www.ThePharmaJournal.com

The Pharma Innovation

Kaveri Chawan
Department of Genetics and
Plant Breeding, College of
Agriculture, UAS, GKVK,
Bengaluru, Karnataka, India

Identification of trait-specific accessions from the best core set of soybean (*Glycine max* L. Merrill) germplasm

Kaveri Chawan

Abstract
The use of genetic resources in breeding programmes will be managed and accelerated by core collection. A core set of crop genetic resources is considered "first-look sources" of trait-specific accessions for use in crop breeding programs. During the summer of 2021, the genetic variability of 2000 soybean germplasm accessions were assessed at ZARS UAS, GKVK, Bengaluru. The standard stratified clustering approach was used to develop 8 core sets from a base collection based on 13 qualitative and 7 quantitative traits. The core sets were validated using various univariate and multivariate statistics to assess their representativeness for the base collection. During the kharif 2021, 300 germplasm accessions (15% core size) were evaluated at two sites, K-Block GKVK, Bengaluru, and KVK Doddaballapur, to identify trait-specific accessions from the best core set. Eight core sets were developed in the current study using the SSC approach. Logarithmic sampling with preferred allocation approach-based core set of 15% size was identified as the best representative of the base collection. Many trait-specific accessions were promising for the combination of desirable traits from the best core set that are suggested for preferential use in crossing programmes to generate variability for developing farmer-acceptable varieties.

Keywords: Core collection, germplasm, *Glycine max*, k-means clustering, standard stratified clustering, trait-specific, variability

Introduction
Soybean (*Glycine max* L. Merrill) is an important floral crop rich in protein (40%) and oil (20%), belongs to family Fabaceae. United States of America, Brazil, Argentina, China, India, Paraguay, Canada, Ukraine, Russia, and Bolivia are among the top 10 soybean producers globally (Dutton *et al*., 2019) [9]. Top production of soybean is mainly reported in the United States, Brazil, and Argentina, with India ranking fourth so far, Table 1 (Global Soybean Market Outlook. 2022). The contribution of India in the world soybean area is 10% but the contribution to total world soybean production is only 4% indicating its relatively lower levels of productivity (1.1 t/ha) as compared to the world average (2.2 t/ha), which is a major cause of concern (Debnath and Babu 2020) [6] and changing climatic condition has altered the yield levels, Table 2 (Ramesh *et al*., 2017) [19].

Over 170,000 germplasm collections of soybean have been collected and maintained in various gene banks throughout the world and form the basis of the diversity available for soybean improvement (Carter *et al*., 2004) [2]. Despite this large quantity of germplasm available only 1,000 have been used in applied breeding programs (Carter *et al*., 2004) [2]. Under present circumstances, the available gene pool plays an important role in selecting and integrating into the changed environment. It is laborious and time consuming to pick the required genetic material out of the available abundant pool, which necessitates creation of core collection for a group of characters and the scientific community shall access the desired genetic material at ease.

Targeting the said objective, an investigation was carried out at UAS Bangalore to develop core sets of soybean using the SSC approach and the best set was selected using various univariate and multivariate validation statistics. As readily available information, trait-specific accessions were finally listed as useful information for the scientific group.

Corresponding Author:
Kaveri Chawan
Department of Genetics and
Plant Breeding, College of
Agriculture, UAS, GKVK,
Bengaluru, Karnataka, India

Table 1: World production and yield from 2018 to 2019 against top international producers (FAOSTAT 2019, 2020)

Country	Production (in Million tons)		
	2018	**2019**	**2020**
USA	120.51	96.67	112.55
Brazil	117.91	114.32	121.80
Argentina	37.79	55.26	48.80
Canada	7.42	6.15	6.36
India	10.93	13.26	11.02
South Africa	1.54	1.17	1.25
Zambia	0.30	0.28	0.30
Nigeria	0.66	0.70	0.60
Malawi	0.18	0.17	0.18
Ghana	0.18	0.18	0.18

Table 2: Climate change impact on soybean

Sl. No.	Climate Variable	Effect on Soybean
1	Increase in temperature	Reduction in growth and yield (Mall *et al.,* 2004) [17].
2	GCM-based scenario	60% increase in water requirement (Curry *et al.,* 1995) [5].
3	5 °C increase in temperature	Reduction in yield (Curry *et al.,* 1995) [5].
4	Doubling of CO_2	Harvest index decreased (Cure and Acock, 1986) [4].
5	Increase in temperature	Alters soybean seed composition (Carver *et al.,* 1986; Gibson and Mullen, 1996; Thomas *et al.,* 2003) [3, 12, 21].
6	Increase in temperature	Reduction in quality and yield (Allen and Boote, 2000) [1].
7	Increase in temperature to 28 °C	Seed oil increased (Piper and Boote, 1999) [18].
8	Increase in temperature from 14 °C to 22 °C	Seed protein decreased (Piper and Boote, 1999) [18].
9	Increase in temperature to 35 °C during seed fill.	Seed oil reduced (Dornbos and Mullen, 1992) [8].
10	Drought stress	Early maturity (Desclaux and Roumet, 1996) [7].

