Plant Molecular Biology Reporter (2025) 43:180–196
https://doi.org/10.1007/s11105-024-01476-z

RESEARCH

Molecular Diversity Studies and Core Development in Sesame Germplasm (*Sesamum indicum* L.) Using SSR Markers

Rasna Maurya[1] · Shivani Singh[1] · Yangala Sudheer Babu[1] · Fatima Nazish Khan[1] · Bhagwat Nawade[1] · Harinder Vishwakarma[1] · Ajay Kumar[1] · Rashmi Yadav[1] · Radhamani Jalli[1] · Mahalingam Angamuthu[2] · Mothilal Alagirisamy[2] · Rajkumar Subramani[1] · Senthilraja Govindasamy[2] · Ashok Kumar[1] · Kuldeep Singh[1,4] · Parimalan Rangan[1,3]

Received: 17 January 2023 / Accepted: 15 July 2024 / Published online: 23 July 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Sesame (*Sesamum indicum* L.), an ancient oilseed crop being cultivated across geographical locations in the tropics, is known for its high-quality oil with a longer shelf life. India, being the center of diversity for this crop, understanding the genetic variability of sesame germplasm being conserved in the national Genebank (NGB) of ICAR-NBPGR, will help identify genotypes for its potential use in broadening the genetic base of the cultivars for sesame crop improvement. We report here the molecular diversity analysis performed using SSR markers on a set of 2,496 sesame germplasm. Hence, the derived data was also subjected to population structure analysis, and a molecular core was generated to assess its phenotypic variability. Parallelly, they were phenotypically characterized for important qualitative and quantitative traits as per the standard descriptor developed by IPGRI, and accessions exhibiting desirable traits were identified. The sesame germplasm used in our study represents collections from 17 countries across the globe and 26 states in India. A total of 140 alleles were obtained using seven polymorphic SSR markers selected from an initial screening comprising 43 SSR markers. The observed heterozygosity was less than the expected heterozygosity since it is a self-pollinated crop (up to 35% outcrossing is reported, categorized as often cross-pollinated). The molecular diversity analysis grouped 2496 accessions into six clusters, while the population structure analysis grouped them into three major clusters or populations. A molecular core developed using the PowerCore software identified 196 accessions, representing all the alleles from the entire 2496 accessions, that can be utilized in breeding programs after phenotypic validation. This study contributes to genetic diversity assessment for sesame germplasm, identifying genetically diverse accessions, and establishing a core set that encapsulates the genetic variability of the sesame germplasm collection. These findings hold relevance for addressing agricultural challenges and enhancing the resilience and productivity of sesame crops in various environmental conditions.

Keywords Sesame · SSR markers · Genetic diversity · Core collection · Population structure

Introduction

Sesame (*Sesamum indicum* L.) is a diploid species ($2n = 2x = 26$) that belongs to the family Pedaliaceae. It is an ancient oilseed crop cultivated for human consumption in tropical regions across the globe and is a high-oil-yielding

Key Message Identified a set of genetically diverse sesame genotypes using molecular core approaches, and such genotypes were phenotypically observed to contain specific traits of interest that are potentially useful for introgression through breeding in a sesame improvement program and simultaneously broaden the genetic base.

✉ Parimalan Rangan
r.parimalan@icar.gov.in

[1] ICAR-National Bureau of Plant Genetic Resources, PUSA Campus, New Delhi 110012, India

[2] Regional Research Station, Tamil Nadu Agricultural University, Tamil Nadu, Vriddhachalam 606001, India

[3] Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane QLD4072, Australia

[4] International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India

(~55%) crop with polyunsaturated fatty acids as the major constituent (Pathak et al. 2014). Additionally, the seeds are also rich in protein (18–25%), making it a highly nutritious oilseed crop that can be consumed directly. Sesame oil, being rich in antioxidants like sesamolin, sesamin, and sesamol, contributes in multifarious ways to human health, like as an antiseptic, bactericide, disinfectant, moth repellant, and anti-tubercular agent (Anilakumar et al. 2010; Bedigian 2010, 2011; Dossa et al. 2017). Despite this fact, global sesame oil production is low compared with other oilseed crops (Yadav et al. 2022).

The genus *Sesamum*, including the crop wild relative (CWR) of cultivated sesame, comprises 31 accepted species and has its origin and distribution in Africa, South and Southeast Asia, Melanesia, and Australia (Manning and Magee 2018; POWO 2022; Pradheep et al. 2021). The Indian subcontinent has been identified as the center of origin for cultivated sesame and its progenitor (Bedigian 2003; Bedigian and Harlan 1986; Bhat et al. 1999; Hiremath 1999). India, China, and Sudan are the largest producers of sesame seeds, with India alone accounting for about 30% of global production (Pandey et al. 2015).

Sesame production is declining due to reduced land area under sesame cultivation as a result of low productivity, and in turn, lesser returns (Duhoon 2004; Ram et al. 2006). The plant genetic resources (PGR) of a crop representing the complete phenotypic variability and diversity being collected and conserved in the Genebanks over a period of time, would be the potential resource for genes underlying variable traits of interest. This would help withstand and adapt to harsh climatic conditions, be tolerant to various biotic stress factors, have high oil yields, etc., which would accelerate the sesame breeding program for higher productivity and enhanced oil yield. For efficient utilization and effective conservation, diversity assessment at the molecular level is necessary when the number of germplasm being conserved is in the tens of thousands of accessions that are practically not feasible to screen phenotypically in a homogenous environment. Molecular diversity assessment is also useful for crop improvement in identifying genetically diverse accessions (Govindaraj et al. 2015). The variability in the existing sesame genetic resources, being collected and conserved in the National Genebank (NGB), is to be assessed to identify newer sources of genetic variability that can be used for phenotypic validation to assess economic traits, productivity, and oil yield, including quality.

Molecular markers have been widely used to assess genetic variation among the germplasm to identify potential accessions that could be utilized in crop improvement programs. Genetic diversity and phylogenetic relationship in sesame accessions were assessed using random amplified polymorphic DNA (RAPD) (Abdellatef et al. 2008; Bhat et al. 1999; Ercan et al. 2004), amplified fragment length polymorphisms (AFLPs) (Laurentin and Karlovsky 2007, 2006; Uzun et al. 2003), sequence-related amplified polymorphism (SRAP) (Zhang et al. 2010), and inter-simple sequence repeats (ISSRs) (Kim et al. 2002; Parsaeian et al. 2011). With the increased availability of sequence information, genetic diversity and molecular characterization studies were performed using SSR markers due to their high polymorphism, reproducibility, high fidelity, chromosome specificity, and co-dominant nature. The different forms of SSRs, like genome sequence-SSRs (gSSRs) (Dixit et al. 2005; Dossa et al. 2016; Uncu et al. 2015), expressed sequence tag-SSRs (EST-SSRs) (Chung Suh et al. 2003; Sehr et al. 2016), or a combination of molecular markers with SSR markers, were reportedly used for genetic diversity assessment (Dar et al. 2017; Pandey et al. 2015; Wu et al. 2014). SSR markers have been used to assess genetic diversity and population structure in sesame (Cho et al. 2011; Dixit et al. 2005; Dossa et al. 2016).

Knowledge of the genetic diversity and population structure of germplasm collections is a key component for effective conservation, crop improvement, and breeding strategies (Thomson et al. 2007). The Indian National Genebank, located at the National Bureau of Plant Genetic Resources, New Delhi, maintains 10,506 (as of September 29, 2022) *Sesamum* accessions. Despite the availability of such a large sesame germplasm collection, the extent of use of this variation by breeders has been limited. Sesame accessions (3129) are being conserved at our NGB with special reference to agronomic traits representing seven quantitative and 12 qualitative traits, and a randomly selected set of 362 accessions were found to be proportionately representing the agro-ecological zones from where they were collected (Bisht et al. 1998). A year later, they studied the various stratified sampling strategies for developing a core set from the set of 3129 accessions with phenotypic data for 19 traits, and depending on the strategy, the representative core was 5–30% (Mahajan et al. 1999). The primary aim of this study was to comprehensively evaluate the genetic diversity within a worldwide collection of sesame germplasm and identify genotypes possessing valuable traits that could significantly contribute to enhancing sesame productivity through breeding programs. Furthermore, the study aimed to establish a core set of genotypes that efficiently represent the entire collection, thereby facilitating more streamlined and effective selection processes. The present study reports on the molecular diversity, population structure, and molecular core development using a set of 2496 sesame accessions from the total *Sesamum* germplasm (10,405 accessions) being conserved at our NGB, representing 17 countries and 26 states of India, using selected seven polymorphic SSR markers from the 43 SSR primers (Dixit et al. 2005; Dossa et al. 2016), screened initially.

Materials and Methods

Plant Material and DNA Extraction

The study material consisted of 2496 sesame accessions, purified through selfing (Ruperao et al. 2024), including indigenous and exotic collections from around 17 countries and 26 states of India, being conserved in the NGB of the ICAR-NBPGR, New Delhi. The passport details for the materials used in this study are given in Table S1. Twenty-one accessions were replicated (seven biological, 14 technical) at random for this experiment and are detailed in Table S1; for details, readers may refer to Ruperao et al. (2024). The seeds of these 2496 accessions were germinated on germination paper. Genomic DNA was extracted from the 5- to 7-day-old seedlings using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) as per the manufacturer's protocol. The quality and quantity of the extracted DNA were assessed using 0.8% agarose gel electrophoresis and a spectrophotometer.

Screening of SSR Marker

A total of 43 SSR primers (Dixit et al. 2005; Dossa et al. 2016), as detailed in Table 1, were used for screening against the 96 samples from 2496 samples, selected based on the preliminary phenotypic and passport information (Table S1). The PCR was performed in a total volume of 10 µL containing the final concentration, × 1 PCR buffer, 0.1 mM dNTPs, 0.25 µM SSR forward and reverse primers, and 0.5U Taq DNA polymerase. The PCR conditions are as follows: initial denaturation for 5 min at 94 °C, followed by 34 cycles of denaturation for 15 s at 94 °C, annealing for 30 s at 48 °C, extension for 30 s at 72 °C, and ending with a final extension for 5 min at 72 °C. The program was set to hold at 4 °C after the completion of the final extension until they were taken for downstream analysis. PCR products were resolved on a 6% PAGE containing acrylamide/bis acrylamide (38:2), × 10 TBE (0.89 M Tris, 0.89 M boric acid, and 0.02 M EDTA, pH 5), and a freshly prepared 10% ammonium per-sulfate solution at 200 V for 2–3 h using a 50-bp ladder (G-Biosciences, USA). Gels were stained with ethidium bromide (EtBr), and images were captured using a gel documentation system. Based on this initial screen, seven highly polymorphic markers (shown in bold letters in Table 1) were selected to assess the molecular diversity, population structure, and molecular core development of 2496 sesame accessions. PCR was performed for the entire 2496 sesame accessions using the selected seven SSR primers (PCR components and program, as detailed above). Hence, amplified PCR products along with a 50-bp DNA ladder (GeneDireX, Taiwan) were electrophoretically resolved on a 2% agarose gel (using × 1 TAE buffer) at 120 V for 1 h, and the images were captured using a gel documentation system (G:Box; Syngene, USA).

