scientific reports

Check for updates

OPEN

A multidimensional core set development of sesame germplasm leveraging agro-morphological traits diversity for genetic improvement and climate resilience

Rashmi Yadav[1]✉, Harinder Vishwakarma[1], Subhrajit Satpathy[2], A. Mahalingam[3], Parimalan Rangan[1,4], Ajay Kumar[1], Sharik Ali[1], Kriti Sharma[1], Kamna Kumari[1], Dev Kumar Arya[1], Indu Rialch[5], Sita Ram Kumhar[6], Vanraj Gohil[7], Rajani Bisen[8], Renu Pandey[9], Vandna Rai[10], Kapil Tiwari[11], Anuj Kumar Singh[11], Ulavappa Basavanneppa Angadi[12], Celia Chalam Vasimalla[1], G. Senthilraja[3], Pankaj Sharma[5], Pardeep Kumar[1], Kailash Narayan Gupta[8], Hemavati Ranebennur[13], Kartar Singh[14], Sapna Langyan[1], Virendra Singh Rana[15], Najam Akhtar Shakil[15], Sanjula Sharma[5], Artika Singh[16], Rakesh Choudhary[16], Kanakasabapathi Pradheep[17], Sunil Shriram Gomashe[18], Senthil Ramachandran[2], Mamta Singh[1], J. Aravind[1], Radhamani Jalli[1], Ashok Kumar[1], Kuldeep Singh[2], Gyanendra Pratap Singh[1], Sanjay Kalia[19] & Abhishek Rathore[2]✉

Sesame (*Sesamum indicum* L.), an important oilseed crop, requires genetic improvement to ensure food security. Exploitation of a small fraction of genetic diversity is a major limiting factor for this crop. ICAR-NBPGR holds a large sesame base collection (6,000 unique accessions), originating from India and other countries. The study aims at the characterization of a sesame germplasm collection and development of working core set representing the variability of entire germplasm collection in the National Genebank. Characterization was carried out in an Augmented Block Design (ABD) under multi-environmental conditions (six locations representing five agro-ecological zones of India) for two consecutive years based on 33 agro-morphological traits, different levels of biotic stress (phyllody, dry root rot), abiotic stress (water logging, drought), and seed quality parameters (oil, protein, lignan, fatty acids). For core set preparation, the coincidence rate (CR%) for range and the variable rate (VR%) for coefficient of variation were maximum. Levene's test and Newman-Keuls tests were performed to compare the entire and core collection. Diversity between the two (entire and core collection) was compared using Shannon–Weaver diversity index. We developed a sesame core set of 773 accessions, which consists of 12.88% of the entire collection. Developed core set exhibited a Coincidence rate (CR)% of 99.8% and a variable rate (VR%) of CV 103.59%. A multipurpose core set of sesame germplasm was assembled to benefit the plant breeding community (for trait introgression), farmers, and researchers to develop climate-resilient sesame cultivars.

Keywords Sesame accessions, Quality traits, Multilocation phenotyping, Multipurpose core collection, Stress tolerance

[1]Indian Council of Agricultural Research-National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi 110012, India. [2]ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), Patancheru, Telangana 502324, India. [3]Regional Research Station, Tamil Nadu Agricultural University, Virudhachalam, Tamil Nadu 606001, India. [4]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD 4072, Australia. [5]Punjab Agricultural University, Ludhiana, Punjab 141027, India. [6]Agricultural Research Station,

Agriculture University Jodhpur, Mandor, Rajasthan 324304, India. [7]Agricultural Research Station, Junagadh Agricultural University, Amreli, Gujarat 365601, India. [8]PC Unit, JNKVV, Jabalpur, Madhya Pradesh 482004, India. [9]Division of Plant Physiology, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India. [10]National Institute for Plant Biotechnology, Pusa Campus, New Delhi 110012, India. [11]Sardarkrushinagar Dantiwada Agricultural University, Sardarkrushinagar, Gujarat 385506, India. [12]Division of Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, Pusa Campus, New Delhi 110012, India. [13]Division of Plant Pathology, ICAR-Indian Indian Agricultural Research Institute, New Delhi 110012, India. [14]ICAR-NBPGR Regional Station Jodhpur, Jodhpur 342003, Rajasthan, India. [15]Division of Agricultural Chemicals, ICAR-IARI, Pusa Campus, New Delhi 110012, India. [16]Rani Lakshmi Bai Central Agricultural University, Jhansi, Uttar Pradesh 284003, India. [17]ICAR-NBPGR, Regional Station, Thrissur, Kerala 680656, India. [18]ICAR-NBPGR, Regional Station, Akola, Maharashtra 444104, India. [19]Department of Biotechnology, Ministry of Science and Technology, Government of India, New Delhi, India. ✉email: rashmi.yadav1@icar.org.in; abhishek.rathore@cgiar.org

Sesame (*Sesamum indicum* L.) is an important oilseed crop, mainly cultivated for its seed. It is the oldest known oilseed crop, and history can be traced back to ~ 5,500 years ago in the Indian sub-continent. *S. indicum* (diploid, 2n = 2x = 26) is the only cultivated species in this genus; along with a few wild relatives such as *S. angustifolium* (tetraploid, 2n = 32) and *S. radiatum* (octaploid, 2n = 64)[1]. Health awareness programs and changes occurring in the consumer lifestyle have increased sesame seed consumption globally. Seed contains oil (44–57%), and proteins (18–25%), along with some nutritional and medicinal values[2]. Sesame seeds have useful antioxidant lignans such as sesamin, sesamol, sesamolin and other nutraceuticals which preserve sesame oil from rancidity[3]. Lignans show health–protecting activities such as anti-cancer, anti-inflammatory, and anti-aging properties and have proven beneficial roles in the human heart and brain[4].

Current total sesame production is 6.74 MT (year 2022) globally. Asia (2.40 MT) and Africa (4 MT) are major contributors (97%) to sesame seed production, while the remaining 3% is from North America and Europe. The top ten sesame seed producing countries (for year 2022) are given here. Sudan (1.23 MT), followed by India (0.79 MT), Myanmar (0.76 MT), Tanzania (0.70 MT), Nigeria (0.45 MT), China (0.44 MT), Burkina Faso (0.21 MT), Chad (0.20), Ethiopia (0.18), and Brazil (0.16) (https://ourworldindata.org/). In India, annual production of Sesame seed in India for the year 2022–23 was 2,62,100 MT. The average planting area (year 2023) in India was ~ 1.35 Mha, with an area under cultivation of 1.13 Mha. The top four states involved with highest area for sesame seed cultivation are Madhya Pradesh (0.4 MHa), followed by Uttar Pradesh (0.34 MHa), Rajasthan (0.23 MHa), and Gujarat (0.05 MHa). India exports seeds to around140 different countries, with its top five are Korea, the European Union, the USA, China, and Russia (https://www.sesameconference.com/WSPC2023/WSPC-2023-Highlights.pdf). Globally, Sesame seed oil production still lags behind other oilseed crops such as rapeseed, groundnut, sunflower, and soybean. This is due to several challenges including biotic and abiotic stresses, limited capacity, and small seed size[2]. Therefore improving oil production or yield in sesame remains a significant challenge. . Sesame germplasm resources (genebanks) may provide a solution to this problem. Some of the world's largest sesame seed repositories include, India National Genebank (10,515, www.nbpgr.ernet.in/), National Agro-biodiversity Center, South Korea (7,698), USDA-ARS, USA (1,190 accessions), Australian Grains Genebank, Australia (453 accs)[1]. The Genesys online platform also (https://www.genesys-pgr.org/,) shows 9,963 sesame accessions from 21 countries. However, the large size of the germplasm collection often restricts breeders from utilizing them effectively. As a result, developing a working core collection would be the best approach to investigate the entire sesame germplasm.

A core set is a representation of an entire collection of a species, capturing maximum genetic diversity with minimum repetition. This provides reliable data that makes it easier to optimize genotype/phenotype/genetic markers association studies. Previous studies have reported that phenotypic core collection have been developed in various species of cereals, legumes, and oilseed crops etc.[5–8] Several software packages are available for developing a core set, such as PowerCore[9], CoreHunter[10], GenoCore[11], and. CH3 (Core Hunter 3), an updated version of CoreHunter[12]. Core set development can be done using stratified and non-stratified methods. Various data sets have been used to generate core sets in sesame and other oilseed crops. Previously, phenotypic traits (both quantitative and qualitative) alone or in combination with molecular markers data were used for sesame core development[13–15]. Marker-assisted selection, which is based on molecular markers, is useful for exploring diverse genotypes for speed breeding program. As a result, diverse molecular techniques such as AFLP (Amplified Fragment Length Polymorphism), cpSSR (Chloroplast SSR), cDNA-SSR (Complementary DNA-SSR), EST-SSR (Expressed Sequence Tags-SSR), GBS (Genotyping by Sequencing), ISSR (Inter-Simple Sequence Repeat), RAPD (Random Amplified Polymorphic DNA), RAD-seq (Restriction site-Associated DNA sequencing), RNA-seq, SLAF-seq (Specific Length Amplified Fragment Sequencing), and WGS (Whole-Genome Sequencing)[16]. Recently, a combination of sdRAD and ssRAD genotype data has been used to assess diverse set of sesame accessions[17,18]. In Safflower, another oilseed crop, a core set was developed using geographical diversity as well as phenotypic and molecular data[7]. Similarly, molecular markers such as DArTs, SSRs, and SNPs were used along with phenotypic data to develop a core collection in Olive, another important oilseed crop[19]. Few efforts were made in the past to study diversity and core development in sesame. Not more than a few hundred accessions were considered for diversity studies, which has limited the knowledge for full utilization of germplasm (Table S1).

The main objective of this research was to develop a multipurpose working core collection set of sesame germplasm based on large-scale phenotyping, biotic and abiotic stress, and seed quality traits. We used the maximum coincidence rate (CR%) and variable rate (VR%) for the coefficient of variation to develop the core collection. Statistical analysis such Levene's test, Newman-Keuls tests, and Shannon–Weaver diversity index, were then used to compare the representativeness and diversity of the developed core set with the entire collection.