Materials and Methods

A set of 2000 germplasm accessions along with three checks were characterized for 13 qualitative traits and 7 quantitative traits during summer 2021 following augmented design at ZARS, UAS, GKVK Bangalore. Each block contained 100 germplasm accessions as well as three checks (replicated twice). Each accession was sown in a single row of 1.5m length with row-to-row spacing of 0.45 m. A basal dose of 25:50:25 kg NPK ha⁻¹ was applied to the experimental plot at the time of sowing. Recommended agronomic and plant protection practices were followed to raise a healthy crop.

On five randomly selected plants, data were collected based on visual observation for 13 qualitative traits such as hilum color, seed coat color, early plant vigour, hypocotyl color, flower color, leaf shape, leaflet color, plant pubescence, plant pubescence color, plant pubescence density, plant pubescence type, stem determination, and pod color. Data was also collected on 7 quantitative traits based on counting or measurement, such as days to 50 per cent flowering, plant height, secondary branches plant⁻¹, pods plant⁻¹, days to 80 per cent maturity, 100 seed weight, and seed yield plant⁻¹. A standard stratified clustering approach based on quantitative trait data was used to develop the core sets in the present study. The method used in the SSC approach for creating core sets is discussed below.

Stratification of the germplasm accessions

Accessions were classified into 10 clusters following Ward's hierarchical clustering algorithm based on adjusted means for 7 quantitative traits. Two sampling strategies, viz., proportional and logarithmic, and two allocation strategies, viz., random and preferred, were followed to determine the number of accessions to be selected from each cluster for inclusion in core sets. Thus, a total of 8 core sets were developed following the SSC approach.

Validation of core sets

The chi-square statistic was used to test the homogeneity of accessions for qualitative trait-based frequency distribution of base and core collections. Retention of qualitative trait classes by the core collection was determined using 'Shannon-Weaver diversity index' (Shannon and Weaver, 1949) [20] and 'class coverage' (Kim *et al*., 2007) [15] statistics. To assess the representativeness of eight core sets, quantitative trait-based validation statistics such as standardized mean difference (SMD %), variance difference (VD %), coincidence rate (CR %), and variable rate (VR %) (Hu *et al*., 2000) [14] were used.

Identification of trait-specific accessions from the core set

During kharif 2021, 300 (15% core size) soybean germplasm accessions and three check entries were evaluated at two locations, K-Block GKVK, Bengaluru, and KVK Doddaballapur, to identify the trait-specific accessions from the best core set. The germplasm accessions were classified following a model-based "k-means' clustering approach (Mac Queen, 1967) [16] to unravel the organization of variability. Based on early flowering time [mean – 1 SD], i.e., ≤ 56.58 days from sowing, less plant height (mean – 1 SD) ≤ 36.45, early maturity (mean – 1 SD) ≤ 109.71 days from sowing, and higher (mean + 2 SD) expression of the accessions for the other four traits, single trait-specific and multiple trait-specific accessions were identified.

Results and Discussion
Representativeness of Core Sets

The classes of 13 qualitative traits from eight distinct core sets were compared to those of the base collection. Except for proportional and logarithmic sampling with random allocation of sizes 10% and 15%, the frequency distribution of qualitative traits was comparable to that of the base collection (chi-square was significant for < 4 traits). All of the core sets in the current study had H' estimates that were comparable to

the base collection and covered more than 80% of the defined qualitative trait classes, indicating their representativeness of the base collection for qualitative traits (Table 3).

