Oil Crop Science 8 (2023) 266–277

Contents lists available at ScienceDirect

Oil Crop Science

journal homepage: www.keaipublishing.com/en/journals/oil-crop-science

Large scale genetic landscape and population structure of Ethiopian sesame (Sesamum indicum L.) germplasm revealed through molecular marker analysis

Muez Berhe [a,b], Jun You [a], Komivi Dossa [c,g], Fetien Abay Abera [d,e], Emmanuel Amponsah Adjei [f], Yanxin Zhang [a,*], Linhai Wang [a,**]

[a] Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, and Rural Afairs, No.2 Xudong 2nd Road, Wuhan, 430062, China
[b] Tigray Agricultural Research Institute, Humera Agricultural Research Center, P.O. Box 62, Tigray, Ethiopia
[c] CIRAD, UMR AGAP Institut, 97170 Petit Bourg, Guadeloupe, France
[d] Department of Dryland Crop and Horticultural Sciences, College of Dryland Agriculture and Natural Resources, Mekelle University, Mekelle, Tigray, Ethiopia
[e] University of California Davis Department of Plant Sciences, MS/11 Shields Ave, Davis, CA, 95616, USA
[f] CSIR-Savanna Agriculture Research Institute, P.O.Box TL52, Tamale, Ghana
[g] UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France

ARTICLE INFO

Keywords:
Alleles
Conservation
Genetic diversity
Molecular markers
Population structure

ABSTRACT

Sesame (*Sesamum indicum* L.) plays a crucial role in Ethiopian agriculture, serving both subsistence and commercial purposes. However, our understanding of the extensive genetic diversity and population structure of Ethiopian sesame remains limited. To address this knowledge gap, we genotyped 368 Ethiopian sesame germplasms, categorizing into four distinct breeding groups: Accessions, landraces, improved varieties, and wild types, using a comprehensive set of 28 polymorphic markers, including 23 simple sequence repeat (SSR) and five Insertion-Deletion (InDel) markers. These markers ensured robust genomic representation, with at least two markers per linkage group. Our results unveiled substantial genetic diversity, identifying a total of 535 alleles across all accessions. On average, each locus displayed 8.83 alleles, with observed and expected heterozygosity values of 0.30 and 0.36, respectively. Gene Diversity and Polymorphic Information Content (PIC) were recorded at 0.37 and 0.35. The percentage of polymorphic loci varied significantly among breeding groups, ranging from 8.00% to 82.40%, indicating high diversity in accessions (82.4%), moderate diversity in improved varieties (31.20%) and landraces (29.60%), and limited diversity in wild types (8.00). Analysis of Molecular Variance (AMOVA) results emphasized significant genetic differentiation among populations, with substantial diversity ($P < 0.001$) within each population. Approximately 8% of the entire genetic diversity could be attributed to distinctions among populations, while the larger proportion of genetic diversity (92%) resided within each individual sesame population, showcasing heightened diversity within each group. Our study's findings received support from both Bayesian clustering and Neighbor-joining (NJ) analysis, reaffirming the credibility of our genetic structure insights. Notably, Population structure analysis at its highest Δk value ($k = 2$) revealed the existence of two primary genetic clusters, further subdivided into four sub-populations at $k = 4$. Similarly, NJ analysis identified two prominent clusters, each displaying additional sub-clustering. In conclusion, our research provides a comprehensive understanding of genetic groups, subpopulations, and overall diversity within Ethiopian sesame populations. These findings underscore the significant genetic diversity and population structure within Ethiopian sesame germplasm collections. This genetic richness holds promise for breeding and conservation efforts, highlighting the importance of preserving genetic diversity to ensure adaptation to changing environments and meet the needs of farmers and consumers.

* Corresponding author.
** Corresponding author.
E-mail addresses: muez.sbn@gmail.com (M. Berhe), komivi.dossa@cirad.fr (K. Dossa), fetien.abay@mu.edu.et (F.A. Abera), emmaadjei1@gmail.com (E.A. Adjei), 13277065857@163.com (Y. Zhang), linhai827@163.com (L. Wang).

https://doi.org/10.1016/j.ocsci.2023.11.003
Received 18 October 2023; Received in revised form 27 November 2023; Accepted 27 November 2023
Available online 22 December 2023

1. Introduction

Sesame (*Sesamum indicum* L.) is a diploid plant with 2n = 2x = 26 chromosomes, belonging to the family Pedaliaceae and the genus Sesamum. The Pedaliaceae family comprises 16 genera and approximately 60 species, with 37 of those species belonging to the Sesamum genus which has average genome size of ~360 MB (Kefale and Wang, 2022). Among these, *Sesamum indicum* L. is the most commonly cultivated variety (Kobayashi et al., 1990). While sesame is primarily considered a self-pollinated plant, there is a reported low percentage of cross-pollination (Kobayashi et al., 1990). Sesame is indigenous to tropical and some temperate regions, thriving at altitudes ranging from 500 to 800 m above sea level (masl) and occasionally up to 1300 masl under ideal conditions, including well-drained soils, moderate fertility, and temperatures between 27 °C and 35 °C (Baraki and Berhe, 2019). Sesame seeds are not only a popular spice but also a rich source of folic acid, oil, protein, unsaturated fatty acids, vitamins, and minerals. Their beneficial properties have led to various applications in the production of insecticides, pharmaceuticals, and skincare products (Anilakumar et al., 2010). Sesame oil, in particular, is known for its stability, thanks to the presence of three natural antioxidants: sesamolin, sesamin, and sesamol, which collectively contribute to about 50–60% of the oil's stability (Lee et al., 2008). The origins and domestication process of cultivated sesame have been subjects of debate, with Africa and Asia identified as probable origins (De Candolle, 1892; Weiss, 2000). Ethiopia has been suggested as a center of origin and diversity for cultivated sesame (Bedigian, 1981; Mahajan et al., 2007), with the country hosting significant morphological and genetic diversity within its sesame populations (Baraki and Berhe, 2019; Teklu et al., 2021). Ethiopia's Ethiopian Biodiversity Institute (EBI) houses one of the largest collections of sesame genetic resources in Africa, preserving around 5000 genetically diverse sesame germplasm resources (Woldesenbet et al., 2015).

In Ethiopia, sesame holds immense economic significance, serving as both a crucial income source and a dietary staple. It plays a substantial role in contributing to Ethiopia's gross domestic product (GDP), with nearly half of the designated land for oil crops in 2020 dedicated to sesame cultivation (Berhe et al., 2023; Teklu et al., 2021). Ethiopia stands as a major contributor to sub-Saharan Africa's sesame exports, with approximately 8.96% of the world's sesame exports originating from the country in 2019 (Berhe et al., 2023). This export volume was valued at 307 million USD, making sesame Ethiopia's second-largest export commodity after coffee (Kassie et al., 2023). Sesame production also holds great potential as a tool for poverty alleviation and rural development in Ethiopia. It is particularly well-suited for smallholder farmers, who can cultivate it on their own land without the need for extensive mechanization or costly inputs. As of 2019, sub-Saharan African nations collectively exported 1,465,493 tons of unprocessed sesame, valued at 1.9 trillion USD, with Ethiopia contributing significantly to this export volume (FAOSTAT, 2020). Till 2020, 45.7% of the allocated land for oil crops in Ethiopia was earmarked for sesame cultivation (CSA, 2020). Recognizing the pivotal role of sesame in their economy, the Ethiopian government has prioritized sesame as a key crop, implementing various policies and initiatives to support its production and export (Diriba, 2018).

Across the world, extensive morphological, physiological, and reproductive variations have been reported in sesame (Dossa et al., 2016; Wang et al., 2014; Wei et al., 2015). Similarly, Ethiopian sesame germplasm, collected from diverse regions within the country, exhibits significant morphological and genetic differences (Abate et al., 2015; Gebremichael and Parzies, 2011; Teklu et al., 2021; Woldesenbet et al., 2015). Genetic diversity studies are essential, and while morphological markers are influenced by environmental factors, molecular markers, such as SSRs (simple sequence repeats), offer a more reliable means of estimating genetic diversity (Ahmar et al., 2020; Sharma et al., 2022). Numerous studies have explored the genetic variations in sesame germplasm collections using various markers, including RAPD (random amplified polymorphic DNA) (Akbar et al., 2011; Ercan et al., 2004), AFLP (amplified fragment length polymorphism) (Uzun et al., 2003; L.-B. Wei et al., 2009), SSRs (Dossa et al., 2016; Park et al., 2014; Teklu et al., 2022), ISSR (inter simple sequence repeats) (El Harfi et al., 2021), and SNPs (single-nucleotide polymorphisms) (Basak et al., 2019; Cui et al., 2017; Wang et al., 2014). Employing those different markers, understanding the genetic diversity and population structure of sesame is crucial for sustainable development and food security, as it allows for the cultivation of crops in various conditions. This knowledge aids in the development of varieties adapted to local environments, ultimately increasing yields and resilience. This plays a vital role in plant breeding, providing the foundation for developing new varieties with improved traits. Additionally, identifying and conserving unique genetic resources through genetic diversity studies can ensure the crop's sustainability. Moreover, such studies can uncover varieties with unique properties that may have higher market demand, enhancing the crop's profitability. However, few studies have focused on large samples of sesame accessions from Ethiopia, the primary center of origin for cultivated sesame and its wild relatives. Hence, there is a need for comprehensive genetic diversity analysis among Ethiopian germplasm collections representing diverse sesame growing areas. Therefore, this study is designed to examine the genetic diversity and population structure of Ethiopian sesame germplasm using SSR and InDel markers to offer valuable insights for conservation, breeding, adaptation, and marketability of this vital crop.