Morphological Characterization of Sesame Germplasm Under Field Condition

The same set of 2496 sesame accessions [1262 indigenous collection (IC); 1234 exotic collection (EC)] was evaluated under field conditions for morphological traits. The experiment was carried out at two locations (the NBPGR farm, Pusa Campus, New Delhi, and the farm of TNAU-RRS, Virudhachalam, Tamil Nadu, India) for two subsequent years (2018–2019 and 2019–2020) in an augmented block design (ABD) with five checks, i.e., RT-346, PB Til-2, GT-10, VRI-1, and TMV-7. Seeds were sown at 45 cm spacing between the rows and 10 cm between plants in a row. Appropriate agricultural practices were followed, and irrigation and pesticides were applied at the appropriate time. A few quantitative and qualitative traits were selected for the study. The traits studied were as follows: plant growth type; plant height; pubescence on leaf, stem, and capsule; number of primary and secondary branches; days to flower initiation; days to maturity; and number of capsules per plant. The morphological data was recorded as per the descriptors for *Sesamum* spp. IPGRI.

Data Analysis

The molecular weight of the amplified product was estimated for the samples using a 50-bp ladder as a reference. Based on molecular weight estimates, the scoring was performed manually in a co-dominant scoring pattern, and the same was entered in an Excel sheet for further analysis. Data analysis was performed only for distinct and reproducible bands. The Excel sheet containing data in the required format was subjected to molecular diversity analysis using GenAlEx v6.5 (Peakall and Smouse 2012). Genetic diversity parameters such as sample size (*N*), number of alleles per locus (Na), number of effective alleles per locus (Ne), Shannon's information index *(I)*, observed (Ho) and expected (He) heterozygosity, unbiased expected heterozygosity (uHe), and fixation index (*F*) were estimated using the GenAlEx v6.5 software with default settings. The pair-wise genetic distance among 2496 accessions was estimated, and a genetic distance matrix was obtained using GenAlEx v6.5 software. Genetic distance estimates were exported from GenAlEx v6.5 in nexus format and run through MEGA X (Kumar et al. 2018) to construct a dendrogram using the UPGMA method. The dendrogram was exported in Nexus format from MEGA X and used in iTOL v6 (Letunic and Bork 2021). A circulogram with additional color-coded strips was generated using iTOL for easy interpretation.

Table 1 Primer sequence for the 43 SSRs (boldfaced list are the selected seven primers)

S. no	Primer name	Forward primer (5′-3′)	Reverse primer (5′-3′)	References
1	*Si* 1	CTTTGATTGGGCCACCCTA	TGTTTGTTCTTCTTCCCCCA	Dossa et al. (2016)
2	*Si* 2	GCTTCCACCTAGCTCGGTTAT	CCAGCAATCATGTCTGCTTAAT	Dossa et al. (2016)
3	*Si* 3	TGTTGTTTGACCGTCTTCCA	TCGGGCTAGAAACCAACAGT	Dossa et al. (2016)
4	*Si* 4	TGGACAAAGACACAATCACACA	TTGAATTTCGATCTTTCCATCA	Dossa et al. (2016)
5	*Si* 5	CTGATGCAAAAACTGCCAAA	ACCGCACTCAAAGGTTCAAT	Dossa et al. (2016)
6	*Si* 6	TCAAAGTGTACCACAAAACGC	TCTTTCTCTCTCAATCATTTGTTTATG	Dossa et al. (2016)
7	*Si* 7	GATGGGGAAAGAGATTGGGT	ATTGAATCGACGTAATTTATCCTT	Dossa et al. (2016)
8	*Si* 8	GCCATTTCTTCATTTGGTGC	GCATTTCAATTTTTCACCCC	Dossa et al. (2016)
9	*Si* 9	CCAGACCCAAACCCAATAGA	TGCATTTAAGGCTGTGCAAC	Dossa et al. (2016)
10	*Si* 10	CGTGTGCCCAATATTTGAGTT	TCAACCTCCTCCCTACACAA	Dossa et al. (2016)
11	*Si* 11	CATTAACACCATTACGCAAACA	TTTGGCAAAACTGCAATGAA	Dossa et al. (2016)
12	*Si* 12	TCAGGAAGAAAGAATTGCTGC	CAACCCAACCATCCTGACTC	Dossa et al. (2016)
13	***Si* 13**	**TGCAAGGACAACCAAAATCA**	**TGCACTGCATTGTCTCCTTT**	Dossa et al. (2016)
14	*Si* 14	TGGTTCATACATCCTCTTTTTGG	CATAATTGATGCCTAAATTTCTCC	Dossa et al. (2016)
15	*Si* 15	TTCTGTCACCAAGAATTGCG	GTCAAAATTGAGGGTTGCGT	Dossa et al. (2016)
16	*Si* 16	TCATTAACCCATCATTGCGA	TGCTCACACATAACAGTTGGG	Dossa et al. (2016)
17	*Si* 17	TTCCAATTCTACAAGCGCAG	CCGATCAAAACTAGTATGGCAA	Dossa et al. (2016)
18	***Si* 18**	**GTCCGTCAACTCGATCACCT**	**TTCAACCAAACCCCATCATT**	Dossa et al. (2016)
19	*Si* 19	AATTAAAATTGGCCCAAGCC	GGCCAAGTGGAAATTTGAAG	Dossa et al. (2016)
20	*Si* 20	ACACATACGGACAGGCACAG	ATATAGCCAGTTTGGCTGCG	Dossa et al. (2016)
21	*Si* 21	AAAATCCTCTTTTTCCGACGA	GATTTTGACACCTTTGCCTGA	Dossa et al. (2016)
22	*Si* 22	CGCTTGAATTAATTGCATCTACC	CCAAGTGAACATAGAAATCTGCC	Dossa et al. (2016)
23	*Si* 23	CCAGCTCTATTGTGCGTTGA	CACTGCTTTCTCTGAAAGGCT	Dossa et al. (2016)
24	***Si* 24**	**CGCCTTTCTCCTCCTTATCC**	**CATTCAGTCTTACGTCCAAATTTCT**	Dossa et al. (2016)
25	***Si* 25**	**ACTGCACCCTCTGCATTTTT**	**GCACGTGTGGGGTACCTTTA**	Dossa et al. (2016)
26	*Si* 26	GGGGTGAGTATTTCGGGAGT	TCCATGCATCTTTTACACTGAA	Dossa et al. (2016)
27	*Si* 27	GGCCAACCCTTTTCAGATTT	GGGCTTCACAACACAAGACA	Dossa et al. (2016)
28	*Si* 28	CAAGTCGCCATCACACTCAT	TCGAGTTGGAATGCAACAAA	Dossa et al. (2016)
29	*Si* 29	TCCCAATCAGTTAGGTCGAG	TTAAGCTTAGGGGTCGGGTT	Dossa et al. (2016)
30	*Si* 30	CAACACCACCAACGCATATC	AGCAACGATTCACGACATTG	Dossa et al. (2016)
31	*Si* 31	TCCTGAATTCAAACGCATTG	TCCTAAACCCTCTGCACCAC	Dossa et al. (2016)
32	*Si* 32	AATTGGACTCCGGCTAGGAT	CGCCCTCATCCTTACAATCT	Dossa et al. (2016)
33	*Si* 33	CCCCTCTCAAATAAGCCCTC	AGGAAGGAGGGTGTCCCTAA	Dossa et al. (2016)
34	*Si* 34	TCATATATAAAAGGAGCCCAAC	GTCATCGCTTCTCTCTTCTTC	Dixit et al. (2005)
35	*Si* 35	GGAGAAATTTTCAGAGAGAAAAA	ATTGCTCTGCCTACAAATAAAA	Dixit et al. (2005)
36	*Si* 36	CCCAACTCTTCGTCTATCTC	TAGAGGTAATTGTGGGGGA	Dixit et al. (2005)
37	*Si* 37	TTTTCCTGAATGGCATAGTT	GCCCAATTTGTCTATCTCCT	Dixit et al. (2005)
38	***Si* 38**	**GCAGCAGTTCCGTTCTTG**	**AGTGCTGAATTTAGTCTGCATAG**	Dixit et al. (2005)
39	***Si* 39**	**CCACTCAAAATTTTCACTAAGAA**	**TCGTCTTCCTCTCTCCCC**	Dixit et al. (2005)
40	***Si* 40**	**GCAAACACATGCATCCCT**	**GCCCTGATGATAAAGCCA**	Dixit et al. (2005)
41	*Si* 41	TTTCTTCCTCGTTGCTCG	CCTAACCAACCACCCTCC	Dixit et al. (2005)
42	*Si* 42	CCATTGAAAACTGCACACAA	TCCACACACAGAGAGCCC	Dixit et al. (2005)
43	*Si* 43	TCTTGCAATGGGGATCAG	CGAACTATAGATAATCACTTGGAA	Dixit et al. (2005)

The population structure for these 2496 samples using the scoring data of seven SSR primers was assessed using a Bayesian model-based clustering program implemented in STRUCTURE v2.3.4 (Falush et al. 2003; Hubisz et al. 2009; Pritchard et al. 2000; Raj et al. 2014). Population structure was evaluated in terms of *K*-values (an assumed fixed number of subpopulations), each characterized by a set of allele frequencies at each locus. The STRUCTURE analysis was performed using K values ranging from 1 to 20, iterated 25 times each, with the length of the burn-in period and the number

of Markov Chain Monte Carlo (MCMC) replications after burn-in was set to 100,000. An ad-hoc statistic ΔK based on Evanno's method was used to determine the most likely number of population (K) (Evanno et al. 2005). After calculating the optimal K-value, which was found to be 3 in our case, we performed the analysis with $K = 3$ (iterated 150 times) with the length of the burn-in period and the number of MCMC replicate values set at 500,000 and 750,000, respectively.