Table 3: Summary of validation statistics to identify representative and best core set (s) of soybean germplasm accessions

| Sl. No	Validation statistic	Core size							
		10% of base collection				15% of base collection			
		Proportional sampling		Logarithmic sampling		Proportional sampling		Logarithmic sampling	
		Random allocation	Preferred allocation	Random allocation	Preferred allocation	Random allocation	Preferred allocation	Random allocation	Preferred allocation
	I Qualitative traits								
1	Significant Chi square (No. of traits)	9	2	8	2	9	2	12	2
2	Shannon-Weaver diversity index (H^1)	1.00±0.48	0.97±0.611	1.00±0.52	1.00±0.50	0.99±0.40	0.99±0.48	1.02±0.43	1.00±0.63
3	Class coverage	96.7	96.7	97.8	96.7	97.8	96.7	97.8	96.7
	II Quantitative traits								
1	Significant 't' test	2	3	2	3	2	4	0	3
2	Significant 'F' test	3	2	3	2	2	4	0	3
3	SMD %	0.6	0.99	1.82	3.37	0.65	1.28	1.42	3.53
4	VD %	10.01	6.89	14.43	13.92	24.52	6.79	9.87	29.79
5	CR %	80.06	90.53	80.96	91.32	84.22	90.11	85.21	92.11
6	VR %	89.31	96.89	92.99	96.95	90.93	96.22	92.99	101.95

(SMD) standardized mean difference; (VD) variance difference; (CR) coincidence ratio; (VR) variable rate

Comparison of SSC Strategies: In the current study, a core size of 15% retained more VD %, CR %, and VR %, indicating greater representativeness of the core sets (Fig. 1a; Table 4). Logarithmic sampling strategy-based core sets better represented the base collection than proportional sampling strategy-based core sets since they retained higher CR %, VR %, and VD % (Fig. 1b; Table 4). Among the two allocation strategies, preferred allocation was superior to random allocation, as evidenced by a lower SMD % and higher CR %, VR %, and VD % (Fig. 1c; Table 4).

Table 4: Comparison of core sizes, sampling strategies, and allocation strategies of developing core sets in Soybean

Statistics	Core sizes		Sampling strategy		Allocation strategy	
	10%	15%	Proportional	Logarithmic	Random	Preferred
SMD %	1.72	1.7	0.88	2.54	2.29	1.12
VD %	16.26	12.8	12.05	17	14.35	14.71
CR%	85.72	88.41	86.23	87.4	82.61	91.02
VR%	95.29	95.52	93.34	96.22	91.56	98
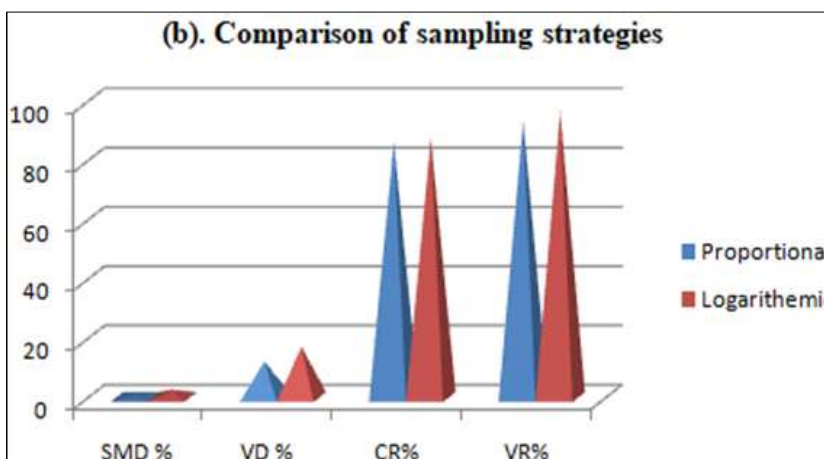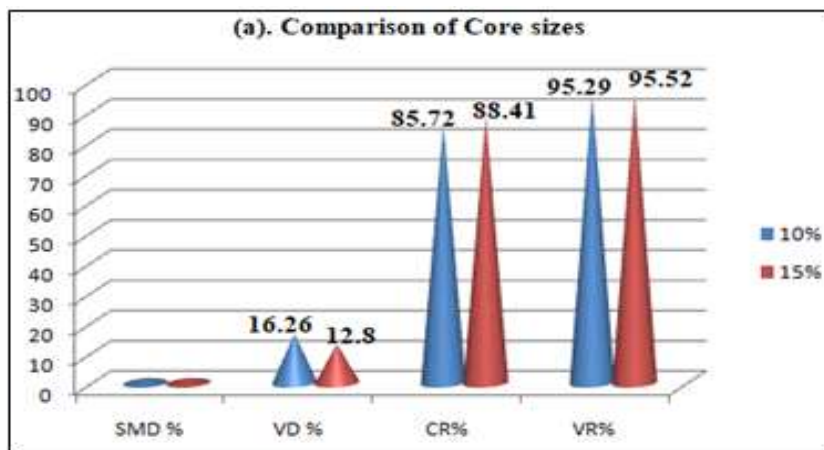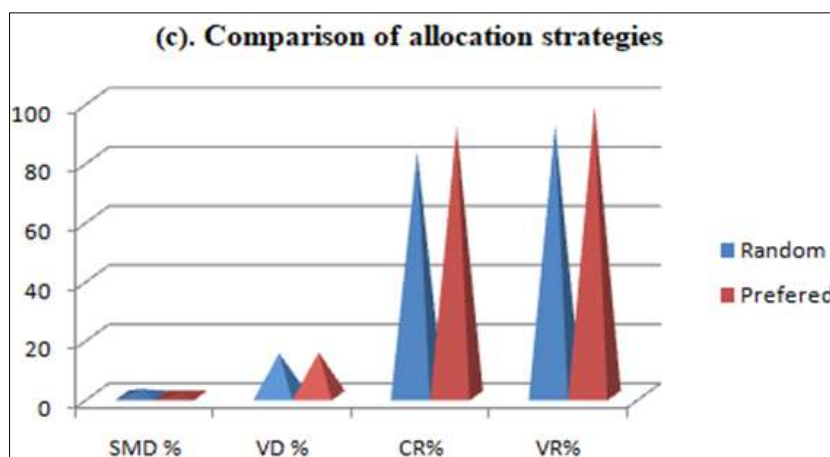
Fig 1: Graph depicting comparison of (a) core sizes, (b) sampling strategies, and (c) allocation strategies for developing core sets in soybean