2. Materials and methods

2.1. Plant materials

A total of 368 sesame germplasms were employed in this study, encompassing diverse populations from four distinct sesame population categories: germplasm collections (Accessions), improved cultivars, landraces, and wild type accessions. These samples were sourced from the comprehensive collection of sesame germplasm housed at the Ethiopian Biodiversity Institute (EBI) and from seeds stored on farms across various regions of Ethiopia. The improved varieties were obtained from repositories such as the Humera Agricultural Research Center (HuARC) in the Tigray Region, Haramaya University (HU) in the Oromia region, and the Gander Agricultural Research Center (GARC) and Srinka Agricultural Research Center in the Amhara regions. To ensure genetic purity, all germplasms underwent two generations of self-pollination at the Humera Agricultural Research Center (HuARC) station (14150 N, 36370 E). Additional details regarding the samples can be found in Supplementary Materials Table S1.

2.2. DNA extraction

Ten sets of bulked leaves from three-week-old seedlings of each genotype were harvested. These seedlings were grown on a plastic array within a controlled environment chamber in the molecular laboratory. The primary goal was to obtain leaf tissues for DNA extraction. The method employed for DNA extraction involved the utilization of the CTAB technique, which relies on cetyltrimethyl ammonium bromide, as described by Zhang et al., in 2012. The extracted DNA's quality and quantity were assessed through spectrophotometry using the NanoDrop 2000 instrument from Thermo Scientific, located in Wilmington, DE, USA. Following extraction, the DNA samples were preserved by storing them at a temperature of −20 °C, ensuring their availability for subsequent usage.

2.3. PCR and electrophoresis

To assess marker polymorphism, a preliminary screening involving 57 potential markers and five randomly chosen accessions was conducted. Ultimately, a set of 28 highly polymorphic markers (23 SSR and 5 InDel) markers were selected based on their substantial polymorphism

information content and their comprehensive coverage of all Linkage Groups within the sesame genome.

The Polymerase Chain Reaction (PCR) was employed to amplify all the 28 polymorphic markers. The PCR reactions were prepared with a total volume of 20 μL, comprising 25 ng of DNA, 4 μmol of forward primers, 4 μmol of reverse primers, 1 × buffer, 0.25 mmol of deoxyribonucleotide triphosphates (dNTPs), and 0.80 U of Taq polymerase. The PCR amplification followed a specific profile: an initial denaturation step at 94 °C for 1 min, primer annealing at temperatures ranging from 45.2 to 53 °C for 1 min, and elongation at 72 °C for 1 min. This cycle was repeated 34 times to facilitate the amplification process. Subsequently, the reaction was concluded with a final extension step at 72 °C for 10 min. It is noteworthy that, except for the annealing temperatures, the PCR conditions remained consistent across all primers used in the study. This standardized approach ensures uniformity in the amplification process. Following the PCR amplification, the products were subjected to size separation using 6% denaturing polyacrylamide gels. The electrophoresis parameters and silver staining of gels were performed according to the protocols established by Zhang et al. (2012). To further analyze and document the amplified products, capillary electrophoresis was conducted using an ABI 3730 automatic sequencer. The resulting marker data, representing fragment sizes, were meticulously recorded and presented in an Excel spreadsheet for subsequent analysis and interpretation. This comprehensive methodology provides a robust and standardized approach to SSR marker analysis, ensuring reliable and reproducible results.

2.4. Scoring and data analysis

Following the acquisition of genetic data through PCR and capillary electrophoresis, the GeneMarker v3.0 software was employed for analysis, generating reports of peak areas and electropherogram images (Holland and Parson, 2011). Manual allele scoring was conducted, based on the presence ("1″) or absence ("0″) of specific size alleles within each germplasm sample. For genetic parameter estimation, POPGENE version 1.32 was utilized, encompassing metrics such as the number of alleles (Na), effective number of alleles (Ne), Nei's Gene Diversity (He), and observed heterozygosity (Ho) (Yeh, 1997). Major Allele Frequency (MAF), Number of private alleles (Np), and Polymorphic Information Content (PIC) were computed using PowerMarker version 3.25 (Liu and Muse, 2005).

To explore genetic structure, Analysis of Molecular Variance (AMOVA) was performed using GENALEX 6.5 to assess variation within and among germplasm categories (Peakall and Smouse, 2006). A genetic distance matrix was generated, and Principal Coordinates Analysis (PCoA) was conducted using GENALEX 6.5 to unravel pairwise genetic relationships among the 368 sesame genotypes. For constructing a Neighbor-Joining (NJ) tree based on Nei's genetic distance, MEGA version 11 was employed (Tamura et al., 2021). Inference of population structure was achieved using the Bayesian clustering method in STRUCTURE 3.2.4 (Pritchard et al., 2010). Ten runs were executed for each k value (ranging from 1 to 10), with the true k determined using the Online STRUCTURE Harvest (OSH) web-based tool (http://taylor0.biology.ucla.edu/struct_harves(Earl and VonHoldt, 2012).

3. Results

3.1. Marker polymorphism in the sesame different populations

Genetic variation across diverse populations of Ethiopian sesame was meticulously investigated through the utilization of a set of 28 polymorphic markers. A comprehensive analysis of 368 sesame germplasms provided detailed insights, and the markers' information is available in Supplementary Material Table S2. Each linkage group was adequately represented by a minimum of two markers, resulting in the identification of 535 alleles. The average allelic richness of 8.83 alleles per genetic locus, as detailed in Table 1 and Supplementary Material Table S3, serves as a baseline measure of genetic diversity within these sesame populations.

The allelic diversity spectrum ranged from two alleles, observed in several markers (ID0046, ZMM1043, ZMM5015, ZMM4664, ZMM5358, ZMM185, and ZMM1691), to a remarkable 54 alleles per locus, notably found in markers ID0041 and ZMM3223. This range, outlined in Supplementary Material Table S3, signifies the extensive genetic variation present in specific genomic regions. Notably, MAF assessment at each genetic locus revealed a range from 0.11 to 1.00, providing insights into the prevalence of major alleles. On average, MAF was found to be 0.75, indicating a balanced distribution of major and minor alleles across the sampled populations.

Furthermore, variability in Gene Diversity and PIC values was observed across the set of markers. Some markers exhibited high values, indicating substantial diversity and informativeness, while others displayed lower values, suggesting regions of reduced diversity. This nuanced information, detailed in Table 1, guides our understanding of the genetic landscape of sesame populations. The computed mean Gene Diversity value of 0.37 reflects the overall genetic diversity, with fluctuations ranging from 0.01 (detected in markers ZMM1851 and ZMM1691) to a notably higher 0.94 (observed in marker ID0041). PIC values, providing a measure of marker informativeness, ranged from 0.01 to 0.94, with an average PIC value of 0.35, as outlined in Supplementary Material Table S3. These values highlight the diverse nature of the markers used in this study, offering researchers a range of tools for further investigations.

Moreover, Ho ranged from 0.15 to 0.50, with a mean of 0.36, representing the proportion of heterozygous individuals in the populations. The He spanned from 0.12 to 0.50, with a mean of 0.30, providing insights into the genetic equilibrium within the populations. These findings, detailed in Table 1, contribute to our understanding of the genetic

Table 1
Genetic parameter estimates of 28 SSR markers using 368 sesame genotypes.

Marker	Product size	MAF	Allele No	Ho	He	Gene Diversity	PIC
ID0041	280	0.52	24.00	0.43	0.48	**0.56**	**0.55**
ID0046	101	0.86	2.00	0.50	0.50	**0.21**	**0.18**
ID0068	199	0.71	5.83	0.30	0.39	**0.41**	**0.37**
ID0145	196	0.71	4.33	0.18	0.19	**0.40**	**0.34**
ID0175	271	0.54	16.00	0.21	0.30	**0.64**	**0.61**
ZMM1033	179	0.67	17.63	0.28	0.38	**0.48**	**0.47**
ZMM1043	184	0.78	6.29	0.48	0.50	**0.31**	**0.27**
ZMM1189	212	0.78	4.75	0.27	0.36	**0.33**	**0.30**
ZMM1353	169	0.72	7.00	0.16	0.19	**0.44**	**0.40**
ZMM1637	265	0.55	9.5	0.25	0.34	**0.59**	**0.52**
ZMM1691	220	0.88	3.00	0.12	0.15	**0.18**	**0.17**
ZMM1700	258	0.98	3.00	0.23	0.32	**0.04**	**0.04**
ZMM1809	256	0.81	5.50	0.37	0.42	**0.29**	**0.26**
ZMM1851	280	0.98	3.00	0.27	0.38	**0.04**	**0.04**
ZMM1935	280	0.65	18.33	0.34	0.41	**0.53**	**0.51**
ZMM2202	276	0.43	24.00	0.26	0.35	**0.74**	**0.71**
ZMM2321	280	0.78	8.50	0.34	0.41	**0.32**	**0.29**
		78					
ZMM2818	279	0.91	9.00	0.30	0.39	**0.18**	**0.18**
ZMM3013	216	0.69	7.00	0.13	0.16	**0.47**	**0.44**
ZMM3223	279	0.52	33.00	0.43	0.48	**0.68**	**0.66**
ZMM3690	244	0.70	13.10	0.46	0.49	**0.43**	**0.40**
ZMM3741	264	0.59	20.5	0.43	0.49	**0.60**	**0.58**
ZMM4645	179	0.81	6.00	0.25	0.33	**0.29**	**0.27**
ZMM4664	184	0.91	3.5	0.38	0.47	**0.13**	**0.12**
ZMM4803	268	0.96	4.00	0.14	0.18	**0.08**	**0.08**
ZMM5015	151	0.85	4.27	0.38	0.48	**0.23**	**0.22**
ZMM5358	164	0.61	7.38	0.33	0.40	**0.49**	**0.45**
ZMM6141	167	0.79	4.00	0.13	0.17	**0.36**	**0.32**
Mean		**0.75**	**8.83**	**0.30**	**0.36**	**0.37**	**0.35**

Allele No = number of alleles per locus by population, Ho = observed average Heterozygosity, He = expected Heterozygosity, MAF = Major allele frequency, PIC = polymorphic information content.

characteristics and potential factors influencing heterozygosity levels in Ethiopian sesame populations.