Results

Core set evaluation based on geographical origin

Based on the two consecutive years' (2020–21, 2021–22) agro-morphological data of 6,000 accessions sown at the multi-environment location, a working core set of 773 sesame accessions was developed. The core set developed represents 12.88% of the entire sesame germplasm collection conserved in the National gene bank. Out of 773 accessions, 276 accessions were exotic collections (EC) from 19 different countries, and 497 accessions were of indigenous collections (IC) representing 22 states and 1 union territory of India (Fig. 1). The passport data of the sesame germplasm accessions included in the core set is given in Supplementary File S1. India, Singapore, and Israel were highly reflected in the core set followed by Bangladesh, USA, and Sri Lanka (Fig. S1). Ghana, Italy, Nigeria, and the Philippines showed 0.26% contribution to the core set while Bolivia, Egypt, Iran, South Africa, Syria, and Turkey showed 0.13% (minimum) contribution to the core set. In India, the state of Madhya Pradesh was highly reflected in the sesame core set diversity, followed by Maharashtra, Rajasthan, and Andhra Pradesh (Fig. S1). Assam, Bihar, Jammu and Kashmir, Karnataka, Mizoram, and Nagaland each contributed 0.2% to the sesame core genetic diversity.

Core set evaluation based on statistical data and tests

The percentage of significant difference between the core set developed and the entire germplasm collection for the mean difference percentage (MD%) and the variance difference percentage (VD%) of traits was 3.41% and 4.45%, respectively. Since this approach focuses on maximizing the VR% and CR% for each trait, it achieves more than 30% representation within trait-specific core sets. The developed sesame core set exhibited similar variability, resembling that of the entire collection. Table 1 presents the range, mean, median, and variance for the traits considered in the sesame core set collection. Means of traits between the entire collection and core set showed no significant difference, indicating their strong resemblance. Similarly, value of median, standard deviation, and variance were also non-significant between the entire collection and the developed core set. The homogeneity of variance test yielded similar results, with a p-value less than equal to 0.05 indicating non-significant difference between the variances of the core set and the entire collection at 5% level of significance. For qualitative (17) or categorical traits, Shannon–Weaver diversity index was carried out (Table 2). Value of the H' index indicated that the extracted core set could effectively represent the diversity present in the entire germplasm collection. Comparison of frequency distribution using the chi-square test revealed that all traits, except for four (plant growth type, leaf shape, capsule hairiness, and capsule dehiscence), showed non-significant differences when compared to the entire collection with the core set (Table 2).

The box plot for the core set and the entire collection of quantitative traits under study showed a similar distribution pattern (Fig. 2). Likewise, the frequency distribution pattern of 17 sesame qualitative traits showed a similar distribution pattern for the entire collection and the core set developed (Tables 2; Fig. 3). The trait association between the core set and the entire germplasm collection was depicted through PCA and correlation analysis. Nearly all the traits, as well as their values, were found preserved in the resultant core set and the entire collection. Seed yield attributing traits such as no. of capsules per plant, seeds per capsule, test weight, and yield per plant were positively correlated with each other. Pearson's correlation coefficient also revealed that plant height was significantly positively correlated to the no. of nodes to first flower, number of primary and secondary branches, in both the core set as well as in the entire collection (Table S3). We observed days to maturity was significantly positively correlated to plant height, no. of primary and secondary branches, it might be due to the indeterminate growth habit of the sesame crop.

Furthermore, spatial distribution of traits in core and entire collection was compared using PCA. PCA analysis revealed that first five principal components (PC1 to PC5) eigen values > 1 (ranged 1.02 to 2.956) in the entire collection, could explain 53.02% (variance ranged from 6.798% to 19.706%) of total genetic variability (Fig. S2, Table S4). Similarly, the first five principal components in the developed core set PC1 to PC5 having eigenvalues > 1 (ranged 1.125 to 2.68), contributed 53.97% (variance ranged from 7.497% to 17.874%) of total genetic variability. Factor loadings also suggested that individual PCs were quite closer in the developed core set and the entire germplasm. E.g., plant height (entire 0.519; entire 0.659), primary branches (entire 0.663; core set 0.657), no. of nodes to first flower (entire 0.787; core set 0.602). Hence, based on the traits correlation and PCA, we could observe that the phenotypic trait variability of the entire germplasm was captured in the prepared core set.

Agro-morphological characterization of sesame core set

High variability was observed in different agro-morphological traits of sesame. For plant growth type, indeterminate growth habit (97–98%) was the major trait among sesame accessions. Nearly 3% of determinate-type sesame accessions (EC346370) were present in the germplasm. Sesame germplasm consists of 6.4% monostem (IC110789) or non-branching accessions. 0.8% core set accessions showed strong leaf pubescence (EC350635), similarly high capsule pubescence (EC370488), and high stem pubescence (IC205780) was observed in 1.5% and 4.8% sesame accessions, respectively (Fig. 4). The days to maturity ranged from 60 (IC16240) to 137 (IC511001) and 65 (IC413204) to 137 (IC205631) in the entire collection and core set, respectively (Table 1). Internode length, capsule width, seeds per capsule were 100% conserved in the developed core set. The plant height ranged from 40 cm (EC346415) to 219 cm (IC306022) and 40 (EC377047) to 200 cm (IC511090) in entire and core set respectively, with a mean of 96.24 cm and 97.73 cm, in the entire collection and core set, respectively. Similarly, mean capsules per plant were 54.99 and 52.37 in the entire and core set respectively.

The entire collection ranged from 12 (EC347031) to 280 (EC347052), with a mean of 54.80 capsules per plant and 12 (EC891108) to 269 (EC346321) capsules per plant in the core set, having a mean of 52.38. Multi-capsular (per node) arrangements were present in 11.6% (EC361718) and 13.9% (EC346798) for 2 capsules per leaf axil in the entire and core set, respectively, while 3 capsules per leaf axil were present 1.4% and 1.9%

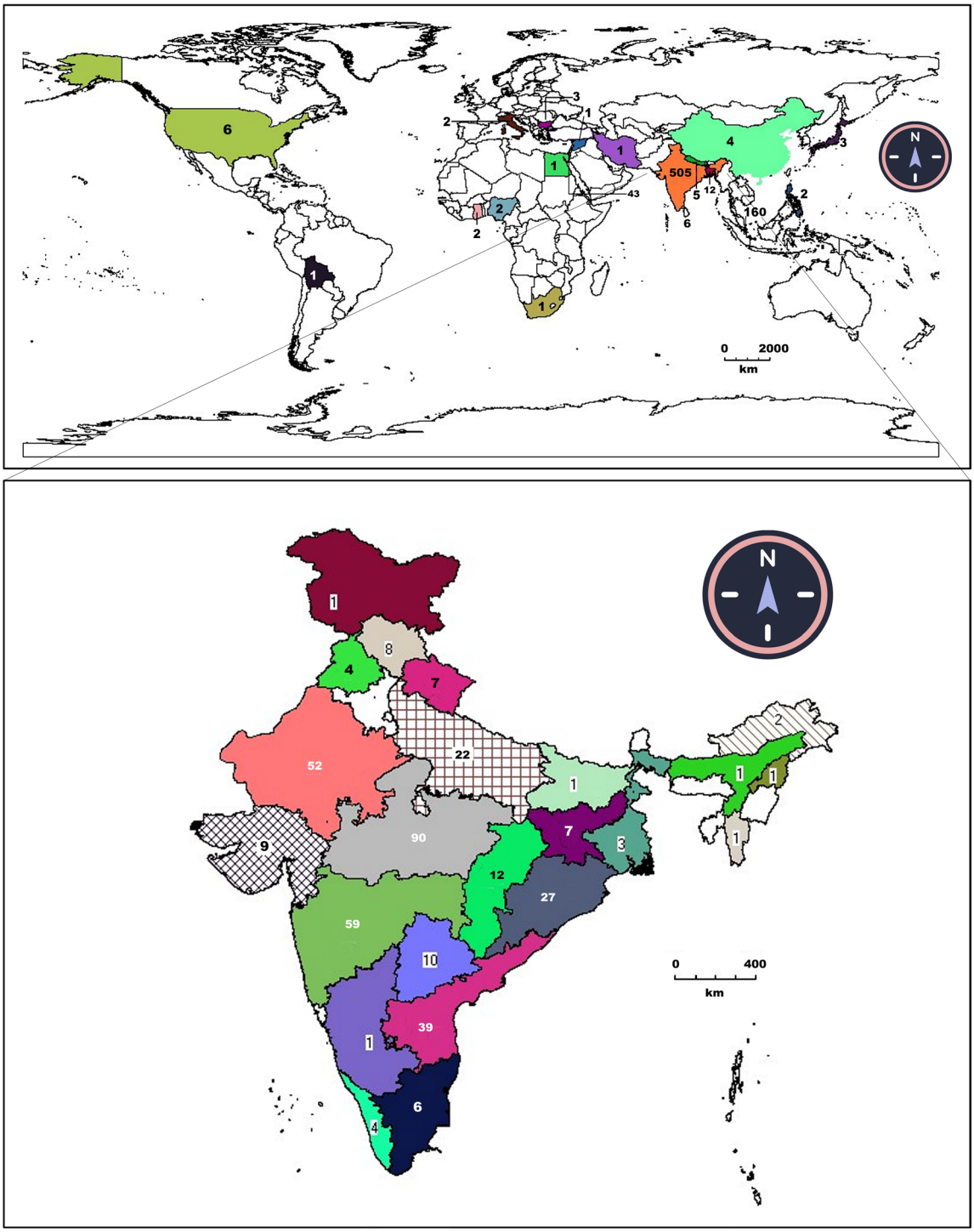


Fig. 1. Sesame core set diversity for indigenous and exotic collection. (**a**) Distribution across different countries of the world, (**b**) Distribution across different states of India. Unknown-due to incomplete or lack of passport data on the location of collection.