A molecular core was generated among the 2496 samples, employing the molecular data obtained using the seven SSR markers. For this, PowerCore v1.0 software was used (Kim et al. 2007). Variability estimates between the complete set of 2496 accessions and the developed core set were compared to understand the level of representation of the alleles in the core from the population using the Shannon–Weaver diversity index and the Nei genetic diversity index.

The phenotypic mean data (2018–2019 and 2019–2020 at two locations) was taken for statistical analysis. The mean and coefficient of variation were calculated as per the standard statistical procedures. PCA (principal components analysis), scree plot, loading plot, and correlogram were prepared using Minitab statistical software. Hierarchical clustering for morphological traits was carried out by considering complete linkage and Euclidean distance and was represented in the form of a dendrogram. A core set was prepared based on phenotypic data from 2496 accessions using CoreHunter v3.0 software (http://www.corehunter.org/) using default parameter settings such as core size, intensity (10%, 12.5%, and 15%), time (5 s), and improvement time (1 s) (De Beukelaer et al. 2018).

Results

Molecular Diversity of Sesame Genotypes

Genetic diversity analysis was conducted using SSR markers to investigate the diversity pattern and develop a representative core-collection from a large collection of sesame germplasm being conserved at our National Genebank. It was performed on a dataset comprising 2496 sesame accessions genotyped using 7 SSR markers. The GenAlEx v6.5 (Peakall and Smouse 2012) software was utilized to estimate several genetic diversity parameters. These parameters offer insights into the diversity present within the sesame accessions. Initially, 43 SSR loci (Table 1) were screened to identify polymorphic SSR loci against a set of 96 accessions for use in molecular diversity analysis among 2496 sesame accessions. All seven polymorphic SSRs were used for molecular diversity analysis against 2496 accessions. These seven SSR primers produced a total of 140 alleles, with an average of 20 alleles per locus, with amplicon sizes ranging from 130 to 500 bp (Table 2). All the amplicon sizes were less than 500 bp for most of the SSR markers. Coincidentally, fragments of this range do not have linear mobility. Mobility primarily depends on the gel concentration (agarose) and electric field strength (Labrie et al. 2000; Stellwagen 2009). At times, mobility is independent of the strength of the electric field, although in general, mobility is estimated to vary in proportion to the square root of the field strength (Duke et al. 1994; Holmes and Stellwagen 1990). Locus Si39 has the maximum number of alleles (27) while locus Si13 has the least number of alleles (7). The number of effective alleles ranged from 2.627 (Si13) to 5.774 (Si38), with an average of 3.64 effective alleles per locus. The expected heterozygosity (He) ranged from 0.619 (Si13) to 0.827 (Si38), with an average of 0.70 per locus. The average observed heterozygosity (Ho) was 0.097 and ranged from 0.000 (Si25) to 0.324 (Si13). The average unbiased expected heterozygosity (UHe) was 0.705 and ranged from 0.619 (Si13) to 0.827 (Si38). The average fixation index F was 0.855, with a range of 0.477 (Si13) to 1.000 (Si25). The Shannon's information index ranged from 1.142 (Si13) to 2.067 (Si38), with an average of 1.622 (Table 2).

Table 2 The genetic diversity index of 2496 accessions using 7 SSR marker

Locus	*N*	Na	Ne	*I*	Ho	He	UHe	*F*
Si18	2457	21.000	4.508	1.770	0.003	0.778	0.778	0.996
Si24	2113	24.000	2.994	1.552	0.015	0.666	0.666	0.977
Si25	2277	12.000	2.967	1.350	0.000	0.663	0.663	1.000
Si38	2460	24.000	5.774	2.067	0.046	0.827	0.827	0.945
Si39	2210	27.000	2.821	1.681	0.100	0.645	0.646	0.844
Si40	2349	25.000	3.775	1.795	0.188	0.735	0.735	0.744
Si13	1959	7.000	2.627	1.142	0.324	0.619	0.619	0.477
Mean	2260.714	20.000	3.638	1.622	0.097	0.705	0.705	0.855
SE	69.304	2.845	0.433	0.116	0.046	0.029	0.029	0.072

Na no. of different alleles, *Ne* no. of effective alleles = 1/(Sum pi^2), *I* Shannon's information index = −1*sum (pi*Ln (pi)), *Ho* observed heterozygosity = no. of Hets/*N*, *He* expected heterozygosity = 1 − sum pi^2, *UHe* unbiased expected heterozygosity = (2*N*/(2*N* − 1))*He, *F* fixation index = (He − Ho)/He = 1 − (Ho/He)

The dendrogram obtained using the genetic distance matrix revealed a complex distribution pattern for the accessions studied, which were grouped into six clusters (Fig. 1). Cluster 1 consisted of 165 accessions, of which 132 were from different states of India and 33 from other countries. Cluster 2 is comprised with 267 accessions, of which 93 were from different states in India and 174 were from other countries. Cluster 3 comprised 397 accessions, of which 156 were from different states of India and 241 were from other countries. Cluster 4 consisted of 584 accessions, of which 258 were from different states of India and 326 were from various countries. Cluster 5 consisted of 519 accessions, of which 314 were from different states of India and 205 were from various countries. Cluster 6 consisted of 564 accessions, of which 309 were from different states of India and 255 were from various countries. Cluster 1, Cluster 3, and Cluster 5 contain a higher number of accessions from P3 of the population structure as compared to P1 and P2. Cluster 4 contains most of the accessions from P2 and P3. The accessions from P1 and P2 were mainly represented in cluster

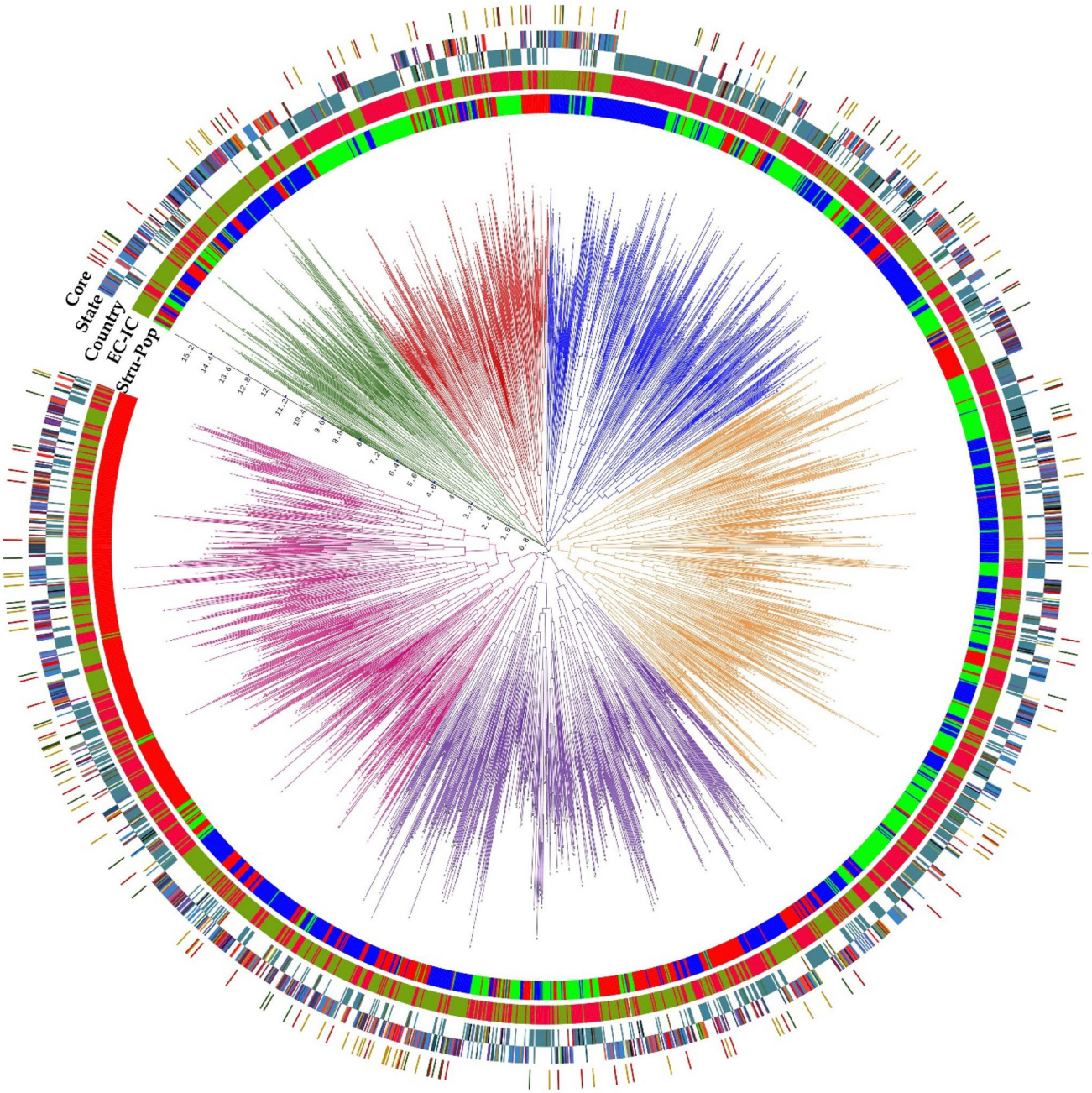

Fig. 1 Dendrogram showing the genetic relationship among the 2496 sesame accessions using seven polymorphic SSR markers

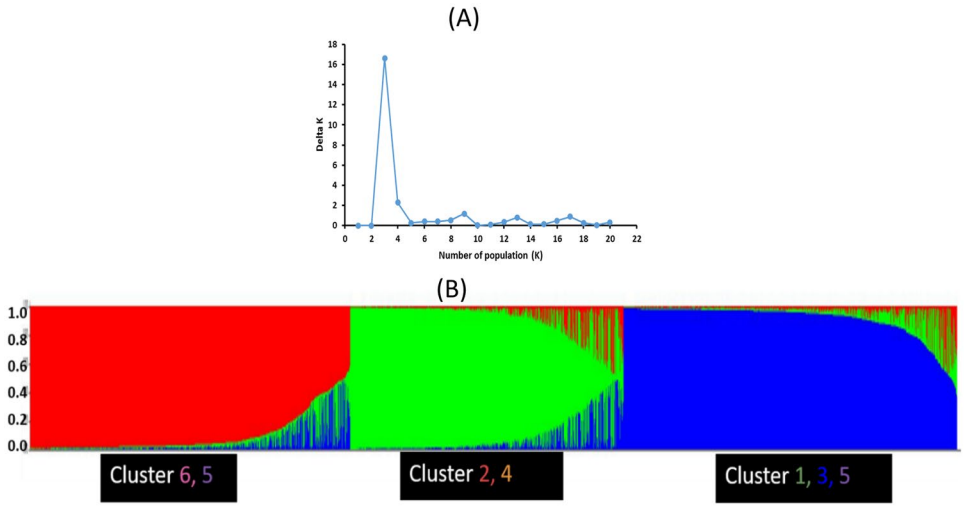

Fig. 2 STRUCTURE analysis for 2496 sesame accessions. **A** Magnitude of ΔK as a function of K. The peak value of ΔK was at $K=3$. **B** Box Plot representing the (sub-)population structure for the 2496 sesame accessions studied. A red, green, and blue color code represents the three populations P1, P2, and P3, respectively, as identified using the STRUCTURE tool

6 and cluster 2, respectively. The results obtained through molecular diversity analysis provide valuable insights into the genetic variability present in sesame accessions, laying the foundation for further genetic studies and conservation initiatives aimed at enhancing sesame productivity.