3.2. Allele variation among across the different Ethiopian sesame population using 368 sesame germplasm and 28 markers

The allelic patterns and genetic diversity (Table 2) reveal the range of genetic variation present among the Ethiopian sesame populations, classified based on their origin or sources. Within the Accession population, a remarkably rich genetic diversity was evident. This is particularly exemplified by a significantly high mean allele number (8.62) observed across the examined loci. Conversely, the improved population shows reduced allele diversity, with a lower mean allele number (2.40) compared to the Accession group. The Landraces population occupies an intermediate position, retaining more diversity than the improved group but less than the Accession population. This is demonstrated by a marginally higher mean allele number than the improved group, yet still below that of the Accession population. The Wild population, in contrast, displays the lowest mean allele number, indicative of a limited genetic diversity. This paucity might arise due to a small population size or genetic isolation.

Na and Ne exhibit a consistent trend across populations, with the Accession group displaying the highest count and the Wild group presenting the least. Both the Improved and Landraces populations share a comparable level of allele diversity at the studied loci, suggesting fewer allele variants per locus in these groups. The Accession population, characterized by the highest Na and Ne, indicates a greater overall allele diversity, reflecting a broader genetic spectrum within this group. In contrast, the Wild group, with the least Na and Ne, demonstrates a more restricted range of alleles, suggesting a narrower genetic diversity. Analysis of MAF values across the four populations reveals a common theme—the prevalence of dominant alleles. However, the degree of dominance varies. In the Accession population, the lowest MAF (0.75) suggests a relatively balanced distribution of alleles, indicative of a diverse genetic composition. In contrast, the Improved and Landraces populations exhibit slightly higher MAF values, pointing to a more dominant allele presence compared to the Accession group. Notably, the Wild population stands out with the highest MAF, signifying an extreme dominance of a single allele. This exceptional dominance in the Wild group may indicate a unique genetic profile or potential selective pressures acting on specific loci. Variation in MAF values provides insights into the prevalence and dominance of specific alleles, offering clues about the genetic composition and potential evolutionary forces shaping each population.

The analysis, detailed in Table 2, illuminates the genetic diversity and marker informativeness across different populations of sesame. A standout observation was the Accession population, which demonstrates the highest genetic diversity, evident through superior gene diversity, PIC, and % PIC values. This suggests that the employed markers effectively capture the rich genetic variation within the Accession group. Moreover, the high percentage of polymorphic loci (82.4%) within the Accession category underscores its significance as a reservoir of diverse genetic material. In contrast, the Improved and Landraces populations exhibit intermediate levels of gene diversity, PIC, and % PIC values. This points towards a moderate genetic diversity in these populations, and the employed markers provide informative insights into their genetic makeup. The moderate percentages of polymorphic loci (31.2% in improved and 29.6% in landraces) suggest a level of shared genetic diversity within these groups. Conversely, the Wild population stands out with the lowest gene diversity, PIC, and % PIC values, indicating a significant dearth of genetic diversity and limited marker informativeness. This aligns with the pronounced scarcity of polymorphic loci (8.0%) within the Wild type sesame populations, highlighting a distinctive genetic profile that sets them apart from the other sesame populations.

Moving beyond these broad assessments, a closer look at the allelic count (Table 3) provides deeper insights. The allelic count exhibits a considerable range, from 10 alleles within the Wild type population to an impressive 103 alleles within the Accession group. This wide range underscores substantial disparities in genetic diversity among populations, with the Accession group representing a genetic reservoir of notable richness. Furthermore, the count of alleles with a frequency equal to or exceeding 5% varies across populations, emphasizing differences in the abundance of shared alleles. Alleles exclusive to a specific population range from 0 in the Improved group to a substantial 55 within the Accession group, highlighting varied levels of distinct genetic variation. Additionally, the count of bands present in less than or equal to 50% of the populations varies, indicating differing prevalence levels for certain bands across populations.

3.3. Euclidean distances and population relationships in Ethiopian sesame

Upon delving into the analysis of relative Euclidean distances among various Ethiopian sesame populations (as detailed in Table 4), discernible patterns emerge that offer valuable insights into the genetic relationships among these groups. The "Wild" population was exhibiting the greatest Euclidean distances when compared to the other groups suggesting the wild sesame population diverges significantly in terms of its frequency distribution of markers in comparison to other population types. The distinctiveness of the wild population is highlighted by its elevated Euclidean distances, suggesting a unique genetic makeup that

Table 3
Total band pattern by population using 368 sesame germplasms and 28 SSR.

Band pattern	Sesame populations			
	Accession	improved	landraces	Wild
No. Bands	103	39	37	10
No. Bands Freq. ≥ 5%	33	39	31	10
No. Private Bands	55	0	1	1
No. LComm Bands (≤ 25%)	0	0	0	0
No. LComm Bands (≤ 50%)	12	10	9	0

No. Bands = No. of Different Bands, No. Bands Freq. ≥ 5% = No. of Different Bands with a Frequency ≥ 5%, No. Private Bands = No. of Bands Unique to a Single Population, No. LComm Bands (≤ 25%) = No. of Locally Common Bands (Freq. ≥ 5%) Found in 25% or Fewer Populations, No. LComm Bands (≤ 50%) = No. of Locally Common Bands (Freq. ≥ 5%) Found in 50% or Fewer Populations.

Table 2
Summary of the allele pattern and genetic diversity of the Ethiopian sesame over loci for each population.

Sesame population	Mean value over loci							
	N	Allele No.	Na	Ne	MAF	Gene diversity	PIC	% of Polymorphic Loci
Accession	323	8.62	1.65	1.07	0.75	0.36	0.33	82.40
Improved	120	2.4	0.62	1.08	0.79	0.29	0.26	31.20
landraces	28	2.70	0.59	1.07	0.80	0.27	0.24	29.60
Wild	5	1.4	0.16	1.03	0.90	0.14	0.12	8.00
Grand Mean over Loci		3.78	0.76	1.06	0.81	0.27	0.24	37.8

"N = population size, Allele No. = number of alleles per locus by population, Na = number of different alleles per locus by population, Ne = number of effective alleles per locus by population, MAF = Major allele frequency, PIC = polymorphic information content."

Table 4
Euclidean Distances between Ethiopian Sesame Populations Based on the SSR marker Frequency Distributions".

OTU	Accession	improved	landraces	Wild
Accession	0.0000	0.1907	0.2085	0.2752
improved	0.1907	0.0000	0.1499	0.2558
landraces	0.2085	0.1499	0.0000	0.2330
Wild	0.2752	0.2558	0.2330	0.0000

OTU= Operational Taxonomic Unit.

sets it apart from the other sesame populations. In contrast, both the "Accessions" and "Improved Varieties" populations display comparatively shorter distances, as depicted in Supplementary Fig. 1. This pattern suggests that these groups potentially share certain similarities in terms of their marker frequency distribution. The reduced Euclidean distances between accessions and improved varieties indicate a closer genetic resemblance, hinting at shared genetic traits or a common genetic background between these two groups.

Notably, the Euclidean distance between "Improved Varieties" and "Landraces" is noticeably lower than the distance between wild types and other populations. This intriguing finding suggests a moderate level of resemblance between improved varieties and landraces in their marker frequency distribution. The closer Euclidean distance between these two cultivated groups may indicate shared genetic elements resulting from human selection and cultivation practices.

3.4. Genetic relationships unveiled through Principal Coordinate Analysis

Principal Coordinate Analysis (PCoA) serves as a powerful statistical tool, allowing for the visual exploration of genetic similarities and differences among a diverse collection of samples. In this study, PCoA was applied to unravel the intricate genetic connections within a set of 368 sesame germplasm samples, aiming to unveil underlying genetic clusters. The graphical representation in Fig. 1 provides a visual encapsulation of the outcomes derived from the PCoA analysis, where each point on the graph represents a distinct sesame genotype. The foundation of PCoA was established using a genetic distance matrix derived from molecular data. The initial set of three axes collectively illuminated 19.47% of the variance present within the molecular dataset. Specifically, the first, second, and third axes contributed to 7.99%, 6.98%, and 4.50% of the variance, respectively. This multivariate approach allowed for a comprehensive view of the genetic landscape, revealing intriguing patterns and relationships among the sesame populations.

The PCoA analysis provided valuable insights, indicating that, while certain subsets within the accessions displayed noticeable differences, a clear separation based on their origins was not immediately evident (Fig. 1). Notably, populations from improved, landraces, and wild variants exhibited overlaps with the accession population, hinting at potential genetic commonalities. The absence of distinct separation suggests that these populations share certain genetic traits, possibly influenced by historical interactions or shared ancestry. Conversely, the improved and landraces populations, marked by their overlap, suggested a degree of genetic affinity existing between the two populations. This overlap may be indicative of shared genetic elements resulting from human selection and cultivation practices. Fig. 1 further displayed distinctive genetic clustering, where a considerable cohort of individuals from the accessions clustered cohesively. In a separate enclave, the wild type was accompanied by some accessions, while none from the improved and landraces populations found themselves in the same group. This clustering pattern suggests a complex interplay of genetic factors, potentially influenced by ecological and historical factors shaping the diversity within sesame populations. Furthermore, the majority of individuals from the landraces and nearly all from the improved populations congregated into a separate cluster. Two individuals from the improved population, in an interesting exception, found their place alongside the accession group. This intricate genetic ballet, as visualized in Fig. 1, uncovers the complex web of genetic relationships and interplay among the diverse sesame populations under scrutiny.

