scientific reports

Check for updates

OPEN

Leveraging genetic resource diversity and identification of trait-enriched superior genotypes for accelerated improvement in linseed (*Linum usitatissimum* L.)

Vikender Kaur[1]✉, Sunil S. Gomashe[2], Shashank K. Yadav[1], Devender Singh[1], Sheela[1], Shubhendra Singh Chauhan[1], Vinay Kumar[1], Balram Jat[1], Nandan Ramesh Tayade[2], Sapna Langyan[1], Nutan Kaushik[3], Mamta Singh[1], Munisha Kheralia[1], Dhammaprakash Pandhari Wankhede[1], J. Aravind[1], Vartika Srivastava[1], Kavita Gupta[1], Ashok Kumar[1] & Gyanendra Pratap Singh[1]

Linseed or flaxseed, native to the Indian subcontinent, had undergone domestication, edaphic selection and evolutionary processes that may have resulted in huge genetic variability in Indian genotypes. To understand the hitherto unexplored genetic diversity for sustainable flaxseed production amid challenges of climate fluctuation and identify trait-specific high-yielding genotypes, 2576 unique linseed accessions were comprehensively evaluated for 36 traits for up to six environments representing two major agroecological zones in India. A wide range of variability was recorded for days to initiation of flowering (42.86–114.99), plant height (43.31–122.88 cm), capsules/plant (64.62–375.87), seed size (6.06–14.44 cm^2), thousand seed weight (2.80–11.86 g), seed yield (2.93–17.28 g/plant), oil content (30.14–45.96%) and fatty acid profile especially the key constituent omega-3 fatty acid (25.4–65.88%). Most of the traits such as plant height, flowering time, seed yield, seed and capsule size showed a high or moderately high level of variance coupled with high broad sense heritability indicating precise capturing of less heritable quantitative traits. The infraspecific classification of the tested collection revealed the seed/oil type (2498 accessions) as the dominant morphotype over dual-purpose/fiber flax (78 accessions) in the conserved collection. Correlation analysis indicated a significant positive association between flowering time, plant height, days to maturity and oil content. Trait-specific superior genotypes for earliness (50% flowering in < 60 days, maturity in < 122 days), bold seeds with high thousand seed weight (> 11 g), capsules/plant (> 350), oil content (> 45%) and fatty acid composition (> 65% alpha-linolenic acid) were identified to aid genetic improvement of linseed and to broaden the narrow genetic base.

Keywords Fiber flax, Genetic diversity, India, Linseed genetic resources, Trait-specific genotypes

Linseed or flaxseed (*Linum usitatissimum* L.) is an annual self-pollinated crop known to be domesticated since prehistoric times for fiber and oil/seed purposes[1]. Linseed oil and seeds are laden with high industrial and nutraceutical value[2,3]. In addition, it is a potential strategic crop for natural fiber used in making linen. Presently,

[1]Indian Council of Agricultural Research-National Bureau of Plant Genetic Resources (ICAR-NBPGR), New Delhi, India. [2]Indian Council of Agricultural Research-National Bureau of Plant Genetic Resources (ICAR-NBPGR), Regional Station-Akola, Akola, Maharashtra, India. [3]Amity Food and Agriculture Foundation, Amity University Uttar Pradesh, Gautam Buddha Nagar, Noida, UP, India. ✉email: kaurvikender@gmail.com; vikender.kaur@icar.gov.in

India ranks seventh in terms of production and eleventh in terms of exports while Canada has been the leading producer and exporter of flax over the past decade[4,5].

In linseed, the plant architecture evolved into two distinct morphotypes—'flax type' (erect statured, 80–120 cm plant height, sparsely branched) and seed/oil type (highly branched bushy plant architecture having low plant height ranging between 40 and 60 cm) owing to preferential cultivation or human involved selection[6–8]. However, recently another intermediate morphotype, the dual-type (where both seeds and stems can be commercially utilized) has been emphasized for commercial exploration for double-purpose[9–11]. The oil-type linseed has been cultivated in India for a long and therefore it is expected to have gone through long-term edaphic selection, domestication and evolutionary process[12]. Northwest India, identified as a central Asiatic center, is the region of origin for linseed, exhibiting significant biological diversity within the genus *Linum*[13,14]. The significance of the Indian flaxseed germplasm lies in its contribution to various economically important characteristics such as thousand seed weight[15,16], high mucilage content[17], flowering time, early maturity[6,16,18,19], low ALA[20,21], efficient root system architecture, drought tolerance[22,23], salt tolerance[24,25], resistance to insect pest bud fly (*Desyneura lini* Barnes)[26,27], and nutritional quality[28], which has gained global recognition. Hence, the collections of linseed germplasm from naturally diverse regions like India are anticipated to contain essential genetic variations, serving as a valuable resource for the exploration of complex traits and enhancement of crop productivity. The National Genebank of India conserves around 2800 *Linum* germplasm accessions which were taken for comprehensive multi-environmental phenotyping in the present study. After removing the duplicate accessions, misidentified and wild germplasm, a total of 2576 unique accessions of cultivated linseed diverse in place of collection, country of origin and biological status were finally retained for field evaluation. These accessions underwent evaluation for agro-morphological characteristics and seed quality traits across various seasons at two distinct locations representing the primary agro-ecological zones for linseed cultivation in India. The specific objectives for the present work were (i) to get insights into the magnitude and nature of prevailing genetic diversity, (ii) infraspecific classification to categorize the germplasm accessions into different morphotypes suited to different end users and stakeholders, and (iii) identify trait-specific promising genotypes that would constitute a useful resource and better guidance for accelerated genetic resource utilization in linseed improvement.

Results

Phenotypic variability in qualitative traits

A broad range of qualitative trait variability was observed in the linseed germplasm collection under evaluation (Fig. 1). The Shannon–Weaver diversity index (H') ranged from 0.398 (flower shape) to 1.478 (anther colour). The growth habit and plant architecture were characterized as bushy in 507 accessions, erect in 314 accessions and semi-erect in 1755 accessions (Table S1). Most of the accessions (66.46%) showed no lodging tendency and only 3.61% were highly prone to lodging. The phenological trait expression was quite fascinating (Fig. 2). The flower colour showed a distinct variation for various tonalities of blue (55.12%), dark blue (7.76%), light blue (0.85%), purple (29.27%), violet (1.55%) and white (5.43%). Similarly high variation was observed for other floral traits such as flower aestivation, shape, flower venation colour, anther and stamen colour (Fig. 2). The most common type of germplasm accessions was characterised by semi-erect plant type (68.13%), funnel shaped flowers (87.31%) with blue colour corolla (55.12%), twisted aestivation (63.78%), blue stamens (58.27%) and cream-coloured anthers (48.12%) (Fig. 1 and Table S1). For seed traits, the predominant seed coat colour was brown (67.24% of accessions) followed by dark brown (25.43%) and light brown in 6.37% of accessions while yellow-coloured seeds were recorded in less than 1% germplasm (25 accessions) only. The seeds were lustrous in most of the germplasm accessions (83.42%). The diversity in seed coat colour and lustre is shown in Fig. 3.

Diversity assessment for quantitative traits, heritability estimates and genetic advance

An analysis for variance (ANOVA) for the respective environment was done independently to see the phenotypic expression in each location/environment. Most of the studied traits except CA, CC, EPV, NDVI, PH and TPH were heterogenous for error variances across different environments for phenotyping of germplasm in the present study. Therefore, data for heterogenous traits was transformed and combined data analysis was done. Pooled ANOVA revealed that the means for each trait were significantly different among the examined collection for all 26 quantitative parameters (Table S2). Also, the mean differences were significant ($P \leq 0.01$) between germplasm accessions and check varieties for all the characteristics except DPM. The mean, range, and coefficients of variation (phenotypic and genotypic) based on adjusted means of 2576 accessions for major quantitative traits for each environment are shown in Table 1. A wide range of variability was recorded among the accessions as evident from descriptive statistics (Table 1) and Violin cum box plots presenting the distribution of quantitative trait variation (Fig. 4). The highest PCV and GCV were recorded for CPP and SY in all the six environments whereas least PCV and GCV were documented for DPM in all environments. The mean and range of trait expression were comparable in all four years in the Delhi and Akola locations which fall under the eco-geographical Zone II and Zone III of linseed cultivation in India, respectively. The pooled data across all tested environments for 26 quantitative traits revealed that PCV and GCV were high for SY, TPH, TSW, CPP, STE, OLE, LIN, EPV, SA and low for CD, DPM, SB, SL and OC (Table 2). The phenological traits showed a wide variation ranging from a minimum of 41.4 days to a maximum of 119.92 days for initiation of flowering (DF5) at the Delhi location while a lower range (29.69–87.82 days) was seen at the Akola location. Similar observations were recorded for other phenological traits such as DF50, flowering completion (DF95) and attainment of physiological maturity. The mean flowering time for DF50 was around 80–85 days at Delhi which was higher than Akola (60–61 days). For the PH trait, the range (34.77–124.35 days) and mean (63.35–75.95) was higher in the Delhi location in comparison to the range (14.96–75.12 cm) and mean PH (38.41–45.64 cm) recorded at Akola location. These results indicate conspicuous differences in plant phenotypic expression between Delhi

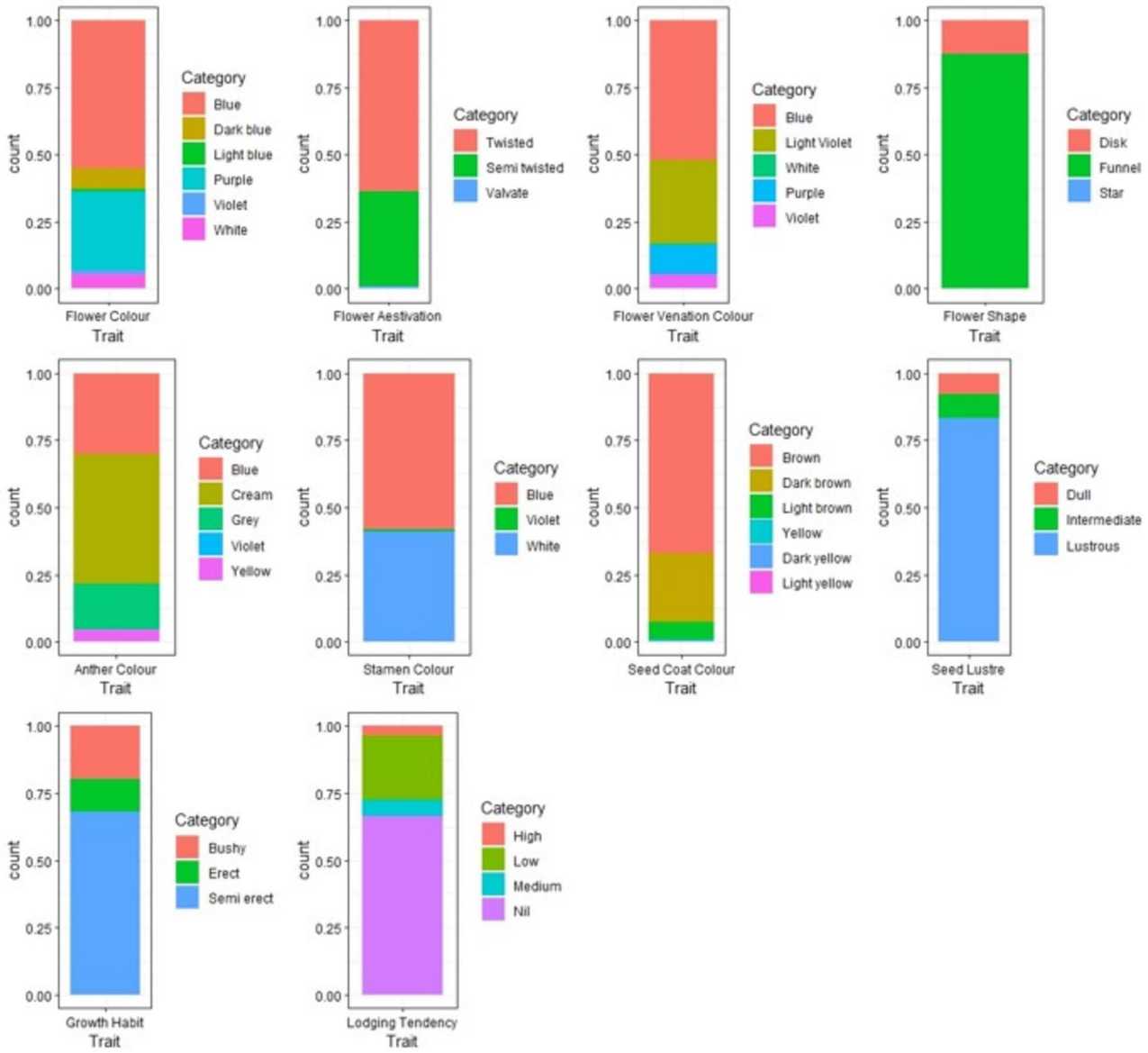

Fig. 1. Stacked bar plots showing frequency distribution for 10 qualitative traits in linseed germplasm.

and Akola regions, owing to their location in different agro-climatic regions. Overall linseed germplasm at the Akola location showed earlier flowering, maturity, shorter PH and lesser yield as indicated by low SY and its attributes (CPP, PBP, TSW) than in Delhi, although, the coefficients of variations were more or less similar in both locations/across all environments. The mean oil content was 41.30% and it ranged from a low of 30.14% in accession (IC0356252) to as high as ~45.96% in IC0510933 and IC0510943. Fatty acid profiling exhibited remarkable diversity in the studied germplasm accessions. The content of the most important omega-3 fatty acid (ALA) ranged from 25.4% (IC0498905) to 65.88% (IC0526089). Similarly, ranges were wide for other FAs such as PAL (3.36–15.61%), STE (2.37–13.65%), OLE (14.31–49.78%) and LIN (1.72–18.83%).