Population Structure Analysis

A Bayesian model-based clustering program, STRUCTURE 2.3.4 (Falush et al. 2003; Hubisz et al. 2009; Pritchard et al. 2000; Raj et al. 2014), was used to infer the genetic relatedness among 2496 sesame accessions. The ad hoc statistic ΔK based on the Evanno method was used to determine the most likely number of (sub-)populations (K) with reference to the total number of accessions (Evanno et al. 2005). The K versus ΔK (calculated as per the Evanno method) was plotted to identify the optimal K and was found to be $K=3$ (Fig. 2A). A red, green, and blue color were labeled to distinguish the three (sub-)populations P1, P2, and P3, respectively, in the STRUCTURE-plot (Fig. 2B). The accession details within each sub-population are provided in Table S1. The sub-population P1 consisted of 862 accessions (34.54%), of which 400 were collected from different states of India (Andhra Pradesh, Arunachal Pradesh, Assam, Gujarat, Madhya Pradesh, Maharashtra, Odisha, Rajasthan, Tamil Nadu, and Uttar Pradesh), representing the indigenous collections. In this sub-population, 360 accessions were imported from other countries, primarily Singapore (308); for details, readers may please refer to Table S1. The second sub-population (P2) comprised 736 accessions (29.49% of the total 2496 accessions), of which 79 were from different states of India (Andhra Pradesh, Madhya Pradesh, Maharashtra, and Rajasthan), and 615 were imported from other countries. The highest number of accessions represented in P1 and P2 were from Singapore (308 and 528). The last sub-population (P3) comprised 898 accessions (35.98%), of which 605 were collected from different states of India (Andhra Pradesh, Madhya Pradesh, Maharashtra, Odisha, Rajasthan, Tamil Nadu, and Uttar Pradesh), and 219 were imported from other countries. The cluster analysis showed that P1 represents accessions from cluster 5 and cluster 6 of the circulogram (Figs. 1 and 2B). The accessions from clusters 2 and 4 of the circulogram are grouped in P2, while P3 has accessions represented from clusters 1, 3, and 5 (Figs. 1 and 2B). The results revealed that the three sub-populations consisted of accessions sourced from diverse origins.

Development of the Core Collection

A core collection was established using molecular diversity data generated using seven SSR loci for the 2496 accessions of sesame. The two-core sets were developed using PowerCore v1.0 (Kim et al. 2007), one using the random (R) option, which resulted in 151 accessions, while the non-random (NR) option resulted in 130 accessions. A composite core comprising of

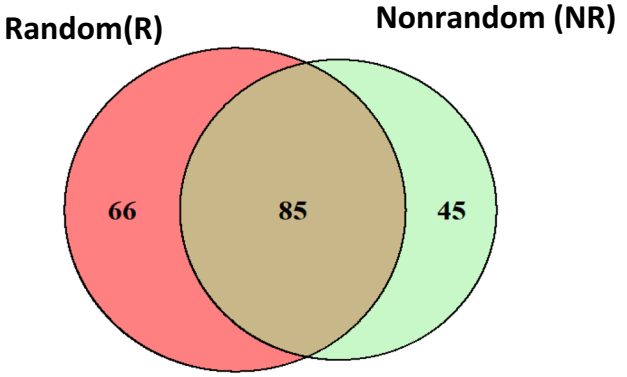

Fig. 3 Venn diagram representing the composite-core collection developed using two different methods, random (R) and non-random (NR), in the PowerCore tool

196 accessions (45-NR + 66-R + 85-common to NR and R) was developed by combining R and NR core sets with overlaps being removed (Figs. 1 and 3). The composite core collection of 196 accessions was found to represent 79 (9.16%), 60 (8.15%), and 57 (6.34%) accessions from populations 1, 2, and 3, respectively (Table 4). The composite core collection accounted for 7.85% of the entire collection and represents 100% of the alleles identified using seven SSR markers in the entire 2496 accessions. The allelic richness between the core and entire collections (2496 accessions) was compared using the Shannon–Weaver (Sh.W.) and Nei calculation indices (Table 3). The distributions of the Sh.W. and Nei indices in the seven SSR markers for the core and the entire collections were very similar (Table 3). A sample gel profile exhibiting allelic polymorphism is provided in Fig. 4. In addition, these 196 accessions' representation in the molecular diversity clusters was also studied (Fig. 1). The highest number, 47 accessions each from clusters 5 and 6, were present in the composite coreset. While 15, 21, 34, and 32 accessions were represented from clusters 1, 2, 3, and 4, respectively (Table 4), they are nearly proportionate with respect to the total number of accessions from each of these clusters.

Morphological Variability of Sesame Germplasm

Contrasting differences were observed in sesame germplasm for all the traits under study, indicating a higher variability in the sesame germplasm studied. The plant height ranged from 41.17 cm to 180 cm. The mean plant height was 111.5 ± 24.18 cm (Table 5). A total of 131 sesame accessions were observed to be dwarf in nature, having plant height below 75 cm (EC358033, IC0041924, IC127265, EC538037, EC334977, EC347071). A total of 82 sesame accessions (EC346539, EC347068, IC306009, IC205548, IC41953, IC205692, IC376292) showed a determinate type of growth habit. Generally, the sesame plants showed moderate branching, but plants without any branch or monostem type could be useful for mechanical harvesting. Monostem nature of plants was observed in a few sesame accessions (IC204133, IC132546, EC378841, IC430439, EC542303, IC204678). The number of primary branches ranged from 0 to 16, while the number of secondary branches ranged from 0 to 28. For growth habit, 1202 accessions (IC129960, IC129354, IC500533, IC129596, IC501066) represented the erect growth habit type; semi-erect growth habit was represented by 518 accessions (IC129989, IC129854, IC129542, IC129697, IC132410); and a few accessions such as EC346926, EC378841, and IC205607 had showed prostrate nature.

Pubescence on the leaf, stem, flower, and capsule of the sesame plant help in overcoming biotic and abiotic stresses (Kouighat et al. 2024). Few sesame accessions (EC346354, IC129902, IC205328) showed high pubescence on leaves, and 70 accessions (eg: IC129989, IC500986, EC342784,

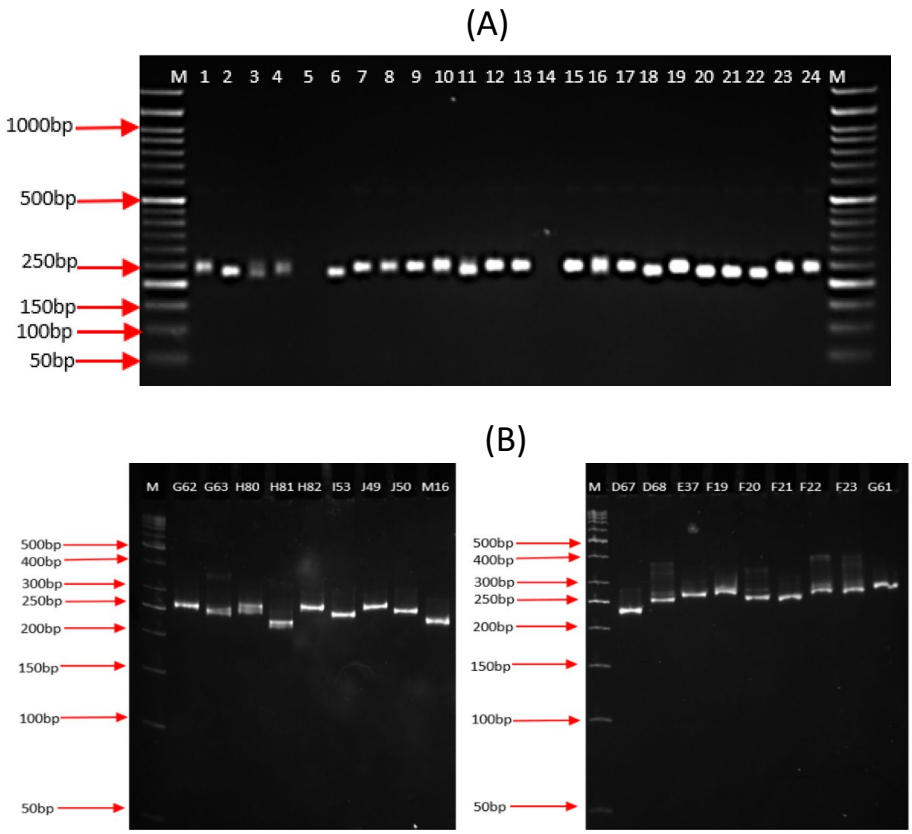

Fig. 4 Representative PCR amplification profile in agarose and polyacrylamide gel electrophoresis. **A** Agarose gel electrophoresis profile for the Si18 primer with 24 samples (B01-B24, sample B23 is a part of the core) from the 2496 sesame accessions using the 50-bp ladder (GeneDireX, Taiwan); **B** polyacrylamide gel electrophoresis profile for the Si24 primer with 18 samples (sample J50 is a part of the core) from the 2496 sesame accessions using the 50-bp ladder (G-Bioscience, USA)

Table 3 Comparison of diversity index between 196 core and entire accessions using seven SSR markers

SSR loci	C-Sh.W	C-Nei	C-allele	E-Sh.W	E-nei	E allele
Si18	2.237	0.84	22	1.831	0.785	22
Si24	2.473	0.844	31	1.791	0.74	31
Si25	1.935	0.809	13	1.529	0.712	13
Si38	2.329	0.848	25	2.179	0.838	25
Si39	3.361	0.927	55	2.116	0.741	55
Si40	3.535	0.944	60	2.326	0.811	60
Si13	2.106	0.846	16	1.77	0.791	16