For a more comprehensive understanding of differentiation patterns, each population underwent further analysis based on their collection from administrative regions. In the case of the accession population, 160 samples were utilized out of the total 323, which were gathered from seven distinct administrative regions. The outcomes, depicted in Fig. 2, displayed how genetic variation is distributed among Ethiopian sesame accessions originating from various administrative regions. The principal findings stem from the first three axes of the PCoA analysis, collectively encapsulating the majority of the variation within the dataset. The primary axis, accounting for 9.38% of the variance, possesses the highest capacity for capturing underlying genetic diversity among the accessions. It encompasses a wide spectrum of genetic distinctions, highlighting the principal source of differentiation among the samples. Subsequently, the second axis, contributing an additional 8.68% of the variance, builds upon the first axis by providing further insights into genetic patterns. The third axis, responsible for 5.35% of the variance, becomes more meaningful when combined with the first two axes, contributing to a cumulative variance of 23.41%. Through the PCoA analysis, a comprehensive view emerges, indicating that although certain population subsets within the various sesame populations exhibit distinguishable differences, a clear separation based on their origins isn't immediately evident (as visualized in Fig. 2).

Concerning the landraces, the outcomes of the PCoA analysis, as depicted in Fig. 3, reveal that the initial three axes effectively capture a substantial portion (approximately 48.62%) of the genetic diversity inherent in the landraces of Ethiopian sesame, organized based on their administrative regions. The primary axis, responsible for a notable 27.02% of the total variance, plays a pivotal role in capturing the

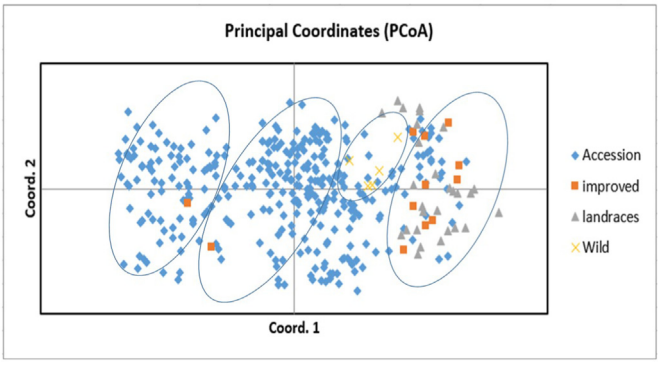

Fig. 1. Principal coordinate analysis of the 368 sesame accessions representing four sesame population groups Each symbol represents one variety from one of the four groups.

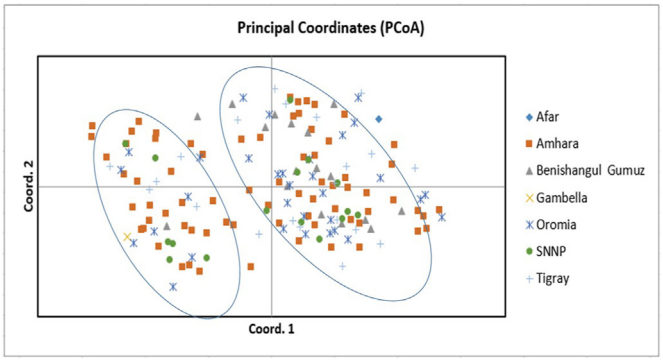

Fig. 2. Principal coordinate analysis of the 160-sesame accession sourced from EBI, which represents seven geographical regions.

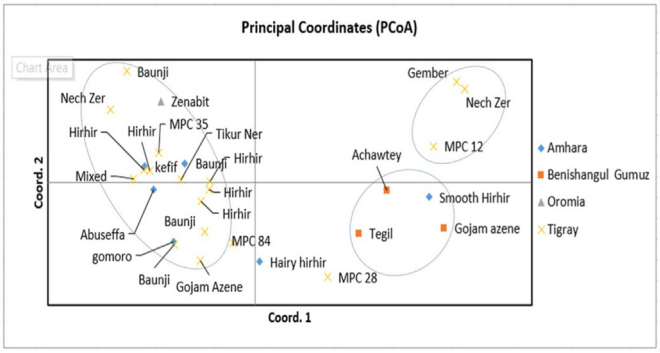

Fig. 3. Principal coordinate analysis of the 28 sesame landraces collected from four geographical regions. Each symbol represents one variety from one of the four studied regions.

fundamental genetic diversity existing among the landraces. Building upon this, the second axis contributes 12.58% of the variance, further delineating the landraces by highlighting region-specific genetic patterns. Moreover, the third axis clarifies 9.03% of the variance and, when combined with the first two axes, contributes to an overall cumulative variance of 48.62%. These axes together offer a meaningful representation of the genetic relationships and distinctions among the diverse landrace populations across various administrative regions. Consequently, they contribute to a deeper comprehension of the genetic makeup of these distinct sesame varieties. As observed in Fig. 3, landraces collected from Benishangul Gumuz, along with a genotype from Amhara, form a distinct subset. In contrast, most of the landraces from the Tigray region group together, constituting another cluster. Additionally, the landraces from the Amhara region overlap with certain genotypes from the Tigray region, implying certain shared genetic attributes between these populations.

Regarding the improved Ethiopian sesame varieties, the PCoA analysis offers compelling insights into the genetic landscape of these varieties released by diverse Agricultural Research Institutes across various administrative regions within Ethiopia (Amhara, Oromia, Tigray) (Fig. 4). The cumulative impact of the first three axes sheds light on a significant portion of the genetic diversity inherent in these varieties. The cumulative percentages provide a lucid perspective on the extent of genetic variation that these three axes elucidate. In this particular case, these axes harmoniously account for 59.85% of the total genetic diversity found within the dataset. Axis 1, with its 28.78% explanation of genetic

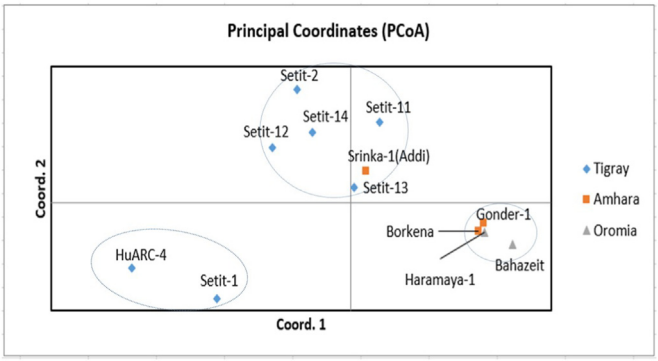

Fig. 4. Principal coordinate analysis of the 12 sesame improved varieties released by four research institutes: Humera Agricultural Research Center (HuARC) located in Tigray region, Haramay Univeristy (HU) situated in Oromia region, Gondar Agricultural Research Center (GARC) and Srinka Agricultural Research Center both of which are located in the Amhara region.
(NB: Setit-12 = Setit-1_1, Setit-12 = , Setit-1_2 = , Setit-13 = , Setit-1_3, Setit-14 = Setit-1_4).

variance, encapsulates the paramount and most influential origin of genetic diversity within the improved sesame varieties sourced from varied research institutes in different administrative regions. This primary axis extracts and highlights fundamental distinctions that define the genetic makeup of these varieties, driven by their unique regional origins. In parallel, Axis 2 contributes a substantial 18.43% towards elucidating genetic variability. The collective influence of the first two axes, totaling 47.20%, underscores that a significant portion of the overall genetic diversity can be attributed to these axes. This suggests the presence of specific and distinct genetic signatures that set apart the improved sesame varieties from different administrative regions, signifying unique patterns inherent to each region's breeding efforts. Furthermore, Axis 3, accounting for 12.64% of genetic variance, provides deeper insights into the nuances of genetic differences manifested among the improved sesame varieties.

The illustrative outcome of the PCoA analysis, as visualized in Fig. 4, underscores the distinctive genetic clustering patterns of the improved sesame varieties. These patterns indicate the discernible connections and disparities among the sesame varieties released by various national and regional research institutes within different administrative regions of Ethiopia. Specifically, two varieties from Amhara and two from Oromia manifest as a distinctive subset, forming a distinct and separate cluster. This clustering implies a level of genetic affinity shared between these sesame varieties from these respective regions. Similarly, the separate cluster formed by five varieties from Tigray and one from Amhara signifies genetic similarities among these varieties. Additionally, two varieties from the Tigray region manifest yet another separate cluster, indicating distinct genetic traits exclusive to this region.

3.5. *Analysis of Molecular variance reveals distinct genetic diversity in Ethiopian sesame populations*

A comprehensive summary of the outcomes derived from analysis of molecular variance (AMOVA) applied to the diverse origins and sources of Ethiopian sesame germplasm is encapsulated in Table 5. The major findings of this analysis distinctly establish the presence of statistically significant genetic variability ($p < 0.001$) among the four distinct Ethiopian sesame populations. This statistical significance reinforces the notion that discernible genetic differentiations exist across accessions, improved varieties, landraces, and wild types.