(IC430421), respectively in the entire and core set. Test weight, i.e., weight of 1000 seeds ranged 1.0 g (IC23296) to 4.9 g (IC205601); 1.03 g (EC370818) to 4.7 g (IC500674) in the entire collection (mean 2.69 g) and core set (2.67), respectively (Table 1). So, considering qualitative and quantitative traits of the sesame core set, the developed core collection here conserves all the beneficial phenotypic traits present in the entire collection. The hierarchical clustering of the sesame core set divided the accession into 6 clusters (Fig. 5a). The major cluster (Cluster 1) consists of 54% accessions followed by cluster 2 (20% accessions), the least contribution was made by cluster 5 (1%) and 6 (0.5%). Similarly, agro-morphological traits were also clustered into major 6 clusters (Fig. 5b), cluster 1 consists of important yield attributing traits viz. no. of sees per capsule, yield per plant, test

	Range (Min–Max)		Mean		Median		Stdev		Variance		Skewness	
Traits	Entire	Core	Entire	Core	Entire	Core	Entire	Core	Entire	Core	Entire	Core
Plant height (cm)	40–219	40–200 (91)[1]	96.24a	97.73a	97a	98a	18.81	17.19	353.77[b]	295.51[b]	0.522	0.56
Primary branches	0–22.33	0–19.33 (87)[1]	4.49a	4.38a	4a	4a	2.39	2.39	5.75[b]	5.715[b]	0.733	0.47
Secondary branches	0–36	0–33 (92)[1]	5.04a	4.86a	5a	4a	3.37	3.25	11.37[b]	10.61[2]	1.22	1.57
Internode length (cm)	0.8–11.5	1.3–11.5 (100)[1]	4.75a	4.06a	4a	3.66a	2.65	2.18	7.02[b]	4.766[b]	1.89	1.95
Capsules per plant	12–280	12–269 (96.1)[1]	54.99a	52.37a	47a	45a	32.28	30.55	1041.92[b]	933.55[b]	1.874	1.73
Capsule length (mm)	13–45	16.4–38.4 (85)[1]	25.05a	24.51a	25a	24.63a	3.08	2.74	9.50[b]	7.55[b]	0.218	0.32
Capsule width (mm)	4.17–23.8	4.3–23.8 (100)[1]	7.50a	7.64a	7.5a	7.6a	1.28	1.12	1.64[b]	1.271[b]	0.77	1.11
Seeds per capsule	13–86	13–86 (100)[1]	37.99a	38.76a	37a	38a	8.70	7.83	75.77[b]	61.43[b]	0.365	0.24
Days to maturity (d)	51–137	65–137 (100)[1]	93.49a	96.09a	92a	95a	12.33	13.04	152.04[b]	170.25[b]	0.375	0.16
Test weight (g)	1.0–4.9	1.03–4.7 (96)[1]	2.69a	2.67a	2.67a	2.65a	0.54	0.53	0.29[b]	0.283[b]	0.021	0.02
Yield per plant (g)	1.69–59.53	1.7–57.93 (97.3)[1]	8.04a	8.20a	7.54a	7.58a	3.97	3.59	15.72[b]	12.91[b]	1.922	1.69

Table 1. Comparative statistics of sesame quantitative traits in entire germplasm and developed core set. cm-centimeter, mm-millimeter, d-days, g-gram, Min-minimum, Max-maximum, Stdev-Standard deviation. a- Means followed by the same letter are statistically non-significant. [1] indicates % range retained in the core collection. [b] indicates equal variances were assumed between the entire collection and core set based on Levene's test ($p \leq 0.05$).

		Shannon–Weaver diversity index		Frequency distribution		
S. no	Descriptor	Entire collection	Core set	No. of classes	χ^2	*p*
1	Plant growth type (PGT)	0.11	0.15	2	4.443	0.035
2	Plant growth habit (PGH)	0.72	0.68	3	2.471	0.65
3	Stem hairiness (pubescence) (SH)	0.78	0.71	4	1.898	0.755
4	Branching pattern (BP)	0.79	0.83	4	4.913	0.296
5	Leaf hairiness (pubescence) (LH)	0.86	0.87	4	5.286	0.809
6	Leaf shape (LS)	1.35	1.26	6	39.553	0.006
7	Lobe incision of basal leaf (LIBL)	0.56	0.58	3	5.16	0.82
8	Flowers per leaf axil (FLA)	0.30	0.36	2	1.63	0.202
9	Corolla hairiness (pubescence) (CoH)	0.74	0.68	3	8.366	0.079
10	Exterior corolla color (ECC)	0.84	0.81	7	14.312	0.939
11	Interior corolla color (ICC)	0.61	0.68	7	17.506	0.489
12	No. of locules per capsule (LPC)	0.11	0.12	4	10.453	0.315
13	Capsule arrangement (CA)	0.25	0.34	2	3.622	0.057
14	Capsule hairiness (CaH)	0.61	0.68	4	26.494	0.002
15	Capsule shape (CS)	0.95	0.79	4	5.055	0.829
16	Capsule dehiscence (CD)	0.69	0.64	3	9.45	0.009
17	Seed coat color (SCC)	1.51	1.56	7	36.439	0.448

Table 2. Shannon–Weaver diversity index (H') and frequency distribution for various qualitative descriptors in the entire collection and core set of sesame germplasm. χ^2 represents the Chi-square value, p represents the value of significance.

weight; cluster 2 mainly consists of capsule and flower related feature such as flower color, pubescence on capsule and flower, capsule and flower arrangement. Cluster 3 consists of stem and leaf related features viz. stem and leaf pubescence, leaf shape. A dendrogram was prepared in which the entire collection (6,000 accessions) marked as grey color while developed core collection (773 accessions) was shown with red color (Fig. 5c).

Promising traits and accessions for breeding program and sesame crop improvement

The developed sesame core set was highly variable for agro-morphological traits. The core set developed was endowed with different agro-morphological traits such as determinate type, multi-locular, multi-capsular, monostem, early maturity (≤ 60–65 d), long capsules (≥ 40 mm), test weight (≥ 4 g), and more no. of capsules per plant (≥ 250) (Tables 3). The core set also consists of resistant accessions against biotic stress (phyllody and dry root rot), abiotic stress (waterlogging and drought) as well as quality traits (high oil, high protein (23.49–27.25%), high lignan content, high fatty acid content (Table 3). The value of oil content, protein content, fatty acid content, and lignan content for the core set developed are given in Supplementary File S2. As a whole, the

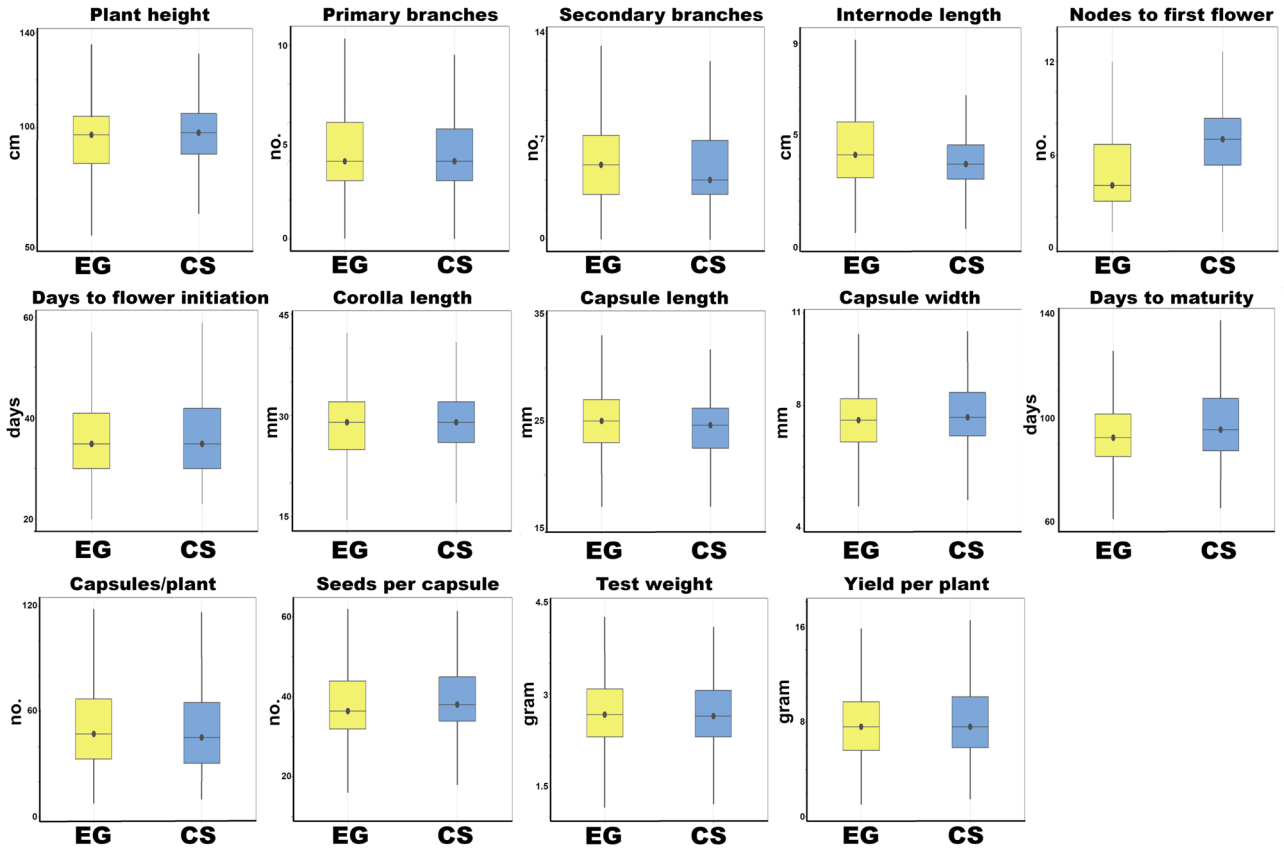


Fig. 2. Box plot representing quantitative traits in sesame's entire collection and developed core set. The trait name is given on top of each box plot, the units are given on the y-axis while x-axis represent the entire/core collection. EG stands for entire germplasm; CS is the core set.

developed core set is highly beneficial for breeders and farmers to select specific traits of their choice. We also observed that many accessions of the core set possessed multiple beneficial traits, for e.g., IC131707 (long, thick capsules, test ≥ 4 g), IC203904 (phyllody tolerant, test weight more than 4 g), IC204107 (drought tolerant, test weight ≥ 4 g), IC205495 (determinate, thick capsules, test weight more than 4 g). So, for further studies we would validate these accessions under different environmental stresses to maximize the yield, and this would be a lead for future research work. Thus, the core set developed here would form the basis of sesame crop improvement for sustainable sesame production in the near future.