The difference between phenotypic and genotypic variation was low for all the traits except seed yield and its components (CPP, SY, TPH and PBP) indicating underlying genetic factors responsible for observed variation and not an environmental role. For days to maturity, the extent of variation was relatively less than that of flowering time as evident from low PCV and GCV across all tested environments. For nutritional composition, the tested germplasm revealed low PCV for oil content (6.99%) while it was medium (10.46%) for ALA to high (20.78%) for STE. Broad sense heritability ($h^2_{(bs)}$) was high for all the traits ranging from 80.13% (CPP) to 99.64% (PAL) except DPW and PBP for which it was recorded in the low (26.36%) and medium (59.35%) categories. The genetic advance was estimated low (< 10) for DPM, medium (10–20) for OC, flower/capsule and seed morphometrical traits and high (> 20) for the remaining traits. The traits such as PH, TPH, SY, flowering time, seed and capsule size showed high or moderately high variance levels in conjugation with high heritability and genetic advance.

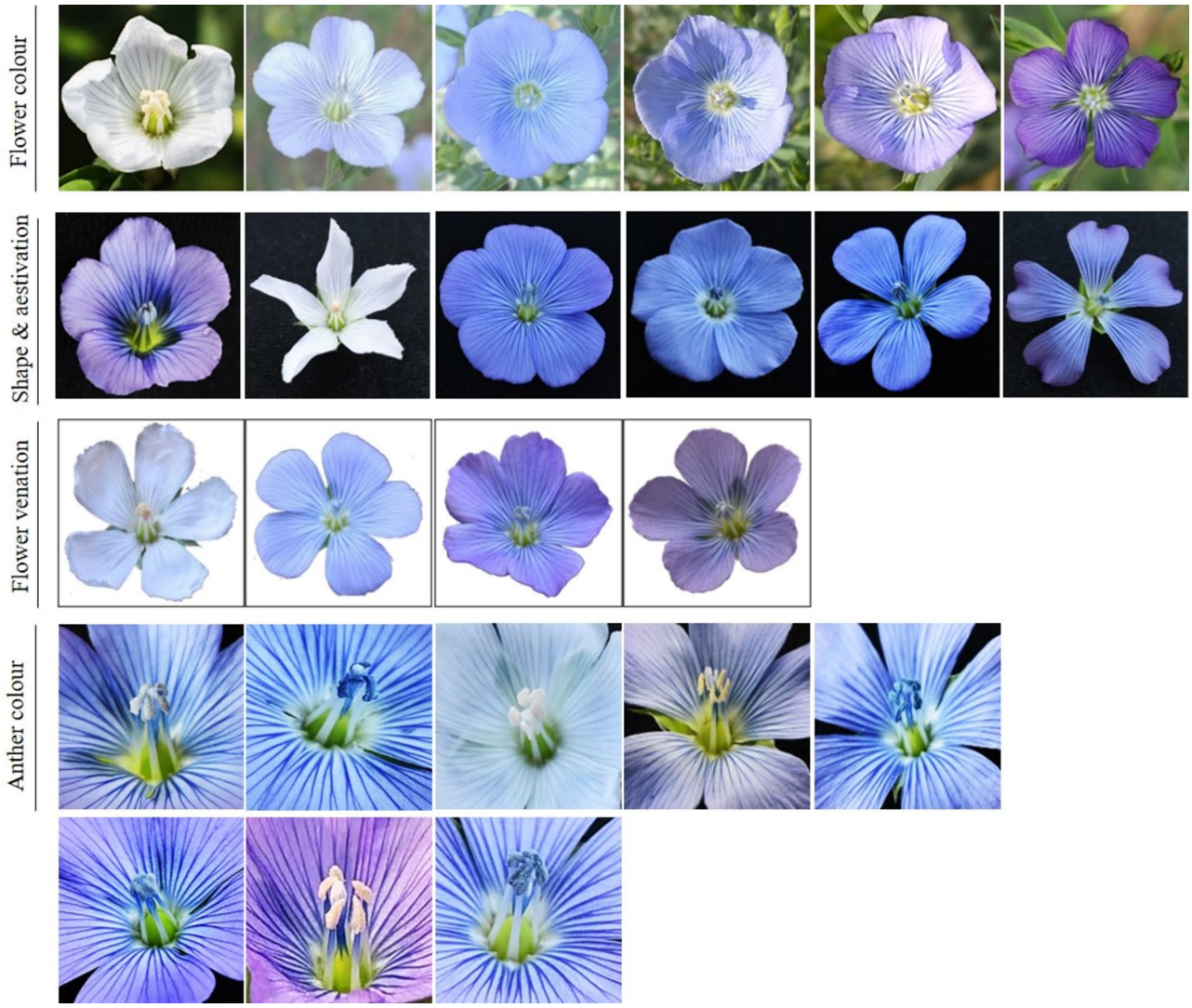

Fig. 2. Variability in linseed germplasm for flower colour—white, light blue, blue, dark blue, purple and violet; Flower shape—Funnel, Star, Disc; Corolla aestivation—twisted, Semi-twisted, Valvate; Flower Venation—white, blue, purple, violet; Anther colour—grey, blue, cream, yellow, violet; Stamen colour—blue, violet white.

Categorization of germplasm into different morphotypes

All the examined 2576 accessions were categorized in two distinct morphotypes based on plant architecture the seed/oil type linseed and dual-purpose fiber flax type. A total of 78 accessions were placed under dual-purpose fiber flax linseed based on average PH > 75 cm and TPH > 50 cm. Out of these 78 accessions, 12 accessions were classified as fiber flax type having erect growth habit, straight stem with few branches restricted to having PH > 100 cm and TPH > 50 cm. Most of the germplasm (2498 accessions) were categorized under seed/oil type linseed. The mean, range and CV (%) of the accessions grouped under these two categories are presented in Table 3. There were conspicuous differences among these two morphotype groups for plant height, phenological traits, seed size, yield and thousand seed weight. The mean plant height and TPH were 91.83 cm and 56.47 cm in dual-purpose/fiber flax type while plants were dwarf (mean PH = 67.40 cm, mean TPH = 31.76 cm) in linseed type. The accessions grouped under the seed/oil type linseed group were bold-seeded and exhibited early flowering and maturity compared to the flax-type group. For instance, the mean days taken for initiation of flowering were 75.46 in oil-type linseed whereas dual-purpose/fiber-type accessions took more than 90 days for DF5. Similarly, the attainment of 50% flowering was earlier (82.02 days for mean DF50) compared to dual/fiber type accessions (98.18 days). Although the completion of flowering was earlier (mean DF95 = 88.11 days) in oil type accessions compared to dual/fiber type (105.02 days), an almost similar number of days were taken to attain physiological maturity by both the groups (141 for oil type and 146.58 days for dual/fiber group), which indicates that dough stage was longer in oil-type accessions. Another characteristic feature of the oil-type linseed group was bold seededness as evident from more SA (9.71 cm^2) and TSW (6.72 g) as compared to dual-purpose/fiber flax accessions which had a mean SA of 8.67 cm^2 and TSW of 5.26 g. The mean SY was also comparatively more (8.51 g) than the dual-purpose/fiber group (5.26 g).

Fig. 3. Seed coat colour and lustre diversity in linseed germplasm accessions.

Identification of trait-specific promising genotypes

The trait-specific superior germplasm accessions were identified based on agronomic evaluation and oil quality profiling. The superior accessions for agro-morphological traits validated for over 4 years in the Delhi location are documented in Table 4. Only one accession (IC0499042) exhibited flowering initiation (DF5 = 42.86) and 50% flowering (DF50 = 49.63) in comparison to best check LSL93 (48.12 and 52.28 days to attain 5% and 50% flowering, respectively). Although, another accession IC0525939 exhibited early flowering initiation (46.87 days for DF5), but not DF50% per se in comparison to check LSL93. Two exotic (EC0041579 and EC0718852) and one

Environment	DF5	DF50	DF95	DPM	CS	PH	PBP	CPP	SY	TSW	SPC	CA	CD	SA	SL	SB
DEL 2018–2019																
Min	42.51	54.69	57.64	123.92	10.43	34.77	1.71	21.87	1.84	2.54	3.34	22.34	5.50	6.14	3.90	1.85
Max	111.63	125.44	129.39	158.17	34.14	120.22	11.93	427.23	22.74	12.46	10.26	53.16	8.23	15.24	6.22	3.40
Mean	76.58	85.00	93.88	143.89	18.81	64.60	5.62	189.55	10.22	6.98	7.76	36.57	6.80	10.29	5.06	2.73
SE	0.18	0.17	0.18	0.09	0.05	0.18	0.03	1.24	0.07	0.04	0.02	0.09	0.01	0.04	0.01	0.01
SD	9.07	8.84	8.90	4.65	2.58	9.21	1.37	62.81	3.68	1.79	1.27	4.81	0.45	1.85	0.44	0.26
PCV (%)	11.63	10.69	9.00	4.89	13.67	14.16	22.65	32.39	35.74	25.25	16.29	13.15	6.54	17.58	8.66	9.29
GCV (%)	10.85	8.54	8.40	4.28	11.98	13.08	17.70	29.22	34.06	24.45	13.87	12.36	5.94	16.99	7.87	6.67
DEL 2019–2020																
Min	41.40	46.81	55.43	119.02	11.03	37.84	1.34	42.89	1.85	2.58	3.59	21.29	5.14	5.85	3.80	2.06
Max	117.40	129.76	131.55	155.42	33.63	124.09	11.88	430.36	23.70	12.35	10.35	57.77	8.57	14.79	6.07	3.29
Mean	77.86	84.53	89.42	142.72	19.52	68.64	6.02	212.96	9.55	6.73	7.85	36.56	6.80	9.51	4.90	2.61
SE	0.22	0.19	0.19	0.12	0.05	0.19	0.03	1.30	0.07	0.03	0.02	0.11	0.01	0.03	0.01	0.01
SD	10.93	9.90	9.50	5.87	2.32	9.75	1.57	66.10	3.48	1.77	1.19	5.34	0.50	1.63	0.44	0.23
PCV (%)	14.15	12.02	10.63	4.11	11.53	14.13	25.29	30.50	35.95	26.24	15.30	14.58	7.23	16.98	8.81	8.66
GCV (%)	13.35	11.46	10.18	3.69	7.51	13.19	20.03	28.47	33.98	25.42	13.16	13.73	6.61	16.49	8.38	7.18
DEL 2020–2021																
Min	44.62	49.74	53.69	117.92	9.12	41.89	2.04	32.25	1.36	2.10	3.34	21.97	5.37	5.60	3.54	2.01
Max	111.99	124.86	126.06	154.42	34.42	123.82	10.91	435.00	23.12	12.04	10.26	56.04	8.43	15.12	6.28	3.26
Mean	74.98	79.94	84.38	141.72	16.15	63.35	5.43	194.96	7.95	6.38	7.76	37.18	6.85	9.28	4.82	2.59
SE	0.18	0.18	0.20	0.12	0.05	0.20	0.03	1.40	0.08	0.03	0.02	0.11	0.01	0.03	0.01	0.01
SD	9.04	9.26	10.00	5.97	2.62	9.92	1.51	71.02	3.86	1.74	1.27	5.78	0.53	1.64	0.44	0.24
PCV (%)	13.69	13.59	13.56	4.29	16.29	15.97	27.62	36.60	38.64	27.63	16.29	15.39	7.63	17.89	9.18	9.18
GCV (%)	13.09	13.16	13.07	3.90	14.61	14.88	24.99	35.23	37.58	27.17	13.87	14.57	6.58	17.10	8.42	7.92
DEL 2021–2022																
Min	42.29	46.10	51.44	121.54	11.16	38.22	1.83	43.48	1.18	2.60	3.17	21.78	5.15	5.62	3.74	1.94
Max	119.92	124.35	141.19	154.79	34.20	124.35	12.27	382.33	19.17	12.16	10.75	56.38	8.45	15.15	6.34	3.32
Mean	74.13	80.48	86.70	139.27	20.23	75.95	5.96	136.07	6.21	6.62	7.43	35.80	6.73	9.65	4.94	2.64
SE	0.21	0.20	0.20	0.18	0.06	0.23	0.03	1.09	0.05	0.03	0.03	0.10	0.01	0.03	0.01	0.01
SD	10.55	9.94	10.08	9.06	2.87	11.85	1.76	55.15	2.51	1.74	1.28	4.86	0.45	1.70	0.44	0.24
PCV (%)	14.85	13.13	11.53	6.86	14.78	15.80	30.76	39.97	40.22	26.21	16.23	13.46	6.69	17.29	8.79	8.95
GCV (%)	14.11	12.35	10.65	6.01	13.51	15.08	26.27	36.17	35.05	25.73	13.13	12.45	5.60	16.56	8.20	7.91
AK 2020–2021																
Min	31.57	34.13	36.91	51.69	8.39	14.96	1.21	6.13	0.50	1.42	1.31	–	–	–	–	–
Max	87.82	93.00	96.41	133.69	30.25	68.83	13.68	254.78	12.57	9.51	9.83	–	–	–	–	–
Mean	56.77	60.11	63.64	108.94	16.62	38.41	4.32	61.50	2.59	4.57	6.81	–	–	–	–	–
SE	0.18	0.18	0.19	0.15	0.05	0.12	0.03	0.50	0.02	0.02	0.03	–	–	–	–	–
SD	9.01	9.17	9.71	7.48	2.53	5.91	1.32	25.50	0.97	1.24	1.31	–	–	–	–	–
PCV (%)	18.09	17.27	16.65	8.74	15.38	15.41	30.73	42.50	37.45	27.63	19.24	–	–	–	–	–
GCV (%)	16.70	15.96	15.46	7.05	11.29	12.50	25.93	36.84	34.59	20.68	17.47	–	–	–	–	–
AK 2021–2022																
Min	29.69	35.39	37.79	76.71	8.11	24.14	2.39	6.47	0.70	1.37	2.18	–	–	–	–	–
Max	85.44	89.64	92.42	148.96	29.49	75.12	15.64	274.64	10.49	9.80	10.78	–	–	–	–	–
Mean	56.63	60.69	63.97	113.28	17.72	45.64	7.24	85.66	2.84	5.22	7.31	–	–	–	–	–
SE	0.18	0.17	0.17	0.18	0.06	0.13	0.04	0.62	0.02	0.03	0.03	–	–	–	–	–
SD	8.95	8.49	8.84	9.26	2.99	6.73	1.95	31.51	1.07	1.36	1.41	–	–	–	–	–
PCV (%)	17.13	15.28	14.63	8.42	17.64	15.02	28.29	40.26	37.68	27.64	19.73	–	–	–	–	–
GCV (%)	14.85	13.32	12.15	6.77	13.20	12.31	23.95	34.71	33.68	18.90	14.18	–	–	–	–	–

