

Genetic Variability and Population Structure of Ethiopian Sesame (*Sesamum Indicum* L.) Germplasm Assessed Through Phenotypic Traits and Simple Sequence Repeats Markers

Desawi Hdru Teklu (✉ desawi.hdru@ata.gov.et)

African Centre for Crop Improvement, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa

Hussein Shimelis

African Centre for Crop Improvement, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa

Abush Tesfaye

International Institute of Tropical Agriculture, Oyo Road, P.M.B. 5320, Ibadan, Nigeria

Jacob Mashilo

African Centre for Crop Improvement, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa

Xiurong Zhang

Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops of the Ministry of Agriculture, Wuhan 430062, China

Yanxin Zhang

Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops of the Ministry of Agriculture, Wuhan 430062, China

Komivi Dossa

Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops of the Ministry of Agriculture, Wuhan 430062, China

Admire Shayanowako

African Centre for Crop Improvement, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa

Research Article

Keywords: Agronomic traits, Ethiopia, genetic diversity, microsatellites, population structure, principal component analysis, *Sesamum indicum*

Posted Date: February 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-226709/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Ethiopia is one of the centres of genetic diversity of sesame (*Sesamum indicum* L.). The sesame genetic resources present in the country should be explored for local, regional and international sesame improvement programs to design high performing and market preferred varieties. This study's objectives were to determine the extent of genetic variation among 100 diverse cultivated sesame germplasm collections of Ethiopia using phenotypic traits and simple sequence repeat (SSR) markers to select distinct and complementary specimens for breeding. One-hundred sesame entries were field evaluated at two locations in Ethiopia for agro-morphological traits and seed oil content using a 10 × 10 lattice design with two replications. Test specimens were profiled using 27 selected polymorphic SSR markers.

Results

The analysis of variance revealed significant ($P \leq 0.05$) entry by environment interaction for plant height, internode length, number of secondary branches, and seed yield. Genotypes such as Hirhir Kebabo Hairless-9, Setit-3, Orofalc ACC-2, Hirhir Humera Sel-6, ABX=2-01-2, and Setit-1 recorded higher grain yield of > 0.73 ton ha⁻¹ with excellent performance in yield component such as oil and seed yield per hectare. Seed yield had positive and significant ($p < 0.01$) associations with oil yield ($r = 0.99$) useful for simultaneous selection for yield improvement in sesame. The SSR markers revealed gene diversity and polymorphic information content of 0.30 and 0.25, respectively, showing that the tested sesame accessions were genetically diverse. Cluster analysis resolved the accessions into two groups, while population structure analysis revealed four major heterotic groups, this enabling selection and subsequent crosses to develop breeding populations for cultivar development.

Conclusions

Based on phenotypic and genomic divergence, the following complementary specimens were selected: Hirhir Humera Sel-6, Setit-3, Hirhir Kebabo Hairless Sel-4, Hirhir Nigara 1st Sel-1, Humera-1 and Hirhir Kebabo Early Sel-1 (from cluster II-a), Hirhir kebabo hairless-9, NN-0029(2), NN0068-2 and Bawnji Fiyel Kolet, (from cluster II-b). The selected genetic resources are recommended for use in sesame production and breeding programs in Ethiopia.

Background

Sesame (*Sesamum indicum* L., $2n=2x=26$) is a multi-purpose high-value oilseed crop. It is a global commodity serving the food, feed, and cosmetic industries. The seed oil content of sesame is about 60%, the highest when compared with other oilseed crops such as sunflower (~45%), rapeseed (~40%) and soybean (~20%) [1-4]. Sesame oil comprises about 85% unsaturated, and 15% saturated fatty acids. The fatty acid contains linoleic acid (~46%), oleic acid (~38%), palmitic acid (~12%), and stearic acid (~4%) [4-7]. Sesame oil is a rich source of protein (~24%), carbohydrate (~13.5%), vitamins (e.g. A and E), lignans (sesamin and sesamol), γ tocopherol, phytosterols (β -sitosterol and Campesterol), policosanols (Docosanol, Tetracosanol, Hexacosanol, and Octacosanol) and lipids [4, 7-9]. These attributes make sesame a 'superfood' comprising all the essential human nutrients in desirable proportions.

Sesame is the second most valuable export crop after coffee (*Coffea arabica* L.) and a major contributor to the gross domestic product in Ethiopia [10]. In the country, the area allocated for sesame production in 2018 was 294,819.49 ha, approximately 39.4% of the total estimated area allocated for oil crops production [11]. Compared with global sesame production, Ethiopia ranks 8th with an annual total production of 301,302 tons after Sudan (981,000 tons), Myanmar (768,858 tons), India (746,000 tons), Nigeria (572,761 tons), Tanzania (561,103 tons), China (433,386 tons) and China Mainland (431,500 tons) [12].

Ethiopia is the center of origin and diversity for the cultivated sesame and its allied species. The Ethiopian Biodiversity Institute (EBI) maintains one of the most extensive core collection of sesame genetic resources in Africa. About 5000 genetically diverse sesame germplasm resources are conserved by the EBI [13]. The germplasm pool can provide an array of unique economic traits, and gene combinations for global sesame improvement. However, the genetic resources kept at the EBI is yet to be explored for local, regional, and international sesame improvement programs to develop high performing and market preferred varieties. The mean sesame yield in Ethiopia is 0.68 ton ha⁻¹, which is relatively low compared with a mean yield of 1 ton ha⁻¹ in sub-Saharan Africa and 1.29 ton ha⁻¹ in Egypt [11, 12]. The low productivity in the country is attributable to a lack of improved and high varieties, and the use of traditional production technologies, among other constraints. Landrace varieties are the main sources of seed for cultivating sesame in Ethiopia. Landraces are inherently low yielders and prone to capsule shattering leading to reduced productivity and economic gains. However, landraces are highly valued for possessing intrinsic farmer-preferred attributes such as unique taste and aroma, adaptation to harsh and local growing conditions which are often low input farming systems, and marginal agricultural lands [8, 14].

The sesame genetic resources maintained at the EBI which comprises mostly of landrace accessions can be explored to search for new source of useful genetic variation for yield and yield- components, resistance to diseases and insect pests, tolerance to abiotic stresses, capsule shattering tolerance, and nutritional quality. This will identify desirable and complementary parents for sesame breeding and genetic analyses programs for gene discovery and traits introgression. This requires rigorous phenotyping and genotyping to establish the degree of genetic polymorphism in the germplasm pool and to delineate the heterotic groups.

Previous studies have reported considerable phenotypic variation for agronomic and quality traits in sesame genetic resources from Ethiopia [15-18]. However, these studies did not fully represent the landrace collections from various parts of Ethiopia. Hence there is a need for a comprehensive assessment of the genetic diversity present in the Ethiopian sesame using a relatively larger number of accessions representing the diverse germplasm resources and sampled from various regions through phenotypic traits and effective molecular markers.

Several molecular markers such as amplified fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), microsatellites or simple sequence repeat (SSR), and single nucleotide polymorphisms (SNPs) markers are widely used in genetic diversity analysis of various crop genetic resources. SSR or microsatellites have been commonly used in genetic variation studies on sesame [19-22]. The SSRs are preferred for their ability to detect higher degrees of polymorphism, higher reproducibility and abundant coverage of the genome [20, 21]. In addition, SSR markers can be used for loci with multiple co-dominant alleles [23]. [20, 21] assessed genetic diversity and population structure present in sesame genetic resources sampled from China and Korea using 44 and 23 SSRs, respectively. The authors reported the presence of two and three major heterotic groups among the Chinese and Korean collections. The level of genetic diversity varies among different germplasm populations and environmental conditions, suggesting that each set of populations must be assessed in a target production environment for selection and genetic grouping. Therefore, this study's objectives were to determine the extent of genetic variation among 100 diverse sesame germplasm collections of Ethiopia using phenotypic traits and simple sequence repeat markers to select and recommend distinct and complementary parents for direct production, breeding and conservation.