Discussion

Germplasm characterization is a prior step for efficient genebank management. Genebanks are resources for crop improvement through *ex-situ* conservation and are used for sharing material with researchers and breeders. Systematic evaluation of sesame is essential for climate resilience and sustainable production. Sesame, an underutilized minor oilseed crop, has the potential to provide essential nutrients (especially oil and protein) and generate income for farmers. However, its productivity is low due to a lack of access to quality seeds. A major factor behind this is the lack of information of sesame accessions conserved in the gene bank. To fulfill this gap, we carried out a unique large-scale study to phenotypically characterize the sesame accessions at multiple environments (six locations) for two consecutive years, and developed a sesame core set of 773 accessions. Sesame germplasm exhibited a broad range of variations for all studied quantitative and qualitative agro-morphological traits (Table 1; Fig. 4). Recent studies have concentrated on sesame characterization using phenotyping data[20], against different biotic and abiotic stresses[21–24], and yield-attributing traits[25]. High genetic diversity was observed in sesame germplasm using agro-morphological traits at all locations (different agro-climatic zones). This might be due to genetic factors, light conditions, prevailing environmental variations, soil moisture conditions, suitable assimilation and translocation, among other factors. Still, there are limitations to estimate genetic diversity under the influence of environmental factors. Sesame is highly prone to shattering losses, a major limitation in sesame production. Recent studies have shown application of paclobutrazol can significantly reduce seed shattering in sesame[26]. Few recent studies, have used chemical mutagens such as sodium azide for sesame crop improvement. Plants raised from sodium azide-treated sesame seeds showed significant variability in quantitative and qualitative traits[27–29]. Genotypes Hirhir, Setit 1, and Setit 2 changed their behavior from complete to partial shattering. The determinate nature observed in a few sesame accessions is an additional advantage to farmers, as these accessions grow to a limited extent and show uniform and early

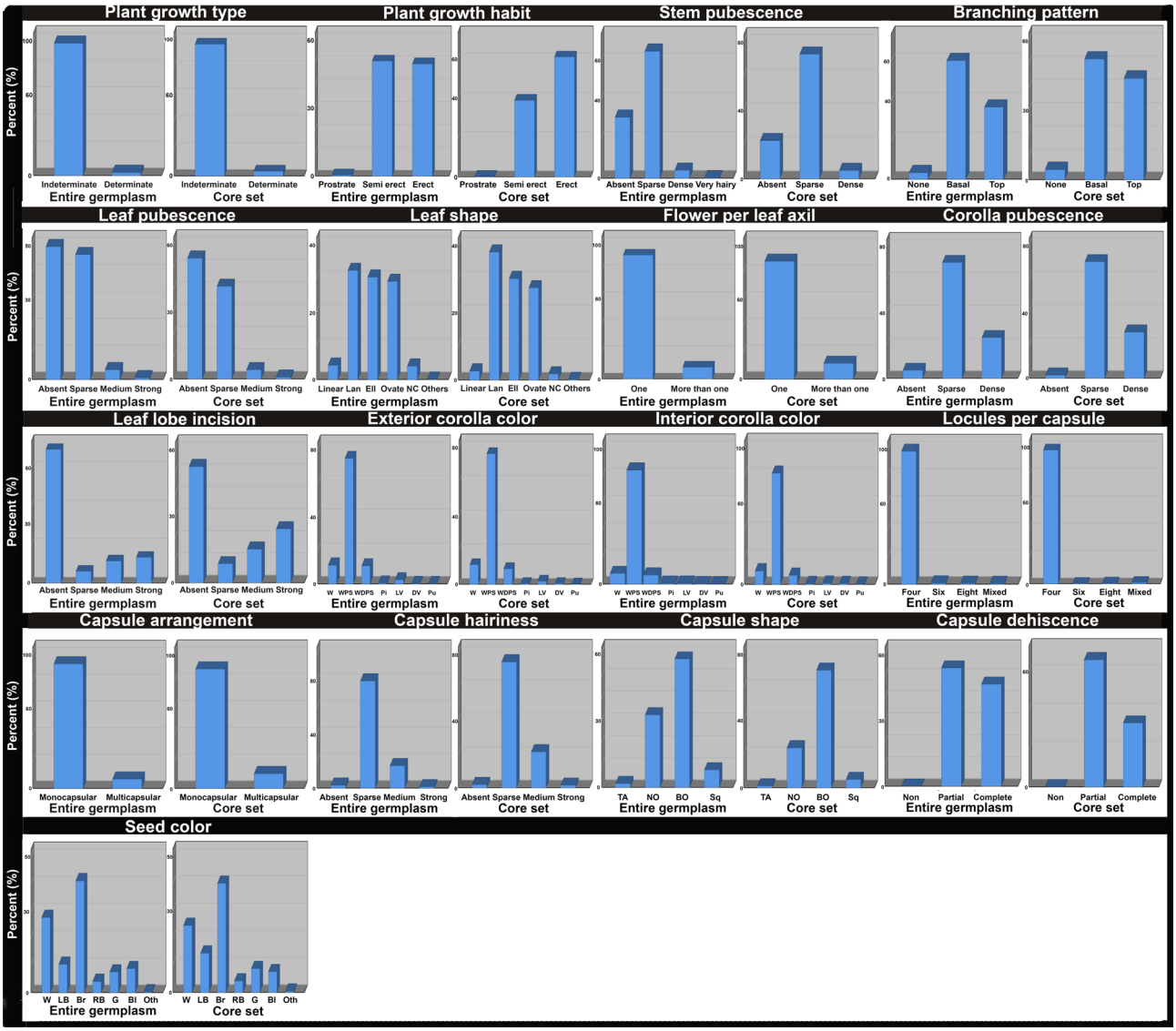


Fig. 3. Bar plot representing relative percentage of 17 qualitative traits in sesame's entire collection and developed core set. Abbreviations used; Leaf shape, Lan stands for lanceolate, Ell-elliptic, NC-narrowly cordate; Corolla color, W-white, WPS-white with pink shade, WDPS-white with deep pink shade, Pi-pink, LV-light violet, DV-deep violet, Pu-purple; Capsule shape, TA-tapered at apex, NO-narrow oblong, BO-broad oblong, Sq-square; Seed color, W-white, LB-light brown, Br-brown, RB-reddish brown, G-grey, Bl-black, Oth-others. The trait name is given on top of each bar plot, y-axis represents the relative percent while x-axis represents the variation in each traits under entire/core set.

maturity. Monostem or non-branching sesame accessions show uniform maturity and are easy for mechanical harvesting. As compared to branching accessions, monostem accessions might allow high planting density and would give higher yields and productivity. Regarding capsule arrangement, one capsule generally exists at each leaf axil, but a few accessions reported in this study had two to three capsules per leaf axil (Table 3). This trait can be desirable to farmers or breeders and has potential application for higher yield and productivity. Similarly, accessions with less internode length and two to three capsules per leaf axil are a preferred choice for breeders to get improved yields. Seed coat color shows linkage with seed quality, disease resistance, and seed biochemistry. In a previous study, black seed coat color was found to be linked with 20 candidate genes[30]. Furthermore, black seeds are a rich source of edible melanin especially for the food industry and other biotechnological applications.

Here, we have chosen accessions with stable traits at all locations for core set development. Such accessions have the genetic inheritance to remain unaffected by environmental variations and would be of great use to plant breeders for developing sesame improvement programs. The relatively high coefficient of variation observed for important agro-morphological traits in sesame shows an opportunity for crop improvement through visual selection. However, crosses made based on extremes (i.e., lowest and highest value of traits) might not always result in high yields or better performance. Previous studies recommended choosing genotypes with considerable genetic distance and good phenotypic characters for developing high yielding varieties [31]. Our