Sh.W. and Nei indicate the diversity index using Shannon and Weaver and the Nei calculation using the PowerCore software

C core collection group, *E* entire group

Table 4 Core accessions present in each cluster and population from entire accessions

Cluster/population	Core acc	Entire acc	%
Cluster 1	15	165	9.09
Cluster 2	21	267	7.87
Cluster 3	34	397	8.56
Cluster 4	32	584	5.48
Cluster 5	47	519	9.06
Cluster 6	47	564	8.33
Average			8.06 (mean of 6 clusters)
Overall avg			7.85% (196/2496)
P1	79	862	9.16
P2	60	736	8.15
P3	57	898	6.34
Average			7.88 (mean)
Overall avg			7.85% (196/2496)

P1 population 1, *P2* population 2, *P3* population 3

IC500944, IC696229) with medium pubescence were observed. Low leaf pubescence was observed in 1561 sesame accessions (eg: IC129960, IC129354, IC129596, IC23267, IC129370), while 623 accessions (eg: IC500533, IC501066, IC273123, EC358046, IC510968) were without any leaf pubescence. Similarly, a few sesame accessions (EC346319, EC346591, EC346354, IC629079) were observed to have high pubescence on the stem surface, while 99 accessions (eg: IC500986, EC342784, IC131929, IC205692, IC132546) exhibited medium pubescence. Low stem pubescence was noted in 1469 sesame accessions (eg: IC129960, IC129354, IC129596, IC23267, IC129370), while a few accessions (IC129829, IC129703, EC370961) were found to be devoid of pubescence on the stem. In relation to pubescence on capsules, 142 sesame accessions (eg: IC205454, IC204175, IC129375, EC370878, IC204313, IC500665, IC52095) had high pubescence on capsule; while 554 accessions (eg: IC129854, IC132573, IC129697,

Table 5 Statistical analysis of agro-morphological traits in sesame (mean of two locations for 2018–2019 and 2019–2020)

Morphological traits	Mean	SE mean	StDev	C.V	Min	Max	Skewness	Kurtosis
Quantitative traits								
Plant height (cm)	111.5	0.637	24.18	21.67	41.17	180.0	0.33	0.07
Primary branches (no.)	4.619	0.055	2.089	45.23	0.00	16.0	0.99	1.61
Secondary branches (no.)	3.611	0.089	3.375	93.46	0.00	28.0	1.44	4.43
Flower initiation (days)	28.350	0.076	2.902	10.24	21.0	39.0	−0.05	−0.51
Days to maturity (days)	72.761	0.178	6.736	9.26	65.0		0.40	−0.71
Capsules per plant (no.)	101.50	1.15	43.76	43.11	22.0	280.0	0.70	0.16
Qualitative traits								
Growth type	1.03	0.004	0.18	17.59	1.00	2.00	5.61	31.85
Stem pubescence	2.81	0.02	0.85	30.58	0.00	7.00	−0.45	2.65
Leaf pubescence	2.62	0.03	1.16	44.26	0.00	7.00	−1.33	1.74
Capsule pubescence	3.72	0.03	1.37	37.02	0.00	7.00	0.41	0.93

For qualitative traits, growth type: 1, indeterminate; 2, determinate; for stem; leaf and capsule pubescence: 0, glabrous; 3, weak; 5, medium; 7, strong

IC132410, EC100043) had exhibited medium pubescence. Low capsule pubescence was observed in 1381 sesame accessions (eg: IC129960, IC129354, IC500533, IC129596, IC23267); while 244 accessions (eg: IC501066, IC129279, IC273123, IC205637, IC132464) did not show any pubescence on the capsule.

Flowering, or anthesis, is an important stage; therefore, it is important to determine the flowering initiation stage. The more earlier the flowering of sesame, the earlier it will reach the maturity stage in determinate types. Generally, the sesame plant is considered to reach physiological maturity when 80% of the capsules are mature and ready to harvest. The flowering starts at 21 d in sesame accessions and ranges up to 39 d. A few very early maturing accessions were: IC554423, IC129282, EC370536, IC129461, and EC346784. On the other hand, the number of days to maturity ranged from 65 to 129 days. 698 sesame accessions matured in less than 70 days; therefore, these accessions can be considered early-maturing accessions (Table 5, Fig. 5). Some of them were EC346816, IC511188, IC511072, EC346192, EC346949, IC129723, IC204037, and EC346370. These early-flowering or early-maturing accessions may be used further for a speed breeding program. The number of capsules per plant is an important trait that determines the yield of the plant, and it ranged from 22 to 280 (IC205601, IC130071, IC43107, EC370754, IC554432) (Table 5).

PCA results revealed 3 major principal components (Fig. 5A). A total of 58.8% genetic variability was contributed by these 3 major components (capsules per plant, plant height, and leaf pubescence). The loading plot and correlogram have been shown in Fig. 5B and C. A total of six clusters were observed in the hierarchial clustering (Fig. 5D). Pearson's correlation coefficient between different agro-morphological traits studied has revealed that plant height is positively correlated with the number of

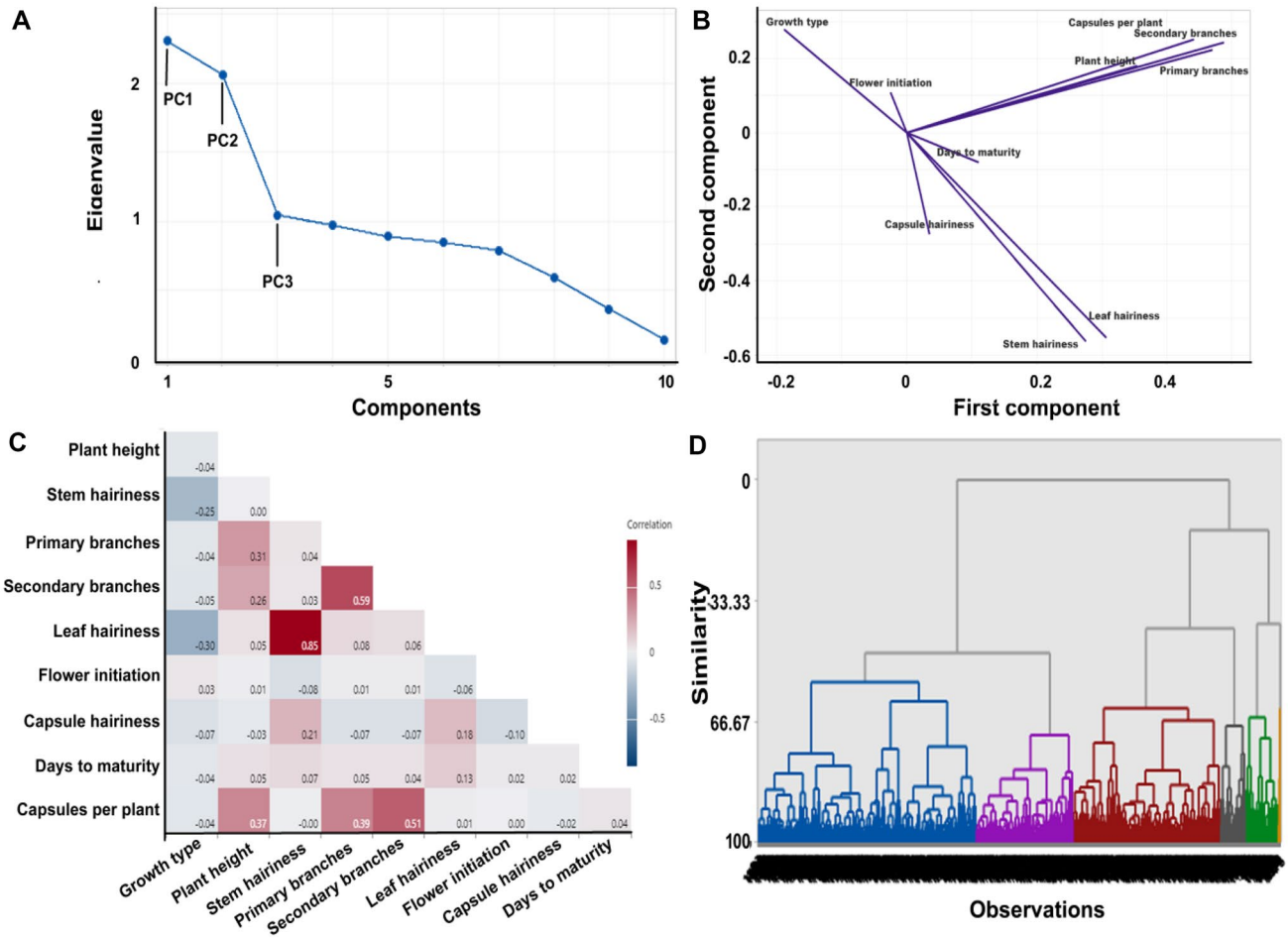

Fig. 5 **A** Principal components analysis (PCA) for the morphological traits studied in the sesame accessions; **B** loading plot analysis performed between the first and second principal component that gives an understanding of the co-relation for the other morphological traits; **C** correlogram representing the correlation (color-coded: red, positively correlated and blue, negatively correlated) among different morphological traits studied; **D** hierarchical clustering of the sesame accessions studied for the morphological traits

Table 6 Pearson's correlation analysis between different agro-morphological traits in sesame

	Plant height	Stem hairiness	No. of primary branches	No. of secondary branches	Leaf hairiness	Days to flower initiation	Capsule hairiness	Days to maturity
Plant height	1	0.002	0.309**	0.259**	0.046	0.006	−0.029	0.052*
Stem hairiness	0.002	1	0.037	0.033	0.853**	−0.076**	0.206**	0.074**
No. of primary branches	0.309**	0.037	1	0.592**	0.078**	0.010	−0.069**	0.045
No. of secondary branches	0.259**	0.033	0.592**	1	0.060*	0.012	−0.066*	0.045
Leaf hairiness	0.046	0.853**	0.078**	0.060*	1	−0.058*	0.182**	0.125**
Days to flower initiation	0.006	−0.076**	0.010	0.012	−0.058*	1	−0.096**	0.020
Capsule hairiness	−0.029	0.206**	−0.069**	−0.066*	0.182**	−0.096**	1	0.021
Days to maturity	0.052*	0.074**	0.045	0.045	0.125**	0.020	0.021	1

*Correlation is significant at the 0.05 level

**Correlation is significant at the 0.01 level

primary and secondary branches and the number of capsules per plant (Table 6). We observed that the secondary branches of the sesame plant do contribute to plant yield with an increased number of capsules. However, an analysis for finding a trade-off threshold between the energy costs for secondary branch formation and capsule formation with filled seeds needs specific attention. Pubescence on the leaf, stem, and capsule were also positively correlated to each other. Stem and leaf pubescence were highly positively correlated to each other (0.85), followed by stem and capsule pubescence and leaf and capsule pubescence.