Efficiency of SSC Approaches: Among the eight representative core sets, logarithmic sampling with preferred allocation approach-based core set of size 15 per cent was identified as the best representative of the base collection since it retained higher CR %, VD %, and VR% based on quantitative traits, H' estimates were comparable to those of the base collection, and "class coverage" statistics covered more than 80 per cent of the defined qualitative trait classes.

Identification of trait-specific accessions from the core set.
Qualitative traits
Plant growth, leaf and floral traits: Genotypes with purple hypocotyl (75.91%) were highest compared to green hypocotyl (24.09%). Purple flowers (53.8%) dominated the collection, followed by white (21.78%), dark purple (12.21%), and light purple (12.21%) flowers. Good plant vigour accessions dominated the collection (80.2%), followed by very good plant vigour (14.19%) and poor plant vigour (5.61%). Indeterminate leaf shape (50.47%) with light green (43.89%) leaves dominated the collection, followed by narrow leaves (23.76%) and broad leaves (22.77%) with white (27.39%), green (15.51%), and dark green (13.2%) leaves. Determinate-type accessions dominated the collection

(59.74%), followed by semi-determinate (21.12%) and indeterminate (19.14%) types (Fig. 2; Table 5a).

Pod and seed traits: Accessions bearing pubescent pods (97.36%) with tawny (64.69%) and light tawny (24.75%) colored pubescence were higher compared to non-pubescent pods (2.64%) and grey colored pubescence (7.92%). The collection was dominated by accessions with normal pubescence density (43.56%), followed by sparse (27.72%), semi-sparse (20.46%), glabrous (3.3%), and dense (2.31%). Accessions with erect (49.5%) type pubescence dominated the collection, followed by semi-appressed (19.8%), appressed (16.17%), curly (10.56%), and retrose tip (1.32%). Dark brown pod accessions dominated the collection (38.61%) followed by brown (31.68%) and light brown (29.7%). Seeds with black colored hilum (45.54%) were higher followed by brown (36.63%), imperfect black (7.59%), grey (5.28%), buff (3.96%) and yellow hilum (0.99%) colored genotypes. Genotypes with yellow-colored seed coats (60.07%) were found to be prominent over yellowish white (9.24%), green (9.24%), reddish brown (5.28%), imperfect black (4.95%), black (2.64%), and buff (0.66%) (Fig. 3; Table 5b).

Table 5a: Variability for plant growth, leaf and floral traits and their frequency in core set of 300 soybean germplasm accessions