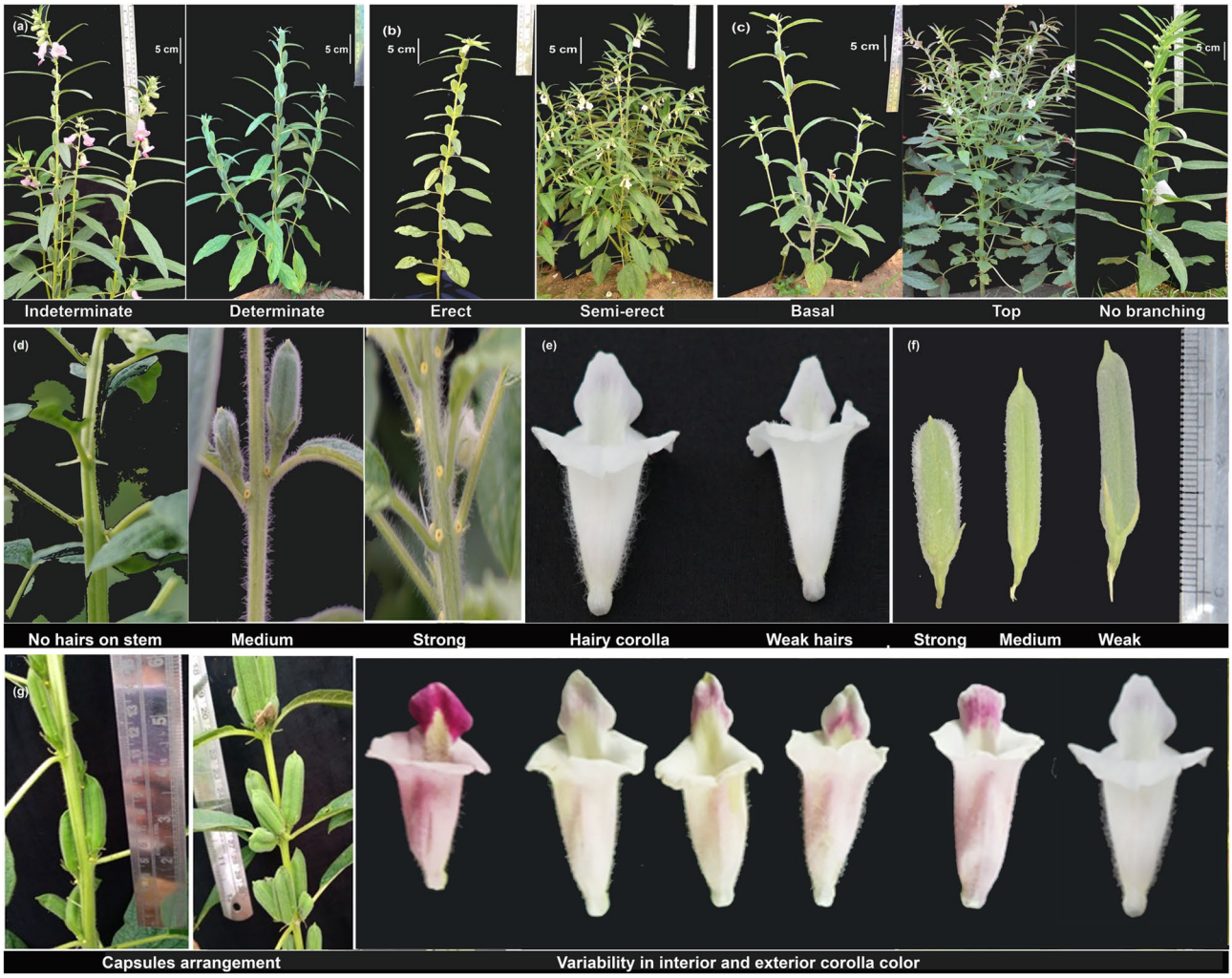


Fig. 4. Variability in different agro-morphological traits observed in core set of sesame. (**a**) Plant growth type, (**b**) plant growth habit, (**c**)branching pattern, (**d**) variability of pubescence on stem, (**e**) corolla pubescence, (**f**) variability in capsule pubescence and capsule length, (**g**) capsule arrangement, (**h**) variability in sesame flower color.

research findings indicated that accessions belonging to cluster II are related to yield-attributing traits such as yield per plant, capsules per plant, and test weight (Fig. 5). Therefore, hybridization programs may be developed with accessions from cluster II for achieving high yield potentials.

The formation of sesame core collection from a large set of germplasm, studied under six environmental locations across India, will provide breeders and researchers with a choice to pick a specific set of germplasm for effective utilization. The developed sesame core set had maximum genetic diversity and representativeness, which is useful for (i) genebank curators, geneticists, and taxonomists, and (ii) breeders[32]. The core set developed here is an excellent resource to fully utilize sesame germplasm. The core set aims to provide sesame breeders with a manageable set of accessions (indigenous as well as exotic collection). Our sesame core set is a good core collection having no redundant entries, representing the whole collection, covering all geographic regions, being smaller in size (12.8% of the whole collection), and easily managed or evaluated for its further utilization. While sesame core collection were also developed by previous researchers, they were restricted to a small number of accessions or traits (Table S1). Most of the earlier core collection studies have used summary statistics, Shannon–Weaver diversity index, H', Chi-square, PCA, and correlations to evaluate the developed core set[33,34]. Here, we have considered all such statistical analyses for the sesame core set evaluation. Hence, we concluded that the developed sesame core set is an excellent representation of the entire sesame germplasm collection currently available in National Genebank of India.

A recently published study by co-authors performed molecular diversity analysis based on SSR markers on a large set of sesame germplasm, and a molecular core of 196 sesame accessions was generated based on phenotypic variability (three major clusters)[35]. The developed molecular core set represented sufficient genetic variation of 2,496 sesame accessions. The 18 common accessions for phenotypic and molecular core are potential and can be used directly for sesame improvement programs and addressing various agricultural challenges such as high yield, easy mechanical harvest, biotic and abiotic stress tolerance. Our study lacks trait-specific investigation

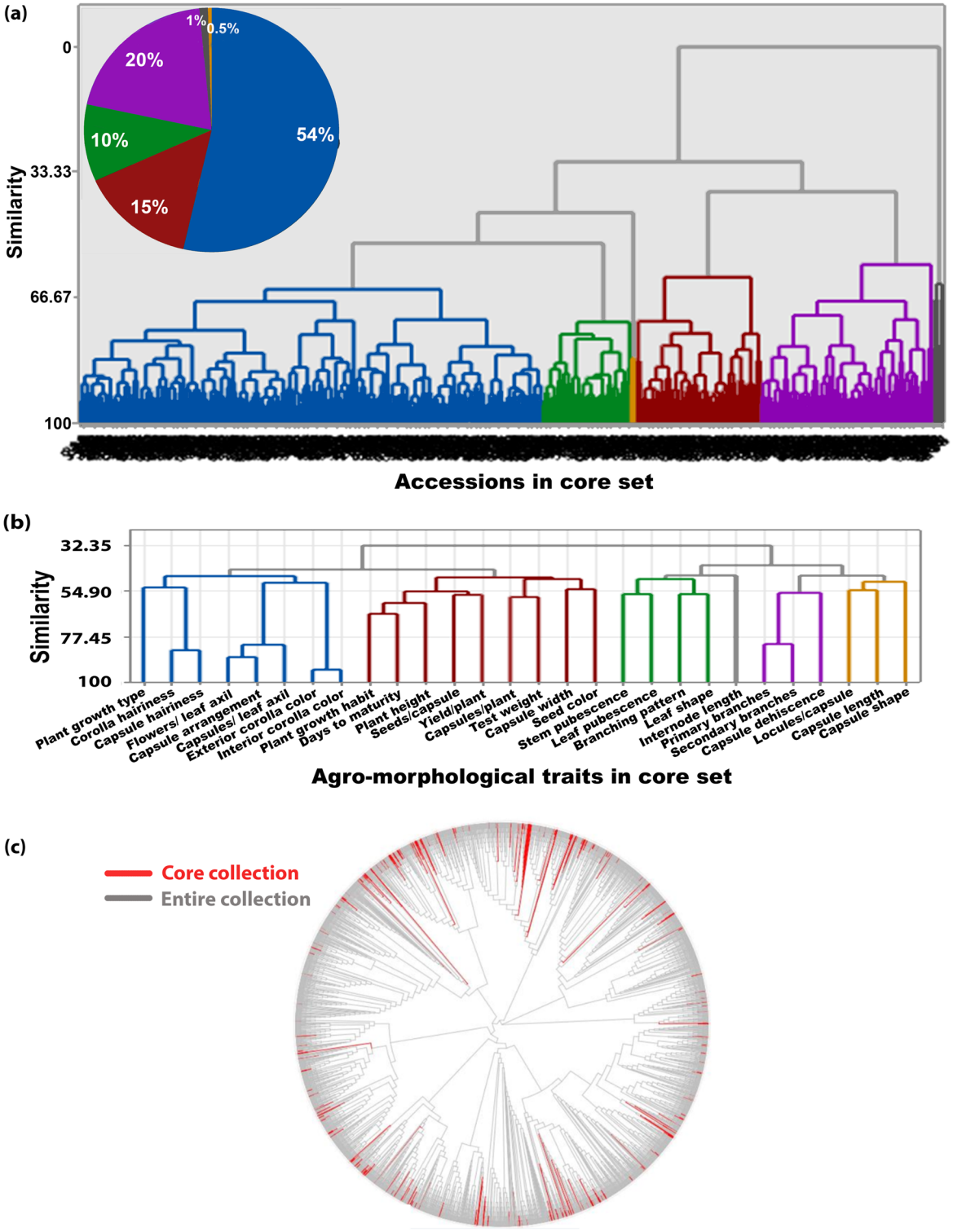


Fig. 5. Hierarchical classification of sesame core set, (**a**) based on core collection accessions, (**b**) based on core collection traits, (**c**) dendrogram representing the entire collection and core collection.

based on molecular markers such as AFLP (Amplified Fragment Length Polymorphism), RAPD (Random Amplified Polymorphic DNA), SSR (Simple Sequence Repeat), and ISSR (Inter Simple Sequence Repeat), which would be carried out in future studies. Further, ddRAD-seq approach generated 64,910 SNPs from 2,496 sesame accessions, led to the core development of 384 accessions in which 2,515 SNPs represented maximum diversity[18]. This would be an effective tool to carry out MAS (Marker Assisted Selection)-based molecular breeding program. SNPs also consist of associated genes with high oil yield, may be used to screen high oil-yielding genotypes. The