A core set was developed based on phenotypic data (plant growth type; plant height; pubescence on leaf, stem, and capsule; number of primary and secondary branches; days to flower initiation; days to maturity; and number of capsules per plant) from 2496 accessions using Core-Hunter v3.0 software (http://www.corehunter.org/) (De Beukelaer et al. 2018). The resultant core set of 376 (15%) accessions showed that 18 sesame accessions were common between phenotypic data and molecular diversity SSR data. The traits observed in these accessions are shown in Table 7.

Table 7 Observed sesame accessions common to SSR-based and phenotyping core

No	Accessions	Trait observed in the field
1	EC370826	Good plant type
2	EC370718	Capsule length, stem pubescence, no. of capsules per plant
3	EC346795	Good plant type
4	EC346449	Good plant type
5	EC370545	Good plant type
6	EC346336	Good plant type
7	EC334975	Stem hairiness
8	EC370828	Good plant type
9	EC346869	Good plant type
10	EC350635	Multicapsular arrangement
11	EC346420	Leaf pubescence
12	EC346393	Good plant type
13	EC346815	Determinate type
14	EC347156	Corolla pubescence
15	EC370432	Good plant type
16	EC346821	8 locules per capsule, 3 capsule per leaf axil, stem and capsule pubescence
17	EC350640	Good plant type
18	EC334967	No. of primary branches

Discussion

Molecular diversity and population structure assessment help identify the genetically distant and diverse genotypes among the germplasm for potential use in breeding programs to broaden the genetic base of cultivars. This is especially helpful when the germplasm collected runs into tens of thousands. A few decades ago, due to the constraints of handling huge germplasm for phenotyping and the molecular tools in its initial stage, randomly selected genotypes with known phenotypic diversity alone were utilized in breeding programs to broaden the genetic base (Bisht et al. 2004). Hence, prior knowledge of genetic diversity in large collections of genetic resources can streamline efficient utilization (Liu et al. 2019). Prior to this study, most reports on genetic diversity assessment using SSR markers had a smaller sample size, not more than a few hundred (Cho et al. 2011; Dar et al. 2017; Dossa et al. 2016; Pandey et al. 2015; Surapaneni et al. 2014; Wei et al. 2014; Wu et al. 2014; Yol and Uzun 2012). In the present study, we screened 43 SSR markers and selected seven highly polymorphic ones to assess molecular diversity estimates and population structure patterns in a representative set of 2496 sesame accessions being conserved in the NGB. The total number of alleles for the seven SSR loci varied from 7 to 27 with an average of 20 alleles per locus. Several studies reported a high level of genetic diversity with a number of alleles per locus (2.3–11.3), heterozygosity (0.42–0.63), and polymorphic information content (0.46–0.52) (Cho et al. 2011; Dossa et al. 2016; Park et al. 2014; Wei et al. 2014). The variability in these values from ours' would be attributed to genetic variability in the samples and variation in the choice of marker. In the present study, the average "He" value (0.70), being greater than 0.50 (Sa et al. 2021), suggests higher allelic diversity among the 2496 sesame accessions studied. Our results corroborate with the previous study (Wei et al. 2014), reporting higher allelic diversity with a similar average He value of 0.68. The higher value of the expected heterozygosity (mean He = 0.705) when compared to the observed heterozygosity (mean Ho = 0.097) underscores the self-pollinating nature of the sesame population and is known to be supplemented with cross-pollination.

Shannon's information index (*I*) measures species' diversity in a population (Shannon 1948; Shannon and Weaver 1964; Spellerberg and Fedor 2003). The higher the value of *I*, the greater the species' diversity in a particular population. The diversity index has been categorized as high when the value is ≥ 0.60, or low if it is between 0.40 and 0.10. The diversity estimates for sesame were derived using Shannon's information index (0.70) and were found to be high, consistent with earlier reports (Dossa et al. 2016; Zhang et al. 2010).

The genetic distance matrix was calculated using GenAlEx version 6.5 (Peakall and Smouse 2012) for all pairwise combinations among 2496 sesame accessions. This matrix was further used in MEGA X (Kumar et al. 2018) to construct a tree revealing a complex distribution pattern with six major clusters among the 2496 accessions studied (Fig. 1). The distribution of sesame accessions within a cluster was found to be mixed across geographical regions from where they were originally collected and conserved. Similar results were reported in rice accessions collected from the same geographic regions clustered differently during diversity analysis (Islam et al. 2018; Yan et al. 2010). The distribution of sesame accessions across different geographical origins can probably be attributed to centuries of seed migrations driven by human activities such as trade and cultivation practices. These movements of seeds across regions likely facilitated gene flow across geographically diverse accessions when planted together. Additionally, the often-cross-pollinated nature of sesame enhances the gene flow across the accessions. These reasons underscore the lack of any direct association between genetic diversity and the accessions' geographical origin. These highlight the significant role of human-mediated activities, like trade and migration, in shaping the genetic structure of sesame across geographical locations. Hence, accessions from the same country frequently form distinct clusters due to their genetically variable nature (Laurentin and Karlovsky 2007, 2006; Wang et al. 2018; Yol and Uzun 2012). This observation suggests that prolonged cultivation may prompt local adaptations among accessions, while those preserved in gene banks experience minimal environmental influence, except during reproduction cycles.

Population Structure and Relationships

Bayesian model-based structure analysis is an alternative method to assess diversity patterns (Falush et al. 2003; Hubisz et al. 2009; Pritchard et al. 2000; Raj et al. 2014), inferring population structure among the accessions studied to understand allelic richness (Basak et al. 2019; Cho et al. 2011; Cui et al. 2017; Eltaher et al. 2018; Porras-Hurtado et al. 2013; Teklu et al. 2022). Population structure analysis helps to identify varying levels of genetic relatedness among subgroups within a sample. This divergence often occurs due to factors like geographic isolation or the sampling of individuals from different locations along a geographical gradient. To accurately infer the divergence between two populations, it is essential to first establish the existence of distinct populations and determine which individuals belong to each (Liu et al. 2020). Previous reports on population structure analysis in sesame were reported, with two to four (sub)populations showing substantial exchange of germplasm (Cho et al. 2011; Cui et al. 2017; Teklu et al. 2022). Molecular diversity analysis grouped these 2496 accessions into six major clusters (Fig. 1), while population structure analysis represented this set of accessions into

three major sub-populations (Fig. 2). The results revealed that the three sub-populations consisted of genotypes sourced from diverse origins. This suggests a significant exchange of alleles among sesame populations, indicating that geographic separation is no hindrance to gene flow between sesame accessions.

The core collection is a representative sample from large collections and is a true representation of the total population, reflecting equivalent genetic diversity with a minimum number of accessions that could enhance effective utilization. The core germplasm resource is an effective way to utilize genetic diversity, management, and conservation (Brown 1989a, b; Frankel 1984; Xu et al. 2020). The proportion of core germplasm would be determined by the size of the initial germplasm resource (Balakrishnan et al. 2000; Zhang et al. 2010). During the construction of a core using molecular markers, allele retention in the core when compared to the entire collection is frequently considered an evaluation criterion. Using the sampling theory of selectively neutral alleles, it was argued that the entries in a core subset are ~ 10% of the entire collection, with a maximum of three thousand entries per species (Brown 1989b). This level of sampling could successfully retain at least 70% of the alleles in the entire collection. The core collections were reported in previous studies for germplasm with large collection sets. For instance, 11.90% (1704/14310) in groundnut (Upadhyaya et al. 2003), 7.52% (1103/14651) in chickpea (Archak et al. 2016), 10.1% (278/2751) in sesame (Park et al. 2014), 10.32% (570/5522) in safflower (Dwivedi et al. 2005), and 5–30% in sesame (Bisht et al. 1998; Mahajan et al. 1999). In our study, the core collection amounted to 7.853% (196/2496 accessions), and these 196 accessions retained all alleles of the entire sesame collection of 2496 accessions (Fig. 3; Table S1). The total and average allele counts were 140 and 20, respectively, for the core set, which is the same as the initial sesame collection. Thus, the representation of alleles, which is essential for maintaining the genetic diversity of a population, is not compromised while designating the core set of accessions in sesame. The allelic richness between the core and entire collections was compared using the Shannon–Weaver (Sh.W.) and Nei indices calculations (Table 3). Compared to the entire sesame collection (2496), the core collection (196) represented the entire range of genetic diversity and maintained 100% of the allelic richness (Table 3). The designated core collection (196) gives a fair and proportionately equal representation of each of the six clusters as well as the three subpopulations (Table 4). Furthermore, the calculated PIC value on the core collection was 0.98, which was not significantly different from the entire sesame germplasm collection used in the study. These results demonstrate that the core collection represents sufficient genetic variation and can be used to assess phenotypic variability for selecting the most appropriate candidates in a crop improvement program. Our study highlights the role of molecular tools in efficiently managing plant genetic resources, especially when the collections were in tens of thousands. We identified the diverse representative set—molecular core collection—from the set of 2496 sesame accessions studied. The major limitations in using SSR markers for germplasm management are the time and cost factors which can be compensated appropriately by using genotyping-by-sequencing approaches (Ruperao et al. 2023, 2024; Yol et al. 2021). The diverse set (molecular core) identified through this study was grown in the field to observe the phenotypic variability for various agro-morphological traits. The phenotypic variability observed in the sesame accessions reported here is potentially useful for trait/resilience introgression through breeding in the sesame improvement program.