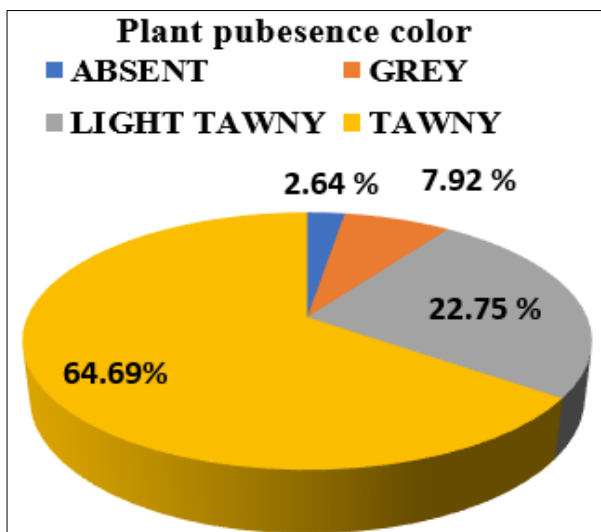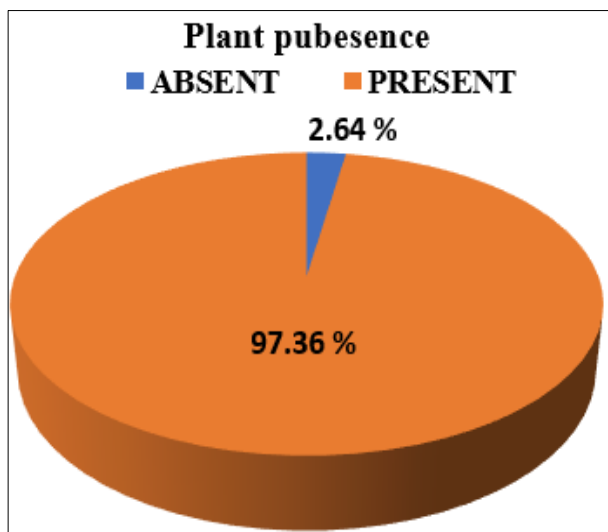| Sl. No.	Traits	Score	Classified as	Frequency	Percentage
1	Hypocotyl color	1	Green	73	24.09
		2	Purple	230	75.91
2	Early plant vigour	1	Poor	17	5.61
		2	Good	243	80.2
		3	Very good	43	14.19
3	Leaf shape	1	Broad	69	22.77
		2	Indeterminate	162	53.47
		3	Narrow	72	23.76
4	Leaflet color	1	White	83	27.39
		2	Light green	133	43.89
		3	Green	47	15.51
		4	Dark green	40	13.2
5	Growth habit	1	Determinate	181	59.74
		2	Semi-determinate	64	21.12
		3	Indeterminate	58	19.14
6	Flower color	1	White	66	21.78
		2	Light purple	37	12.21
		3	Purple	163	53.8
		4	Dark purple	37	12.21
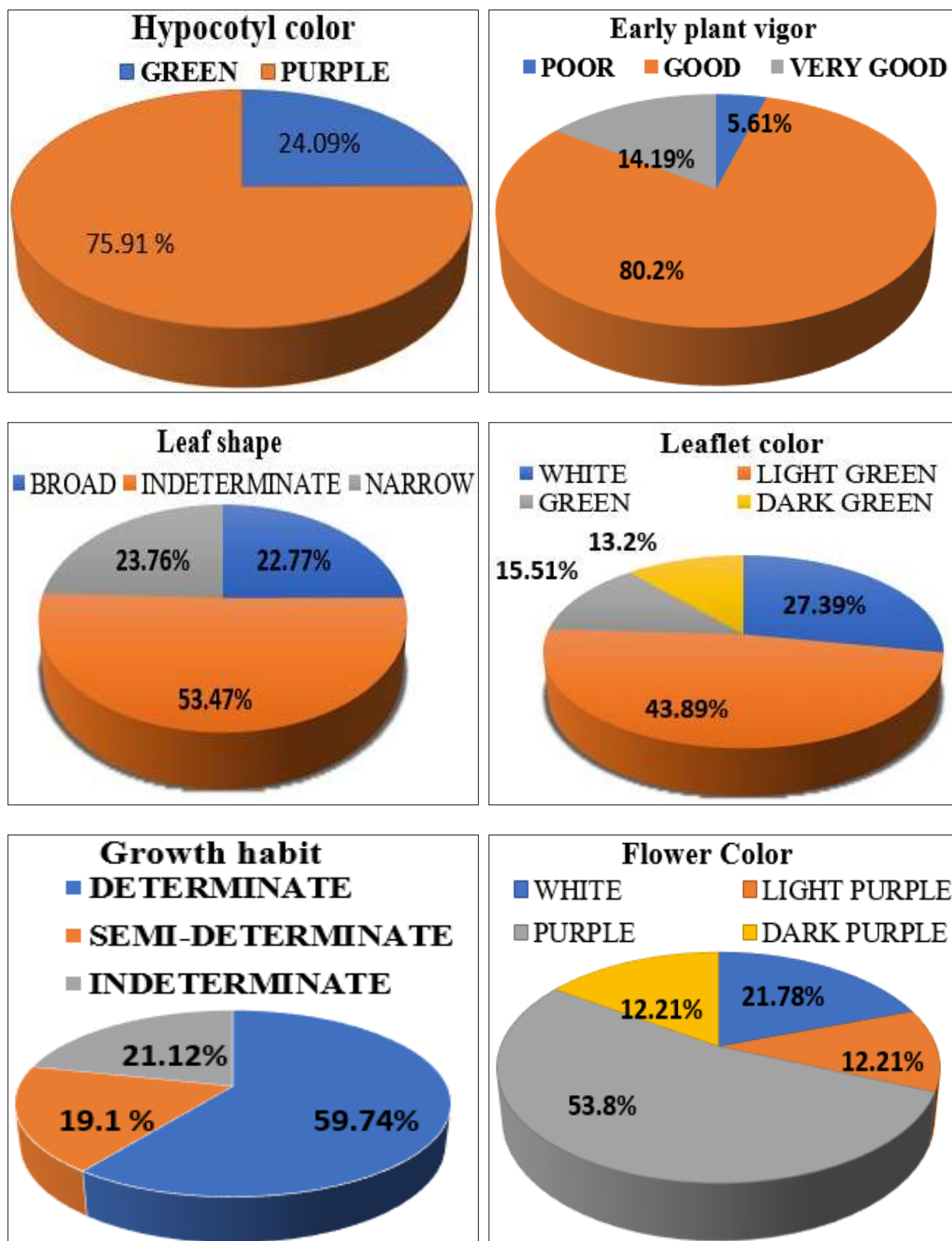
Fig 2: Pie diagram depicting variability for growth habit, leaf and floral traits among core set of 300 soybean germplasm accessions