S. no	Specific traits observed in the core set	Accessions	Beneficial to the community
	Determinate	EC346370, EC346378, EC347154, EC361734, EC370617, IC127395, IC132346, IC203997, IC205495, IC43011, IC500322, IC500665, IC511132, IC511136	For synchronized maturity and harvesting
	Multicapsular	EC346663, EC346679, EC346784, EC346798, EC346802, EC347097, EC347109, EC370424, EC370639, EC542291, EC542297, IC127316, IC129647, IC43205	May be considered for more yield
	Multilocular	EC132836 (6–12), EC343405 (6), EC351884 (8), EC370367 (6,8,10), EC370818 (8), EC376975 (6,8,10), EC542295 (8,10,12,18), IC127420 (8), IC131566 (6,8), IC131812 (6,8),	Contribution towards more yield per plant
	Monostem	EC346693, EC377001, EC377081, EC378841, IC204079, IC205631, IC205715, IC306012, IC554437	For easy manual/machine harvesting and high-density planting with reduced spacing
	Early maturity (< 65 d)	IC205342, EC346949, EC347015, IC131794, IC41910, IC430435, IC511140, IC96076	For intercropping with other crops and can be planted 2–3 times/year
	Long capsules (> 40 mm)	EC346920, EC377073, IC131707, IC500873, IC52095, IC52594	More seeds per capsules for more yield
	Test weight (1000 seed weight > 4 g)	EC370338, EC376992, EC377116, EC378841, IC129676, IC130026, IC130071, IC131648, IC131707, IC132353, IC203904, IC203960, IC204107, IC204509,	Can be targeted for more seed yield
	More capsules per plant (207–240)	IC205371 (240), EC346680 (230), IC511085 (225), EC133843 (224), EC346298 (220), EC346669 (210), EC370760 (210), EC377249 (210), EC520262 (207),	Can be targeted for more yield per plant
	More seeds per capsules	IC306041 (88.4), IC500248 (82.4), EC376972 (81.1), IC204219 (80.8)	Can be targeted for more yield
	Yield per plant (g)	IC132397 (57.93 g), IC204817 (33.5 g), IC205245 (32.45 g), EC346680 (31.83 g), IC398050 (30.11 g)	More yield per plant for more oil content
	Resistance against biotic stresses- phyllody and dry root rot	Phyllody- IC500237, EC346707, EC370609, IC129276, IC129724, IC132584, IC203904, IC204757, IC205166 DRR- IC430488, IC43205, EC346321, EC346774, EC351866, EC355667, EC377073, IC129276, IC129293, IC129607, IC130047, IC131586, IC131676, IC204069,	For growing in areas having phyllody and dry root rot occurrence
	Resistance against abiotic stresses- waterlogging and drought	WL- EC334965, EC334970, EC100042, EC334981, EC346273, EC346798, EC377285, IC127268, IC127316, IC129812, IC131682, IC131790, IC204160, IC204663, Drought- EC370700, EC346378, EC346693, EC370521, EC370901, EC961169, IC127420, IC129356, IC129738, IC204065, IC204107, IC204165, IC204214, IC205524,	For growing in areas having waterlogging and drought occurrence
	Quality traits	High oil content- IC129498 (55.34%), EC377230 (54.57%), IC205669 (53.01), High protein content- IC418034 (27.23%), EC346837 (27.30%), IC413196 (25.18), IC129607 (24.64%), IC500944 (24.09), High lignan content- IC23335-1 (12.88%), IC205352 (11.08%), EC370691 (7.95%), EC346616 (6.18%), Fatty acids- Oleic acids- EC334978 (> 50%), IC204775 (> 50%); Linoleic acid- EC346209 (> 49%), EC370924 (> 49%); Stearic + palmitic acid- IC500428 (< 11%), EC346297 (< 11%)	Higher oil protein, lignans, fatty acids in sesame for nutritional benefits to human consumption

Table 3. Promising traits observed in the developed core collection for plant breeders, farmers, and community benefits. *d* days, *mm* Millimetre, *g* Gram, *WL* water logging, *DRR* dry root rot.

MDPI panel consists of 2,515 SNPs can be utilized across the globe in identifying most diverse set of sesame germplasm, which is cost-effective as compared to GBS (Genotyping By Sequencing approaches). The identified core set based on multi-approach would be applicable for other crop systems for the utilization of a large set of germplasm.

Previous studies have developed core collection in different plant species such as millets, wheat, lentil[36]. Recent studies on linseed initially prepared four to five core sets and finally narrowed down to one, which would represent the maximum genetic diversity of the overall germplasm[3]. These core sets mainly focused on the agro-morphological features only, but the sesame core set developed here shows trait specific accessions based on agro-morphological traits, accessions tolerant against biotic stress (phyllody and dry root rot), abiotic stress (waterlogging stress and drought), seeds quality traits (oil content, protein content, lignans, fatty acids). So, the developed core set is a multipurpose core set, first of its kind, having features for effective utilization of sesame crop improvement. Another criterion for core-collection evaluation is to examine spatial distribution of entries through principal components analysis (PCA). The first five components in our entire collection contributed to substantial genetic variability, which was reflected completely in the developed core set. We discussed here that principal components were preserved in our sesame core collection, and it has been reported in several recent studies[33]. In a previous report on sesame, first three principal components contributed to 56.18% genetic variability. A total of 3,129 sesame accessions were characterized based on 19 agro-morphological trait descriptors. Using principal component strategy, accessions were narrowed down to 100 that could represent whole spectrum of the entire variability. In addition, similar trends of core collection representation have been observed in mungbean, sesame, safflower, and barley[33,34,37,38]. Previous studies in sesame developed a mini core set containing 184 accessions based on 14 agro-morphological traits[4]. Previous studies were limited to one of more aspects in developing core set of sesame. There have been no reports on working core collection set development from a large set of sesame germplasm conserved in India's National Genebank. The developed core set shows high genetic diversity and maximum representation of the entire collection. Further, this working core collection set can be used to identify key genes involved in trait-specific pathways, stress tolerance genes, and their linked pathways as well as high-quality traits to improve oil, protein, and lignan content.

Methods

Plant material and experimental sites

In this study, we used sesame germplasm conserved at the Indian National Genebank (INGB) of the Indian Council of Agricultural Research -National Bureau of Plant Genetic Resources (ICAR-NBPGR), New Delhi, India. The collection consisted of 6,000 sesame accessions, including landraces, traditional cultivars, and primitive cultivars. The germplasm collection comprises sesame accessions from India (indigenous, 64.29%) and other countries (exotic, 35.71%). The germplasm accessions were purified through single-plant selection at the Tamil Nadu Agricultural University, Regional Research Station (RRS), Vridhachalam, Tamil Nadu, India. After removing any duplicates, 6,000 unique accessions were used for further phenotypic characterization. The information regarding sesame-based collection conserved in the National gene-bank is available at the Plant Genetic Resources (PGR) portal (http://pgrportal.nbpgr.ernet.in/). The progeny seeds from purified single plants were then phenotypically characterized for two consecutive years (2020–21, 2021–22) at six different locations across India including ICAR-NBPGR, New Delhi; Punjab Agricultural University (PAU), Ludhiana; Project Coordinating Unit, Jawaharlal Nehru Krishi Vishwavidyalaya (PCU, JNKVV), Jabalpur; Agriculture University (AU), Mandor, Jodhpur, Rajasthan; Agricultural Research Station (ARS), Amreli, Gujarat; Regional Research Station (RRS), Vridhachalam, Tamil Nadu, India to get the complete expression of germplasm genetic potential at different environmental conditions.

The selected six experimental locations represent five distinct agro-climatic zones of India, with the following weather conditions during the sowing period. For trans-Gangetic plains (Zone 6- Delhi and Punjab), the semi-arid regions is characterized by hot and humid conditions. Temperature typically ranged from 25℃–42℃, with an annual rainfall of 70–125 cm. The western plateau and hills zone (Zone 9- Jodhpur) experienced temperatures between 24℃ and 41℃, and an annual rainfall of 25–75 cm. Gujarat plains (Zone 13- ARS Amreli) have mix of hot and humid weather, with intermittent thunderstorms. Average temperature ranged from 27 °C to 34 °C, and rainfall averaged ~ 24 cm. Central plateau and hills zone (Zone 8- PCU JNKVV, Jabalpur), the region is semi-arid to sub-humid, with temperatures ranging from 26℃–40℃, and an annual rainfall of 50–100 cm. East coast plains and hills (Zone 11- RRS, Vridhachalam) have hot and humid conditions with average temperature of 26℃–37℃, rainfall with notably low at 8 cm, while humidity remained high at ~ 70%.

Experimental design and field experiment

The experiment was conducted across all locations using an Augmented Block Design (ABD). This design involved 600 blocks, each containing thirteen entries: three check varieties and ten sesame accessions. Six check varieties, viz. RT346 (Rajasthan), PBTil2 (Punjab), GT10 (Gujarat), VR1 (Tamil Nadu), VRI2 (Tamil Nadu), and TMV7 (Tamil Nadu) were randomized and assigned to each block. The crop was sown in the month of July at all locations, with fifty seeds planted per row. Each row was 2 m long, with a row-to-row distance of 45 cm. After seedling establishment, manual weeding was performed, and plant-to-plant distance of 10 cm was maintained within rows. Standard agricultural practices, including water and nutrient management, were followed. Irrigation was applied three times, i.e., at 0–15 DAS (days after sowing), 35–65 DAS, 65–85 DAS as per standard practices. The crop was harvested manually when approximately 80% of the plants in a row attained physiological maturity. The harvested replicates were threshed separately.

Descriptor of sesame phenotyping used in the study

For phenotypic characterization of sesame germplasm, a total of thirty-three agro-morphological traits (16 quantitative traits, 17 qualitative traits) were recorded during the life cycle of sesame using IPGRI descriptors for sesame (https://cgspace.cgiar.org/server/api/core/bitstreams/2bf9f6db-ece0-4ce4-9471-e2b8661b85b8/content). To record the data for all agro-morphological traits listed in Table S2, the 'Field Book' app on tablets was used. Five healthy plants were tagged per row (P1 to P5) for detailed measurement. Measurements for traits such as plant height (cm), internode length (cm), corolla length (mm), capsule length (mm), and capsule width (mm), were taken using a scale and a digital vernier caliper. The number of seeds were counted, and test weight was taken using a seed counter with an inbuilt weighing balance.

Screening sesame accessions for nutritional quality traits

A comprehensive screening of sesame germplasm (5500 accessions) was conducted over two consecutive years (2020–21, 2021–22) across three distinct laboratories to evaluate key quality traits. Oil and protein content estimations were performed at ICAR-NBPGR, New Delhi, while lignan content was analyzed at ICAR-Indian Agricultural Research Institute (ICAR-IARI), New Delhi. Fatty acid composition was determined at PAU, Ludhiana. Prior to analysis, sesame seeds were oven-dried at 50℃ for 24 h to reduce the moisture content. The dried seeds were then ground into a fine powder, and powdered samples were stored in air-tight containers within a dessicator to preserve their integrity until nutritional quality parameters could be analyzed. All samples were analyzed in duplicate following established official and standard methods, with results expressed as the mean of these replicates. Total protein content was determined via total nitrogen estimation using the micro-Kjeldahl method[39]. A DUMAS MAX N EXCEED nitrogen analyzer was employed, and a Jone's factor of 5.3 was applied for calculation. Oil content was extracted over 24 h using a Soxhlet extractor[39] with petroleum ether (40–60 °C) as the non-polar solvent. For lignan content estimation, 2 g of powdered sesame seeds were dissolved in 100 mL of hexane and sonicated for 20 min. The mixture was then filtered, and hexane was removed using a rotary evaporator. A 1000 ppm extract was prepared, followed by filtration through a 0.22 µm syringe filter. Standard curves were prepared using three different concentrations (100, 50, and 25 ppm) of sesamol, sesamin and sesamolin. Lignan content in the samples was quantified using an Ultra-Fast Liquid Chromatography (UFLC) system (Shimadzu, Japan), equipped with a LC-20 AD pump and SPD-M20A diode array detector. A Shimadzu Shim-pack reverse phase C-8 column (4.6 mm × 250 mm × 5 µm) was used. A 10 µL of sample was

injected and eluted in an isocratic mode with methanol:water (70:30) at a flow rate of 1 mL/min for 20 min, with detection at 200 nm. Data was analyzed using LabSolutions Post-run processing software (Shimadzu, Japan), and the concentrations of sesamol, sesamin, and sesamolin (in ppm) were calculated from their respective peak areas using the established standard curves.