Materials And Methods

Plant materials

The study used a mini-core collection of 100 sesame entries originally collected from the Amhara, Tigray, Afar, Oromia, and Gambela regions in Ethiopia. The test specimens were obtained from the sesame and groundnut breeding program of Werer Agricultural Research Centre of the Ethiopian Institute of Agricultural Research (EIAR). The collection comprised of 95 accessions, one landrace (farmer variety), and four released varieties. The landrace variety "Hirhir" is widely cultivated by farmers in the study areas. The four released varieties (i.e., Setit-1, -2, -3, and Humera-1) were developed by the Humera Agricultural Research Center (HuARC) through mass selection amongst the local germplasm collections. The details of the germplasm collections used in the study are summarised in Supplementary Table 1.

Phenotyping

Description of the study environments

The study was conducted in the north western Ethiopia in two selected locations, namely, Humera (14°15' N, 36°37' E) and Kebabo (13°36' N, 36°41' E). Humera and Kebabo are agricultural research stations of the Humera Agricultural Research Centre of EIAR and the Tigray Agricultural Research Institute (TARI), respectively. The two sites represent the major sesame production environments in Ethiopia. Humera and Kebabo are situated at an altitude of 609 and 696 meters above sea level and receive a total rainfall of 576.4 and 888.4 mm, respectively. The mean minimum and maximum temperatures at Humera site range from 20.3 to 36.5 °C. Kebabo has mean minimum and maximum temperatures of 16.9 and 31.7 °C. The two sites have predominantly clay soil [34].

Field experiments and agronomic management

The experiment was conducted under field condition and laid out using a 10 × 10 simple lattice design, with two replications, at each site. Each entry was planted in four rows plots measuring four meters in length, with an inter-row and intra-row spacings of 0.4m and 0.1m, respectively. The trials were maintained following the standard agronomic practices of sesame production [34].

Phenotypic data collection

Data were collected based on a whole plot basis from the two central rows on plant or plot basis during plant growth and at harvest. Phenotypic data were collected from 10 randomly selected and tagged plants per entry at physiological maturity. Plant height (PH) was measured from the base to the tip of the plant, stem height from the base to the 1st branch (SHB) was measured from the base of the plant to 1st emerged primary branch using a ruler and expressed in cm. Internode length (INL) was measured between two consecutive nodes situated in the middle of the plant. The number of primary branches per plant (NPB) was counted from the main stem of the plant, while number of secondary branches per plant (NSB) was counted from the main branch of the plant. Distance from the base of the lowest branch to the 1st capsule (DFLBC) was measured as the distance between the lowest situated primary branch to the 1st emerged capsule on the main stem and expressed in cm.

The number of days to flowering (DF) was recorded by counting the number of days from planting to the date when 50% of the plants showed flowers, while days to maturing (DM) was recorded as the number of days from planting to the date when 75% of the plants showed physiological maturity. The number of capsules per plant (NCP) and number of seeds per capsule (NSP) were counted from a composite of three capsules per plant at harvest. Thousand seed weight (TSW) was measured from a random sample of 1000 seeds of each entry. Seed yield (SY) was measured in grams per plot and later converted into ton (t) per hectare (ha⁻¹). Oil content was determined using the Near-Infrared Spectroscopy (NIR) (FOSS, model DS2500, Denmark). Oil yield per hectare was calculated and expressed in tons per hectare as the product of seed yield and percent oil content.

Phenotypic data analysis

The phenotypic data were subjected to analysis of variance (ANOVA) using the lattice and general linear model (GLM) procedures of the SAS software version 9.4 [35]. A combined analysis of variance across the two locations was performed after Bartlett's homogeneity test of variance. The correlation among traits was performed using R software version 4.0 [36] to determine the magnitude of associations among the studied traits. Multivariate analysis using the principal components was performed using the SAS version 9.4 [35].

Genotyping

DNA extraction, primer selection, polymerase chain reaction, and electrophoresis

The above 100 sesame entries (Supplementary Table 1) were planted at the Oil Crops Research Institute (OCRI) - the Chinese Academy of Agricultural Sciences (OCRI-CAAS), China. Ten seeds per entry were sown in a plastic tray in a growth room. From each entry, three two-weeks old healthy plants were randomly selected, and fresh young leaves were collected and ground in liquid nitrogen for DNA extraction. The DNA was extracted following the Cetyl-tetramethyl ammonium bromide (CTAB) method. Approximately 200 mg of ground plant tissue combined with 500 μ L of CTAB buffer was incubated in a water bath at 65 °C, 4 times for 10 minutes, and subjected to centrifugation at 12,000 rpm for 10 min at 4 °C. The supernatant was then transferred into new 5 ml micro-tubes, and 400- μ l chloroform: iso-amyl alcohol (24:1) was added into the tubes and mixed gently. After a minute of centrifugation (centrifuged at 12,000 rpm for 10 min at 4 °C), the supernatant was transferred into new 5 ml micro-tubes, and 400- μ l isopropanol was added into the tubes, mixed gently and kept at -20 °C for 30 minutes and subjected to centrifugation at 12,000 rpm for 10 min at 4 °C. The precipitated DNA was washed by 75% ethanol three times. The resulting pellet was dried under vacuum and dissolved in 100 μ l DD H₂O. DNA concentrations were measured using the Quantus™ Fluorometer (Promega). Microsatellites from 13 linkage groups were designed and used for the following experiments. The 27 primers were selected because of their suitability in discriminating sesame specimens. The presently used primers were initially selected amongst 160 candidate primers based on their higher polymorphic information content and for providing clear and informative amplicon profiles in sesame genetic analysis [37].

The polymerase chain reaction (PCR) conditions were maintained as follows, each PCR reaction was carried out in a 20 μ L solution containing 25 ng of DNA, 4 μ mol of forward primers, 4 μ mol of reverse primers, 1 \times buffer, 0.25 mmol of dNTPs, and 0.80 U Taq polymerase. The temperature profile used for PCR amplification comprised of a denaturation step at 94 °C for 1 minute, followed by primer annealing temperature at 45.2 -53 °C for 1 minute, and elongation at 72 °C for 1 minute. After 34 cycles, the reaction was terminated with a 10 min final extension time at 72°C.

The PCR reaction conditions were the same for all the primers, except for the annealing temperatures. The PCR products were electrophoresed on 6% Acrylamide gel (= 200 ml of 5 \times T.B.E., 420 gm of Urea [H₂NCONH₂], 75 gm of Acrylamide, 3 gm of Bis-Acrylamide and 400 ml of distilled water) at a voltage value of 2000, current 300 A, power 80 W for 1:30 hours. After silver staining, the bands on the gels were recorded, and a total of 27 markers (Supplementary Table 2) with high polymorphism were used for capillary electrophoresis. 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.

Genotypic data analysis

The fragment sizes were determined using the ABI 3730 automatic sequencer. Data were analysed using the software GeneMarker V 2.2.0 to determine peak detection threshold levels that ranged from the minimum intensity of 500 and max intensity of 30,000. The 27 primers were used to detect the band sizes based on the peak detection thresholds, which were then scored using 1 to denote presence and 0 for absence. Genetic parameters, such as major allele frequency (M.A.F.), observed heterozygosity (Ho), expected heterozygosity (He), and the polymorphic information content (PIC) were calculated using Power Marker v3.2. Cluster analysis was carried out using a neighbour-joining (NJ) algorithm using the unweighted pair group method (UWPGM) in R software version 4.0 [36].

The population structure of the 100 sesame accessions was investigated using the Bayesian clustering method in STRUCTURE version 2.3.4 [38]. The length of the burn-in period and Markov Chain Monte Carlo (MCMC) were set at 10,000 iterations [39]. To obtain an accurate estimation of the number of populations, ten runs were performed for each K-value (assumed number of subpopulations), ranging from 1 to 10. Further, Delta K values were calculated, and the appropriate K value was estimated by implementing the [39] method using CLUMPK.

Results

Genetic variation and mean performance of sesame accessions

The combined ANOVA revealed significant ($P \leq 0.05$) entry \times environment interaction for plant height, internode length, number of primary branches, number of secondary branches, distance from base of lowest branch to 1st capsule and seed yield per hectare (Table 1). Entries showed significant ($p \leq 0.05$) differences for days-to-50% flowering, days-to-75% maturity, plant height, internode length, number of secondary branches, number of seeds per capsule, distance from base of lowest branch to 1st capsule and seed yield per hectare.