Table 1. Descriptive statistics of 2576 linseed accessions for major quantitative traits based on multi-environment phenotyping. DF5, Days to 5% flowering; DF50, Days to 50% flowering; DF95, Days to completion of flowering; DPM, Days to 80% maturity, CS, Size of corolla (mm); PH, Plant height (cm); PBP, No. of primary branches per plant; CPP, No. of capsules per plant; SY, Seed yield/plant (g); TSW, Seed weight of 1000 seeds (g); SPC, No. of seeds/capsule; CA, Capsule area (mm^2); CD, Capsule diameter (mm); SA, Seed area (mm^2); SL, Seed length (mm); SB, Seed breadth (mm); PCV (%), Phenotypic coefficient of variation; GCV (%), Genotypic coefficient of variation; DEL 2018–19, Delhi location (Year 2018–2019); DEL 2019–20, Delhi location (Year 2019–2020); DEL 2020–21, Delhi location (Year 2020–2021); DEL 2021–22, Delhi location (Year 2021–2022); AK 2020–21, Akola location (Year 2020–2021); AK 2021–22, Akola location (Year 2021–2022). Seed and capsule morphometrical parameters were recorded for 4 seasons in Delhi location only. Adjusted means derived from ANOVA for each environment are shown here.

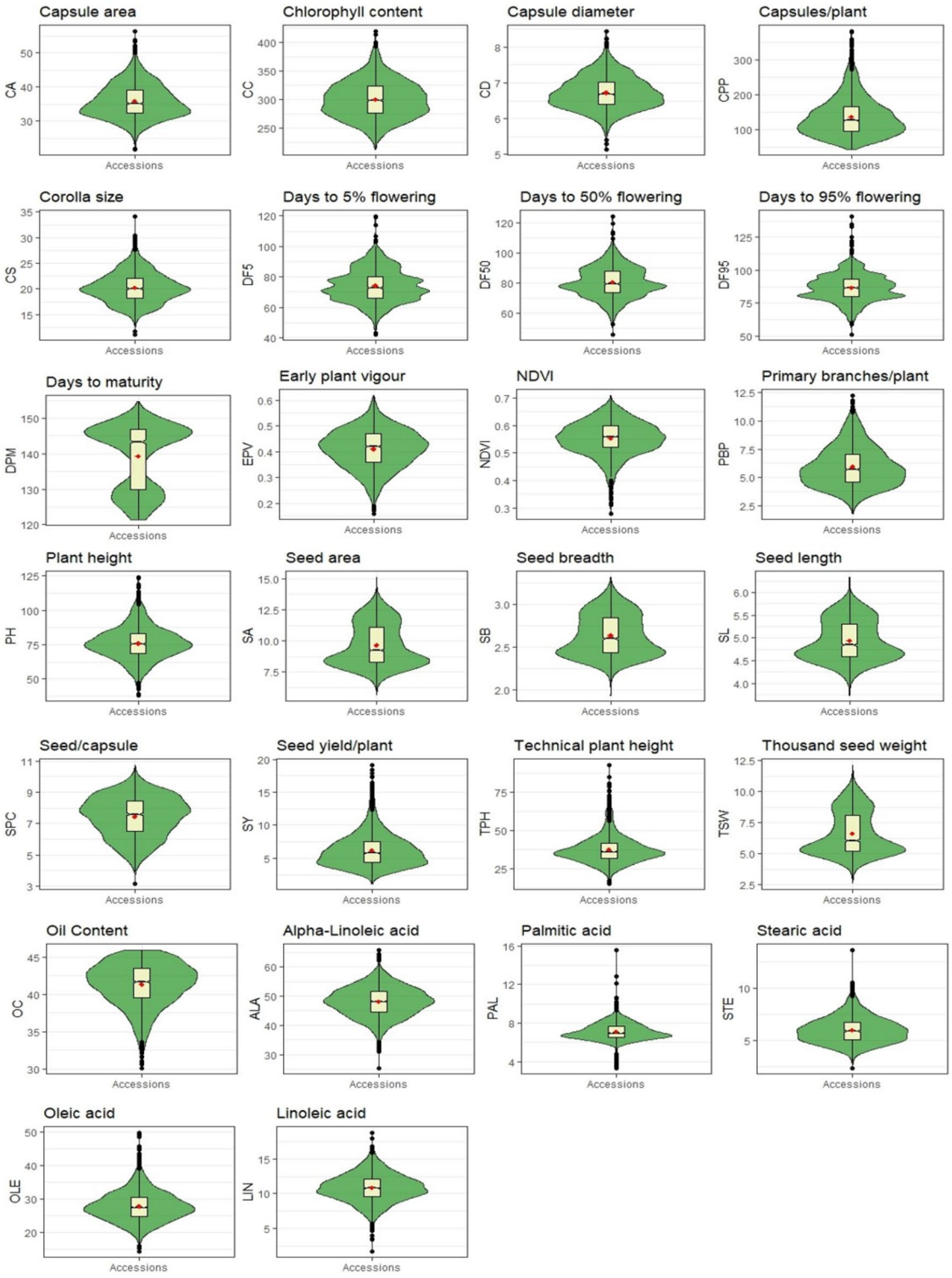

Fig. 4. Violin cum Boxplots showing the distribution of 26 quantitative traits in linseed germplasm (CA, Capsule area (mm^2); CC, Chlorophyll content (mg/m^2); CD, Capsule diameter (mg/m^2); CPP, No. of capsules per plant; CS, Corolla size (mm); DF5, Days to 5% flowering; DF50, Days to 50% flowering; DF95, Days to completion of flowering; DPM, Days to 80% maturity, EPV, Early plant vigour; NDVI, Normalized difference vegetation index; PBP, No. of primary branches per plant; PH, Plant height (cm); SA, Seed area (mm^2); SB, Seed breadth; SL (mm); Seed length (mm); SPC, No. of seeds/capsule; SY, Seed yield/plant (g); TPH, Technical plant height (cm); TSW, Seed weight of 1000 seeds (g); OC, Oil content (%); PAL, Palmitic acid (%); STE, Stearic acid (%); OLE, Oleic acid (%); LIN, Linoleic acid (%); ALA, α-linolenic acid (%). The adjusted means of data recorded for quantitative traits in the Delhi location (Year 2021–2022) are shown in plots.

Traits	Min	Max	Mean ± SE	SD	PCV (%)	GCV (%)	h^2bs (%)	h^2bs category	GA (%)	GAM (%)	GAM category
CA	23.02	53.91	36.53 ± 0.10	4.85	13.23	12.77	93.19	High	9.29	25.43	High
CC	211.40	418.37	299.50 ± 0.66	33.48	12.65	12.02	90.23	High	70.49	23.55	High
CD#	25.59	38.32	31.70 ± 0.04	2.07	6.52	6.13	88.33	High	3.76	11.88	Medium
CPP#	3.57	12.58	7.00 ± 0.02	1.27	18.50	16.56	80.13	High	2.13	30.59	High
CS#	8.16	26.30	12.56 ± 0.02	1.27	10.39	9.40	81.98	High	2.21	17.57	Medium
DF5#	11.45	33.08	19.65 ± 0.05	2.33	12.81	12.57	96.26	High	5.00	25.44	High
DF50#	12.95	34.24	20.63 ± 0.04	2.24	12.04	11.81	96.11	High	4.93	23.88	High
DF95#	15.26	41.06	23.78 ± 0.05	2.41	10.50	10.28	95.92	High	4.94	20.77	High
DPM#	24.14	35.55	27.82 ± 0.02	1.03	3.90	2.00	26.36	Low	0.59	2.12	Low
EPV	0.15	0.59	0.38 ± 0.00	0.06	17.33	16.19	87.27	High	0.12	31.21	High
NDVI	0.31	0.63	0.48 ± 0.00	0.05	10.77	9.33	75.14	High	0.08	16.69	Medium
PBP#	3.06	11.86	6.83 ± 0.02	1.07	15.33	11.81	59.35	Medium	1.28	18.77	Medium
PH	38.63	98.89	59.46 ± 0.13	6.81	11.40	10.94	92.05	High	12.87	21.65	High
SA#	13.45	31.72	21.39 ± 0.07	3.51	16.32	16.03	96.43	High	6.95	32.47	High
SB#	17.10	26.33	21.61 ± 0.04	1.86	8.46	7.86	86.39	High	3.26	15.07	Medium
SL#	24.19	39.21	31.64 ± 0.05	2.66	8.36	8.10	93.74	High	5.12	16.17	Medium
SPC#	5.65	13.71	9.53 ± 0.02	1.16	12.10	10.85	80.51	High	1.91	20.09	High
SY#	2.94	12.14	6.49 ± 0.03	1.39	21.39	19.34	81.78	High	2.34	36.08	High
TPH	13.97	89.95	32.51 ± 0.15	7.43	23.13	21.58	87.05	High	13.51	41.54	High
TSW#	6.33	25.57	13.60 ± 0.06	3.25	24.14	23.80	97.19	High	6.58	48.40	High
OC$	30.14	45.96	41.30 ± 0.06	2.90	6.99	6.78	94.04	High	5.61	13.57	Medium
ALA$	25.40	65.88	48.12 ± 0.10	5.03	10.46	9.85	88.64	High	9.20	19.12	Medium
PAL$	3.36	15.61	7.10 ± 0.02	0.89	12.58	12.56	99.64	High	1.83	25.86	High
STE$	2.37	13.65	5.98 ± 0.02	1.27	20.78	19.99	92.51	High	2.37	39.66	High
OLE$	14.31	49.78	27.72 ± 0.09	4.51	16.19	15.10	87.01	High	8.05	29.06	High
LIN$	1.72	18.83	10.84 ± 0.04	2.00	18.46	18.27	98.05	High	4.05	37.33	High

Table 2. The range, mean and estimates of genetic diversity based on pooled adjusted means of linseed germplasm accessions. CA, Capsule area (mm^2); CC, Chlorophyll content (mg/m^2); CD, Capsule diameter (mm); CPP, No. of capsules per plant; CS, Size of corolla (mm); DF5, Days to 5% flowering; DF50, Days to 50% flowering; DF95, Days to completion of flowering; DPM, Days to 80% maturity, EPV, Early plant vigour; NDVI, Normalized difference vegetation index; PBP, No. of primary branches per plant; PH, Plant height (cm); SA, Seed area (mm^2); SB, Seed breadth (mm); SL, Seed length (mm); SPC, No. of seeds/capsule; SY, Seed yield/plant (g); TPH, Technical plant height (cm); TSW, Seed weight of 1000 seeds (g); OC, Oil content (%); ALA, α-linolenic acid (%); PAL, Palmitic acid (%); STE, Stearic acid (%); OLE, Oleic acid (%); LIN, Linoleic acid (%); PCV, Phenotypic coefficient of variation (%); GCV, Genotypic coefficient of variation (%); h^2bs, Heritability in broad sense; GA, Genetic advance; GAM, Genetic advance over mean. $Data was recorded for one environment (DL2019-20) only. #The values of adjusted means calculated over transformed data for heterogenous traits are shown here.