Table 5b: Variability for pod and seed traits and their frequency in core set of 300 soybean germplasm accessions.

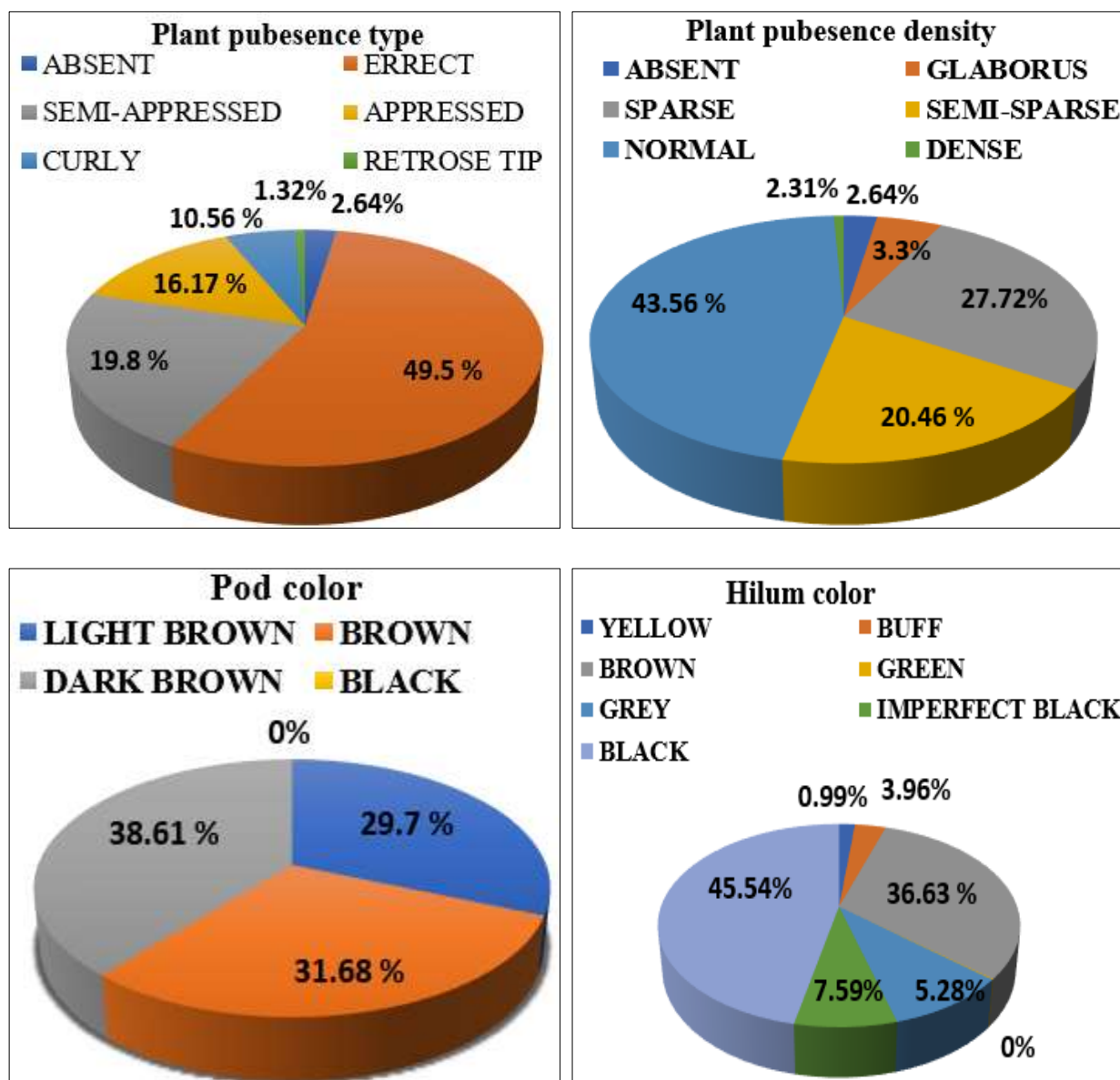| Sl. No.	Traits	Score	Classified as	Frequency	Percentage
1	Plant pubescence	0	Absent	8	2.64
		1	Present	295	97.36
2	Plant pubescence color	0	Absent	8	2.64
		1	Grey	24	7.92
		2	Light tawny	75	24.75
		3	Tawny	196	64.69
3	Plant pubescence density	0	Absent	8	2.64
		1	Glabrous	10	3.3
		3	Sparse	84	27.72
		5	Semi-sparse	62	20.46
		7	Normal	132	43.56
		9	Dense	7	2.31
4	Plant pubescence type	0	Absent	8	2.64
		1	Erect	150	49.5
		3	Semi-appressed	60	19.8
		5	Appressed	49	16.17
		7	Curly	32	10.56
		9	Retrose tip	4	1.32
5	Pod color	1	Light brown	90	29.7
		2	Brown	96	31.68
		3	Dark brown	117	38.61
		4	Black	0	0
6	Hilum color	1	Yellow	3	0.99
		2	Buff	12	3.96
		3	Brown	111	36.63
		4	Green	0	0
		5	Grey	16	5.28
		6	Imperfect black	23	7.59
		7	Black	138	45.54
7	Seed coat color	1	Yellowish white	38	9.24
		2	Yellow	182	60.07
		3	Green	28	9.24
		4	Buff	2	0.66
		5	Reddish brown	16	5.28
		6	Grey	0	0
		7	Imperfect black	15	4.95
		8	Black	22	2.64