Based on seed yield response, the top 10 best performing and the five bottom performing accessions are summarized in Table 2. The mean seed yield across locations was 0.48 ton ha⁻¹ and mean thousand-seed weight was 2.9 g. The highest seed yield was recorded in entries such as: Hirhir Kebabo Hairless-9 (1.01 ton ha⁻¹), Setit-3 (0.84 ton ha⁻¹), Orofalc ACC-2 (0.80 ton ha⁻¹), Hirhir Humera Sel-6 (0.78 ton ha⁻¹), ABX=2-01-2 (0.74 ton ha⁻¹) and Setit-1 (0.73 ton ha⁻¹). These specimens expressed high oil yields of 0.40, 0.40, 0.40, 0.39, 0.36 and 0.39 ton ha⁻¹, respectively. The accessions Bawnji Fiyel Kolet, NN0056, Hirhir Humera Sel-8, NN-0068-1, and ACC-NS-010 had the highest oil content of 55.6, 55.2, 54.7, 54.6, and 54.10%, respectively. The five bottom performing accessions in terms of seed yield were NN-0183-3 (0.17 ton ha⁻¹), NN-0020 (0.24 ton ha⁻¹), NN-0108-2 (0.26 ton ha⁻¹), NN00136-1 (0.26 ton ha⁻¹) and NN-0143 (0.28 ton ha⁻¹) with low oil yield of 0.07, 0.12, 0.04, 0.13 and 0.15 ton ha⁻¹, in that order. These accessions yielded below average seed and oil yields.

Correlations of yield and yield components

Phenotypic correlation coefficients for the studied traits are presented in Table 3. Seed yield was significantly and positively correlated with oil yield ($r = 0.99$; $p < 0.01$). Significant and positive correlations were also observed between seed yield and internode length ($r = 0.35$; $p < 0.01$), number of secondary branches ($r = 0.21$; $p < 0.01$), number of capsules per plant ($r = 0.18$; $p < 0.01$), number of seeds per capsule ($r = 0.17$; $p < 0.01$), stem height from base to 1st branch ($r = 0.16$; $p < 0.01$) and thousand-seed weight ($r = 0.23$; $p < 0.01$).

Principal component analysis

Principal component analysis (PCA) was determined to show the contribution of each trait to the overall observed variation. The analysis revealed six principal components (PCs) with Eigenvalues > 1.0. The six PCs cumulatively accounted for 66.20 % of the total variation among the entries (Table 4). Principal component one (PC1) explained 17.26% to the total variation and was positively correlated with NSPC, SYH and OYH, and negatively correlated with DF and DM. Principal component two (PC2) accounted for 14.59% of the total variation, and positively correlated with some of the yield component traits, such as DF, DM and DFLBC. Principal component three (PC3) accounted for 10.46% of the total variation and positively correlated with NSB and OC. PC4, PC5 and PC6 accounted for 9.11, 8.10 and 6.67% of the total variation, respectively. PC4 was negatively correlated with PH and OC, but positively correlated with NPB and NCPP, whereas PC5 was positively associated with NSB and SHB, and PC6 was negatively correlated with INL and NSB, but positively correlated with PH and OYH.

Genetic polymorphism of the SSR markers

The summary statistics describing the SSR markers are presented in Table 5. The major alleles frequency per locus ranged from 0.52 to 0.96, with a mean of 0.78 alleles per locus. The observed heterozygosity varied from 0.08 to 0.96, with a mean of 0.43. The unbiased expected heterozygosity (gene diversity) of the markers ranged from 0.08 to 0.5, with a mean of 0.30. The PIC values ranged from 0.07 (for markers ID0041, ID0175, and ZMM2818) to 0.37 (ZMM3261 and ZMM1189) with a grand mean value of 0.25.

Population structure analysis

The structure analysis revealed four populations amongst the 100 sesame entries (Fig. 1., Table 6). A total of 63 accessions were allocated to the four populations, whereas 37 accessions were admixtures with no specific membership (Table 6). Population I consisted of 24 accessions collected from the following regions: Amhara (17 collections), Tigray (3), Afar (3), and Oromia (1). Population II had 13 accessions originally collected from the Amhara region (8), Afar (4) and Tigray (1). Population III comprised nine accessions, sourced from the Amhara (5 accessions) and Tigray (4) regions. Population IV consisted of 17 accessions sourced from Tigray (10 accessions), Amhara (6), and Afar (1) regions.

Entries allocated in population I had good branching ability (e.g. NN-0052, NN-0029-1, and GXT=85(28-2)), higher number of seeds per capsule (e.g. Gojam Azene (Aleka), and ABXT-85-Sel-2-1), with a relatively higher thousand-seed weight varying from 3.2 to 3.3 gram (e.g. NN0016-1, Hirhir Nigara 1st Sel-2 and Hirhir Kebabo Hairless Sel-6). Intermediate seed yield varying from 0.55 to 0.65 ton ha⁻¹ were recorded for ACC-203-020, Hirhir Kebabo Hairless Sel-6 and NN-0026, while a relatively better oil yield of 0.32 ton ha⁻¹ (NN-0026), and higher seed oil content of 51.0 to 51.2% were achieved in NN0064-1 and NN0071.

Accessions in population I such as GXT=85(28-2), NN0064-1, and NN00713 had relatively higher oil content of 51.0, 51.1 and 51.2%, respectively. Population II accessions were early maturing with tall plants and possessed relatively high number of seeds per capsule, seed and oil yields and oil content. In this population some accessions such as NN0025 and ABX=2-01-2 had the highest number of seeds per capsule of 63, and 65, in that order. Cluster II entries such as Orofal ACC-2, and ABX=2-01-2 had relatively the highest seed yield of 0.80, and 0.74 ton ha⁻¹, respectively. In the same population, the accession Orofal ACC-2 had relatively higher oil yield of 0.40 ton ha⁻¹. In addition, population II comprised of accessions such as NN0056, Hirhir Baeker-Sel-3, and Orofal ACC-2 that expressed relatively higher oil content of 55.2, 53.4, and 52.5%, respectively. Specimens allocated in population III were early maturing with taller plants, and possessed higher thousand-seed weight of 3.5 gram (e.g. Hirhir Filwha Large Seeded), highest seed yield of 1.01 ha⁻¹ (e.g. Hirhir Kebabo Hairless-9), higher oil yield of 0.40 ton ha⁻¹ (e.g. Hirhir Kebabo Hairless-9) and higher seed oil content of 55.6% (e.g. Hirhir Kebabo Hairless-9).

Specimens allocated in population IV were early maturing with taller plants, and possessed relatively high number of seeds per capsule, better thousand-seed weight, with higher seed and oil yields and oil content. Entries Setit-3 and NN-0020 allocated in population IV had relatively higher number of seeds per capsule of 63 and 62, respectively. Accessions Hirhir Kebabo Hairless-Sel-7, Hirhir Kebabo Hairless Sel-4, and Setit-1 allocated in population IV had higher thousand-seed weight of 3.4, 3.3, and 3.3 gram, in that order. The majority of the test entries in this population recorded seed yield varying from 0.55 to 0.84 ton ha⁻¹. Specimens Setit-3, Hirhir Humera Sel-6, and Setit-1 had relatively higher seed yields of 0.84, 0.78, and 0.73 ha⁻¹, respectively. In population IV, specimens such as Setit-3, and Hirhir Kebabo Hairless-Sel-7 expressed relatively higher oil yields of 0.40, and 0.37 ton ha⁻¹, in that order. Furthermore, accessions Hirhir Kebabo Hairless Sel-4, Hirhir Humera Sel-6, Setit-1, Hirhir Kebabo Hairless Sel-4, Morgo-Sel-P=13, Gojam Azene (Yohans Sel-1), and Hirhir Kebabo Hairless-Sel-7 which were grouped in population IV had relatively the highest oil content of 54.7, 53.9, 53.8, 53.5, 53.2, 53.1 and 53.0%, respectively. To develop new breeding populations possessing desirable economic traits new crosses could be developed between selected parents. Hence accessions Orofal ACC-2 (from population II), Hirhir Filwha Large Seeded (population III), and Setit-3, Hirhir Humera Sel-6 (population IV) are ideal candidates with complementary traits.