native germplasm (IC0499136) were found to be promising for earliness (DPM < 122 days) in comparison to the best check exhibiting earliness (LSL93 for DPM = 132.19 days). For fiber flax, the genotypes namely EC0041481, EC0041495-1, EC0718827, IC0499135 and IC0597268 exhibited tall PH (> 103.06 cm) in comparison to flax varieties Tiara (90.86 cm) for and dual-purpose variety Rashmi (82.42 cm) used as checks were identified. Additionally, four of them also exhibited more technical height, a measure of stem length available for fiber extraction (> 75 cm) in EC0041495-1, EC0718827 and IC0597268 while yellow-seeded flax landrace IC0499135 from Gadchiroli district of Indian state Maharashtra had TPH > 70 cm, had significantly higher TPH compared to checks Tiara (45.41 cm) and Rashmi (46.03 cm). For fatty acid composition, the superior genotypes validated from the seed harvested from two seasons (2019–2020 and 2020–2021) at the Delhi location are shown in Table 5. The promising genotypes identified based on mean oil content from seeds harvested from 4 environments (Delhi location: Year 2019–2020 and Year 2020–2021 and Akola location: Year 2020–2021 and Year 2021–2022) are presented in Table 6. In addition, the superior accessions exhibiting high trait value for multiple traits have been identified. Table S3 enlists the multiple trait-enriched superior genotypes along with mean trait values from four environments. Six accessions (IC0629090, IC0096510, IC0499053, IC0585312, IC0096568, IC0096530) having high oil content (> 44%) also showed more capsules/plant (> 190–258) and high seed yield/plant (> 9 g), which is significantly higher than the national check T397. One genotype EC0041737 exhibited superiority for many traits such as thousand seed weight, capsules/plant, seed yield, plant height, capsule size, seed size, and moderately high oil content. Similarly genotypes IC0447870 and IC0620507 were found promising for a high number of capsules/plant, primary branches, and seed yield. Accession IC0597268 had more primary branches (> 9) and capsules/plant (190) in addition to tall plant height and technical plant height. Big capsule size and bold

Trait	Seed/Oil type linseed (n = 2498)					Dual purpose/ Fiber flax type (n = 78)				
	Min	Max	Mean	SD	CV (%)	Min	Max	Mean	SD	CV (%)
PH	43.31	101.29	67.40	7.83	11.61	77.65	122.88	91.83	7.81	8.50
TPH	13.92	52.52	31.76	5.92	18.65	50.12	89.68	56.47	7.34	13.00
CA	22.82	53.93	36.63	4.84	13.21	24.75	44.70	33.34	3.91	11.72
EPV	0.15	0.59	0.38	0.06	16.86	0.22	0.52	0.37	0.07	18.87
NDVI	0.31	0.64	0.48	0.05	10.87	0.36	0.61	0.49	0.05	9.96
CCI	34.85	95.75	61.41	6.68	10.88	48.38	78.95	61.52	6.09	9.90
CC	212.54	419.37	299.53	32.62	10.89	245.96	389.50	297.57	26.99	9.07
OC[$]	30.14	45.96	41.29	2.89	6.99	32.08	45.82	42.01	3.00	7.14
PAL[$]	3.36	15.61	7.10	0.90	12.66	5.59	8.83	7.04	0.72	10.28
STE[$]	2.37	13.65	6.00	1.24	20.70	2.94	9.76	5.36	1.17	21.82
OLE[$]	14.31	49.78	27.65	4.43	16.03	18.63	45.54	30.06	5.48	18.22
LIN[$]	1.72	18.83	10.87	1.98	18.19	5.00	16.52	9.92	2.49	25.07
ALA[$]	25.40	65.88	48.13	5.00	10.38	33.02	64.13	47.46	6.07	12.79
CD	5.47	8.22	6.80	0.45	6.55	5.66	7.50	6.53	0.38	5.75
CD[#]	25.59	38.32	31.69	2.07	6.52	26.51	35.06	30.47	1.71	5.62
CPP	64.62	375.87	183.46	39.10	21.31	95.73	309.52	184.26	42.97	23.32
CPP[#]	2.81	16.50	8.07	1.74	21.53	4.21	13.44	7.97	1.87	23.48
CS	11.88	34.10	18.61	1.85	9.97	17.18	25.41	20.77	1.74	8.40
CS[#]	9.08	26.29	14.24	1.46	10.27	13.16	19.55	16.04	1.34	8.33
DF5	42.86	114.99	75.46	8.68	11.51	70.12	110.57	90.60	7.72	8.52
DF5[#]	13.03	34.67	22.81	2.58	11.32	21.24	33.35	27.33	2.33	8.53
DF50	51.14	124.85	82.02	8.32	10.14	78.70	122.10	98.18	7.72	7.86
DF50[#]	14.71	36.07	23.67	2.42	10.24	22.76	35.29	28.37	2.25	7.92
DF95	55.80	130.55	88.11	8.25	9.37	89.03	127.22	105.02	8.21	7.82
DF95[#]	18.09	41.86	28.37	2.63	9.28	28.57	40.90	33.79	2.65	7.85
DPM	121.96	153.06	141.48	4.74	3.35	131.08	154.06	146.58	4.00	2.73
DPM[#]	28.52	36.07	33.21	1.17	3.52	30.58	36.20	34.42	1.01	2.93
PBP	2.28	9.82	5.75	0.99	17.26	3.84	9.51	6.11	1.16	19.00
PBP[#]	2.62	12.08	7.06	1.23	17.40	4.78	11.58	7.45	1.43	19.20
SA	6.22	14.44	9.71	1.60	16.46	6.06	10.94	8.67	1.00	11.51
SA[#]	13.70	31.72	21.46	3.52	16.40	13.45	24.29	19.19	2.21	11.53
SB	2.13	3.21	2.64	0.23	8.53	2.09	2.84	2.52	0.16	6.39
SB[#]	17.49	26.33	21.64	1.85	8.54	17.10	23.27	20.61	1.32	6.38
SL	3.76	6.13	4.94	0.42	8.44	4.05	5.35	4.69	0.28	5.94
SL[#]	24.19	39.21	31.69	2.66	8.40	25.93	34.43	30.11	1.81	6.00
SY	3.19	17.65	8.51	1.96	22.98	3.95	13.44	7.69	2.19	28.49
SY[#]	2.93	17.28	8.22	1.97	23.98	3.71	13.62	7.32	2.19	29.91
TSW	2.80	11.86	6.72	1.65	24.59	2.89	8.77	5.26	1.03	19.54
TSW[#]	7.45	31.69	17.91	4.42	24.68	7.70	23.42	13.99	2.70	19.28
SPC	3.54	10.05	7.66	1.04	13.62	6.44	9.35	8.34	0.70	8.40
SPC[#]	3.56	10.03	7.66	1.04	13.63	6.45	9.35	8.36	0.70	8.33

Table 3. Comparative evaluation of seed/oil type linseed and dual purpose/fiber flax type morphotypes. PH, Plant height (cm); TPH, Technical plant height (cm); CA, Capsule area (mm^2); EPV, Early plant vigour; NDVI, Normalized difference vegetation index; CCI, Chlorophyll content index; CC, Chlorophyll content (mg/m^2); OC, Oil content (%); PAL, Palmitic acid (%); STE, Stearic acid (%); OLE, Oleic acid (%); LIN, Linoleic acid (%); ALA, α-linolenic acid (%); CD, Capsule diameter (mm); CPP, No. of capsules per plant; CS, Size of corolla (mm); DF5, Days to 5% flowering; DF50, Days to 50% flowering; DF95, Days to completion of flowering; DPM, Days to 80% maturity, PBP, No. of primary branches per plant; SA, Seed area (mm^2); SB, Seed breadth (mm); SL, Seed length (mm); SY, Seed yield/plant (g); TSW, Seed weight of 1000 seeds (g); SPC, No. of seeds/capsule; SD, Standard deviation; CV, Coefficient of variation; [$]Data was recorded for one environment (DL2019-20) only; [#]The values of adjusted means were calculated over transformed data for heterogenous traits.

Genotype	Environment				Average (non-transformed)	Average (transformed)
	DEL 2018–2019	DEL 2019–2020	DEL 2020–2021	DEL 2021–2022		
Capsule area (mm^2)#						
EC0041737-Sel	46.32	57.77	49.95	56.38	52.61	–
IC0525943-Sel	52.75	53.10	56.04	53.83	53.93	–
Sharda (Check)	39.01	39.19	40.50	38.41	39.28	–
No. of capsules per plant						
IC0629219-Sel*	350.64	365.82	300.05	357.95	343.62	14.94
IC0567363-Sel	344.32	360.69	398.76	249.70	338.37	14.88
IC0498706-Sel	343.76	421.64	415.41	200.61	345.36	15.18
T397 (Check)	167.43	165.14	173.95	187.26	173.45	7.06
Days to 5% flowering						
IC0096496-Sel	52.20	52.21	50.12	56.29	52.71	16.03
IC0499042-Sel	42.51	41.40	44.62	42.92	42.86	13.03
IC0525939-Sel	46.36	45.71	50.12	45.29	46.87	14.32
Shekhar (Check)	72.50	73.75	69.32	67.79	70.84	21.41
LSL93 (Check)	-	49.86	48.96	45.54	48.12	14.82
Days to 50% flowering						
IC0096496-Sel	55.36	55.81	53.99	58.72	55.97	16.22
IC0499042-Sel	57.94	50.76	49.74	46.10	51.14	14.71
IC0525939-Sel	54.69	59.81	53.99	52.72	55.30	16.10
Shekhar (Check)	74.79	78.14	74.89	74.11	75.48	21.91
LSL93 (Check)	–	54.82	51.71	48.32	51.62	16.29
Days to 80% maturity						
IC0096496-Sel	128.00	126.04	130	122.54	126.65	30.35
EC0041579-Sel*	122.00	123.92	118.92	124.41	122.31	28.67
EC0718852-Sel	121.00	127.42	117.92	121.54	121.97	28.58
IC0499136-Sel$	121.23	126.42	117.92	121.54	121.78	28.52
T397 (Check)	140.00	140.93	141.04	141.75	140.93	33.19
LSL93 (Check)	–	130.00	134.00	132.57	132.19	36.04
Plant height (cm)#						
EC0041481-Sel	99.28	110.30	102.84	116.68	107.28	–
EC0041495-1-Sel	99.41	110.30	108.34	124.35	110.60	–
EC0718827-Sel	94.24	99.30	101.34	117.35	103.06	–
IC0499135-Sel	120.22	124.090	123.82	123.39	122.88	–
IC0597268-Sel	103.74	101.30	110.34	118.68	108.52	–
Tiara (Check)	93.93	89.00	88.50	92.00	90.86	–
Rashmi (Check)	82.20	84.60	81.00	81.87	82.42	–
Seed area (mm^2)						
EC0041478-B-Sel	14.93	13.04	13.99	13.02	13.75	30.29
EC0041737-Sel	14.51	12.99	11.97	15.15	13.66	30.13
IC0096489-Sel	14.51	13.59	13.910	13.39	13.85	30.55
IC0096733-Sel	14.57	14.16	14.10	13.40	14.06	31.10
IC0499144-Sel	13.67	13.04	14.47	14.39	13.89	30.56
JLS67 (Check)	11.78	12.17	11.78	12.18	11.98	24.64
Thousand seed weight (g)						
EC0041478-B-Sel	12.33	11.79	10.41	10.31	11.21	29.67
IC0096529-Sel	10.60	12.35	12.04	11.81	11.70	31.33
IC0499143-Sel	11.18	10.59	11.47	12.16	11.35	30.58
IC0499144-Sel	12.46	11.35	11.89	11.74	11.86	31.69
IC0525957-Sel	11.66	12.26	10.89	11.02	11.46	30.57
JLS67 (Check)	8.99	9.10	8.99	9.10	9.05	24.21
Seed yield/plant (g)						
IC0629219-Sel*	12.01	14.96	11.61	14.07	13.16	12.62
IC0620507-Sel	15.53	9.52	11.30	12.56	12.23	11.37
EC0041737-Sel	12.45	9.28	12.48	9.49	10.95	10.83
Continued						

Genotype	Environment				Average (non-transformed)	Average (transformed)
	DEL 2018–2019	DEL 2019–2020	DEL 2020–2021	DEL 2021–2022		
IC0499144-Sel	12.11	11.62	8.59	7.77	10.02	9.92
T397 (Check)	7.75	7.34	8.13	7.25	7.62	6.92

Table 4. Performance of trait-specific promising genotypes for key agro-morphological traits. CA, Capsule area (mm^2); CPP, No. of capsules per plant; DF5, Days to 5% flowering; DF50, Days to 50% flowering; DPM, Days to 80% maturity; PH, Plant height (cm); SA, Seed area (mm^2); TSW, Seed weight of 1000 seeds (g); SY, Seed yield/plant (g); DEL 2018–19, Delhi location (Year 2018–2019); DEL 2019–20, Delhi location (Year 2019–2020); DEL 2020–21, Delhi location (Year 2020–2021); DEL 2021–22, Delhi location (Year 2021–2022). #represents the traits (Capsule area and Plant height) homogenous for error variance. *denotes white flowered germplasm and $indicates yellow-seeded linseed. The performance of superior genotypes is shown in comparison to the overall best check and or best Zonal check for the respective trait.