Approximately 8% of the entire genetic diversity is attributable to distinctions among these populations, underscoring the existence of notable genetic variations. This observed variability not only affirms the unique genetic makeup of each population but also emphasizes that these differences are statistically significant. It suggests that the diverse sesame populations in Ethiopia exhibit clear genetic distinctions, providing a foundation for understanding the population structure and facilitating targeted breeding efforts. Furthermore, the analysis highlights that the larger proportion of genetic diversity (92%) resides within each individual sesame population. This intriguing finding underscores the richness of genetic diversity within each group—accessions, improved varieties, landraces, and wild types. The notably heightened genetic diversity within populations signifies the presence of diversity within every group, indicating that each population harbors unique genetic traits and variants.

3.6. *Population structure*

The analysis involved the use of a Bayesian clustering model-based method to study genotypic data from 28 polymorphic markers and 368 sesame genotypes. In order to accurately classify genetic groups, the Δk value was employed. Importantly, the highest Δk value, observed at $k = 2$ (Fig. 5), indicated the presence of two primary genetic clusters, namely Cluster 1 (G1) and Cluster 2 (G2), within the 368 sesame germplasms. For a more comprehensive understanding of differentiation based on specified origins or sources, the primary clusters were further divided

Table 5
Analysis of Molecular Variance (AMOVA) between the difference sesame origins.

Source	df	SS	MS	Est. Var.	%	Value	P (rand ≥ data)
Among Pops	3	303.749	101.250	2.643	8%	0.083	0.001
Within Pops	364	10589.129	29.091	29.091	92%		
Total	367	10892.878		31.734	100%		

Est.Var. = estimated variance.

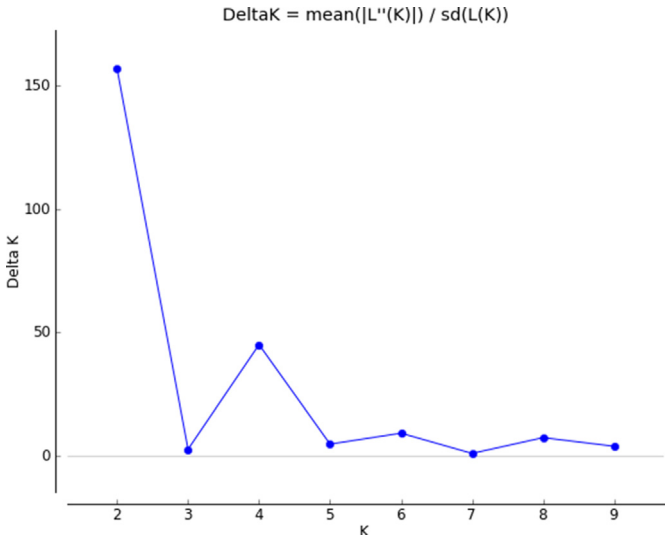

Fig. 5. Estimated Δk of the 204 sesame accessions over 10 runs for each K value.

into four sub-populations (P1–P4) using STRUCTURE analysis at k = 4. This choice emerged as the most suitable number of sub-populations following the initial division at k = 2. It was also noted, as shown in Fig. 6, that at Δk = 6 and Δk = 8, exploring additional subdivisions of the main clusters into six (P1–P6) and eight (P1–P8) sub-populations could be considered.

The assessment of membership probability (≥0.6) unveiled that 94.3% (347) of the 368 germplasms could be confidently assigned to specific groups, highlighting clear genetic distinctions (Table 6). However, 19 genotypes did not exhibit clear membership to any particular group, emphasizing the complexity of certain genetic profiles (Table S4). Notably, nearly all total germplasms (367 out of 368) demonstrated a probability of membership ≥0.6 within the context of k = 2 (Table S4), indicating clear differentiation among distinct groups. Only one genotype, identified as an admixture, displayed a membership probability below 0.6 and was shared between the two primary groups, suggesting a degree of genetic overlap (see Table 7).

The Fixation Index (Fst) values provided valuable insights into genetic differentiation. A Fst value of 0.34 associated with Cluster 2 indicated a moderate degree of genetic differentiation, while the lower Fst value of 0.26 in Cluster 1 suggested a relatively lower level of genetic differentiation. The lower Fst value in Cluster 1 hinted at greater genetic similarities and heightened gene flow. In contrast, the higher Fst value in Cluster 2 indicated pronounced genetic distinctions, contributing significantly to overall genetic variation.

For deepen understanding, an assessment of average genetic distance among the subpopulations was detailed in Table 8. The observed average genetic distances among individuals within the same sub-population were relatively modest, manifesting in the low values of average distances (expected heterozygosity). However, it is noteworthy that P1 and P2 demonstrated relatively higher average distances when contrasted with P2 and P3. The mean Fst values, which gauge the extent of genetic differentiation across sub-populations, spanned a range from 0.12 to 0.49. These values indicated a range of moderate to high degrees of genetic differentiation among the subpopulations (Table 8). Notably, P1 displayed the most pronounced Fst value (0.49) within the sub-populations, signifying a significant genetic differentiation within the populations that compose this sub-population. In contrast, P2 showcased a lower mean Fst value (0.12) in comparison to the other sub-populations (Table 8). Meanwhile, P3 and P4 exhibited intermediate Fst values, indicative of a moderate level of genetic differentiation.

The analysis of the proportion of individuals in each cluster (% membership) revealed significant variability, with Sub-Population P2 standing out with the highest proportion at 37%, while Sub-Population P1 exhibited the lowest proportion at 17% (Table 8). A closer examination of these sub-populations showed distinct patterns and shed light on the genetic composition and distribution within the Ethiopian sesame germplasm. The sub-population P1 was characterized by a relatively modest overall membership, predominantly comprising individuals sourced from the accession groups. Notably, it shows limited representation from the improved and landraces groups and excludes individuals from the wild group (Table 5). The findings suggest a focused genetic makeup predominantly associated with specific cultivation practices, potentially reflecting a distinct set of genetic traits within the accession group. In contrast, Sub-Population P2 stands out with the highest total membership count and minimal genetic divergence. It is exclusively composed of individuals from the accession group, making it the largest sub-population within the analysis (Table 5). This dominance of accession individuals implies a strong genetic homogeneity within this sub-population, possibly influenced by common breeding or cultivation practices. Comprising individuals originating from accession, Sub-Population P3 underscores the prevalence of genetic elements associated with the accession group. This sub-population adds to the understanding of the distinct genetic patterns within the accession category. On the other hand, sub-Population P4, revealed a diverse membership composition, encompassing individuals from all the different groups—accessions, landraces, improved, and even the wild group (Table 5). The significant concentration of accession and landraces representation within P4 suggests the preservation of traditional cultivars and a noteworthy blend of genetic diversity. Additionally, individuals from the "Improved" group and the "Wild" group are well represented, indicating a broader genetic spectrum within this sub-population.

3.7. Cluster analysis of 368 sesame germplasm and four sesame population groups

The cluster analysis conducted on sesame populations categorized them into four distinct groups: accessions, landraces, improved varieties, and wild types. The methodology employed Neighbor-joining (NJ) analysis using 28 primers, visually depicted in Fig. 7, revealing two prominent clusters labeled as (I) and (II), each further branching into several sub-clusters. Cluster (I) encompassed 46.7% of the sesame germplasm, prominently featuring about 75% of the improved varieties, 80% of the wild types, and an impressive 92.9% of the landraces. Within Cluster (I), a single sub-cluster gathered nearly all wild type specimens, a significant portion of sesame landraces, and a notable subset of improved cultivars. On the other hand, Cluster (II) comprised 53.3% of the entire sesame germplasm pool, further divided into three sub-clusters. Notably, a substantial 96.9% of sesame accessions sourced from the Ethiopian Biodiversity Institute were primarily affiliated with Cluster (II). This comprehensive cluster analysis effectively partitioned the sesame

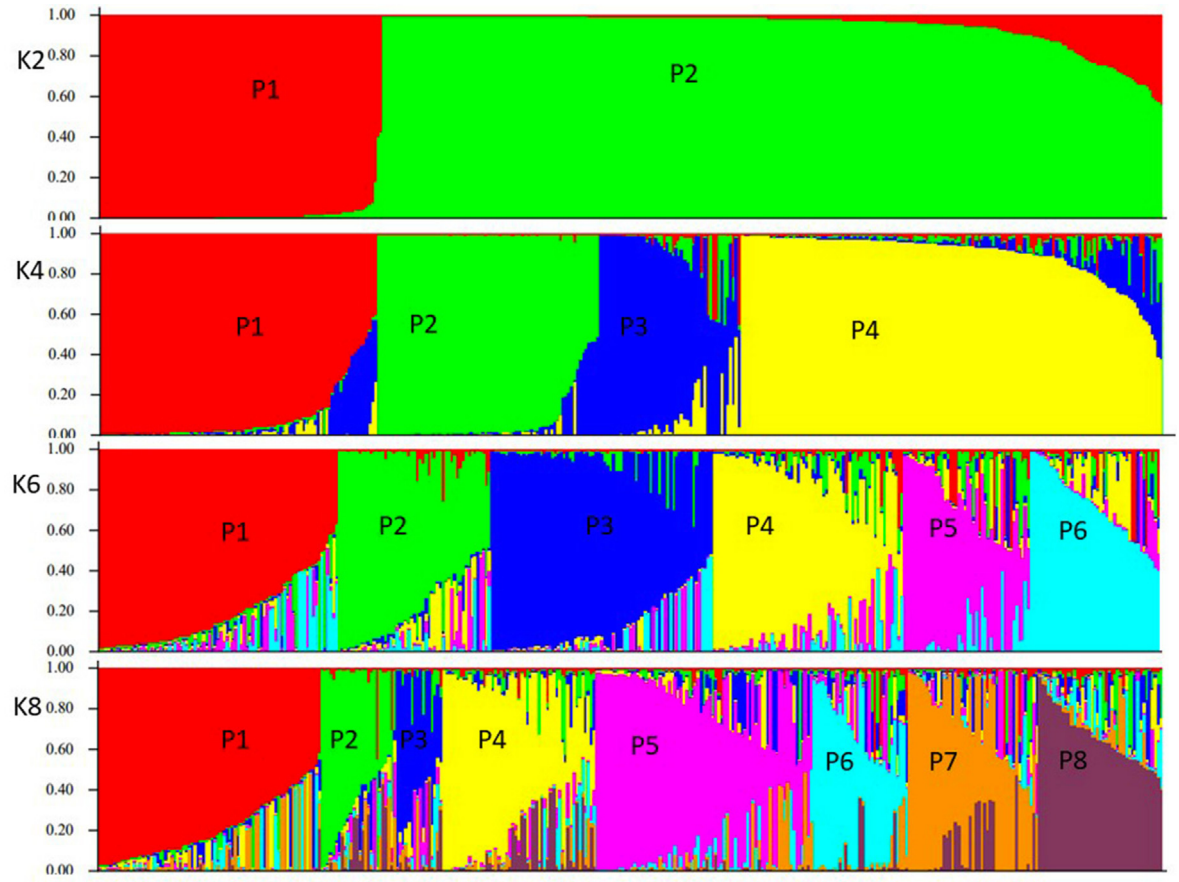

Fig. 6. Estimated population structure of sesame accessions assessed by STRUCTURE. A thin vertical bar, partitioned into up to k colored segments, represents each individual.