Cluster analysis of 100 sesame accessions

The cluster analysis involving 100 sesame specimens resolved two clusters, and each cluster was further partitioned into two sub-clusters (Fig. 2.). Cluster I consisted of 49 accessions and one improved variety sourced from the following regions: Amhara (37 accessions), Tigray (5 accessions and one improved variety), Afar (6 accessions) and Oromia (1 accession). Cluster II contained 50 diverse specimens of which 28 accessions were from Amhara, while 13 accessions, one landrace and 3 improved varieties from Tigray, 2 accessions (from Afar), 2 accessions (Oromia), and 1 accession (Gambela).

Accessions allocated in Cluster I-b had good branching ability (e.g. NN-0052, NN-0029-1, Teiahir Sanja Sel-6 and GXT=85(28-2)). These specimens had relatively better seed yield varying from 0.55 to 0.74 ton ha⁻¹ (e.g. ABX=2-01-2, Hirhir Baeker-Sel-3, NN-0029-1 and Hirhir Kebabo Hairless Sel-6). Cluster I accessions had relatively higher oil yields of 0.32 to 0.37 ton ha⁻¹ (Shwarobit (83), ABX=2-01-2, Hirhir Baeker-Sel-3 (selected from sub cluster I-b), NN-0044-2, and NN-0026 (from sub cluster I-a). Higher thousand-seed weight varying from 3.2 to 3.3 g were recorded in Hirhir Kebabo Hairless Sel-6, NN0016-1, and NN0032 (from sub Cluster I-b), Hirhir Nigara 1st Sel-2 and NN-0143) (from sub Cluster I-a). Cluster I entries were distinguished with higher number of seeds per

capsule of 61 to 67 and with relatively higher seed oil contents of 53.1 to 54.6% (e.g. Gonjam Azene (Aleka), ABX=2-01-2, NN0025, Shwarobit (83) (sub Cluster I-b) and ABXT-85-Sel-2-1) (sub Cluster I-a).

Accessions grouped in sub Cluster I-b such as NN-0068-1, NN0015, Tejareb Girar, and Hirhir Baeker-Sel-3 had the highest oil content of 54.6, 54.0, 53.8 and 53.4%, respectively. Cluster II accessions (except accession ACC-202-358) were early maturing with tall plants, and possessed higher high number of seeds per capsule, thousand-seed weight, seed and oil yields and oil content. In Cluster II-b some specimens such as Hirhir and NN-0029(2) had higher number of seeds per capsule varying from 65 to 72. Accessions Bawnji Gobate, Hirhir Filwha Large Seeded, NN-0029(2) (selected from sub Cluster II-b) Hirhir Kebabo Hairless-Sel-7, and Setit-1 (from sub Cluster II-a) had higher thousand-seed weight of 3.5, 3.5, 3.4, 3.4, and 3.3 gram, in that order. Cluster II-a entries such as Setit-3, Orofalc ACC-2, Hirhir Humera Sel-6, and ACC-NS-007(2), [selected from sub cluster II-a], and Hirhir Kebabo Hairless-9, and ACC 205-180 (from sub Cluster II-b) had the highest seed yields of 1.01, 0.84, 0.80, 0.78, 0.77, and 0.72 ton ha⁻¹, respectively. In the same cluster, the specimens Setit-3, Orofalc ACC-2, Hirhir Humera Sel-6, and ACC-NS-007(2) (sub cluster II-a) and Hirhir Kebabo Hairless-9 (sub cluster II-b) expressed relatively higher oil yields of 0.40, 0.40, 0.40, 0.39 and 0.39 ton ha⁻¹, in that order. In addition, Cluster II comprised of accessions such as Bawnji Fiyel Kolet and ACC 205-180 (sub cluster II-b), and NN0056, Hirhir Humera Sel-8, Hirhir Kebabo Early Sel-1, and Hirhir Kebabo Hairless Sel-4 (sub cluster II-a), with the highest oil contents of 55.6, 54.1, 55.2, 54.7, 53.7, and 53.5%, respectively. Hence specimens Setit-3, Orofalc ACC-2, Hirhir Humera Sel-6, and ACC-NS-007(2) (selected from sub cluster II-a), and Hirhir Kebabo Hairless-9, and ACC 205-180 (from sub cluster II-b) were ideal candidates with complementary traits for sesame breeding.

Discussion

Genotypic variation and mean performance for seed and oil yields, and yield-component traits

Assessment of genetic diversity among crop genetic resources is essential to identify candidate accessions possessing desirable traits including yield and quality attributes. The current study evaluated the genetic variation present among 100 accessions of sesame through rigorous field phenotyping and polymorphic SSR markers as a preliminary step to select genetically complementary parental accessions for breeding.

The test specimens showed significant ($P \leq 0.05$) variation for seed yield and yield components (Table 1). This suggests that the germplasm pool contains vital phenotypic traits for sesame improvement through hybridization and selections. The test specimens were sourced from five historically sesame-growing regions in Ethiopia. Given the long agricultural history and sesame production of the collection areas, it is expected that the test specimens have adapted and evolved under local conditions through natural selection. This caused genetic differentiation of the studied sesame accessions for seed and oil yields, and important yield-contributing agronomic traits. For example, the present study identified and selected sesame specimens such as Hirhir Kebabo Hairless-9 and Setit-3 with high seed yields of ≈ 0.8 tons ha⁻¹ and higher oil yields of 0.40 ton ha⁻¹. The selected specimens which are locally referred to as Humera types are known for their unique quality associated with product aroma and taste [24]. The selected specimens expressed higher seed yield which is above the mean yield of 0.68 tons ha⁻¹ currently recorded in Ethiopia using traditional varieties.

Traits associations

Sesame seed and oil yields are low in Ethiopia due to a lack of high yielding varieties. These results in low financial returns for producers and processors across the sesame value chains. To improve selection response and genetic gains for economic traits, selection of highly heritable yield-contributing traits associated with seed and oil yields may be targeted in sesame improvement programmes. The strong and positive correlation between seed and oil yield among the studied sesame specimens implied both traits could be improved simultaneously in the present population. Poor correlations observed between seed yield with yield-related traits including internode length, number of secondary branches, number of capsules per plant, stem height from base to 1st branch and thousand seed weight would provide low selection response for seed yield. Similarly, oil yield exhibited low correlations with internode length, number of secondary branches per plant, and thousand-seed weight implying reduced selection response for seed yield via these traits. Oil content showed poor associations with agro-morphological traits hindering direct selection. Despite the low and poor associations between seed and oil yields, and oil content with yield-related agronomic traits the present study revealed wide phenotypic variation among the studied sesame populations for several phenotypic traits. These are valuable traits for future sesame phenotypic analysis, selection and improvement in Ethiopia. Also, the assessed germplasm was diverse for seed and oil yields, and oil content. This aided identification and selection of sesame specimens such as Hirhir Kebabo Hairless-9, Setit-3, Orofalc, Hirhir Humera Sel-6, Setit-1 with high seed and oil yields as useful germplasm to design and develop improved cultivars. Also, sesame specimens with relatively higher oil content including Hirhir Humera Sel-6, Setit-1, ACC 205-180 and Orofalc ACC-2 are suitable candidates to develop new breeding populations possessing higher oil yield and content.

The traits identified through the PCA accounted for much of the variation among the test accessions necessitating their value in future sesame selection programs. Nevertheless, 33.80% of the total variation was not explained by the PCA probably due to the limited number of test locations used in the study. Hence, there is need to assess the test accessions across multiple test environments and using effective molecular markers to complement the phenotypic data.