Fatty acid composition# (%)								
Genotype	Palmitic acid	Stearic acid	Oleic acid	Linoleic acid	α-Linolenic acid	Growth habit	Flower colour	Seed coat colour
IC0096572-Sel	7.08	7.62	43.60	10.76	30.94	Semi-erect	Purple	Brown
IC0498613-Sel	6.63	4.34	17.64	7.68	63.72	Bushy	Blue	Brown
IC0498640-Sel	4.03	3.31	15.81	12.97	63.89	Semi-erect	White	Brown
IC0498664-Sel	6.29	3.88	18.44	7.71	63.68	Semi-erect	Dark blue	Brown
IC0498706-Sel	4.29	3.37	23.91	4.77	63.66	Semi-erect	Purple	Dark brown
IC0498759-Sel	5.79	3.48	20.94	5.65	64.13	Semi-erect	Purple	Brown
IC0498905-Sel	15.61	10.41	36.74	11.83	25.40	Erect	White	Dark brown
IC0498952-Sel	4.48	3.80	20.78	6.67	64.28	Semi-erect	Purple	Dark brown
IC0526089-Sel	5.06	2.85	19.00	7.22	65.88	Erect	Blue	Brown
T397 (Check)	6.12	6.54	37.83	8.70	40.34	Bushy	Blue	Brown

Table 5. Promising genotypes validated for fatty acid composition in different morphological background. #Indicates fatty acid composition validation from seeds harvested from two environments in Delhi location (Year 2019–2020 and Year 2020–2021).

Genotype	Oil content$	Growth habit	Flower colour	Seed coat colour
IC629090-Sel	45.24	Semi erect	Purple	Light brown
IC096510-Sel	44.41	Semi erect	White	Brown
IC499020-Sel	45.82	Semi erect	Blue	Dark brown
IC118868-Sel	44.57	Semi erect	Purple	Dark brown
IC499053-Sel	44.60	Bushy	Purple	Brown
IC096510-Sel	44.20	Semi erect	White	Brown
IC498659-Sel	43.84	Semi erect	Purple	Dark brown
EC041601-A-Sel	44.09	Semi erect	Dark blue	Dark yellow
IC585360-Sel	45.54	Semi erect	Purple	Yellow
IC585312-Sel	45.12	Bushy	Blue	Brown
IC268341-Sel	44.76	Bushy	Purple	Dark brown
IC385355-Sel	44.53	Erect	Purple	Brown
IC096568-Sel	45.79	Semi erect	Purple	Brown
IC096530-Sel	44.32	Semi erect	White	Brown
T397 (Check)	42.29	Bushy	Blue	Brown

Table 6. Promising genotypes validated for oil content in multiple environments. $Indicates average oil content (%) from seeds harvested from four environments (Delhi location: Year 2019–2020 and Year 2020–2021 and Akola location: Year 2020–2021 and Year 2021–2022).

	OC	PAL	STE	OLE	LIN	ALA	CA	CPP	CS	DF50	DPM	PH	SA	SPC	SY	TPH
PAL	−0.09***															
STE	−0.20***	0.47***														
OLE	0.10***	0.03	0.05*													
LIN	−0.11***	0.40***	0.29***	−0.31***												
ALA	0.03	−0.46***	−0.48***	− 0.76**	−0.24***											
CA	−0.26***	0.02	0.41***	−0.12***	0.06**	− 0.03*										
CPP	0.05**	− 0.02	−0.13***	0.06**	− 0.02	− 0.01	−0.25***									
CS	0.04*	− 0.06**	− 0.04	0.05*	−0.13***	0.02	0.16***	− 0.05*								
DF50	0.19***	− 0.03	−0.19***	0.14***	−0.13***	− 0.01	−0.41***	0.13***	0.16***							
DPM	0.13***	− 0.01	−0.11***	0.12***	−0.10***	− 0.04*	−0.25***	0.10***	0.15***	0.68**						
PH	0.04*	−0.08***	−0.12***	0.12***	−0.11***	− 0.01	0.03	0.06**	0.43***	0.37***	0.35***					
SA	−0.29***	0.04*	0.43***	−0.12***	0.04*	− 0.03	0.90**	−0.26***	0.12***	−0.46***	−0.27***	− 0.02				
SPC	0.23***	− 0.02	−0.33***	0.11***	− 0.02	0.01	−0.50***	0.15***	0.05**	0.38***	0.25***	0.19***	−0.66***			
SY	− 0.01	− 0.02	0.04*	0.04	0.01	− 0.04*	0.20***	0.60***	0.03	−0.11***	−0.03***	0.12***	0.14***	0.03*		
TPH	0.07***	− 0.05*	−0.10***	0.12***	−0.13***	− 0.02	− 0.05**	− 0.01	0.40***	0.51***	0.41***	0.79**	−0.09***	0.19***	0.01	
TSW	−0.30***	0.00	0.42***	−0.12***	0.03***	− 0.02	0.89***	−0.25***	0.09***	−0.52***	−0.31***	−0.07***	0.93***	−0.63***	0.19***	− 0.15***

Table 7. The correlation matrix for the key quantitative traits in linseed germplasm accessions. PAL, Palmitic acid (%); STE, Stearic acid (%); OLE, Oleic acid (%); LIN, Linoleic acid (%); ALA, α-linolenic acid (%); CA, Capsule area (mm^2); CPP, No. of capsules per plant; CS, Size of corolla (mm); DF50, Days to 50% flowering; DPM, Days to 80% maturity; PH, Plant height (cm); SA, Seed area (mm^2); SPC, No. of seeds/capsule; SY, Seed yield/plant (g); TPH, Technical plant height (cm); TSW, Seed weight of 1000 seeds (g); OC, Oil content (%); *significance at $P \leq 0.05$; **significance at $P \leq 0.01$; ***significance at $P \leq 0.001$.

seeds with high thousand seed weight (> 11 g) were recorded in the genotypes namely EC0041478-B, IC0096529, IC0499143, IC0499144, and IC0525957.

Correlation

Pairwise correlation coefficients having significant relationships are shown in (Table 7 and Fig. S1). As expected, the capsule area was highly correlated ($r = 0.90$, $P \leq 0.01$) with the seed area indicating bold capsule had bold seeds. Seed size displayed a negative correlation with CPP ($r = -0.26$, $P \leq 0.01$), DPM ($r = -0.27$, $P \leq 0.01$), DF50 ($r = -0.46$, $P \leq 0.001$) and oil content ($r = -0.29$, $P \leq 0.001$). A similar association was visualized for capsule size with these traits. PH and TPH exhibited a significant association of moderate magnitude ranging from $r = 0.35–0.51$ with corolla size, days to attain 50% flowering and maturity. For SY and related components (CA, SA, CPP) and PH showed a positive correlation. As expected TSW showed a strong positive correlation ($r = 0.89$, $P \leq 0.01$) with CA and SA ($r = 0.93$, $P \leq 0.001$) and a low to moderate correlation of negative magnitude with DF50, CPP, SPC. Oil content was positively correlated with DF50, DPM and SPC while a negative correlation of low to moderate magnitude was observed with CA, SA, TSW, STE and LIN. The most important fatty acid constituent, ALA had a negative correlation of magnitude ranging from $r = -0.24$ to $r = -0.76$ with other fatty acids.

Principal component analysis (PCA)

Principal Component Analysis of 26 agro-morphological and seed quality traits showed no significant clustering of the accessions in the 3-dimensional space on plotting of the first three principal coordinate data (Fig. S2). However, when the accessions were grouped according to their collection sites into the different zones of cultivation, Zone II (556 accessions from Indian states Delhi, Bihar, Jharkhand, West Bengal, Assam, Nagaland, Meghalaya, Sikkim, Uttar Pradesh excluding Bundelkhand region) and Zone III (901 accessions from Madhya Pradesh, Bundelkhand region of Uttar Pradesh, Rajasthan, Chhattisgarh, Odisha, Maharashtra, Karnataka, Telangana, Andhra Pradesh and Tamil Nadu) could be distinguished from each other despite significant overlap. Accessions of Zone I (98 accessions from Himachal Pradesh, Punjab, Haryana, Jammu and Kashmir and Uttarakhand), exotic accessions (143) as well as those of unknown origin (859 Indian and 19 exotic) could not be distinguished into clusters. Around 75% of quantitative trait variation is contributed by the first seven principal components

Principal component	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7
Eigenvalue	7.95	3.38	2.27	1.85	1.61	1.31	1.23
Proportion of variance	30.58	13.02	8.75	7.13	6.20	5.03	4.73
Cumulative proportion	30.58	43.59	52.34	59.47	65.66	70.69	75.43
CA	0.307	0.201	− 0.021	0.002	0.056	− 0.041	0.082
CC	− 0.055	− 0.032	0.069	0.015	0.179	− 0.095	0.440
CD	0.303	0.210	− 0.019	0.003	0.054	− 0.036	0.088
CPP	− 0.101	− 0.015	0.000	0.589	0.174	− 0.015	0.152
CS	− 0.001	0.285	− 0.097	− 0.054	− 0.172	0.228	0.290
DF5	− 0.254	0.291	0.056	− 0.085	0.186	− 0.151	− 0.090
DF50	− 0.262	0.298	0.050	− 0.091	0.172	− 0.136	− 0.079
DF95	− 0.264	0.304	0.042	− 0.089	0.152	− 0.108	− 0.080
DPM	− 0.182	0.284	0.052	− 0.077	0.127	− 0.109	0.002
EPV	0.080	0.059	− 0.098	0.304	− 0.191	0.240	− 0.393
NDVI	0.048	0.127	− 0.007	0.251	− 0.027	0.079	− 0.586
PBP	− 0.137	0.129	0.005	0.263	0.342	− 0.257	− 0.108
PH	− 0.077	0.377	− 0.077	0.025	− 0.200	0.359	0.153
SA	0.324	0.179	− 0.017	− 0.027	0.060	− 0.090	0.019
SB	0.314	0.189	− 0.018	− 0.045	0.079	− 0.114	0.002
SL	0.322	0.175	− 0.029	− 0.023	0.035	− 0.077	0.036
SPC	− 0.237	− 0.042	0.002	0.043	− 0.093	0.210	0.188
SY	0.048	0.073	− 0.019	0.596	0.095	0.065	0.281
TPH	− 0.111	0.397	− 0.052	− 0.039	− 0.164	0.308	− 0.009
TSW	0.326	0.134	− 0.031	0.004	0.031	− 0.084	0.045
OC	− 0.115	− 0.022	− 0.055	0.011	− 0.162	− 0.067	0.076
ALA	− 0.015	− 0.054	− 0.586	− 0.092	0.283	0.092	− 0.026
PAL	0.026	− 0.021	0.486	− 0.038	0.157	0.271	− 0.019
STE	0.166	0.071	0.428	− 0.031	0.126	0.032	− 0.063
OLE	− 0.059	0.089	0.321	0.142	− 0.532	− 0.362	0.054
LIN	0.046	− 0.102	0.292	− 0.056	0.344	0.466	− 0.005

Table 8. Relative proportion of variance and traits explained by principal component analysis in the evaluated linseed collection. CA, Capsule area (mm^2); CC, Chlorophyll content (mg/m^2); CD, Capsule diameter (mm); CPP, No. of capsules per plant; CS, Size of corolla (mm); DF5, Days to 5% flowering; DF50, Days to 50% flowering; DF95, Days to completion of flowering; DPM, Days to 80% maturity, EPV, Early plant vigour; NDVI, Normalized difference vegetation index; PBP, No. of primary branches per plant; PH, Plant height (cm); SA, Seed area (mm^2); SB, Seed breadth (mm); SL, Seed length (mm); SPC, No. of seeds/capsule; SY, Seed yield/plant (g); TPH, Technical plant height (cm); TSW, Seed weight of 1000 seeds (g); OC, Oil content (%); ALA, α-linolenic acid (%); PAL, Palmitic acid (%); STE, Stearic acid (%); OLE, Oleic acid (%); LIN, Linoleic acid (%). PC_1 to PC_7 denote principal components 1–7.

(PCs) with PC_1 and PC_2 contributing 30.58% and 13.02% variation, respectively (Fig. S3 and Table 8). The component wise PC loading in Fig. S3 revealed that the largest proportion of variation was explained by capsule and seed morphometrical traits (CA, CD, SA, SB, SL) and TSW in PC_1. The maximum variation towards PC_2 was impacted by phenological traits (DF5, DF50, DF95, DPM), PH, TPH, CS and NDVI. All the fatty acids except ALA contributed to PC_3 whereas seed and oil content (OC, CPP and SY) were explained in PC_4.