Table 6
Summary of proportion of membership of sesame population to each subpopulation.

Clustering based on Δk = 2					Clustering based on Δk = 4				
S/N	Administrative region	Group (clusters)			Sub-populations				
		G1	G2	Total	P1	P2	P3	P4	Total
1	Accession	95	227	**322**	40	140	88	36	**304**
2	Improved	2	10	**12**	2	0	2	8	**12**
3	Landraces	0	28	**28**	2	0	0	24	**26**
4	Wild	0	5	**5**	0	0	0	5	**5**
Total		**97**	**270**	**367**	**44**	**140**	**90**	**73**	**347**

Table 7
Evaluation of Genetic Divergence (net nucleotide distance), Proportion of Membership, Within-Population Expected Heterozygosity, and Mean Fst Values in the Study of Population Structure of 368 Sesame Accessions and Genotypes Utilizing 28 Markers based on Δk = 2.

Population	net nucleotide distance		% membership	expected heterozygosity	Mean fixation index (Fst)
	P1	P2			
P1		0.013	0.299	0.11	0.27
P2	0.013		0.701	0.10	0.34

populations based on their molecular characteristics and origins, offering insights into the genetic relationships and diversity among different cultivars. The genetic codes representing each genotype in Fig. 7 were thoughtfully provided for clarity in Table S1.

Additionally, employing the unweighted pair group method with arithmetic mean (UPGMA), another cluster analysis was conducted on the four sesame populations categorized by their origins and sources. This analysis revealed three primary clusters, with wild sesame types and accessions occupying separate clusters, indicating substantial genetic differentiation between these two groups. In contrast, both improved varieties and landraces were grouped within a common cluster, suggesting a closer genetic relationship between them (Supplementary Fig. 1). This observation aligns with Euclidean distance findings (Table 4), emphasizing that the wild sesame population exhibits more pronounced divergence compared to other population types, while both accessions and improved varieties show relatively shorter genetic distances when compared with the accession and wild populations.

4. Discussion

The comprehensive analysis of genetic diversity in Ethiopian sesame populations from four distinct breeding groups, namely Accessions, Landraces, Improved varieties, and Wild types, revealed several key

Table 8
Evaluation of Genetic Divergence (net nucleotide distance), Proportion of Membership, Within-Population Expected Heterozygosity, and Mean Fst Values in the Study of Population Structure of 368 Sesame Accessions and Genotypes Utilizing 28 Markers based on $\Delta k = 4$.

Population	net nucleotide distance				% membership	Expected heterozygosity	Mean fixation index (Fst)
	P1	P2	P3	P4			
P1		0.02	0.03	0.03	0.17	0.11	0.49
P2	0.02		0.01	0.01	0.37	0.07	0.12
P3	0.03	0.01		0.02	0.25	0.09	0.21
P4	0.03	0.01	0.02		0.21	0.11	0.32

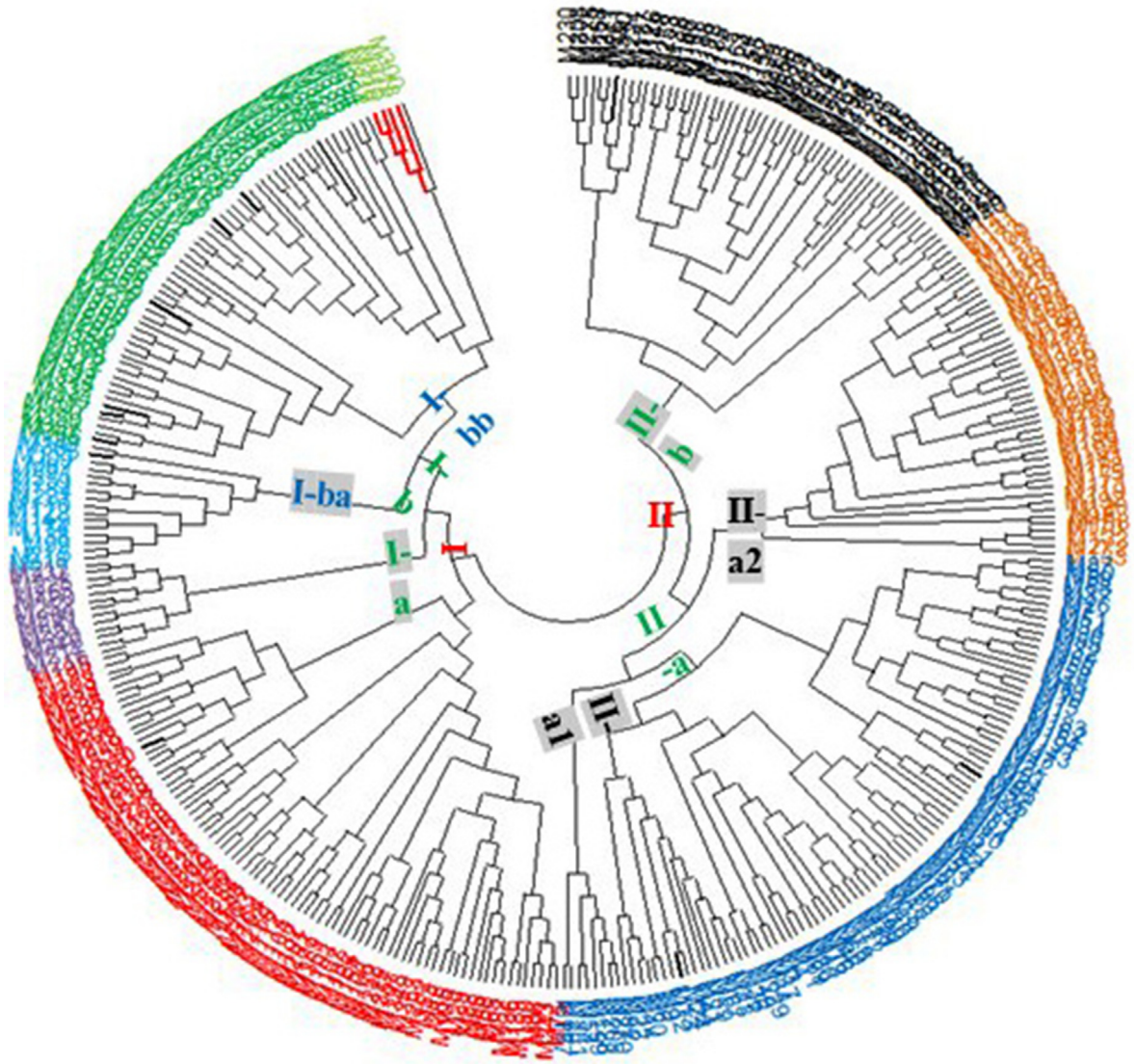

Fig. 7. Dendrogram based on Neighbor-joining (NJ) method showing the genetic relationship among 368 sesame entries using 28 markers.

findings and insights. Firstly, the choice of 28 markers, including 23 SSR and five InDel markers, provided a robust and cost-effective means to assess genetic diversity (Dossa et al., 2016; Khlestkina et al., 2004). These markers exhibited variations in parameters such as Major Allele Frequency (MAF), allele count, gene diversity, observed heterozygosity (Ho), expected heterozygosity (He), and Polymorphic Information Content (PIC). The variability in marker characteristics which was consistent with prior research findings (Dossa et al., 2016; Teklu et al., 2022; Wei et al., 2015) likely reflects differences in selective pressures, mutation rates, or recombination rates across the genome (Tiffin and Ross-Ibarra, 2014). Notably, the number of alleles per locus in this study exceeded those reported in previous research (Cho et al., 2011; Dossa et al., 2016; Wei et al., 2014), indicating the richness of genetic diversity in the sampled populations. Additionally, the MAF values in our study were notably higher compared to the reports of Dossa et al. (2016) and Wei et al. (2014), and relatively consistent with the findings of Teklu et al. (2022), who examined 100 accession genotypes using 27 SSR markers. Regarding the PIC value, our study yielded higher results than Teklu et al. (2022) but lower values compared to Cho et al. (2011), Dossa et al. (2016), and Wei et al. (2014). These observed disparities among studies could be attributed to differences in accession selection, sampling methodologies, and the number of markers utilized (Dossa et al., 2016;

Tesfaye et al., 2022).