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

SSR markers are amongst the useful genomic resources to complement phenotypic data for effective selection. The present study recorded a mean major alleles frequency per locus of 0.78 among the sesame population (Table 5), which was much higher than values of 0.41 and 0.17 reported by [21, 25] using 23 and 21 SSR among 129 Korean and 25 Ghanaian sesame specimens, respectively. Variation in alleles frequency is attributable to genotypic differences and number of SSR markers used in genetic analysis [26-28]. The mean observed heterozygosity of 0.43 reported on the present study is lower than the value of 0.56 reported by [25] when assessing 25 sesame specimens using 21 SSR markers. The observed heterozygosity in this study was higher than values of 0.23, 0.01, and 0.12 reported by [19, 21, 22] when assessing 50, 129 and 36 sesame specimens using 10, 23 and 10 SSR markers, respectively. The mean expected

heterozygosity ($H_e = 0.30$) recorded in the present study (Table 5) was lower than values of 0.72 and 0.34 reported by [21, 22] when evaluating 129 and 36 sesame accessions using 23 and 10 SSR markers, respectively. The higher heterozygosity recorded in the present study suggested that Ethiopian sesame population have high level of genetic variation for selection.

The genetic variability was confirmed by population structure analysis which revealed four distinctive populations comprising of specimens collected from different regions of Ethiopia. Most released entries (Humera-1, Setit-1 and Setit-3) were grouped in sub-population 4. [20, 21] reported two and three populations among 94 and 129 sesame accessions sampled from China and Korea collections and when assessed with 44 and 23 SSR markers. The higher gene fixation index of 0.39 for population I which comprised of accessions collected from Amhara, Tigray, Afar, and Oromia regions suggested higher genetic differentiation attributable to high gene flow among these regions. Conversely, the low gene fixation index observed for population III which comprised of accessions sourced from the Amhara and Tigray regions indicated low differentiation among the groups probably due to gene flow through germplasm exchange between sources of collections. The exchange of planting material regardless of geographical distances might be attributed to a low degree of differentiation in sesame populations observed in the current study.

Cluster analysis identified two major clusters and four sub-clusters revealing the existence of genetic variation among the assessed sesame entries (Fig. 2.). [21] grouped 129 sesame specimens into two clusters using 23 SSR markers. In the present study, the clustering patterns of the specimens did not correspond to the predefined population structure based on the regions of collection. This may be because specimens gathered from similar regions belong to the same gene pool or may have similar ancestral relationships [29]. Conversely, [30] reported that genetic dissimilarity among test specimens could arise due to the diverse ancestral origin, high gene flow caused by cross-pollination and possible gene or chromosomal mutation. In this study, some sesame specimens collected from different regions were grouped in the same cluster such as Hirhir Kebabo Hairless Sel-6 (Tigray) and Gojam Azene (Yohans Sel-1) (Amhara), and ACC-NS-007(2) (Oromia) and GA-002(3) (Gambela) which were found in in cluster I and II. In agreement with the current study, [31] reported that geographical separation did not affect genetic distance among 24 sesame specimens. [32] reported that geographical separation does not affect the genetic differentiation of germplasm. Therefore, a key indicator of genetic diversity is not necessarily the geographical origin of germplasm collections. The exchange of genetic materials among farmers and traders in the regions contributes to high gene flow and a lack of genetic differentiation. [33], suggested that farmers' selections and management practices affect the patterns of genetic diversity.

To develop new breeding populations possessing desirable agronomic traits, especially high seed and oil yields, crosses could be made between distantly related and complementary specimens selected from different clusters. For instance, for improved seed and oil yields the following entries were selected such as Setit-3, Orofalc ACC-2, Hirhir Humera Sel-6, ACC-NS-007(2), Hirhir Kebabo Hairless-9, and ACC 205-180. These specimens are localised in sub cluster II-a and sub cluster II-b. The two clusters contained candidates with excellent seed and oil yields.

Conclusion

The current study determined the extent of genetic variation among 100 diverse sesame germplasm collections of Ethiopia using phenotypic traits and simple sequence repeat (SSR) markers to select distinct and complementary parents for breeding. The test specimens exhibited significant phenotypic variation for key agronomic traits, oil content, and oil yield, which were underpinned by their genetic diversity. The sesame specimens were differentiated into four major populations based on the model-based population structure analysis. The moderate heterozygosity and fixation index among the accessions suggests that the accessions have distinct heterotic groups desirable for breeding. Specimens Hirhir Kebabo Hairless-9 and Setit-3 had high mean performance for seed and oil yield and were clustered in sub clusters II-b and II-a, respectively, which provides an opportunity for selection as divergent parental lines for sesame breeding. Based on wide genetic divergence, the following specimens were selected for use in future sesame breeding programs: Hirhir Humera Sel-6, Setit-3, Hirhir Kebabo Hairless Sel-4, Hirhir Nigara 1st Sel-1, Humera-1 and Hirhir Kebabo Early Sel-1 (selected from subgroup II-a), Hirhir kebabo hairless-9, NN-0029(2), NN0068-2 and Bawnji Fiyel Kolet, (from subgroup II-b). Progeny development and field evaluation through combining ability analysis are recommended among the selected parents to establish the heterotic groups and to establish sesame pre-breeding.

Declarations

Ethics approval and consent to participate

The field studies were conducted using sesame genetic resources complying with the guidelines of the Ethiopian Institute of Agricultural Research Institute (EIAR).

The sesame genetic resources were kindly supplied by the Ethiopian Biodiversity Institute (EBI) for exclusive use for this research.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was financially supported by the Ethiopian Agricultural Transformation Agency (ATA), 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) and the Ethiopian Institute of Agricultural Research (EIAR), Humera Agricultural Research Centre (HuARC).

Authors' contributions

DHT: The main author for the current study, carried out the phenotyping and molecular studies, performed both data analysis, Conceptualized, Data curation, Writing-original draft. HS: Lead supervision, Conceptualization, designed the study, Writing-review & editing the manuscript. AT: In-country co-supervision, review & editing the manuscript. JM: Writing-review & editing the manuscript. AS: Molecular data analysis, interpretation of results, revised the initial manuscript. XZ: sourced fund and carried out the molecular marker analysis. YZ: carried out the molecular marker analysis KD: carried out the molecular marker analysis. The author (s) read and approved the final manuscript.

Corresponding author

Correspondence to Desawi Hdru Teklu

Email; desawi.hdru@ata.gov.et

Acknowledgments

The Ethiopian Agricultural Transformation Agency (ATA), the Ethiopian Institute of Agricultural Research (EIAR), staffs of the Humera Agricultural Research Centre (HuARC) and Melka Werer Agricultural Research Centre (WARC) are greatly appreciated for germplasm supply and the overall research support.

Authors' information

¹African Centre for Crop Improvement, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa

Desawi Hdru Teklu, Email; desawi.Hdru@ata.gov.et

Professor Hussein Shimelis, Email;- Shimelish@ukzn.ac.za

Dr. Jacob Mashilo, Email;- jacobmashilo.jm@gmail.com

Dr. Admire Shayanowako, Email;- shayanowako@gmail.com

²Ethiopian Agricultural Transformation Agency, P.O. Box 708, Addis Ababa, Ethiopia

Desawi Hdru Teklu, Email;- desawi.Hdru@ata.gov.et

³International Institute of Tropical Agriculture, Oyo Road, P.M.B. 5320, Ibadan, Nigeria.

Dr. Abush Tesfaye, Email;- At.Abebe@cgjar.org

⁴Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops of the Ministry of Agriculture, Wuhan 430062, China

Professor Xiurong Zhang, Email;- zhangxr@oilcrops.cn

Dr. Yanxin Zhang, Email;- zhangyanxin@caas.cn

Dr. Komivi Dossa, Email;- komiri.dossa@ucad.edu.sn

⁵Laboratory of Genetics, Horticulture and Seed Sciences, Faculty of Agronomic Sciences, University of Abomey-Calavi, 01 BP 526 Cotonou, Republic of Benin

Dr. Komivi Dossa, Email;- komiri.dossa@ucad.edu.sn

References

1. De la Vega AJ, Hall AJ. Effect of planting date, genotype, and their interactions on sunflower yield: II. Components of oil yield. *Crop Sci.* 2002;42:1202-1210.
2. Zheljzkov VD, Vick BA, Ebelhar MW, Buehring N, Baldwin BS, Astatkie T, Mille JF. Yield, oil content, and composition of sunflower grown at multiple locations in Mississippi. *Agron J.* 2008;100(3): 635-639. DOI:10.2134/agronj2007.0253.
3. Wei W, Zhang Y, Lv H, Li D, Wang L, Zhang X. Association analysis for quality traits in a diverse panel of Chinese sesame (*Sesamum indicum*) germplasm. *J. Integr. Plant Biol.* 2013;55(8):745-58. DOI: 10.1111/jipb.12049.