Discussion

Presently, 59,786 germplasm accessions of *L. usitatissimum* and 1129 accessions belonging to different wild *Linum* species are conserved in genebanks/institutes worldwide[29]. The National Genebank of India conserves one of the largest base collections of *Linum* with around 2800 germplasm accessions, of which 2576 are the unique accessions of cultivated linseed which have been comprehensively phenotyped for 10 qualitative and 26 quantitative traits in multi-year-location environments in the work described here. This large-scale characterisation provided the first insights into the prevailing genetic diversity in predominantly Indian origin germplasm as around 94% of the evaluated germplasm (2414 accessions) are of Indian origin (Fig. 5, Table 3). Prior to the present report, such large-scale characterisation has been reported mainly for the Plant Gene Resources of Canada (PGRC) repository that conserves around 3551 accessions of cultivated flax and 152 accessions of wild *Linum* species[30]. Many diversity assessment studies were conducted in PGRC flax collection under numerous projects conducted over the years from 1998 to 2008 and a wide range of variations for important traits, such as the onset of flowering (37–69 days), plant height (17–130 cm) and thousand seed weight (2.8–11.5 g), were reported[15,30–32]. Similarly, Worku et al.[33] evaluated the Ethiopian collection, Zhuchenko and Rozhmina[34] evaluated flax landraces for fiber

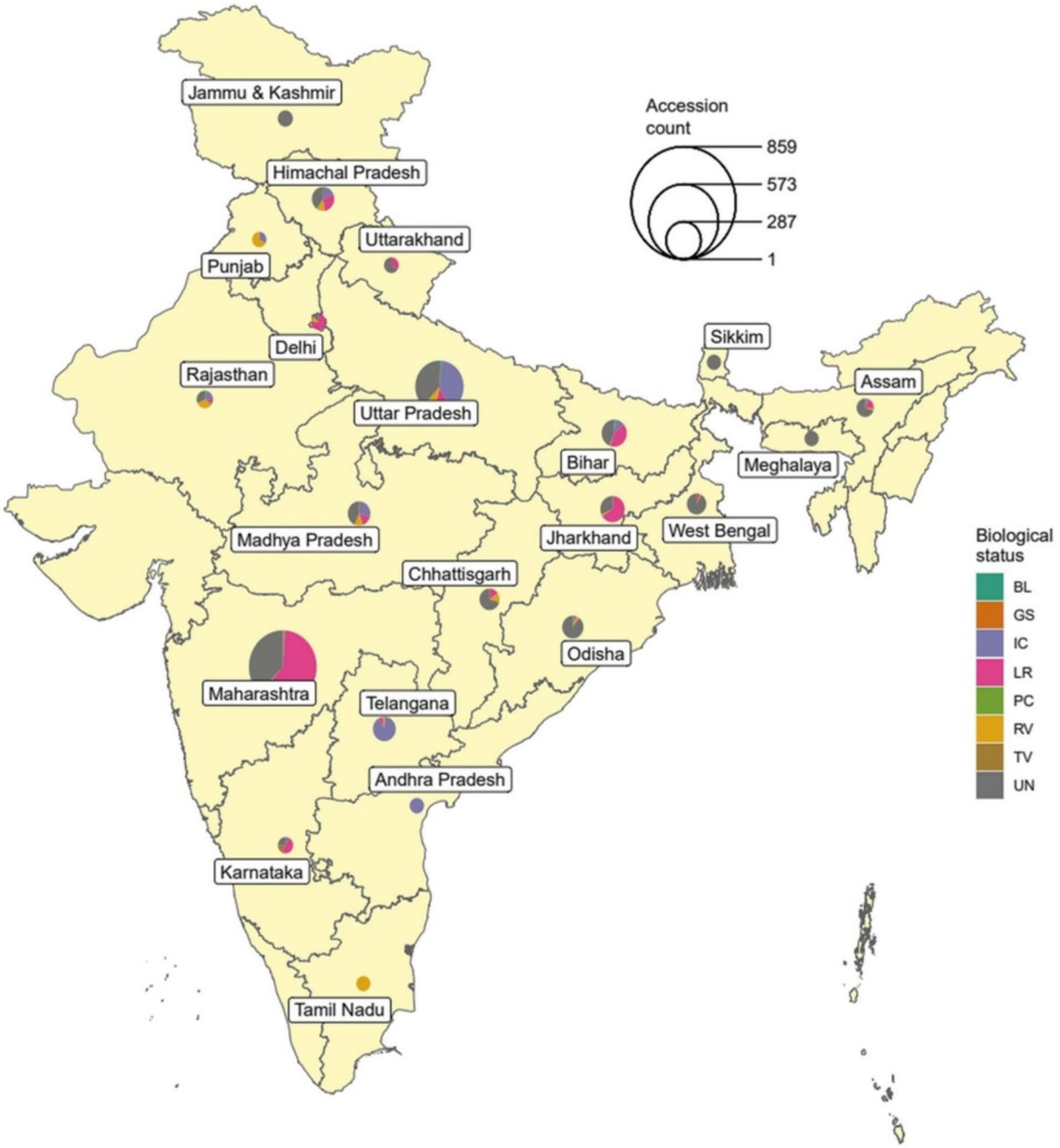

Fig. 5. Collection sites and biological status of Indian linseed germplasm. The map was generated using the R packages ggplot2 v3.4.0 (https://CRAN.R-project.org/package=ggplot2), scatterpie v0.1.8 (https://CRAN.R-project.org/package=scatterpie) and sf v1.0-9 (https://CRAN.R-project.org/package=sf).

quality and You et al.[35] evaluated the PGRC flax core collection for agronomic, seed quality, fiber, and disease resistance traits in multilocation-year environments and reported significant phenotypic variation in both fiber and oil accessions. The examined germplasm collection revealed a wide range of phenotypic variability available for linseed improvement. The range of variation was broader for days to initiation of flowering (43.31–122.88) compared to the global collection (37–69 days) than the PGRC collection as reported by Diederichsen et al.[30]. The range of other traits such as thousand seed weight was comparable among the two collections (2.80–11.86 g in the Indian collection and 2.80–12.40 g in the PGRC collection), while for trait PH the PGRC collection had a minimum of 17 cm and a maximum of 130 cm (data from non-replicated trials) which was different from the minimum PH of 43.31 cm and maximum PH 122.88 cm observed in the collection from NGB India. For oil content, the mean OC in the studied collection was higher (41.30%) compared to PGRC (38.3%). Among the different fatty acid constituents, ALA content is the key fatty acid present in the highest proportions in linseed. For ALA content, the evaluated germplasm exhibited a minimum of 25.40% to a maximum of 65.88% with a

mean of 48.12%. Earlier the ALA content varied from 39.5% in germplasm accession IC564687 to 57.1% in IC564631[20] and to as low as 33.14%[21] in Indian germplasm while in PGRC flax collection (2243 accessions) it was reported to be 39.6–66.7% with a mean of 52.6%[36]. In India, earlier studies on agro-morphological characterization and diversity analysis (although on a smaller number of germplasm accessions) have shown a broad range of phenotypic expression and many trait-specific accessions for early flowering, early maturity, oil content, bold seeds, high-test weight, and ALA were identified[6,18,19,37–40]. The lower SY and related components such as CPP, PBP, and TSW at Akola in comparison to Delhi may be attributed to the shorter duration of phenophases and flowering time duration translating to lower phenotypic expression (PH) and lesser yield at the former location. Earlier our group evaluated a smaller subset of 131 linseed germplasm for earliness, PH and TSW in 5 environments (including Delhi and Akola) and reported a similar range of variation and phenotypic expression as described in the present work[16,18,19].

The genetic variability estimates in the PGRC Core Collection[35] indicated a lower variation for TSW, DF5, PH, DPM, PAL, STE, OLE, and LIN compared to that presented here. The broad sense heritability (hBS) of a trait represents the extent to which the genotypes are influenced by environmental effects. Here, the hBS measured on pooled data from all six environments (Table 2) indicated a high level of variance coupled with heritability (>75%) and genetic advance for all the traits except PBP and DPM for which hBS was 59.35% and 26. 36%, respectively. This suggests that most of these traits are amenable to selection and genetic improvement via conventional or molecular plant breeding approaches. On the other hand, DPM and PBP had low estimates of variability and genetic advance indicating more influence of environmental conditions on these traits. Previous studies in India have also reported low hBS for the DPM trait[6,18,19,41]. In flax, high estimates of hBS for phenological and agronomic traits such as capsule and seed size, flowering time, traits, oil content and TSW were also reported earlier[6,19,37]. The phenotypic variability analysis of PGRC collection by You et al.[35] revealed low hBS for agronomic traits, particularly TSW, PH, DF5, STE and OC, while high hBS for seed quality traits (all five fatty acids and oil content), although the heritability estimates for these traits in the Indian collection are higher than the PGRC collection.

Linseed is broadly classified into two major morphotypes: the seed/oil type or fiber type[13]. Around 82% of the germplasm in PGRC collection is convariety *usitatissimum* and 12.4% is classified as fiber flax under convariety *elongatum* Vav. Et Elladi[30,42]. In the present study, around 97% (2498 accessions) of germplasm was categorised as seed/oil type while only 12 very tall and erect accessions were grouped under fiber flax germplasm. This is because the Indian sub-continent is one of the proposed centres of origin of the linseed type morphotype[13,14] and this crop has been primarily cultivated for seed or oil purposes here since ancient times[5,9,29,43]. Recently another intermediate morphotype, the dual-type (where both seeds and stems can be commercially utilized) has been emphasized for commercial exploration for double-purpose[9–11]. In the present work, 78 accessions exhibiting the desirable plant architecture for the double-purpose linseed were classified as dual-purpose germplasm. The dual-purpose flax varieties grown in warm climates have been reported to produce inferior quality fiber than those in cold climates[11]. In India, linseed is mainly cultivated as an annual cool season (October–March) oilseed crop, therefore dual-purpose linseed is gaining attention and popularity among farmers in India intending to increase farmers' income. All India Coordinated Research Project on Linseed (AICRP-Linseed) in India has developed a few dual-purpose linseed varieties however standardization of agronomic production technologies and commercialization are still lagging.

Germplasm repositories have been proposed to serve not only as a storage facility but also to offer a structured methodology for enhancing the utilization of preserved genetic diversity. In this aspect, several trait-specific superior genotypes have been identified for agro-morphological (Table 4) and seed quality traits (Tables 5 and 6) in the evaluated germplasm in the present work which may have further implications for use as donors in research and breeding for linseed improvement. In this context, the accessions with certain superior trait values as well as high or comparable trait values for other important and yield-contributing traits have also been identified (Table S3), such as high oil content coupled with seed yield and associated components. These superior genotypes with multiple desirable traits are expected to facilitate breeders to choose donor lines as per the breeding objectives without potential negative trade-offs. Despite this, a germplasm accession with desirable traits may also have an adverse correlation with other undesirable traits and therefore could be a constraint in a breeding program. In such a situation, selection for both traits simultaneously is expected to be more effective than the selection based on a single trait. Further, the genetic dissection of the individual trait would help in developing an allele model to study optimal strategies under different gene actions for long-term genetic gain[44].

Globally, several research reports have emphasized the importance of Indian linseed germplasm for key traits of economic importance such as high yield, thousand seed weight, high mucilage content, flowering time, early maturity, low ALA, efficient root system architecture, drought and salt tolerance[6,15–19,22–25,29,45]. The advantage of the present study is that the entire collection of conserved linseed germplasm has been extensively characterized for 36 traits in up to 6 environments for assessment of genetic diversity and trait variability. The identified genotypes provide a smaller and more manageable set of diverse representative germplasm accessible to breeders and may also aid in bridging the conservation and utilization gap. The present multi-environment phenotyping data could also be utilized in genome-wide association studies if the genebank accessions are genotyped, which would facilitate the genetic dissection of important traits, identifying QTLs and candidate genes. Further, making use of DNA level variability, and in combination with phenotypic variability, a molecular and composite core collection, respectively could be developed to enable researchers/breeders to make use of smaller and diverse sets for further trait discovery and allele mining.

Conclusion

In the present study, large-scale characterization of linseed germplasm assembly at NGB, India in up to 6-year-location environments led to the assessment of genetic diversity, infraspecific classification of conserved collection and identification of promising trait-enriched genotypes. A wide spectrum of variability was unravelled for the seed/oil-type, dual purpose and fiber flax germplasm. Infraspecific classification revealed the predominance of seed/oil-type morphotypes over the fiber flax. The identification of trait-specific superior genotypes for earliness, plant height, bold seeds, grain yield, oil content and fatty acid profile in oilseed type linseed and fiber flax may be utilized in future selection and breeding to cater to different agro-ecologies and end-uses. This work will stimulate the linseed breeders and germplasm curators to have deep and meaningful insights into the prevailing genetic diversity in large germplasm repositories to enable expedited access to economically important diverse material for trait introgression.