Remarkably, the high percentage of polymorphic loci (82.4%) within the accessions category reflects a substantial level of genetic diversity present in this group. This diversity likely arises from the inclusion of diverse source materials, resulting in a wide array of genetic variations. Conversely, the relatively modest percentage of polymorphic loci (31.2%) observed in the improved varieties group suggests that this population may have undergone selective breeding or genetic enhancement, potentially leading to a reduction in genetic diversity (McLean--Rodriguez et al., 2021). Improved varieties often result from targeted breeding for specific traits, which can result in a decrease in overall genetic variation. Similarly, the landraces exhibit a relatively modest percentage of polymorphic loci (29.6%). This trend may be attributed to the traditional cultivation and conservation practices applied to landraces. These practices tend to limit genetic exchange, contributing to the observed decrease in genetic diversity within this group (McLean-Rodríguez et al., 2022; Van de Wouw et al., 2010). In contrast, the wild type sesame populations display an exceptionally low percentage of polymorphic loci (8.0%), indicating a significant lack of genetic diversity within this category. Factors such as isolation or historical bottleneck events could potentially explain this reduced genetic variability over time (Bohra et al., 2022; Tirnaz et al., 2022).

The overall scope of the diversity analysis suggests that the genetic diversity within Ethiopian sesame populations can be classified as ranging from moderate to high. This classification aligns with prior research using various marker types, including SSR, SNP, and ISSR markers (Dossa et al., 2016; Teklu et al., 2022; Tesfaye et al., 2022; Wei et al., 2015; Woldesenbet et al., 2015). Previous studies have consistently reported moderate to high levels of genetic diversity in Ethiopian sesame populations (Teklu et al., 2022; Woldesenbet et al., 2015). Notably, regional variations in genetic diversity have been observed, with some areas exhibiting higher diversity than others (Woldesenbet et al., 2015). These findings collectively emphasize the rich genetic diversity within Ethiopian sesame populations, with implications for breeding, conservation, and agricultural practices. Understanding the extent of genetic diversity can guide breeding strategies and help preserve unique genetic resources. Additionally, regional variations highlight the influence of environmental factors and cultivation practices on genetic diversity.

PCoA conducted in this study has provided valuable insights into the complex patterns of genetic diversity and phylogenetic relationships among various Ethiopian sesame populations. While clear distinctions were observed within specific subsets of the accessions group, no strict separation based on population origin or breeding groups was evident. This suggests that genetic variation in Ethiopian sesame populations is not solely determined by their breeding history or geographical origin. Notably, overlaps were identified between the improved, landraces, and wild variant populations, with the accessions population, indicating the presence of shared genetic traits among these groups. These findings align with previous reports of overlapping genetic characteristics among sesame populations from diverse origins (Gebremichael and Parzies, 2011). Of particular interest is the observed overlap between the improved and landraces populations, suggesting a notable degree of genetic similarity between these two groups. In contrast, the wild type sesame population formed a distinct cluster, with certain accessions showing similarity to the wild type. Additionally, clear genetic differentiation emerged when populations were categorized based on their administrative regions, indicating the influence of regional concentrations of sesame cultivation on genetic diversity. Factors such as environmental conditions, cultivation practices, and the presence of distinct sesame varieties, such as Humera type, Gondar type, and Wollega type which represents the major sesame seed-producing regions in Tgray, Amhara and Ormoia of Ethiopia respectively (Zerihun, 2012). According to this author, each of these varieties possesses unique traits that render them suitable for various applications such as baking or oil production. For instance, the Humera variety, commonly cultivated in Tigray, is renowned for its uniform white seeds, aroma, and sweetness, making it ideal for bakery products. However, it is susceptible to shattering losses that can impact yields. The Gondar variety, grown in Amhara, is also fit for the bakery market but may harbor different genetic attributes than the Humera variety, thus contributing to its distinct genetic profile. The Wollega variety, prevalent in Oromia, boasts smaller seeds, elevated oil content, and reduced sweetness, rendering it advantageous for oil production. The presence of these unique sesame varieties, each characterized by distinct genetic traits, likely contributes to the observed genetic variation within sesame populations from the Amhara, Oromia, and Tigray regions. Additionally, differences in seed selection, and other human interventions can also contribute to genetic diversity (Van de Wouw et al., 2010). These findings highlight the complex interplay between genetics and environmental factors in shaping the genetic diversity of sesame populations which is resulted in variations in their DNA (Dossa et al., 2016).

AMOVA emphasized significant genetic distinctions among population origins and sources. It revealed that approximately 92% of the overall genetic diversity resides within individual sesame populations, emphasizing the wealth of genetic variability within each specific group. This observation magnifies the inherent richness of genetic diversity within each specific group. The heightened diversity within populations suggests that accessions, improved varieties, landraces, and wild types each encompass a broad array of genetic traits. This augmented proportion of genetic diversity within every distinct population accentuates the wealth of genetic variability inherent to Ethiopian sesame populations. This aspect is of notable value as a foundational resource for breeding endeavors and preservation initiatives. However, it is pertinent to note that a separate study reported greater genetic diversity among Ethiopian sesame populations (Admas et al., 2013). Possible explanations for the comparatively lower level of genetic differentiation among populations in this study encompass factors like elevated gene flow or shared ancestry (Dossa et al., 2016; Tesfaye et al., 2022). Additionally, the sample size utilized in this analysis might not have been sufficient to detect significant genetic disparities among populations. However, the observed significant genetic variability among populations ($P < 0.001$), coupled with the patterns identified in the PCoA analysis, accentuates the coexistence of both shared attributes and distinctive genetic characteristics within and across these populations.

The Bayesian clustering method (Pritchard et al., 2010) derived from the structure analysis and the cluster analysis employed the Neighbor-Joining (NJ) method (Yeh, 1997) revealing two main clusters that further branched into multiple sub-clusters. The alignment between the cluster analysis and structure analysis in this study yields complementary insights into the genetic makeup of Ethiopian sesame populations. The concurrence between these approaches underscores the effectiveness of the methodologies utilized in capturing the intricate genetic relationships and distinctions among the populations. This congruence bolsters the credibility of the findings and enhances our confidence in discerning these observed patterns. Evaluating these analyses collectively provides a more comprehensive understanding of the distinct genetic groups, subpopulations, and the broader genetic diversity inherent within Ethiopian sesame populations.

5. Conclusion and recommendations

In light of the findings concerning the genetic diversity and population structure of 368 Ethiopian sesame genotypes using a comprehensive set of 28 markers encompassing the entire genome, several key takeaways emerge. The variability observed in genetic parameters among these markers implies that specific markers may offer greater diversity insights within populations. In a broader context, the genetic diversity within Ethiopian sesame populations can be classified as moderate to high, and markers displaying heightened variability may prove to be particularly valuable for genetic research and practical applications. Consequently, this study lays the groundwork for future investigations, emphasizing the importance of selecting markers characterized by

substantial diversity, thus increasing the likelihood of uncovering significant genetic distinctions among individuals or populations. Based on the results of this study on the genetic diversity and population structure of Ethiopian sesame germplasm, some future research directions can be recommended.

- Firstly, research can be conducted to identify the origin and domestication history of Ethiopian sesame populations. This could involve the use of genomic data and comparative analyses with wild sesame populations and other cultivated sesame varieties from different regions.
- Finally, efforts can be made to conserve and sustainably utilize the genetic resources of Ethiopian sesame populations. This could involve the establishment of gene banks and conservation programs to preserve the genetic diversity of these populations at each region, as well as the development of strategies for the utilization of this diversity in breeding programs and crop improvement.

Funding

The research work presented in this submission has been supported by a diverse array of funding sources, each of which has played a crucial role in advancing our research initiatives. We hereby disclose the following funding sources and delineate their respective contributions:

This research was funded by the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2021-OCRI), China Agriculture Research System (CARS-14), Hubei International Science and Technology Cooperation Project (2022EHB034), the Science and Technology Innovation Project of Hubei Province (2021-620-000-001-035), and the Fundamental Research Funds for Central Non-profit Scientific Institution (2013).

Conflicts of interest

We maintain our unwavering commitment to complete transparency and declare the absence of any conflicts of interest associated with any of the funding sources. Linhai Wang is an Editorial Member for OIL CROP SCIENCE and was not involved in the editorial review or the decision to publish this article.

CRediT authorship contribution statement

Muez Berhe: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jun You:** Writing – review & editing, Software, Conceptualization. **Komivi Dossa:** Writing – review & editing, Conceptualization. **Fetien Abay Abera:** Writing – review & editing. **Emmanuel Amponsah Adjei:** Writing – review & editing, Conceptualization. **Yanxin Zhang:** Writing – review & editing, Software, Methodology, Conceptualization. **Linhai Wang:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization.

Acknowledgements

This research was funded by the Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2021-OCRI), China Agriculture Research System (CARS-14), Hubei International Science and Technology Cooperation Project (2022EHB034), the Science and Technology Innovation Project of Hubei province (2021-620-000-001-035) and the Fundamental Research Funds for Central Non-profit Scientific Institution (2013).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ocsci.2023.11.003.

References

Abate, M., Mekbib, F., Ayana, A., Nigussie, M., 2015. Assessment of genetic diversity in Ethiopian sesame (Sesamum indicum L.) germplasm using ISSR markers. Br. Biotechnol. J. 8, 1–13.

Admas, A., Yohannes, P., Kassahun, T., 2013. Genetic distance of sesame (Sesamum indicum L.) cultivars and varieties from northwestern Ethiopia using inter simple sequence repeat markers. East African Journal of Sciences 7 (1), 31–40.