4. Dossa K, Wei X, Niang M, Liu P, Zhang Y, Wang L, Liao B, Cissé N, Zhang X, Diouf D. Near-infrared reflectance spectroscopy reveals wide variation in major components of sesame seeds from Africa and Asia. *Crop J.* 2018;6(2):202-206.
5. Uzun B, Ülger S, Çağırğan, Mİ. Comparison of determinate and indeterminate types of sesame for oil content and fatty acid composition. *Turk J Agric For.* 2002;26:269-274.
6. Uzun, B, Arslan Ç, Furat, Ş. Variation in fatty acid compositions, oil content and oil yield in a germplasm collection of sesame (*Sesamum indicum*). *J Am Oil Chem Soc.* 2008; 85(12): 1135-1142.
7. Anastasi U, Sortino O, Tuttobene R. Agronomic performance and grain quality of sesame (*Sesamum indicum*) landraces and improved varieties grown in a Mediterranean environment. *Genet Resour Crop Evol.* 2017;64:127–137.
8. Were BA., Onkware AO, Gudu S, Welander M, Carlsson AS. Seed oil content and fatty acid composition in East African sesame (*Sesamum indicum*) accessions evaluated over 3 years. *Field Crops Res.*2006; 97(2):254-260.
9. Gharby S, Harhar H, Bouzoubaa Z, Asdadi A, El Yadini A, Charrouf Z. Chemical characterization and oxidative stability of seeds and oil of sesame grown in Morocco. *J. Saudi Soc. Agric. Sci.* 2017;16(2):105-111.
10. Gebremedhn MB, Tessema W, Gebre GG, Mawcha KT, Assefa MK. Value chain analysis of sesame (*Sesamum indicum*) in Humera district, Tigray, Ethiopia. *Cogent Food Agric.* 2019;5(1):1705741.
11. Central Statistic Authority (CSA). Ethiopian agricultural sample enumeration: Report on the primary results of area, production and yield of temporary crops of private peasant holdings in Meher Season, Addis Ababa, Ethiopia. 2018.
12. FAOSTAT online statistical service. Food and Agriculture Organization of the United Nations, Rome, Italy. 2018;<http://www.fao.org/faostat/en/#data/QC>. Accessed 14 April 2020.
13. Woldesenbet DT, Tesfaye K, Bekele E. Genetic diversity of sesame germplasm collection (*Sesamum indicum*): implication for conservation, improvement and use. *Int. J. Biotechnol. Mol. Biol. Res.* 2015;6(2):7-18. doi: 10.5897/IJBMBR2014.021.
14. Nyongesa BO., Were BA., Gudu S., Dangasuk OG., Onkware AO. Genetic diversity in cultivated sesame (*Sesamum indicum*) and related wild species in East Africa. *J. Crop Sci. Biotech.* 2013.16(1):9-15.
15. Gidey T, Kebede SA, Gashawbeza GT. Extent and pattern of the genetic diversity for morpho-agronomic traits in Ethiopian sesame landraces (*Sesamum indicum*). *Asian J. Agric.* 2012;6:118-128.
16. Teklu DH, Kebede SAI, Gebremichael DE. Assessment of genetic variability, genetic advance, correlation, and path analysis for morphological traits in sesame genotypes. *Asian J. Agric. Res.* 2014;7:118-128, DOI:10.3923/ajar.
17. Hika G, Geleta N, Jaleta Z. Correlation and divergence analysis for phenotypic traits in sesame (*Sesamum indicum*) Genotypes. *Sci. Technol. Arts Res. J.* 2014;3(4):01-09, DOI: <http://dx.doi.org/10.4314/star.v3i4.1>.
18. Hika G, Geleta N, Jaleta Z. Genetic variability, heritability and genetic advance for the phenotypic sesame (*Sesamum indicum*) Populations from Ethiopia. *Sci. Technol. Arts Res. J.* 2015;4(1):20 26, DOI: <http://dx.doi.org/110.4314/star.v4i1.3>.
19. Gebremichael DE Parzies HK. Genetic variability among landraces of sesame in Ethiopia. *Afr. Crop Sci. J.* 2011;19 (1): 1-13.
20. Wei W, Zhang Y, Wang L, Li D, Gao Y Zhang X. Genetic diversity, population structure, and association mapping of 10 agronomic traits in sesame. *Crop Sci.* 2016;56:331-343. doi: 10.2135/cropsci2015.03.0153.
21. Asekova S, Kulkarni KP, Oh KW, Lee MH, Oh E, Kim JI, Yeo U, Pae US, Ha TJ, Kim SU. Analysis of molecular variance and population structure of sesame (*Sesamum indicum*) genotypes using SSR markers. *Plant Breed. Biotech.* 2018;6(4):321-336. doi.org/10.9787/PBB.2018.6.4.321.
22. Araújo E.D.S., Arriel NHC, Santos R.C.D., Lima L.M.D. Assessment of genetic variability in sesame accessions using SSR markers and morpho agronomic traits. *Australian Journal of Crop Science. AJCS.* 2019;13(01):45-54. doi: 10.21475/ajcs.19.13.01. p1157.
23. Gupta PK, Varshney, R. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica.* 2000;113:163–185.
24. Wijnands, J.H.M., J. Biersteker and E.N. Van Loo. Oilseeds business opportunities in Ethiopia. Public private partnership in oil seed. 2009.
25. Adu-Gyamfi R, Prempeh R, Zakari I. Diversity assessment of some sesame (*Sesamum indicum*) genotypes cultivated in northern Ghana using morphological and SSR markers. *Adv Agric.* 2019; Article ID 6067891: 10. <https://doi.org/10.1155/2019/6067891>.
26. He Q, Li XW, Liang GL, Ji K, Guo QG. Genetic diversity and identity of Chinese loquat cultivars/accessions (*Eriobotrya japonica*) using apple SSR markers. *Pla Mol Biol Repo.* 2011;29: 197-208.
27. Baraket G, Chatti K, Saddoud O, Abdelkarim AB, Mars M. Comparative assessment of SSR and AFLP markers for evaluation of genetic diversity and conservation of Fig (*Ficus carica*) genetic resources in Tunisia. *Pla Mol Biol Repo.* 2011;29:171-184.

28. Jifar H, Tesfaye K, Dagne K, Assefa K, Tadele Z. Genetic diversity and population structure of tef [*Eragrostis tef* (Zucc.) Trotter] as Revealed by SSR Markers. *Adv Crop Sci Tech.* 2020;8: 438.
29. Mulualem T, Mekbib F, Shimelis H, Gebre E, Amelework B. Genetic diversity of yam (*Dioscorea spp.*) landrace collections from Ethiopia using simple sequence repeat markers. *Aust J Crop Sci.* 2018;12(8):1223-1230.
30. William TS, Hussein S, Mark L, Isack M, & Admire ITS. Assessment of the genetic diversity and population structure of rice genotypes using SSR markers. *ACTA AGR SCAND B-S P.* 2019; <https://doi.org/10.1080/09064710.2019.1670859>.
31. Zhang H, Wei L, Miao H, Zhang T, Wang C. Development and validation of genetic SSR markers in sesame by RNA-seq. *BioMed Central.* (2012;13:316–317).
32. Ganesamurthy K, Punitha D, Elangovan M. Genetic diversity among the land races of sorghum collected in Tamil Nadu. *Electron J Plant Breeding.* 2010;1:1375–1379.
33. Barnaud A, Trigueros G, McKey D, Joly HI. High outcrossing rates in fields with mixed sorghum landraces: how are landraces maintained? *Heredity.* 2008;101:445-452.
34. Annual Research Report for the Period 2009/10. Humera Agricultural Research Centre, Tigray, Ethiopia. 2010.
35. SAS Institute. Statistical Analysis Software. Version 9.4, SAS Institute Inc., Cary, NC., USA. 2018.
36. R Core Team R.A language and environment for statistical computing. R Foundation for Computing, Vienna, Austria. 2020.
37. Wei X, Wang L, Zhang Y, Qi X, Wang X, Ding X, Zhang J, and Zhang X. Development of Simple Sequence Repeat (SSR) Markers of Sesame (*Sesamum indicum*) from a Genome Survey. *Molecules.* 2014;19(4):5150-5162; <https://doi.org/10.3390/molecules19045150>.
38. Pritchard J, Stephens M, Rosenberg N, Donnelly P. Association mapping in structured populations. *Am J Hum Genet.* 2000;67:170–180.
39. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 2005;14:2611–2620.

