Scandinavica, Section B — Soil Plant Sci. 70, 76–86.
Woldesenbet, D.T., Tesfaye, K., 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.
Yen, G.C., 1990. Influence of seed roasting process on the changes in composition and quality of sesame (Sesame indicum L.) oil. J. Sci. Food Agric. 50, 563–570.
Yol, E., Toker, R., Golukcu, M., Uzun, B., 2015. Oil content and fatty acid characteristics in mediterranean sesame core collection. Crop Sci. 55, 2177–2185. https://doi.org/10.2135/cropsci2014.11.0771.
Zhang, H., Wei, L., Miao, H., Zhang, T., Wang, C., 2012b. Development and validation of genetic SSR markers in sesame by RNA-seq. BMC Genomics 13, 316.
Zhang, Y.X., Zhang, X.R., Che, Z., Wang, L.H., Wei, W.L., Li, D.H., 2012a. Genetic diversity assessment of sesame core collection in China by phenotype and molecular markers and extraction of a mini-core collection. BMC Genet. 13, 102.