Fig 3: Pie diagram depicting variability for pod and seed traits among core set of 300 soybean germplasm accessions

Quantitative traits

ANOVA exhibited a highly significant mean sum of square values for all quantitative traits except days to 80 per cent maturity, while mean sum of square values due to checks were significant for all traits except pods plant[-1] at KVK, Dodballapura (Table 6). Analysis of variance revealed highly significant mean squares attributable to "germplasm accessions" for all traits, while mean sum of square values due to checks were significant for all traits, except pods plant[-1] at K-Block GKVK, Bengaluru. These results indicated differential performance of accessions and checks at both locations.

Table 6: Analysis of variance of a core set of 300 soybean germplasm accessions for seven quantitative traits

Sources of variation	df	Mean sum of squares (MSS)							
		Days to 50% flowering		Plant height (cm)		Secondary branches plant[-1]		Pods plant[-1]	
		KVK Dodballapura	K Block GKVK	KVK Dodballapura	K Block GKVK	KVK Dodballapura	K Block GKVK	KVK Dodballapura	K Block GKVK
Blocks	9	127.09**	128.89**	141.37**	136.10**	0.95**	1.18**	241.13**	252.56**
(Accessions + Checks)	302	60.83**	62.66**	82.42**	81.62**	0.71**	0.63**	134.78**	122.64**
Accessions	299	57.39**	58.60**	68.40**	68.40**	0.54**	0.45**	122.13**	121.49**
Checks	2	123.30**	163.29**	2430.18**	2439.97**	3.33**	3.33**	29.95	29.98
Accessions *vs* Checks	1	2067.59**	2207.29**	760.41**	499.59**	52.21**	58.16**	6207.20**	2891.30**
Error	18	2.63	5.65	9.85	2.91	0.15	0.07	6.94	3.84