Phenotypic Trait Variability in Sesame

Phenotypic variability was studied for the following agro-morphological traits: plant growth type; plant height; pubescence on leaf, stem, and capsule; number of primary and secondary branches; days to flower initiation; days to maturity; and number of capsules per plant. Dwarf-type sesame accessions are generally more tolerant to waterlogging stress than taller ones (Adu-Gyamfi et al. 2019). Plant height is the most variable trait, followed by the number of capsules per plant in sesame (Bisht et al. 2004; Pham et al. 2010). The most contributing traits for genetic variability in this study were primary and secondary branches, plant height, and capsules per plant, while the least contributing traits were plant growth type, leaf pubescence, capsule pubescence, and stem pubescence (Fig. 6). Phenotypic characterization is important as it provides real-time on-field information directly to the breeders to make choices for accessions for crop improvement. The importance of documenting phenotypic variability for traits associated with seed yield and oil yield in sesame was well documented (Nawade et al. 2022; Yadav et al. 2022). Our study found potential sesame accessions based on a genotypic and phenotypic core data set (Table 7). Accessions such as EC370718 exhibit traits such as capsule length, stem pubescence, and the number of capsules per plant. EC350635 displays characteristics related to the multicapsular arrangement, while EC334967 demonstrates traits like the number of primary branches. These attributes hold potential for breeders and researchers to enhance yield in sesame plants. To meet the continuously increasing labor costs involved in sesame cultivation, efforts are directed toward mechanization with higher yields. The use of monostem accessions would give higher outputs through high-density planting and mechanization. Further, the sesame accessions with dense pubescence are more resistant to biotic (leaf curl, insect pests) and abiotic stresses (drought). Therefore, pubescence on sesame could be evaluated for developing resistant sesame plants against

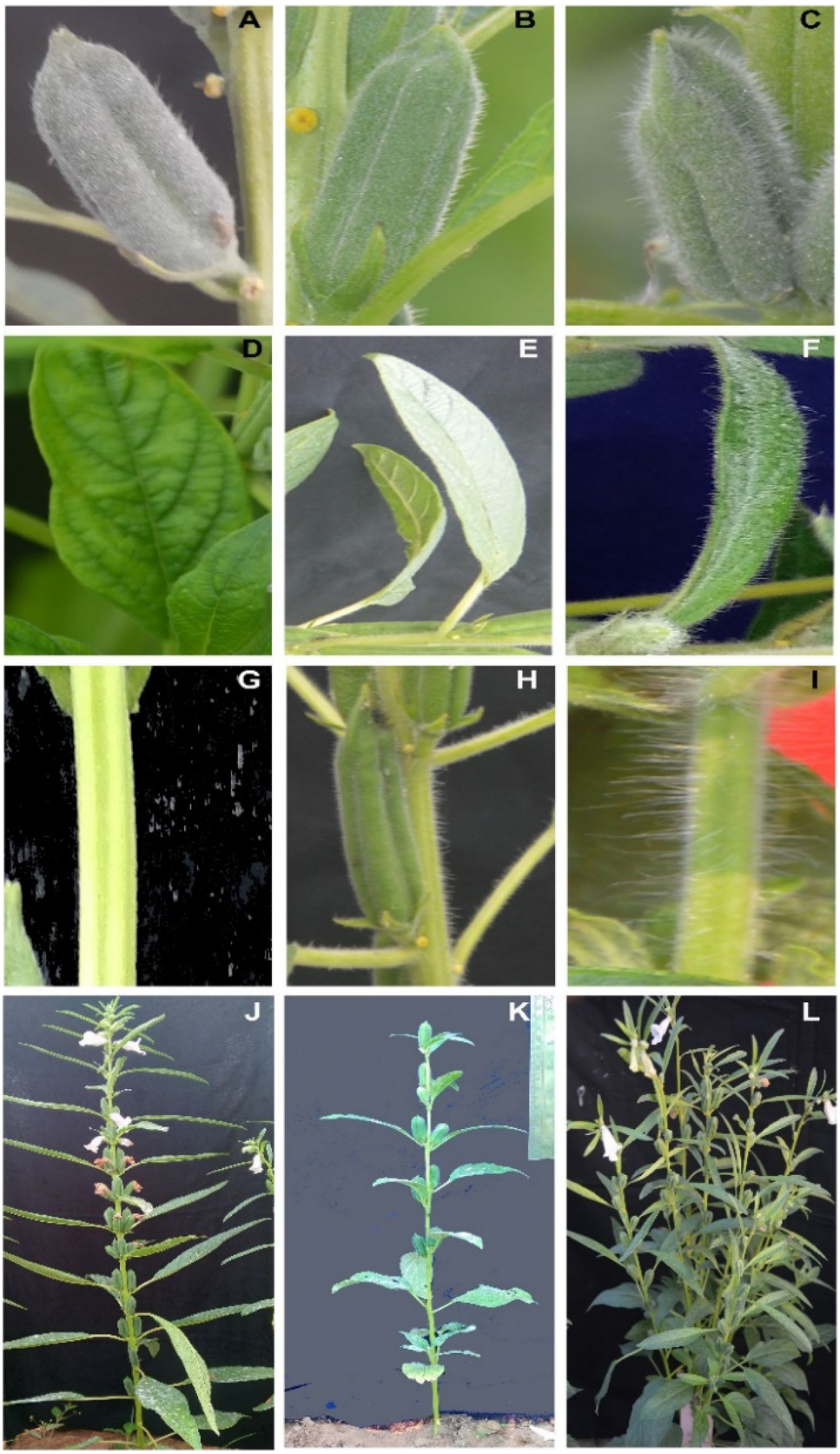

Fig. 6 Representative pictures of the sesame accessions exhibiting variability for the agro-morphological traits. **A**–**C** Variability for capsule pubescence trait; **D**–**F** variability for leaf pubescence trait; **G**–**I** variability for stem pubescence trait; **J** monostem with indeterminate growth and 2–3 capsules per leaf axil; **K** monostem with determinate growth; and **L** good plant type (more no. of primary and secondary branches, more no. of capsules, indeterminate type)

biotic and abiotic factors. The development of improved sesame lines (multicapsular arrangements) is ideal for targeting seed and yield productivity in sesame. For sesame breeders, knowledge of diversity in sesame germplasm for desired traits would contribute to designing breeding approaches for accelerating crop improvement. These findings highlight the role of comprehensive molecular diversity analysis in both understanding and conserving genetic resources for sesame breeding and improvement programs. Population structure analysis and the construction of a representative core collection emerge as essential strategies for capturing genetic diversity within sesame germplasm. These approaches assist in the efficient utilization of genetic resources through the core collection, a smaller set when compared to the total collections that run to tens of thousands. The significance of phenotypic and genotypic variability in sesame germplasm is emphasized, highlighting specific traits and promising accessions with potential for its use in breeding programs. Additionally, the identification of potential sesame accessions will further contribute to breeding strategies aimed at enhancing yield and resilience in sesame crops. Overall, these findings have far-reaching implications for sesame breeding programs, offering avenues for the selection of diverse and promising accessions and the effective utilization of genetic resources.

Conclusion

Molecular markers and agro-morphological variability are effective tools in assessing the genetic variability estimates in crop plants for their utilization in breeding programs. Here, we have assessed the genetic variability estimates in sesame germplasm (2496 accessions) using SSR markers and agro-morphological traits, which would be very helpful in identifying genotypes that are genetically distant and possess unique valuable traits. These identified genotypes are potentially useful in a sesame improvement program for addressing various agricultural challenges like mechanical harvest and high-density planting (monostem types are preferred), biotic stress tolerance (pubescence types are preferred), and abiotic stress tolerance (dwarf types are preferred). The core set developed based on SSR markers represented sufficient genetic variability in the population (2496 accessions). The trait-specific accessions identified for different traits, and the 18 common accessions between the molecular and phenotypic cores would be the key accessions that can be validated phenotypically and directly used for crop improvement programs.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11105-024-01476-z.

Author Contribution Conceptualization, Supervision, and hypothesis: PR, RY; Funding acquisition: PR, RY, KS; Investigation and formal analysis: RM, HV, SS, YSB, FNK, BN, PR; Germplasm and other resources: RY, RJ, RS, AsK, KS; Germplasm selfing, purification and characterization: MaA, MoA, SG, PR, RY, HV; Writing – original draft: RM, PR, HV, AjK; Writing – review & editing: PR, RM, HV.

Funding The authors are thankful to the Department of Biotechnology, Government of India, for the financial support through the research grant number-code 16113200037–1012166.

Data Availability The additional data pertaining to the study reported here were provided in the supplementary information.

Declarations

Competing Interests The authors declare no competing interests.

References

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

Adu-Gyamfi R, Prempeh R, Zakaria I (2019) Diversity assessment of some sesame (Sesamum indicum L.) genotypes cultivated in Northern Ghana using morphological and simple sequence repeat (SSR) markers. Adv Agric 2019:6067891

Anilakumar KR, Pal A, Khanum F, Bawa AS (2010) Nutritional, medicinal and industrial uses of sesame (Sesamum indicum L.) seeds-an overview. Agriculturae Conspectus Scientificus 75:159–168

Archak S et al (2016) Characterization of chickpea germplasm conserved in the indian national genebank and development of a core set using qualitative and quantitative trait data. Crop J 4:417–424

Balakrishnan R, Nair N, Sreenivasan T (2000) A method for establishing a core collection of Saccharum officinarum L. germplasm based on quantitative-morphological data. Genet Resour Crop Evol 47:1–9

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

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

Bedigian D (2003) Evolution of sesame revisited: domestication, diversity and prospects. Genet Resour Crop Evol 50:779–787

Bedigian D (2010) Characterization of sesame (Sesamum indicum L.) germplasm: a critique. Genet Resour Crop Evol 57:641–647

Bedigian D (2011) Cultivated sesame and wild relatives in the genus Sesamum L. In: Bedigian D (ed) Sesame: the genus Sesamum. Medicinal and aromatic plants–industrial profiles series. CRC Press, Boca Raton, FL, USA, pp 33–77

Bhat KV, Babrekar PP, Lakhanpaul S (1999) Study of genetic diversity in Indian and exotic sesame (Sesamum indicum L.) germplasm using random amplified polymorphic DNA (RAPD) markers. Euphytica 110:21–34

Bisht IS, Mahajan RK, Loknathan T, Agrawal R (1998) Diversity in Indian sesame collection and stratification of germplasm accessions in different diversity groups. Genet Resour Crop Evol 45:325–335

Bisht I, Bhat K, Lakhanpaul S, Biswas B, Pandiyan M, Hanchinal R (2004) Broadening the genetic base of sesame (Sesamum indicum L.) through germplasm enhancement. Plant Genet Resour 2:143–151

Brown A (1989b) Core collections: a practical approach to genetic resources management. Genome 31:818–824

Brown A (1989a) The case for core collections The use of plant genetic resources:136–156