Fatty acid composition was determined by gas chromatography (GC) of fatty acid methyl esters (FAMEs), prepared according to the method described by Appelqvist[40]. The analysis was performed on an Agilent 7820A GC system (Agilent Technologies, Palo Alto, CA, USA), which was equipped with a flame ionization detector (FID) and a CP-Sil 88 FAME column (25 m × 0.25 mm × 0.20 µm). The oven, detector, and injector temperatures were maintained at 210 °C, 240 °C, and 230 °C, respectively. A 2 µL sample was injected with a 10:1 split ratio. Fatty acid composition was expressed as the percentage of each individual fatty acid relative to the total fatty acids present.

Screening sesame accessions for stress (biotic, abiotic) tolerance

A large-scale preliminary screening of 5,500 diverse sesame germplasm accessions for biotic stress tolerance against disease phyllody and dry root rot diseases was conducted under field conditions for two consecutive years (2020–21 and 2021–22). This screening was done at four distinct locations in India i.e., RRS, Vridhachalam; Punjab Agricultural University, Ludhiana; PCU JNKVV, Jabalpur; and ICAR-NBPGR Regional Station, Jodhpur, Rajasthan. Two specific phytoplasma groups, 16Sr II-D and 16Sr I-B were utilized for screening the germplasm. The field experiment was established using an augmented block design (ABD). Accessions exhibiting resistance or susceptibility to phyllody and dry root rot were subsequently scored under artificial conditions for two consecutive years (2020–21 and 2021–22) at ICAR-NBPGR, New Delhi, and ICAR-IARI, New Delhi.

For abiotic stress tolerance i.e., waterlogging and drought stress, 5500 sesame accessions were screened in a preliminary large-scale experiment over two years (2020- 21, 2021–22). The screening was carried out at three locations in India- Rani Lakshmi Bai Central Agricultural University, Jhansi, India; Sardarkrushinagar Dantiwada Agricultural University, Gujarat, India; ICAR-NBPGR, Regional Station, Akola, Maharashtra, India. Selected accessions showing resistance or susceptibility to waterlogging and drought stress were further evaluated under artificial conditions for two years (2020–21, 2021–22) at ICAR-IARI, New Delhi, and ICAR-National Institute of Plant Biotechnology (ICAR-NIPB), New Delhi.

Core set preparation

A core set of sesame germplasm was established based on phenotypic data (qualitative and quantitative traits), passport data, stress tolerance, and quality traits. The methodology for preparing the core set focused on two primary criteria: maximizing the coincidence rate (CR%) for range and the variable rate (VR%) for coefficient of variation. CR% indicates the adequate representation of each variable within the core set, while VR% assesses how efficiently the coefficient of variation is represented. The CR (%) and VR (%) were calculated using the following formula:

$$CR\% = \frac{1}{m}\sum_{j=1}^{m}\frac{R_c}{R_c} \times 100$$

where, Re = Range of entire collection, Rc = Range of core collection, and m = number of traits.

$$V \cdot R(\%) = \frac{1}{m}\sum_{j=1}^{m}\frac{CV_c}{CV_e} \times 100$$

where, CVe = Coefficient of variation of entire collection, CVc = Coefficient of variation of core collection, and m = number of traits.

Additionally, the mean difference percentage (MD%) and variance difference percentage (VD%) between the core set and the entire collection were calculated by the following equations as described by Hu et al.[41]:

$$MD\% = \frac{1}{m}\sum\nolimits_{j=1}^{m}\frac{\left|\underline{X}_{core} - \underline{X}_{total}\right|}{\underline{X}_{total}} \times 100$$

$$VD\% = \frac{1}{m}\sum\nolimits_{j=1}^{m}\frac{\left|\sigma^2_{core} - \sigma^2_{total}\right|}{\sigma^2_{total}} \times 100$$

The objective was to create a working core collection that accurately represents the genetic variability of the entire sesame collection. To assess dissimilarities in traits between the core and initial collections, an F-test for variances and a t-test for means were performed at a significance level (α) of 0.05. The resulting p-values quantified the significance of differences between the developed core set and the entire population. Furthermore, chi-square (χ^2) tests were performed to evaluate the similarity of distribution frequencies between the developed core set and the entire collection.

Statistical analysis

Descriptive statistics, including minimum, maximum, standard deviation, standard error, and coefficient of variation, for both the core and entire collection were computed using SPSS software (https://www.ibm.com/spss). The adjusted mean values of all quantitative traits were subsequently standardized to eliminate scale

differences, thereby ensuring equitable contribution to the analysis. These standardized quantitative data, along with qualitative traits and passport information, were then utilized for the identification of the core collection. Box plots were generated for quantitative traits, and frequency distribution plots were prepared for qualitative traits using SPSS software to facilitate and compare between the entire collection and the developed core set. The homogeneity of variance between the entire collection and the developed core set was assessed using Levene's test for equality of variance. The Shannon–Weaver diversity index was carried out to compare the diversity levels of the entire collection and developed core set. Means of the entire collection and the core set were compared using the Newman-Keuls test. A map illustrating the Indian states and foreign countries contributing to the sesame core set diversity was created using DIVA-GIS software (https://www.diva-gis.org/). Principal component analysis (PCA) was performed using SPSS software to identify the maximum variability among sesame traits, with only components possessing eigenvalues greater than 1 selected as principal components. Finally, hierarchical clustering was conducted using Minitab software (https://www.minitab.com/en-us/).

Conclusions

This study summarizes the large-scale screening of sesame germplasm conserved at the National Genebank of India. Here, a multipurpose sesame core collection of 773 accessions was developed, based on 6 multi-location environments (five agro-climatic zones). The developed core set was based on agro-morphological traits, seed quality traits, and stress tolerance, which reflects the strength of this study. CR% and VR% for CV were maximum for core set preparation. The developed core set has captured maximum genetic diversity and representativeness of the entire sesame germplasm collection. The core set provides information about important and novel agro-morphological traits and includes accessions tolerant to biotic and abiotic stresses. It also comprises accessions with desirable seed quality parameters, such as high oil content, high protein content, high lignans, and high fatty acid content. The study lacks marker-based (AFLP, RAPD, SSR, and ISSR) trait-specific investigation, which would be carried out in future studies. Furthermore, DNA/RNA sequencing of the developed core collection, along with multiple 'Omics' tools, would reveal the key genes associated with trait-specific accessions. The significant findings from this core set will be utilized by farmers, breeders, and researchers in the field of oilseed crops, especially sesame, to develop new cultivars and improve productivity. This would help address global food security concerns by contributing to the SDGs (Sustainable Development Goals- no poverty, zero hunger, good health and well-being) in the near future.

Recommendations for future research work

- Novel traits identified in the core set would be effectively utilized by researchers, breeders and farmers.
- Development of trait-specific mapping, RIL (Recombinant Inbred Lines) population and advanced genetic studies.
- Breeding approaches should be focused on developing non-shattering sesame varieties, early maturing, high-yielding, determinate growth type, and easy mechanical harvesting.
- Pan-genome sequencing approach in sesame accessions showing specific-traits.
- Omics-based (proteomics, transcriptomics, and metabolomics) platform should be utilized more to get deep insight into the important agro-morphological traits of the sesame. CRISPR/Cas-9 is a new technique in the field of sesame to develop potential genotypes.

In-depths into key genes and pathways involved in biotic and abiotic stress tolerance in sesame.

Data availability

The data is submitted to IBDC (Indian Biological Data Centre) under project ID PPJ_100097. The corresponding author can be requested for the data.

Received: 28 April 2025; Accepted: 16 September 2025
Published online: 23 March 2026

References

1. Yadav, R. et al. Current research trends and prospects for yield and quality improvement in sesame, an important oilseed crop. *Front. Plant Sci.* **13**, 9863521. https://doi.org/10.3389/fpls.2022.863521 (2022).
2. Myint, D., Gilani, S. A., Kawase, M. & Watanabe, K. N. Sustainable sesame (Sesamum indicum L.) production through improved technology: An overview of production, challenges, and opportunities in Myanmar. *Sustainability.* **12**, 3515. https://doi.org/10.3390/su12093515 (2020).
3. Yadav, R. K., Mishra, J., Shakil, N. A. & Rana, V. S. Nutraceutical uses and phytochemistry of *Sesamum indicum*: a review. *Med. Plants* **15**, 1–14. https://doi.org/10.5958/0975-6892.2023.00005.9 (2023).
4. Ruankham, W. et al. Sesamin and sesamol attenuate H_2O_2-induced oxidative stress on human neuronal cells via the SIRT-1-SIRT3-FOXO3a signaling pathway. *Neutr. NeuroSci.* **24**, 90–101. https://doi.org/10.1080/1028415X.2019.1596613 (2019).
5. Oliveira, M. F., Nelson, R. L., Geraldi, I. O., Cruz, C. D. & Toledo, J. F. Establishing a soybean germplasm core collection. *Field Crops Res.* **119**, 277–289. https://doi.org/10.1016/j.fcr.2010.07.021 (2010).
6. Archak, S. et al. 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**(5), 417–424. https://doi.org/10.1016/j.cj.2016.06.013 (2016).
7. Kumar, S. et al. Utilization of molecular, phenotypic, and geographical diversity to develop the compact composite core collection in the oilseed crop, safflower (Carthamus tinctorius L.) through maximization strategy. *Front. Plant Sci.* **7**, 1554. https://doi.org/10.3389/fpls.2016.01554 (2016).
8. Nanjundan, J. et al. Development of Indian mustard [Brassica juncea (L) Czern] core collection based on agro-morphological traits. *Genet. Resour. Crop Evol.* **69**, 145–162. https://doi.org/10.1007/s10722-021-01211-7 (2021).