Materials and methods

Plant materials

A total of 2576 unique accessions of cultivated linseed under long-term conservation at the National Genebank (NGB), India were evaluated in multi-environment trials in the present study. The tested germplasm comprised 2414 indigenous collections (IC) assembled from 21 Indian states/Union Territories and 162 exotic collections (EC) mainly from Argentina, Australia, Russia and the USA. The geographic distribution and biological status of Indian and exotic germplasm accessions are revealed in Figs. 5 and 6, respectively. The passport data of the evaluated germplasm can be accessed website of ICAR-NBPGR. (http://pgrportal.nbpgr.ernet.in/)

Field evaluation experiment details and data recording

The seeds multiplied through single plant progenies from each accession were sown for agro-morphological evaluation in the last week of October spanning over four consecutive seasons (2018–2019 to 2021–2022) at ICAR-NBPGR, experimental farm, IARI, New Delhi (28° 38′ 53.7″ N, 77° 09′ 05.4″ E and 218 m above mean sea level) and for two seasons (2020–2021 and 2021–2022) at ICAR-NBPGR, Regional station-Akola (20° 42′ 1.44″ N, 77° 6′ 9″ E and 288 m above mean sea level). The accessions were sown in an Augmented Block Design (ABD)[46] across 28 blocks, incorporating ten standard varieties as a national check (T-397) and zonal checks such as Shekhar, Sheela, Sharda, Kartika, JLS-67, LSL-93, JLS-95, Tiara, and Rashmi recommended for linseed with oil/fiber/dual purposes. Each block consisted of 102 entries comprising 92 accessions and 10 checks, which were randomized in every block. Each accession was sown in a 2-m row with inter-row and plant to plant spacing of 45 × 10 cm. Initial fertilization of 60:40 kg/ha of N: P_2O_5 and 20 kg/ha each of S and Zn with complete P, S, and Zn alongside half of N fertilizer was applied in the field during sowing, and the remaining N was top-dressed post-first irrigation. After germination, the first irrigation was provided at 30–40 days post-sowing (DAS) and the subsequent at the flowering stage (65–70 DAS). The threshing and harvesting of each accession was done manually. The data was recorded on 10 qualitative and 26 quantitative traits. The details of all 36 traits, descriptor state and growth stage of recording of observation have been elaborated in Table 9. The field data for all the agro-morphological parameters was recorded digitally using the Fieldscorer™ App. (https://www.katmandoo.org.au)

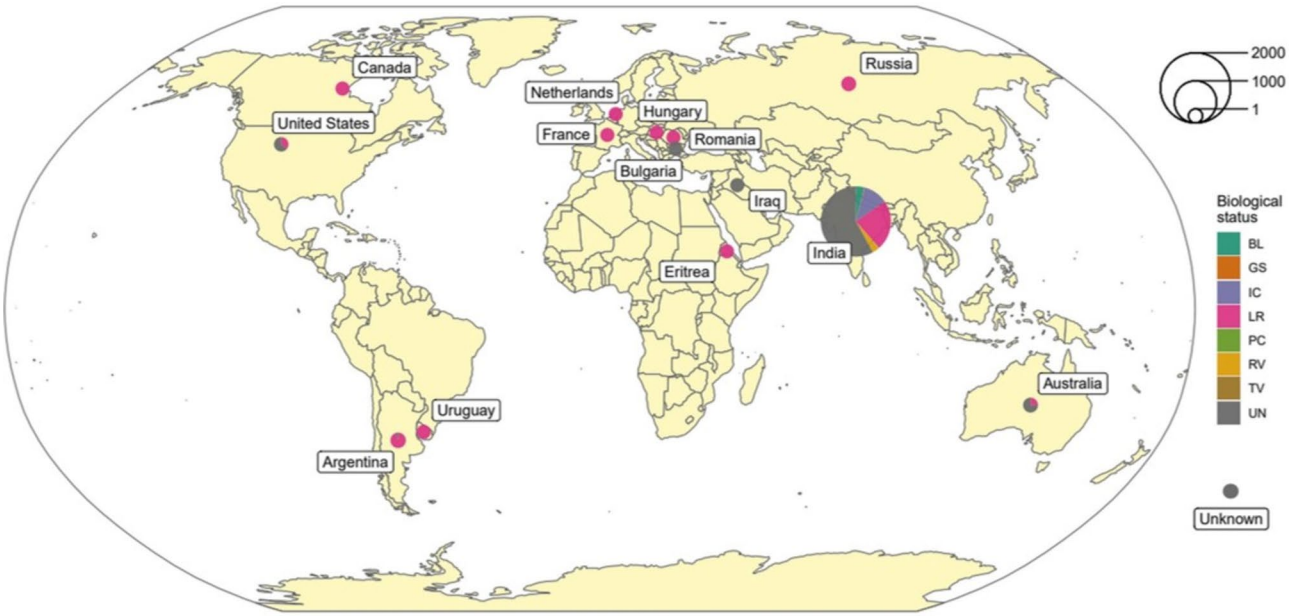

Fig. 6. Geographic distribution and biological status of exotic accessions of linseed germplasm. The map was generated using the R packages ggplot2 v3.4.0 (https://CRAN.R-project.org/package=ggplot2), scatterpie v0.1.8 (https://CRAN.R-project.org/package=scatterpie) and sf v1.0-9 (https://CRAN.R-project.org/package=sf).

S. No	Trait	State	Stage of observation
Qualitative traits			
1	Flower shape	Funnel	Recorded on individual plants at 50% flowering stage, observation done before noon
		Star	
		Disk	
2	Flower colour	Dark Blue	Recorded on individual plants at 50% flowering stage
		Blue	
		Light Blue	
		Purple	
		Violet	
		White	
3	Flower venation colour	White	Recorded on individual plants at 50% flowering stage
		Light-violet	
		Violet	
		Blue	
		Purple	
4	Flower aestivation	Twisted	Observed on individual plants at 50% flowering stage, recorded before noon
		Semi-Twisted	
		Valvate	
5	Stamen colour	White	Colour of the distal part of the filament of individual flowers at 50% flowering stage
		Violet	
		Blue	
6	Anther colour	Cream	Colour of anthers of individual flowers at 50% flowering stage
		Grey	
		Violet	
		Blue	
		Yellow	
7	Plant growth habit	Bushy	Assessed on a group of plants at completion of flowering
		Semi-erect	
		Erect	
8	Seed colour	Dark Brown	Recorded at maturity
		Brown	
		Light brown	
		Dark Yellow	
		Yellow	
		Light Yellow	
9	Seed lustre	Lustrous	Recorded at maturity
		Intermediate	
		Dull	
10	Resistance to lodging		Recorded at maturity on a scale of 1–9, where a score of 1 represents upright plants
Quantitative traits			
1	Early plant vigour	Recorded on a plot basis after 30–35 days of sowing	
2	Days to 5% flowering	Visually assessed when 5% of plants are showing flowers in the plot	
3	Days to 50% flowering	Visually assessed when 50% of plants are showing flowers in the plot	
4	Completion of flowering	Visually assessed when 95% of plants are showing flowers in the plot	
5	Size of corolla (mm)	Measurement of distance from petal to petal in 5 randomly selected individual plants at 50% flowering stage	
6	Plant height (cm)	Recorded by measuring 5 randomly selected individual plants at dough stage (from base of the plant to the tip of the main shoot)	
7	Technical plant height (cm)	Recorded by measuring 5 randomly selected individual plants at dough stage (from base of plant to the point where primary branches arise)	
8	No. of primary branches per plant	Measured on 5 randomly selected individual plants at dough stage	
9	No. of capsules per plant	Counted on 5 randomly selected individual plants per accession for fully developed capsules	
10	No. of seeds/capsule	Counting of seeds in five capsules from each of the 5 randomly tagged plants per accession and then calculating the average	
11	Days to 80% maturity	Number of days from sowing at maturity when 80% of the capsules have attained brown colour	
12	Seed length (mm)	Recorded as average from the seed harvested from 15 randomly selected capsules taken from 5 tagged plants per accession	
13	Seed breadth (mm)		
14	Seed area (mm^2)		
Continued			

S. No	Trait	State	Stage of observation
15	Capsule diameter (mm)	Average of three random samples of 15 capsules from each of the 5 tagged plants per accession	
16	Capsule area (mm^2)		
17	Grain yield/plant (g)	Recorded by weighing the fully dried seeds per plant (average of 5 randomly selected plants)	
18	Thousand seed weight (g)	Weight of 1000 seeds (randomly taken)	
19	Normalized difference vegetation index (NDVI)	Recorded on plot basis at 80–90 DAS	
20	Chlorophyll content (mg/m^2)	Recorded as the average of the five randomly tagged plants per accession at 80–90 DAS	
Oil content and quality			
21	Oil content (%)	Estimated in fully dried seeds at maturity	
22	Palmitic acid (%)		
23	Stearic acid (%)		
24	Oleic acid (%)		
25	Linoleic acid (%)		
26	Linolenic acid (%)		

Table 9. The details of 10 qualitative and 26 quantitative traits recorded in linseed germplasm.

Early Plant Vigour (EPV) and Normalized Difference Vegetation Index (NDVI) were recorded using a handheld crop sensor (GreenSeeker; Trimble). Chlorophyll content (CC; mg/m^2) was measured by the Chlorophyll Content Meter (CCM-300; Opti-Sciences, Inc.). EPV, NDVI and CC were recorded for two seasons (2020–2021 and 2021–2022) in the Delhi location only. Oil content (OC; %) was estimated by the Soxhlet extraction system using petroleum ether as a solvent as per AOAC, 1970[47]. Fatty acid profiling was done as per Kaushik[48]. Estimation of oil content and fatty acid composition was done for only one season (2019–2020) and promising accessions were validated from seed harvested in multiple location-year environments. The morphometry of capsules and seeds was assessed from scanned images using a flat-bed scanner and analyzed utilizing Grain Analysis software (version 1.3). The weight of a thousand seeds (TSW; g) was measured employing a seed counter (Aidex Co. Ltd. Japan; IC VAI). The meteorological data for all tested environments at Delhi and Akola were also recorded and are presented in Table S4.

Statistical analysis

The Analysis of Variances (ANOVA) was performed on data of all 26 quantitative traits from each of the six environments (4 seasons at the Delhi location and 2 seasons at the Akola location) as per Augmented Block Design and adjusted means were derived using the 'R' package augmentedRCBD[49]. Bartlett's Chi-square test was utilized to examine the uniformity of error variances among environments. In cases where traits displayed heterogenous error variances, Aitken's transformation was applied to combine the data for further analysis of adjusted means using pooled data. Visualization of violin cum box plots for all the analyzed quantitative traits was achieved using the 'ggplot2' package in 'R' software. The adjusted means were utilized to calculate genetic variability parameters, heritability estimates, correlation coefficients, and perform Principal Component Analysis (PCA). The phenotypic coefficient of variation (PCV) and genotypic coefficient of variation (GCV) were determined for each quantitative trait in accordance with[50], categorizing the range as specified by Sivasubramanian and Madhavamenon[51]. Broad sense heritability ($h^2_{(bs)}$) was assessed as the ratio of genotypic to phenotypic variances as proposed by Lush[52] and further classified into low, medium, and high categories following Robinson's[53] criteria (1966). The Expected Genetic Advance (EGA) was computed based on the method by Johnson et al.[54], with the standardized selection differential (k = 2.06 at 5% selection intensity). Genetic advance expressed as percentage of the mean was calculated as $GA(\%) = EGA/mean \times 100$. Shannon–Weaver diversity index (H') and evenness were calculated using phenotypic frequencies of qualitative characters (Shannon and Weaver, 1949)[55]. Correlation among all traits was analyzed using PAST software (v4.04)[56]. Principal Component Analysis (PCA) was conducted to explore relationships among measured traits using the 'factoextra' package in 'R'. The infraspecific classification of the tested 2576 accessions was done in two distinct morphotypes based on plant height (PH) and technical plant height (TPH) and plant architecture/growth habit following Singh and Chopra[11].

Conclusions

This report describes the large-scale evaluation of the entire linseed germplasm collection at the National Genebank of India in up to 6-year-location environments. The results described here unravel the prevailing phenotypic variability and genetic diversity that will stimulate the researchers to have more insights into the *ex-situ* germplasm collections for crop improvement. Infraspecific classification revealed the predominance of oilseed/linseed type morphotype over the fiber flax. In addition, trait-specific superior genotypes for varied utility have been identified which will facilitate the effective use of novel germplasm for linseed breeding. These findings are expected to provide better guidance for accelerated genetic resource utilization in linseed improvement and broadening the genetic base for climate resilience.

Data availability

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author. Linseed is not listed in the list of Annexure I crops under the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGFRA), hence not falling within the purview of the Multilateral System (MLS) governing access and benefit sharing. The germplasm can be shared or accessed by linseed breeders in India upon formal request through the National Genebank of India, ICAR-NBPGR, located in New Delhi, adhering to the specified terms, conditions (http://www.nbpgr.ernet.in/Downloadfile.aspx?EntryId=7379) and tariffs (http://www.nbpgr.ernet.in/Portals/6/services/Fee_Stru_2023.pdf). The exchange of linseed germplasm for international collaborations can be facilitated following the Material Transfer Agreement (http://www.nbpgr.ernet.in/Downloadfile.aspx?EntryId=7380) within the framework of Collaborative Research Programs/Projects.