Cho Y-I et al (2011) Evaluation of the genetic diversity and population structure of sesame (Sesamum indicum L.) using microsatellite markers. Genes Genom 33:187–195
Chung Suh M et al (2003) Comparative analysis of expressed sequence tags from Sesamum indicum and Arabidopsis thaliana developing seeds. Plant Mol Biol 52:1107–1123
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
Dar AA, Mudigunda S, Mittal PK, Arumugam N (2017) Comparative assessment of genetic diversity in Sesamum indicum L. using RAPD and SSR markers 3. Biotech 7:1–12
De Beukelaer H, Davenport GF, Fack V (2018) Core Hunter 3: flexible core subset selection. BMC Bioinformat 19:203
Dixit A et al (2005) Development of polymorphic microsatellite markers in sesame (Sesamum indicum L.). Mol Ecol Notes 5:736–738
Dossa K et al (2016) Analysis of genetic diversity and population structure of sesame accessions from Africa and Asia as major centers of its cultivation. Genes 7:14
Dossa K et al (2017) The emerging oilseed crop Sesamum indicum enters the "Omics" era. Front Plant Sci 8:1154
Duhoon S (2004) Exploitation of heterosis for raising productivity in sesame. In: Clements R, Shorter R (eds) New directions for a diverse planet: Proceedings of the 4th international crop science congress, Brisbane, Australia, 2004. Australian Society of Agronomy, p 1899
Duke T, Viovy JL, Semenov AN (1994) Electrophoretic mobility of DNA in gels. I. new biased reptation theory including fluctuations. Biopolym Orig Res Biomole 34:239–247
Dwivedi SL, Upadhyaya HD, Hegde DM (2005) Development of core collection using geographic information and morphological descriptors in safflower (Carthamus tinctorius L.) germplasm. Genet Resour Crop Evol 52:821–830
Eltaher S et al (2018) Genetic diversity and population structure of F3: 6 Nebraska winter wheat genotypes using genotyping-by-sequencing. Front Genet 9:76
Ercan AG, Taskin M, Turgut K (2004) Analysis of genetic diversity in Turkish sesame (Sesamum indicum L.) populations using RAPD markers. Genet Resour Crop Evol 51:599–607
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620
Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587
Frankel O (1984) Genetic perspectives of germplasm conservation. Genet Manip Imp Man Soc 161:161–170
Govindaraj M, Vetriventhan M, Srinivasan M (2015) Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. Genet Res Int 2015:431487
Hiremath S (1999) Genome homology and putative progenitor of sesame (Sesamum indicum L.). J Cytol Genet 34:69–74
Holmes DL, Stellwagen NC (1990) The electric field dependence of dna mobilities in agarose gels: a reinvestigation. Electrophoresis 11:5–15
Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mole Ecol Resour 9:1322–1332
Islam M et al (2018) Diversity and population structure of red rice germplasm in Bangladesh. PLoS ONE 13:e0196096
Kim D, Zur G, Danin-Poleg Y, Lee S, Shim K, Kang C, Kashi Y (2002) Genetic relationships of sesame germplasm collection as revealed by inter-simple sequence repeats. Plant Breed 121:259–262
Kim K-W et al (2007) PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. Bioinformatics 23:2155–2162
Kouighat M, Kettani R, El Fechtali M, Nabloussi A (2024) Exploring mechanisms of drought-tolerance and adaptation of selected sesame mutant lines. J Agric Food Res 15:100911
Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA x: molecular evolutionary genetics analysis across computing platforms. Mole Biol Evol 35:1547–1549
Labrie J, Mercier JF, Slater GW (2000) An exactly solvable ogston model of gel electrophoresis attractive gel-analyte interactions and their effects on the ferguson plot. Electrophoresis 21:823–833
Laurentin HE, Karlovsky P (2006) Genetic relationship and diversity in a sesame (Sesamum indicum L.) germplasm collection using amplified fragment length polymorphism (AFLP). BMC Genet 7:10
Laurentin H, Karlovsky P (2007) AFLP fingerprinting of sesame (Sesamum indicum L.) cultivars: identification, genetic relationship and comparison of AFLP informativeness parameters. Genet Resour Crop Evol 54:1437–1446
Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296
Liu F-M, Zhang N-N, Liu X-J, Yang Z-J, Jia H-Y, Xu D-P (2019) Genetic diversity and population structure analysis of dalbergia odorifera germplasm and development of a core collection using microsatellite markers. Genes 10:281
Liu C-C, Shringarpure S, Lange K, Novembre J (2020) Exploring population structure with admixture models and principal component analysis. In: Dutheil JY (ed) Statistical population genomics (Methods Mol Biol 2090), vol 2090. Methods in molecular biology, vol 2090. Humana Press, New York, USA, pp 67–86
Mahajan R, Bisht I, Gautam P (1999) Sampling strategies for developing Indian sesame core collection. Ind J Plant Genet Resour 12:1–9
Manning JC, Magee AR (2018) Additional new combinations in Sesamum L.(Pedaliaceae: Sesameae) Bothalia 48:a2363
Nawade B, Kumar A, Maurya R, Subramani R, Yadav R, Singh K, Rangan P (2022) Longer duration of active oil biosynthesis during seed development is crucial for high oil yield - lessons from genome-wide in silico mining and RNA-seq validation in sesame. Plants In Press, In Press
Pandey SK, Das A, Rai P, Dasgupta T (2015) Morphological and genetic diversity assessment of sesame (Sesamum indicum L.) accessions differing in origin. Physiol Mol Biol Plants 21:519–529
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 Genet Resour 12:112–119
Parsaeian M, Mirlohi A, Saeidi G (2011) Study of genetic variation in sesame (Sesamum indicum L.) using agro-morphological traits and ISSR markers. Rus J Genet 47:314–321
Pathak N, Rai AK, Kumari R, Thapa A, Bhat KV (2014) Sesame crop: an underexploited oilseed holds tremendous potential for enhanced food value. Agric Sci 5:46023
Peakall R, Smouse PE (2012) GENALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. Bioinformatics 28:2537–2539
Pham TD, Thi Nguyen T-D, Carlsson AS, Bui TM (2010) Morphological evaluation of sesame (Sesamum indicum L.) varieties from different origins. Aust J Crop Sci 4:498–504
Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu MV (2013) An overview of STRUCTURE: applications, parameter settings, and supporting software. Front Genet 4:98

POWO (2022) Plants of the world online. Facilitated by the Royal Botanic Gardens, Kew. Published on the internet; http://www.plantsoftheworldonline.org/. Royal Botanic Gardens, Kew. http://www.plantsoftheworldonline.org/. Accessed 02 September 2022

Pradheep K, suma A, Rangan P, Yadav R, Swathy Subhash K, Reshma R (2021) Genus Sesamum L. in India: an illustrated guide for species identification. ICAR-National Bureau of Plant Genetic Resources, New Delhi

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197:573–589

Ram SG, Sundaravelpandian K, Kumar M, Vinod K, Bapu J, Raveendran T (2006) Pollen–pistil interaction in the inter-specific crosses of Sesamum sp. Euphytica 152:379–385

Ruperao P et al (2023) A pilot-scale comparison between single and double-digest RAD markers generated using GBS strategy in sesame (Sesamum indicum L.). PloS One 18:e0286599

Ruperao P et al. (2024) Double-digest restriction-associated DNA sequencing-based genotyping and its applications in sesame germplasm management. Plant Genome e20447

Sa KJ et al (2021) Construction of a core collection of native perilla germplasm collected from South Korea based on SSR markers and morphological characteristics. Sci Rep 11:23891

Sehr EM et al (2016) Assessment of genetic diversity amongst Ugandan sesame (Sesamum indicum L.) landraces based on agro-morphological traits and genetic markers. J Crop Sci Biotechnol 19:117–124

Shannon CE (1948) A mathematical theory of communication. Bell Sys Tech J 27:379–423

Shannon CE, Weaver W (1964) The mathematical theory of communication. The University of Illinois, Urbana

Spellerberg IF, Fedor PJ (2003) A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' Index. Glob Ecol Biogeogr 12:177–179

Stellwagen NC (2009) Electrophoresis of DNA in agarose gels, polyacrylamide gels and in free solution. Electrophoresis 30:S188–S195

Surapaneni M, Yepuri V, Vemireddy LR, Ghanta A, Siddiq E (2014) Development and characterization of microsatellite markers in Indian sesame (Sesamum indicum L.). Mole Breed 34:1185–1200

Teklu DH, Shimelis H, Tesfaye A, Shayanowako AIT (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

Thomson MJ, Septiningsih EM, Suwardjo F, Santoso TJ, Silitonga TS, McCouch SR (2007) Genetic diversity analysis of traditional and improved Indonesian rice (Oryza sativa L.) germplasm using microsatellite markers. Theor Appl Genet 114:559–568

Uncu AÖ, Gultekin V, Allmer J, Frary A, Doganlar S (2015) Genomic simple sequence repeat markers reveal patterns of genetic relatedness and diversity in Sesame. Plant Genome 8:0087

Upadhyaya HD, Ortiz R, Bramel PJ, Singh S (2003) Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. Genet Resour Crop Evol 50:139–148

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 Breeding 122:95–97

Wang L, Bai P, Yuan X, Chen H, Wang S, Chen X, Cheng X (2018) Genetic diversity assessment of a set of introduced mung bean accessions (Vigna radiata L.). Crop J 6:207–213

Wei X et al (2014) Development of simple sequence repeat (SSR) markers of sesame (Sesamum indicum) from a genome survey. Molecules 19:5150–5162

Wu K, Yang M, Liu H, Tao Y, Mei J, Zhao Y (2014) Genetic analysis and molecular characterization of Chinese sesame (Sesamum indicum L.) cultivars using insertion-deletion (InDel) and simple sequence repeat (SSR) markers. BMC Genet 15:35

Xu J et al (2020) Development of a core collection of six-rowed hulless barley from the Qinghai-Tibetan plateau. Plant Mole Biol Rep 38:305–313

Yadav R et al (2022) Current research trends and prospects for yield and quality improvement in sesame, an important oilseed crop. Front Plant Sci 13

Yan W, Agrama H, Jia M, Fjellstrom R, McClung A (2010) Geographic description of genetic diversity and relationships in the USDA rice world collection. Crop Sci 50:2406–2417

Yol E, Basak M, Kızıl S, Lucas SJ, Uzun B (2021) A high-density snp genetic map construction using ddRAD-seq and mapping of capsule shattering trait in sesame. Front Plant Sci 12:679659

Yol E, Uzun B (2012) Geographical patterns of sesame accessions grown under Mediterranean environmental conditions, and establishment of a core collection. Crop Sci 52:2206–2214

Zhang Y-x, Zhang X-r, Hua W, Wang L-h, Che Z (2010) Analysis of genetic diversity among indigenous landraces from sesame (Sesamum indicum L.) core collection in China as revealed by SRAP and SSR markers. Genes Genom 32:207–215

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.