9. Kim, K. W. et al. Powercore: a program applying the advanced m strategy with a heuristic search for establishing core sets. *Bioinformatics* **23**(16), 2155–2162. https://doi.org/10.1093/bioinformatics/btm313 (2007).
10. Thachuk, C. et al. Core hunter: an algorithm for sampling genetic resources based on multiple genetic measures. *BMC Bioinform.* https://doi.org/10.1186/1471-2105-10-243 (2009).
11. Jeong, S. et al. GenoCore: A simple and fast algorithm for core subset selection from large genotype datasets. *PLoS ONE* **12**(7), e0181420. https://doi.org/10.1371/journal.pone.0181420 (2017).
12. De Beukelaer, H. & Davenport, G. F. Corehunter: Multi-purpose core subset selection. r package version 3.2.1. https://cran.r-project.or g/pack- age=corehunter (2018).
13. Xiurong, Z. et al. Establishment of sesame germplasm core collection in China. *Genet. Resour. Crop Evol.* https://doi.org/10.1023/A:1008767307675 (2000).
14. Zhang, Y. et al. Genetic diversity assessment of sesame core collection in China by phenotype and molecular markers and extraction of a mini-core collection. *BMC Genet* https://doi.org/10.1186/1471-2156-13-102 (2012).
15. Park, J. H. et al. Development and evaluation of core collection using qualitative and quantitative trait descriptor in sesame (Sesamum indicum L.) germplasm. *Korean J. Crop Sci.* **60**(1), 75–84. https://doi.org/10.7740/kjcs.2014.60.1.075 (2015).
16. Weldemichael, M. Y. & Gebremedhn, H. M. Research advances and prospects of molecular markers in sesame: a review. *Plant Mol. Biol. Rep.* **17**, 585–602. https://doi.org/10.1007/s11816-023-00853-6 (2023).
17. Ruperao, P. et al. A pilot-scale comparison between single and double digest rad markers generated using GBS strategy in sesame (Sesamum indicum L). *PLoS ONE* **18**(6), e0286599. https://doi.org/10.1371/journal.pone.0286599 (2023).
18. Ruperao, P. et al. (2024) Double-digest restriction-associated DNA sequencing-based genotyping and its applications in sesame germplasm management. *Plant Genome* **1**, 16. https://doi.org/10.1002/tpg2.20447 (2024).
19. Belaj, A. et al. Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. *Tree Genet. Genomes* **8**, 365–378. https://doi.org/10.1007/s11295-011-0447-6 (2012).
20. Mohanty, T. A., Singh, U. K., Singh, S. K., Singh, D. & Kushwaha, N. Assessment of genetic diversity in sesame (Sesamum indicum L) based on agro-morphological traits. *Curr. J. Appl. Sci. Technol* **39**(25), 100–107. https://doi.org/10.9734/cjast/2020/v39i2530890 (2020).
21. Sehgal, A. et al. Effects of drought, heat and their interaction on the growth, yield and photosynthetic function of lentil (*Lens culinaris* Medikus) genotypes varying in heat and drought sensitivity. *Front. Plant Sci.* **8**, 1776. https://doi.org/10.3389/fpls.2017.01776 (2017).
22. Li, D. et al. Biochemical and physio-morphological indices for sesame drought resistance germplasm selection. *Oil Crop Sci.* **3**(2), 99 (2018).
23. Gopika, K. et al. Physiological traits and indices to identify tolerant genotypes in sesame (Sesamum indicum L) under deficit soil moisture condition. *Plant Physiol. Rep.* **27**, 744–754. https://doi.org/10.1007/s40502-022-00701-9 (2022).
24. Anusha, L. P. et al. Phenotypic characterization of sesame (Sesamum indicum L) revealed promising genotypes for moisture stress conditions. *Plant Genet. Resour.* **21**(3), 218–228. https://doi.org/10.1017/S1479262123000667 (2023).
25. Baraki, F., Gebregergis, Z., Belay, Y., Berhe, M., Teame, G., Hassen, M. et al. 2020 Multivariate analysis for yield and yield related traits of sesame (*Sesamum indicum* L.) genotypes. *Heliyon* **6**, 1–8. https://doi.org/10.1016/j.heliyon.2020.e05295
26. Ahmed, J. et al. Shattering and yield expression of sesame (Sesamum indicum L.) genotypes influenced by paclobutrazol concentration under rainfed conditions of Pothwar. *BMC Plant Biol.* **23**, 137. https://doi.org/10.1186/s12870-023-04145-7 (2023).
27. Weldemichael, M. Y. et al. Effect of sodium azide on quantitative and qualitative stem traits in the M2 generation of Ethiopian sesame (Sesamum indicum L.) genotypes. *Sci. World J.* **6660711**, 13. https://doi.org/10.1155/2021/6660711 (2021).
28. Weldemichael, M. Y. et al. Generating better leaf traits in M2 lines of fourteen Ethiopian sesame (Sesamum indicum L.) genotypes through the treatment of their seeds with sodium azide. *Heliyon* **8**(12), e11984. https://doi.org/10.1016/j.heliyon.2022.e11984 (2022).
29. Weldemichael, M. Y. et al. Treatment of seeds with sodium azide for quantitative and qualitative capsule traits at M2 generation of fourteen Ethiopian sesame (Sesamum indicum L.) genotypes. *Heliyon* **10**(9), e12912. https://doi.org/10.1016/j.heliyon.2023.e12912 (2023).
30. Wang, L. et al. Transcriptome dynamics during black and white sesame (Sesamum indicum L.) seed development and identification of candidate genes associated with black pigmentation. *Genes* **11**(12), 1399. https://doi.org/10.3390/genes11121399 (2020).
31. Pavan, S. et al. Recommendations for choosing the genotyping method and best practices for quality control in crop genome-wide association studies. *Front. Genet.* **11**, 00447. https://doi.org/10.3389/fgene.2020.00447 (2020).
32. Marita, J. M., Rodriguez, J. M. & Nienhuis, J. Development of an algorithm identifying maximally diverse core collections. *Genet. Resour. Crop Evol.* **47**, 515–526. https://doi.org/10.1023/A:1008784610962 (2000).
33. Kaur, V. et al. Multi-environment phenotyping of linseed (Linum usitatissimum L.) germplasm for morphological and seed quality traits to assemble a core collection. *Industrial Crop Prod.* **206**, 117657. https://doi.org/10.1016/j.indcrop.2023.117657 (2023).
34. Schafleitner, R. et al. The AVRDC– the world vegetable center mungbean (*Vigna radiata*) core and mini core collections. *BMC Genomics* **16**, 334. https://doi.org/10.1186/s12864-015-1556-7 (2015).
35. Maurya, R. et al. Molecular diversity studies and core development in Sesame germplasm (Sesamum indicum L) using SSR markers. *Plant Mol. Biol. Rep.* **43**, 180–196. https://doi.org/10.1007/s11105-024-01476-z (2024).
36. Upadhyaya, H. D., Dwivedi, S. L., Singh, S. K., Vetriventham, M. & Sharma, S. Forming core collections in Barnyard, Kodo, and little millets using morpho-agronomic descriptors. *Crop Sci.* **54**, 2673–2682. https://doi.org/10.2135/cropsci2014.03.0221 (2014).
37. Mahajan, R., Bisht, I. & Dhillon, B. Establishment of a core collection of world sesame (Sesamum indicum L.) germplasm accessions. *SABRAO J. Breed. Genet.* **39**(1), 53–64 (2007).
38. Kang, C. W. et al. Selection of a core collection of Korean Sesame germplasm by a stepwise clustering method. *Breed. Sci.* **56**(1), 85–91. https://doi.org/10.1270/jsbbs.56.85 (2006).
39. AOAC.http://www.aoac.org/aoac_prod_imis/AOAC/Publications/Official_Methods_of_Analysis/AOAC_Member/Pubs/OMA/AOAC_Official_Methods_of_Analysis.aspx
40. Appelqvist, L. A. Rapid methods of lipid extraction and fatty acid methyl ester preparation for seed and leaf tissue with special remarks on preventing the accumulation of lipid contaminants. *Ark Kemi* **28**, 551–570 (1968).
41. Hu, J., Zhu, J. & Xu, H. M. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Genet.* **101**, 264–268. https://doi.org/10.1007/s001220051478 (2000).

Acknowledgements

The authors would like to thank the Director, NBPGR, New Delhi, India, for providing laboratory facilities to conduct the research

Author contributions

R.Y.: conceptualization, writing, review, editing, project administration. H.V.: Writing original draft, Data analysis. S.S.: Data analysis. A.M.: Sesame study under field conditions. P.R.: review and editing. A.K.: reviewing. S.A., K.S., K., D.K.A.: data recording. I.R., S.R.K., V.G., R.B.: sesame study under field conditions. R.P., V.R.: artificial screening of sesame. K.T., A.K.S.: sesame study under drought stress conditions. U.B.A.: review and editing.

C.C.V., G.S., P.S., P.K., K.N.G., G.P.R., H.R., K.S.: sesame study for biotic stress. S.L., V.S.R., NAS, S.S.: oil, protein, lignan, fatty acid estimation. A.S., R.C., S.S.G.: sesame study under abiotic stress. K.P.: review and editing. S.R.: sesame phenotyping and waterlogging stress. M.S., A.J, R.J.: provided seeds from genebank. A.K., K.S., G.P.S., S.K.: project administration. A.R.: data analysis, review, and editing. All authors reviewed the manuscript.

Funding

This work was supported by the Department of Biotechnology, Ministry of Science and Technology, Government of India under Grant BT/Ag/Network/Sesame/2019–20 to RY.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-20647-w.

Correspondence and requests for materials should be addressed to R.Y. or A.R.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026