Tables

Table 1. Analysis of variance showing mean square values and level of significance for the studied agro-morphological characters, and oil yield of 100 sesame collections evaluated in two locations in Ethiopia.

Traits												
Source of variation	DF	DM	PH	INL	NPB	NSB	NCPP	NSPC	SHB	DFLBC	TSW	
d.f.												
Rep (Loc)	2	171.27	205.65	714.76	0.00	121.72	28.62	14855.85	522.58	1730.39	374.90	17.61
Block (Loc*Rep)	36	5.05	16.04	240.28	1.16	1.15	0.14	276.21	249.16	70.61	111.98	6.36
Entry (Gen)	99	6.93*	16.46*	360.33*	2.87**	0.61 ns	1.24**	145.00 ns	189.08**	70.52 ns	122.26**	5.43 ns
Location (Env)	1	201.64**	60.06*	423.33ns	852.35**	39.06**	35.40**	116.64ns	2787.84**	11306.07**	36898.56**	1.32ns
Gen x Env	99	5.05 ns	14.03 ns	336.30*	3.43**	0.67*	0.68*	179.46ns	136.61 ns	48.88 ns	107.58**	5.49 ns
Error	162	4.69	11.08	243.30	1.54	0.47	0.40	143.65	112.01	52.94	65.88	5.37

Note: *Gen × Env*= entry by environment interaction, * and ** denote significance difference at the 5% and 1% levels of probability, respectively; *ns* non-significant, *DF* degrees of freedom, *DF* days-to- 50% flowering, *DM* days-to-75% maturity, *PH* plant height (cm), *INL* inter node length (cm), *NPB* number of primary branches per plant, *NSB* number of secondary branches per plant, *NCPP* number of capsules per plant, *NSPC* number of seeds per capsule, *SHB* stem height to 1st branch (cm), *DFLBC* distance from lowest branch to 1st capsule (cm), *TSW* thousand-seed weight (g/1000 seed), *SYH* seed yield per hectare (ton ha⁻¹), *OYH* oil yield per hectare.

Table 2. Mean values for agronomic traits of 100 sesame specimens of Ethiopia showing the top 10 and bottom 5 ranked entries based on seed yield (ton ha⁻¹) across two sites.

No	Entry name or designation	Traits													
		DF	DM	PH	INL	NPB	NSB	NCPP	NSPC	SHB	DFLBC	TSW	SYH	OYH	OC
Top 10 entries															
1	Hirhir Kebabo Hairless-9	41	90	123.9	9.8	4	1	41	58	18.1	42.8	3.0	1.01	0.40	50.9
2	Setit-3	43	91	131.6	10.0	3	2	52	63	18.0	26.5	3.1	0.84	0.40	49.1
3	Orofalc ACC-2	42	89	103.9	8.3	3	2	36	56	13.8	22.4	3.1	0.80	0.40	52.5
4	Hirhir Humera Sel-6	43	92	114.8	9.7	4	4	48	61	27.2	38.1	2.9	0.78	0.39	53.9
5	ABX=2-01-2	44	95	133.2	8.7	4	3	40	65	18.9	40.9	2.7	0.74	0.36	48.9
6	Setit-1	40	88	125.6	7.9	4	2	39	56	21.9	38.9	3.3	0.73	0.39	53.8
7	ACC 205-180	45	97	126.9	10.5	3	2	40	55	32.5	42.0	3.1	0.72	0.38	53.1
8	ACC 203-616	45	94	106.8	8.5	3	1	37	54	25.5	39.0	2.6	0.69	0.35	51.7
9	NN-0029(2)	40	88	118.5	9.5	4	1	44	65	28.2	41.9	3.4	0.68	0.36	50.7
10	GA-002(3)	42	90	119	10.3	4	2	49	63	27.7	34.8	2.8	0.67	0.33	49.3
Bottom 5 entries															
1	NN-0183-3	43	92	117.1	8.1	3	2	34	48	20.8	42.6	2.3	0.17	0.07	45.8
2	NN-0020	43	98	125.8	8.2	4	2	34	62	29.0	44.4	2.6	0.24	0.12	49.3
3	NN-0108-2	39	90	135.2	10.2	4	0	32	47	20.6	47.2	2.9	0.26	0.04	47.3
4	NN00136-1	43	97	99.6	10.2	3	2	42	44	26.1	47.4	3.0	0.26	0.13	47.5
5	NN-0143	42	94	132.2	10.3	4	3	30	55	27.2	39.7	3.2	0.28	0.15	48.3
	Mean	43	92	119.6	9.2	3.5	2	41	51	24.4	39.8	2.9	0.48	0.24	49.7
	CV (%)	5.03	3.61	13.30	13.52	19.21	38.58	29.42	20.92	29.74	20.44	78.06	35.26	90.13	NA
	R ² (%)	72.87	70.33	68.59	86.27	85.16	82.97	77.76	70.65	77.22	85.95	60.98	82.15	66.42	NA
	LSD (p ≤0.05)	3.03	4.65	21.78	1.73	0.96	0.88	16.74	14.67	10.16	11.33	3.23	0.24	0.32	NA

Note: CV coefficient of variation, R² coefficient of determination, LSD least significant difference, NA not available, DF days-to- 50% flowering, DM days-to-75% maturity, PH plant height (cm), INL inter node length (cm), NPB number of primary branches per plant, NSB number of secondary branches per plant, NCPP number of capsules per plant, NSPC number of seeds per capsule, SHB stem height to 1st branch (cm), DFLBC distance from lowest branch to 1st capsule (cm), TSW thousand-seed weight (g/1000 seed), SYH seed yield per hectare (ton ha⁻¹), OYH oil yield per hectare, OC oil content (%).

Table 3. Phenotypic correlations coefficients for assessed agro-morphological traits, oil content and oil yield of 100 sesame collections evaluated across two locations in Ethiopia.

Traits	DM	PH	INL	NPB	NSB	NCPP	NSPC	SHB	DFLBC	TSW	OC	OYH	SYH
DF	0.31**	0.05ns	0.14**	0.09ns	0.09ns	0.16**	-0.21**	0.05ns	-0.13**	-0.02ns	-0.11*	0.08ns	0.09ns
DM	1.00	0.07ns	0.08ns	-0.20ns	0.07ns	-0.10*	-0.12*	0.09ns	0.00ns	-0.15**	-0.06ns	-0.10*	-0.09ns
PH		1.00	0.09ns	0.06ns	0.12*	0.07ns	0.05ns	0.01ns	0.07ns	0.10*	-0.00ns	0.05ns	0.05ns
INL			1.00	-0.17**	0.21**	0.08ns	-0.08ns	0.38**	-0.45**	0.00ns	0.03ns	0.35**	0.35**
NPB				1.00	0.06ns	0.30**	0.08ns	-0.14**	0.25**	0.20**	0.05ns	-0.05ns	-0.06ns
NSB					1.00	-0.05ns	0.04ns	0.34**	-0.166**	-0.122*	0.051ns	0.214**	0.207**
NCPP						1.00	0.119*	-0.00ns	-0.01ns	0.16**	-0.01ns	0.17**	0.18**
NSPC							1.00	-0.07ns	0.21**	0.04ns	-0.08ns	0.15**	0.17**
SHB								1.00	-0.26**	-0.13**	0.02ns	0.16**	0.16**
DFLBC									1.00	-0.01ns	-0.02ns	-0.29**	-0.29**
TSW										1.00	0.05ns	0.23**	0.23**
OC											1.00	0.16**	0.06ns
OYH												1.00	0.99**

Note: * and ** denote significance difference at the 5% and 1% levels of probability, respectively; *ns* non-significant, *DF* days-to- 50% flowering, *DM* days-to-75% maturity, *PH* plant height (cm), *INL* inter node length (cm), *NPB* number of primary branches per plant, *NSB* number of secondary branches per plant, *NCPP* number of capsules per plant, *NSPC* number of seeds per capsule, *SHB* stem height to 1st branch (cm), *DFLBC* distance from lowest branch to 1st capsule (cm), *TSW* thousand-seed weight (g/1000 seed), *OC* oil content (%), *OYH* oil yield per hectare, *SYH* seed yield per hectare (ton ha⁻¹).