Received: 26 April 2024; Accepted: 23 August 2024
Published online: 31 August 2024

References

1. Zohary, D. & Hopf, M. *Domestication of Plants in the Old World* Vol. 3, 125–132 (Oxford University Press, Oxford, 2000).
2. Langyan, S. *et al.* Trends and advances in pre- and post-harvest processing of linseed oil for quality food and health products. *Crit. Rev. Food Sci. Nutr.* **30**, 1–24. https://doi.org/10.1080/10408398.2023.2280768 (2023).
3. Awasthi, S., Kaushik, N., Plaha, N. S., Kaur, V. & Kumar, A. Exploring lipid health indices and protein quality in ninety Indian linseed varieties by comprehensive analysis of fatty acid composition, lignan content, and amino acid composition. *Ind. Crops Prod.* **212**, 118366. https://doi.org/10.1016/j.indcrop.2024.118366 (2024).
4. FAO. https://www.fao.org/faostat/en/#data/QCL (Accessed 5 June 2022) (2022).
5. Yadav, B. *et al.* Integrated omics approaches for flax improvement under abiotic and biotic stress: Current status and future prospects. *Front. Plant Sci.* **13**, 931275. https://doi.org/10.3389/fpls.2022.931275 (2022).
6. Kaur, V. *et al.* Analysis of genetic diversity in Indian and exotic linseed germplasm and identification of trait specific superior accessions. *J. Environ. Biol.* **39**, 702–709. https://doi.org/10.22438/jeb/39/5/MRN-849 (2018).
7. Guo, R. *et al.* Germination, growth, chlorophyll fluorescence and ionic balance in linseed seedlings subjected to saline and alkaline stresses. *Plant Prod. Sci.* **17**, 20–31. https://doi.org/10.1626/pps.17.20 (2014).
8. Zuk, M., Dorota, R., Matuła, J. & Szopa, J. Linseed, the multipurpose plant. *Ind. Crops Prod.* **75**, 165–177. https://doi.org/10.1016/j.indcrop.2015.05.005 (2015).
9. You, F. M. *et al.* Pedigrees and genetic base of the flax varieties registered in Canada. *Can. J. Plant Sci.* **96**, 837–852. https://doi.org/10.1139/cjps-2015-0337 (2016).
10. Irvine, R. B. *et al.* Impact of production practices on fiber yield of oilseed flax under Canadian prairie conditions. *Can. J. Plant Sci.* **90**, 61–70. https://doi.org/10.4141/CJPS08233 (2010).
11. Singh, P. K. & Chopra, P. Double purpose linseed: A viable option for doubling farmers' income in the north-western Himalayan region. *Indian Farming* **68**, 49–54 (2018).
12. Soto-Cerda, B. J., Cloutier, S., Gajardo, H. A., Aravena, G. & Quian, R. Identifying drought-resilient flax genotypes and related-candidate genes based on stress indices, root traits and selective sweep. *Euphytica* **215**, 41. https://doi.org/10.1007/s10681-019-2362-0 (2019).
13. Vavilov, N. I. The origin, variation, immunity and breeding of cultivated plants. *Chron. Bot.* **13**, 1–366 (1951).
14. Zeven, A. C. & Zhukovsky, P. M. *Dictionary of Cultivated Plants and Their Centres of Diversity* (Centre for Agricultural Publishing and Documentation, Wageningen, 1975).
15. Diederichsen, A. & Raney, J. P. Seed colour, seed weight and seed oil content in *Linum usitatissimum* accessions held by plant gene resources of Canada. *Plant Breed.* **125**, 372–377. https://doi.org/10.1111/j.1439-0523.2006.01231.x (2006).
16. Saroha, A. *et al.* Genetic dissection of thousand-seed weight in linseed (*Linum usitatissimum* L.) using multi-locus genome-wide association study. *Front. Plant Sci.* **14**, 1166728. https://doi.org/10.3389/fpls.2023.1166728 (2023).
17. Diederichsen, A., Raney, J. P. & Duguid, S. D. Variation of mucilage in flax seed and its relationship with other seed characters. *Crop Sci.* **46**, 365–371. https://doi.org/10.2135/cropsci2005.0146 (2006).
18. Saroha, A. *et al.* Identification of QTNs associated with flowering time, maturity, and plant height traits in *Linum usitatissimum* L. using genome-wide association study. *Front. Genet.* **13**, 811924. https://doi.org/10.3389/fgene.2022.811924 (2022).
19. Saroha, A. *et al.* Agro-morphological variability and genetic diversity in linseed (*Linum usitatissimum* L.) germplasm accessions with emphasis on flowering and maturity time. *Genet. Resour. Crop Evol.* **69**, 315–333. https://doi.org/10.1007/s10722-021-01231-3 (2022).
20. Sivaraj, N. *et al.* Variability in linseed (*Linum usitatissimum*) germplasm collections from peninsular India with special reference to seed traits and fatty acid composition. *Indian J. Agric. Sci.* **82**, 102–105. https://doi.org/10.56093/ijas.v82i2.15265 (2012).
21. Pali, V. & Mehta, N. Evaluation of oil content and fatty acid compositions of flax (*Linum usitatissimum* L.) varieties of India. *J. Agric. Sci.* **6**, 198–207. https://doi.org/10.5539/jas.v6n9p198 (2014).
22. Sertse, D. *et al.* Loci harboring genes with important role in drought and related abiotic stress responses in flax revealed by multiple GWAS models. *Theor. Appl. Genet.* **134**, 191–212. https://doi.org/10.1007/s00122-020-03691-0 (2021).
23. Soto-Cerda, B. J. *et al.* Integrating multi-locus genome-wide association studies with transcriptomic data to identify genetic loci underlying adult root traits responses to drought stress in flax (*Linum usitatissimum* L.). *Environ. Exp. Bot.* **202**, 105019. https://doi.org/10.1016/j.envexpbot.2022.105019 (2022).
24. Li, X. *et al.* Genome-wide association study of salt tolerance at the seed germination stage in flax (*Linum usitatissimum* L.). *Genes* **13**, 486. https://doi.org/10.3390/genes13030486 (2022).
25. Yadav, N. *et al.* Effect of GA_3 and Calcium on growth, biochemical and fatty acid composition of linseed under chloride-dominated salinity. *Environ. Sci. Pollut. Res. Int.* **31**, 16958. https://doi.org/10.1007/s11356-024-32325-x (2024).
26. Nair, B. *et al.* Multi-environment screening of *Linum* germplasm collection for dissecting the potential of bud fly (*Dasyneura lini* Barnes) resistance and assembling a reference set for efficient utilization in genetic improvement. *Ind. Crops Prod.* **207**, 117743. https://doi.org/10.1016/j.indcrop.2023.117743 (2023).
27. Kumar, V. *et al.* Biophysical and morphological basis of resistance against linseed (*Linum usitatissimum* L.) bud fly (*Dasyneura lini* Barnes). *Indian J. Plant Genet. Resour.* **36**, 382–386 (2023).
28. Plaha, N. S. *et al.* Comparison of nutritional quality of fourteen wild *Linum* species based on fatty acid composition, lipid health indices, and chemometric approaches unravelling their nutraceutical potential. *Heliyon* **9**, e21192. https://doi.org/10.1016/j.heliyon.2023.e21192 (2023).
29. Kaur, V. *et al.* Diversity of *Linum* genetic resources in global genebanks: From agro-morphological characterisation to novel genomic technologies: A review. *Front. Nutr.* **10**, 1165580. https://doi.org/10.3389/fnut.2023.1165580 (2023).

30. Diederichsen, A., Kusters, P. M., Kessler, D., Bainas, Z. & Gugel, R. K. Assembling a core collection from the flax world collection maintained by plant gene resources of Canada. *Genet. Resour. Crop Evol.* **60**, 1479–1485. https://doi.org/10.1007/s10722-012-9936-1 (2013).
31. Diederichsen, A. & Fu, Y. Phenotypic and molecular (RAPD) differentiation of four infraspecific groups of cultivated flax (*Linum usitatissimum* L. subsp. usitatissimum). *Genet. Resour. Crop Evol.* **53**, 77–90. https://doi.org/10.1007/s10722-004-0579-8 (2006).
32. Diederichsen, A. *et al.* Diversity in the flax collection at plant gene resources of Canada in *Proceedings of the 59th Flax Institute of the United States*. 138–143 (Doublewood Inn, Fargo, North Dakota, USA. Flax Institute, Fargo, USA, 2002).
33. Worku, N., Heslop-Harrison, J. S. & Adugna, W. Diversity in 198 Ethiopian linseed (*Linum usitatissimum*) accessions based on morphological characterization and seed oil characteristics. *Genet. Resour. Crop Evol.* **62**, 1037–1053. https://doi.org/10.1007/s10722-014-0207-1 (2015).
34. Zhuchenko, A. A. & Rozhmina, T. A. Mobilization of flax genetic resources. (2000).
35. You, F. M. *et al.* Genetic variability of 27 traits in a core collection of flax (*Linum usitatissimum* L). *Front. Plant Sci.* **8**, 1636. https://doi.org/10.3389/fpls.2017.01636 (2017).
36. Diederichsen, A. Ex situ collections of cultivated flax (*Linum usitatissimum* L.) and other species of the genus Linum L. *Genet. Resour. Crop Evol.* **54**, 661–678. https://doi.org/10.1007/s10722-006-9119-z (2007).
37. Chandrawati, S. N. *et al.* Genetic diversity, population structure and association analysis in linseed (*Linum usitatissimum* L.). *Physiol. Mol. Biol. Plants* **23**, 207–219. https://doi.org/10.1007/s12298-016-0408-5 (2017).
38. Dhirhi, N. & Mehta, N. Estimation of genetic variability and correlation in F2 segregating generation in linseed (*Linum usitatisimum* L.). *Plant Arch.* **19**, 475–484 (2019).
39. Dikshit, N. & Sivaraj, N. Analysis of agro-morphological diversity and oil content in Indian linseed germplasm. *Grasas Aceites* **66**, e060. https://doi.org/10.3989/gya.0689141 (2015).
40. Nizar, M. A. & Mulani, R. M. Genetic diversity in indigenous and exotic linseed germplasm (*Linum usitatissimum* L.). *Electron. J. Plant Breed.* **6**, 848–854 (2015).
41. Kaur, V. *et al.* Phenotypic characterization, genetic diversity assessment in 6778 accessions of barley (*Hordeum vulgare* L. ssp. vulgare) germplasm conserved in National Genebank of India and development of a core set. *Front. Plant Sci.* **13**, 771920. https://doi.org/10.3389/fpls.2022.771920 (2022).
42. Diederichsen, A. & Ulrich, A. Variability in stem fibre content and its association with other characteristics in 1177 flax (*Linum usitatissimum* L.) genebank accessions. *Ind. Crop Prod.* **30**, 33–39. https://doi.org/10.1016/j.indcrop.2009.01.002 (2009).
43. Kaur, V. *et al.* Multi-environment phenotyping of linseed (*Linum usitatissimum* L.) germplasm for morphological and seed quality traits to assemble a core collection. *Ind. Crops Prod.* **206**, 117657. https://doi.org/10.1016/j.indcrop.2023.117657 (2023).
44. Wu & Sanchez. Further, the genetic dissection of the individual trait would help in developing an allele model to study optimal strategies under different gene actions for long-term genetic gain (2011).
45. Kaur, V., Yadav, R. & Wankhede, D. P. Linseed (*Linum usitatissimum* L.) genetic resources for climate change intervention and its future breeding. *J. Appl. Nat. Sci.* **9**, 1112–1118. https://doi.org/10.31018/jans.v9i2.1331 (2017).
46. Federer, W. T. Augmented (or hoonuiaku) designs. *Hawaii Plant Record* **55**, 191–208 (1956).
47. AOAC. *Official Methods of Analysis of the Association of Official Agricultural Chemicals*, 10th ed. 744–745 (AOAC:Rockville, MD, USA, (1970).
48. Kaushik, N. Determination of azadirachtin and fatty acid methy esters of *Azadirachta indica* seeds by HPLC and GLC. *Anal. Bioanal. Chem.* **374**, 1199–1204. https://doi.org/10.1007/s00216-002-1638-7 (2002).
49. Aravind, J., Sankar, S. M., Wankhede, D. P. & Kaur, V. Augmented RCBD: Analysis of augmented randomised complete block designs (R package version 0.1.6) (2023).
50. Burton, G. W. Quantitative inheritance in grasses. *Proc. Int. Grassl. Cong.* **1**, 277–283 (1952).
51. Sivasubramanian, S. & Madhavamenon, P. Genotypic and phenotypic variability in rice. *Madras Agric. J.* **60**, 1093–1096 (1978).
52. Lush, J. L. Intrusive collection of regression of offspring on dams as a method of estimating heritability of characters. *Proc. Am. Soc. Anim. Prod.* **33**, 293–301 (1940).
53. Robinson, H. F. Quantitative genetics in relation to breeding of the centennial of mendelism (The impact of mendelism on agriculture, biology and medicine). *Indian J. Genet.* **26A**, 171–187 (1966).
54. Johnson, H. W., Robinson, H. F. & Comstock, R. Estimates of genetic and environmental variability in soybeans[1]. *Agron. J.* **47**, 314–318. https://doi.org/10.2134/agronj1955.00021962004700070009x (1955).
55. Shannon, C. E. & Weaver, W. *The Mathematical Theory of Communication* 1–117 (The University of Illinois Press, Urbana, 1949).
56. Hammer, O., Harper, D. A. T. & Ryan, P. D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **4**, 1–9 (2001).

Acknowledgements

Funding from the ICAR-NBPGR institutional project 'Characterization, Evaluation and Documentation of Oilseeds Germplasm' (PGR/GEV-BUR-DEL.01.04) for the initial two years of field evaluation is duly acknowledged. The contributions of several scientists and technicians in the exploration and conservation of linseed germplasm at ICAR-NBPGR and the contributions of field staff in carrying out this large-scale characterization are greatly acknowledged.

Author contributions

V.K.: Conceptualization, supervision, writing—original draft; funding acquisition; project administration; SSG: investigation; supervision; S.K.Y., D.S., Sh, S.S.C., ViK (Vinay Kumar), B.J., N.R.T.: Data collection and curation; S.L.: investigation (oil content); N.K.: investigation (fatty acid composition); M.S.: investigation; M.K.: data curation and compilation; D.P.W.: conceptualization; formal analysis; J.A.: resources; software; V.S., K.G.: investigation; A.K., G.P.S.: funding acquisition; project administration. All the authors have read the manuscript critically.

Funding

This work was supported by funding for the project (No. BT/Ag/Network/Linseed/2019–20) from the Department of Biotechnology (DBT), Government of India.

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/s41598-024-71044-8.

Correspondence and requests for materials should be addressed to V.K.

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.