Ahmar, S., Gill, R.A., Jung, K.-H., Faheem, A., Qasim, M.U., Mubeen, M., Zhou, W., 2020. Conventional and molecular techniques from simple breeding to speed breeding in crop plants: recent advances and future outlook. Int. J. Mol. Sci. 21 (7), 2590.

Akbar, F., Rabbani, M.A., Masood, M.S., Shinwari, Z.K., 2011. Genetic diversity of sesame (Sesamum indicum L.) germplasm from Pakistan using RAPD markers. Pakistan J. Bot. 43 (4), 2153–2160.

Anilakumar, K.R., Pal, A., Khanum, F., Bawa, A.S., 2010. Nutritional, medicinal and industrial uses of sesame (Sesamum indicum L.) seeds-an overview. Agric. Conspectus Sci. 75 (4), 159–168.

Baraki, F., Berhe, M., 2019. Evaluating performance of sesame (Sesamum indicum L.) genotypes in different growing seasons in Northern Ethiopia 2019. Int. J. Agron. 2019, 1–7.

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.

Bedigian, D., 1981. Origin, diversity, exploration and collection of sesame. FAO Plant Prod. Protect. Pap. 12, 116–164. *no. 29*.

Berhe, M., Subramanyam, B., Demissie, G., Chichaybelu, M., Abera, F.A., Mahroof, R., Harvey, J., 2023. Effect of storage duration and storage technologies on pest infestations and post-harvest quality loss of stored sesame seeds in Ethiopia. J. Stored Prod. Res. 103, 102161.

Bohra, A., Kilian, B., Sivasankar, S., Caccamo, M., Mba, C., McCouch, S.R., Varshney, R.K., 2022. Reap the crop wild relatives for breeding future crops. Trends Biotechnol. 40 (4), 412–431.

Cho, Y.-I., Park, J.-H., Lee, C.-W., Ra, W.-H., Chung, J.-W., Lee, J.-R., Lee, M.-C., 2011. Evaluation of the genetic diversity and population structure of sesame (Sesamum indicum L.) using microsatellite markers. Genes & genomics 33, 187–195.

CSA, 2020. He federal democratic republic of Ethiopia central statistical agency: agricultural sample survey 2018/19 (2011 E.C.): report on area and production for major crops (private peasant holdings, meher season)., CSA, addis ababa, Ethiopia. Statistical Bulletin No. 586 I (69).

Cui, C., Mei, H., Liu, Y., Zhang, H., Zheng, Y., 2017. Genetic diversity, population structure, and linkage disequilibrium of an association-mapping panel revealed by genome-wide SNP markers in sesame. Front. Plant Sci. 8, 1189.

De Candolle, A., 1892. In: Origin of Cultivated Plants, 2 ed., vol. 48. Appleton, New York, NY, USA.

Diriba, G., 2018. Agricultural and rural transformation in Ethiopia. Ethiopian Journal of Economics 27 (2), 51–110.

Dossa, K., Wei, X., Zhang, Y., Fonceka, D., Yang, W., Diouf, D., Zhang, X., 2016. Analysis of genetic diversity and population structure of sesame accessions from Africa and Asia as major centers of its cultivation. Genes 7 (4), 14.

Earl, D.A., VonHoldt, B.M., 2012. Structure harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4, 359–361.

El Harfi, M., Charafi, J., Houmanat, K., Hanine, H., Nabloussi, A., 2021. Assessment of genetic diversity in Moroccan sesame (Sesamum indicum) using ISSR molecular markers. OCL 28, 3.

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

FAOSTAT, 2020. Statistical Databases: Food and Agriculture Organization of the United Nations: Rome, Italy. https://doi.org/10.1016/s0308-521x(01)00076-2. Available from: Online: http://Www.Fao.Org/Faostat/En/#data/QC. (Accessed 10 September 2023).

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

Holland, M.M., Parson, W., 2011. GeneMarker® HID: a reliable software tool for the analysis of forensic STR data. J. Forensic Sci. 56 (1), 29–35.

Kassie, G.T., Worku, Y., Bachewe, F.N., Asnake, W., Abate, G.T., 2023. Scoping Study on Ethiopian Sesame Value Chain, vol. 3. Intl Food Policy Res Inst.

Khlestkina, E., Huang, X., Quenum, F.J.-B., Chebotar, S., Röder, M., Börner, A., 2004. Genetic diversity in cultivated plants—loss or stability? Theor. Appl. Genet. 108, 1466–1472.

Kobayashi, T., Kinoshita, M., Hattori, S., Ogawa, T., Tsuboi, Y., Ishida, M., Saito, H., 1990. Development of the sesame metallic fuel performance code. Nucl. Technol. 89 (2), 183–193.

Kefale, H., Wang, L., 2022. Discovering favorable genes, QTLs, and genotypes as a genetic resource for sesame (Sesamum indicum L.) improvement. Front. Genet. 13, 1002182.

Lee, J., Lee, Y., Choe, E., 2008. Effects of sesamol, sesamin, and sesamolin extracted from roasted sesame oil on the thermal oxidation of methyl linoleate. LWT–Food Sci. Technol. 41 (10), 1871–1875.

Liu, K., Muse, S.V., 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21 (9), 2128–2129.

Mahajan, R., Bisht, I., Dhillon, B., 2007. Establishment of a core collection of world sesame (Sesamum indicum L.) germplasm accessions. Sabrao Journal of Breeding and Genetics 39 (1), 53–64.
McLean-Rodríguez, F.D., Camacho-Villa, T.C., Almekinders, C.J., Pè, M.E., Dell'Acqua, M., Costich, D.E., 2022. The abandonment of maize landraces over the last 50 years. In: Morelos, Mexico: a Tracing Study Using a Multi-Level Perspective *Rethinking Food System Transformation*. Springer, pp. 1–18.
McLean-Rodriguez, F.D., Costich, D.E., Camacho-Villa, T.C., Pe, M.E., Dell'Acqua, M., 2021. Genetic diversity and selection signatures in maize landraces compared across 50 years of in situ and ex situ conservation. Heredity 126 (6), 913–928.
Park, J.-H., Suresh, S., Cho, G.-T., Choi, N.-G., Baek, H.-J., Lee, C.-W., Chung, J.-W., 2014. Assessment of molecular genetic diversity and population structure of sesame (Sesamum indicum L.) core collection accessions using simple sequence repeat markers. Plant Genetic Resources 12 (1), 112–119.
Peakall, R., Smouse, P.E., 2006. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6 (1), 288–295.
Pritchard, J.K., Wen, X., Falush, D., 2010. Documentation for Structure Software: Version 2.3. University of Chicago, Chicago, IL, pp. 1–37.
Sharma, A., Sharma, S., Kumar, N., Rana, R.S., Sharma, P., Kumar, P., Rani, M., 2022. Morpho-molecular genetic diversity and population structure analysis in garden pea (Pisum sativum L.) genotypes using simple sequence repeat markers. PLoS One 17 (9), e0273499.
Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. 38 (7), 3022–3027.
Teklu, D.H., Shimelis, H., Tesfaye, A., Mashilo, J., Zhang, X., Zhang, Y., 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 (6), 1129.
Teklu, D.H., Shimelis, H., Tesfaye, A., Shayanowako, A.I.T., 2022. Analyses of genetic diversity and population structure of sesame (Sesamum indicum L.) germplasm collections through seed oil and fatty acid compositions and SSR markers. J. Food Compos. Anal. 110, 104545.
Tesfaye, T., Tesfaye, K., Keneni, G., Ziyomo, C., Alemu, T., 2022. Genetic diversity of Sesame (Sesamum indicum L) using high throughput diversity array technology. Journal of Crop Science and Biotechnology 25 (4), 359–371.
Tiffin, P., Ross-Ibarra, J., 2014. Advances and limits of using population genetics to understand local adaptation. Trends Ecol. Evol. 29 (12), 673–680.
Tirnaz, S., Zandberg, J., Thomas, W.J., Marsh, J., Edwards, D., Batley, J., 2022. Application of crop wild relatives in modern breeding: an overview of resources, experimental and computational methodologies. Front. Plant Sci. 13, 1008904.
Uzun, B., Lee, D., Donini, P., Çağirgan, M., 2003. Identification of a molecular marker linked to the closed capsule mutant trait in sesame using AFLP. Plant Breed. 122 (1), 95–97.
Van de Wouw, M., Kik, C., van Hintum, T., van Treuren, R., Visser, B., 2010. Genetic erosion in crops: concept, research results and challenges. Plant Genetic Resources 8 (1), 1–15.
Wang, L., Yu, S., Tong, C., Zhao, Y., Liu, Y., Song, C., Hua, W., 2014. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol. 15 (2), 1–13.
Wei, Liu, K., Zhang, Y., Feng, Q., Wang, L., Zhao, Y., Zhu, X., 2015. Genetic discovery for oil production and quality in sesame. Nat. Commun. 6 (1), 8609.
Wei, Wang, L., Zhang, Y., Qi, X., Wang, X., Ding, X., Zhang, X., 2014. Development of simple sequence repeat (SSR) markers of sesame (Sesamum indicum) from a genome survey. Molecules 19 (4), 5150–5162.
Wei, L.-B., Zhang, H.-Y., Zheng, Y.-Z., Miao, H.-M., Zhang, T.-Z., Guo, W.-Z., 2009. A genetic linkage map construction for sesame (Sesamum indicum L.). Genes & Genomics 31, 199–208.
Weiss, E.A., 2000. In: Oilseed Crops, 2 ed. Blackwell Science, Malden, MA, USA.
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 (2), 7–18.
Yeh, F., 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg. J. Bot. 129, 157.
Zerihun, J., 2012. Sesame (sesame indicum L.) crop production in Ethiopia: trends, challenges and future prospects. Russ. J. Genet. 1 (3), 1–7.