Table 4. Eigenvectors, eigenvalues, percentage and cumulative variation explained by principal components (PCs) for agro-morphological traits, oil content and oil yield of 100 sesame germplasm collections.

Traits	PC1	PC2	PC3	PC4	PC5	PC6
DF	-0.16	0.51	0.21	0.04	-0.17	-0.09
DM	-0.19	0.48	0.06	-0.18	0.06	-0.01
PH	0.19	0.10	-0.24	-0.39	-0.11	0.30
INL	0.17	-0.02	-0.19	-0.27	0.29	-0.61
NPB	-0.09	-0.25	-0.24	0.53	0.30	0.04
NSB	0.14	0.20	0.33	0.10	0.32	-0.40
NCPP	0.25	0.28	0.00	0.37	-0.12	-0.05
NSPC	0.41	0.20	-0.24	0.13	-0.12	-0.19
SHB	0.03	0.16	-0.19	-0.12	0.67	0.13
DFLBC	0.04	0.38	-0.39	-0.09	0.12	0.27
TSW	0.27	-0.21	0.19	-0.16	0.22	0.23
OC	0.15	-0.09	0.50	-0.36	0.04	0.11
OYH	0.44	0.12	0.07	0.19	0.07	0.35
SYH	0.50	-0.01	0.21	0.15	-0.01	0.00
Eigen-values	2.59	2.19	1.57	1.37	1.21	1.00
Proportion variance (%)	17.26	14.59	10.46	9.111	8.10	6.67
cumulative variation (%)	17.26	31.85	42.31	51.43	59.52	66.20

DF days-to- 50% flowering, *DM* days-to-75% maturity, *PH* plant height (cm), *INL* inter node length (cm), *NPB* number of primary branches per plant, *NSB* number of secondary branches per plant, *NCPP* number of capsules per plant, *NSPC* number of seeds per capsule, *SHB* stem height to 1st branch (cm), *DFLBC* distance from lowest branch to 1st capsule (cm), *TSW* thousand-seed weight (g/1000 seed), *OC* oil content (%), *OYH* oil yield per hectare, *SYH* seed yield per hectare (ton ha⁻¹).

Table 5. Genetic parameters estimated for 100 sesame specimens using 27 SSR markers.

Locus	Product size (bp)	Genetic parameters			
		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

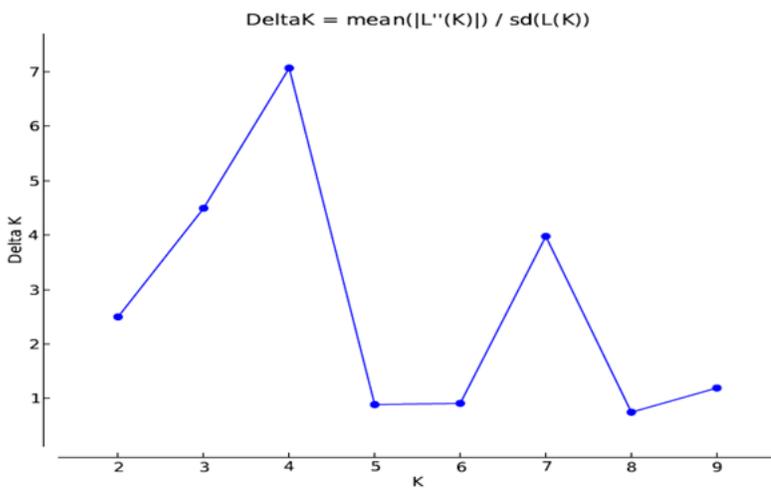
Note: *MAF* major allele frequency, *He* unbiased expected heterozygosity (gene diversity), *Ho* observed heterozygosity, *PIC* polymorphic information content.

Table 6. Genetic clusters and their member entries, the proportion of the membership, mean expected heterozygosity, and fixation index based on structure analysis of 63 sesame entries with 27 SSR markers.

Cluster	Sub-cluster	Entries	Membership %	Expected heterozygosity	Mean fixation index
I	I-a	ACC NS-031E63, Hirhir Nigara 1st Sel-2, Hirhir Kebabo Hairless Sel-6, NN-0183-3, GXT=85(28-2), ABXT-85-Sel-2-1, ABXC-50402, NN-0129-2, ACC-203-020, Gonjam Azene (Aleka), NN-0026, Tejareb Kokit Sel-3, NN0027, NN0009, NN-0088-2 Bawnji Sel-2, G-02, Endelemi kirem sel-2, NN0016-1, NN0038-1, NN-0052, NN0071, NN0064-1	24	0.15	0.39
	I-b	Hirhir Baeker-Sel-3, NN0025, BCS-0041, Orofal ACC-2, ABX=2-01-2, Bering Bawany, ACC-203-612, NN-0143, NN-0146, NN-0044-2, NN-0018-2, NN-0029-1, NN0056	13	0.22	0.23
II	II-a	Hirhir kebabo hairless sel-2, Hirhir kebabo hairless-9, ACC-200-064-1, HIRHIR NIGARA 1ST SEL-1, Bawnji Fiyel Kolet, NN0104, Bawnji Maksegnt, Bawnji Flwaha Sel-2, Hirhir Filwaha Large Seeded	9	0.29	0.02
	II-b	Hirhir Humera Sel-6, Humera-1, Setit-1, Setit-3, Hirhir Sel-2, Hirhir Kebabo Hairless-Sel-7, Hirhir Humera Sel-8, NN0036-1, Hirhir Kebabo Early Sel-1, Hirhir Adgeshu Sel-8, Morgo-Sel-P=13, NN-0020 ACC-203-610, NN0001-2, NN01-13, Gojam Azene(Yohans Sel-1), NN0031	17	0.20	0.29

Figures

(a)



(b)

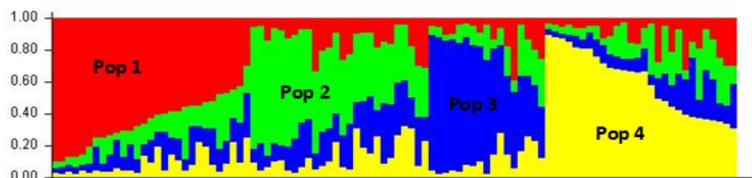


Figure 1

Population structure of 100 sesame entries revealed by 27 SSR markers resolving four sub-populations: (a) Delta K estimation based on the Evanno procedure and (b) Sub-populations for the best delta K value of four. Pop 1, 2, 3 & 4 denote Populations 1, 2, 3 and 4, respectively.

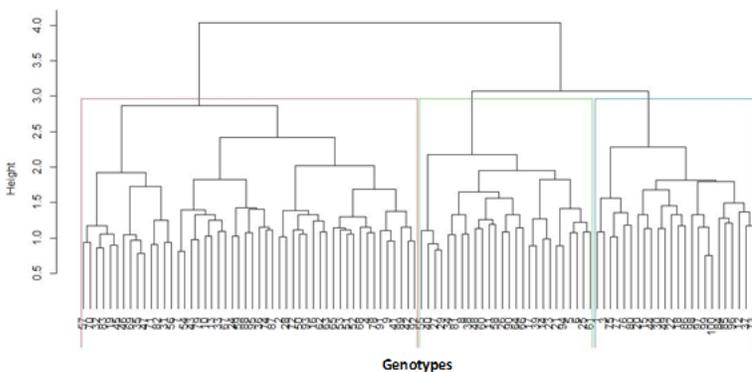


Figure 2

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.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Desawietal.SupplementaryTables.docx](#)
- [Desawietal.Supplementaryinformation.docx](#)