Sources of variation	df	Mean sum of squares (MSS)					
		Days to 80% maturity		100 seed weight (g)		Seed yield plant^{-1} (g)	
		KVK Dodballapura	K Block GKVK	KVK Dodballapura	K Block GKVK	KVK Dodballapura	K Block GKVK
Blocks	9	325.51**	308.72**	3.34**	3.87**	2085.06**	2080.55**
(Accessions + Checks)	302	88.47	92.77**	4.86**	4.61**	721.85**	725.40**
Accessions	299	89.55	89.55**	3.95**	3.95**	721.45**	721.45**
Checks	2	389.95*	573.41**	101.36**	53.65**	3246.26**	3783.76**
Accessions *vs* Checks	1	1842.62**	2666.16**	109.60**	134.20**	14264.23**	13851.44**
Error	18	42.42	3.96	0.12	0.18	31.52	8.28

The accessions were highly variable for pods plant^{-1}, 100 seed weight, and seed yield plant^{-1} traits. Broad-sense heritability was higher (> 60%) for all the traits. Estimates of expected GAM were higher for all the traits except days to 80 per cent maturity (10.89%) (Table 7).

Organization of variability among 300 soybean germplasm accessions
With the exception of secondary branches plant^{-1}, the quantitative trait mean differences and variances between the ten clusters were significant for all traits (Tables 8 and 9). The estimates of the means of 7 quantitative traits were highest among the accessions included in cluster X and cluster VII and were least among the accessions included in cluster IX. It is desirable to choose germplasm accessions from among those included in cluster X and cluster VII for various applications in soybean breeding research.

Trait-specific accessions
In the present study, some of the germplasm accessions were comparable to or superior to the check JS-335 with respect to seven quantitative traits (Table 10). The accessions, such as TGX 1895-33 F, AMS 243-7-3, EC 39049, VLS 62, LOCAL 1, SQL 6, SQL 95, EC 572020, EC 389157, SQL 18, JS 20-93, EC 329157, P 1120, EC 100031, TGX 824 32D, TGX 814-148F, SOJA SAVANNA 6, MAUS 60, MAUS 81, EC 287468, EC 343311, AGS 74, TGX 825-3D, EC 589400, and GP-434, were promising for combinations of desirable traits (Table 11).

Table 7: Descriptive statistics for seven quantitative traits in a core set of 300 soybean germplasm accessions

Sl. No.	Traits	Mean±Std. Error	Range		Standardized Range	Coefficient of variability		Broad-sense h² (%)	GAM (%)
			Min	Max		GCV%	PCV%		
1	Days to 50% flowering	64.23±0.44	50	80	30	10.970	11.480	91.34	21.61
2	Plant height (cm)	44.78±0.48	33	69	36	17.370	17.860	94.60	34.8
3	Secondary branches plant^{-1}	3.97±0.04	3	6	3	17.730	18.170	95.21	35.63
4	Pods plant^{-1}	40.67±0.64	14	70	56	25.450	26.360	93.25	50.63
5	Days to 80% maturity	119.16±0.54	97	138	41	6.400	7.750	68.16	10.89
6	100 seed weight (g)	9.59±0.12	3.9	15	11.1	19.690	20.120	95.72	39.68
7	Seed yield plant^{-1} (g)	70.5±1.54	21.6	131.2	109.6	36.590	36.980	97.90	74.58

Table 8: Estimates of quantitative traits means of a core set of soybean germplasm accessions belonging to different clusters

Sl. No.	Size of the cluster	Mean of clusters										F Statistic	Probability
		C$_1$	C$_2$	C$_3$	C$_4$	C$_5$	C$_6$	C$_7$	C$_8$	C$_9$	C$_{10}$		
		25	38	23	30	36	53	21	19	38	20		
1	Days to 50% flowering	74.15	61.61	57.1	74.46	61.87	61.3	68.9	59.26	59.78	62.9	45.39	1×10^{-4}
2	Plant height (cm)	52.66	42.41	39.78	50.34	43.29	39.34	53.51	43.55	39.93	54.97	29.49	1×10^{-4}
3	Secondary branches plant^{-1}	4.02	3.79	3.74	4.1	3.96	3.84	4.02	3.92	3.93	4.07	0.83	0.58
4	Pods plant^{-1}	36.86	43.05	40.26	29.54	48.63	32.75	53.93	54.88	26.02	61.58	322.65	1×10^{-4}
5	Days to 80% maturity	126.22	118.2	113.23	128.4	116.09	114.41	126.6	115.39	112.7	123.26	30.04	1×10^{-4}
6	100 seed weight (g)	8.81	9.8	9.59	8.8	9.84	10.36	8.83	9.79	9.62	9.42	2.39	1×10^{-2}
7	Seed yield plant^{-1} (g)	67.51	80.27	71.66	40.18	92.41	50.86	101.63	104.5	33.54	117.4	505.52	1×10^{-4}

Table 9: Estimates of quantitative traits variances among a core set of soybean germplasm accessions belonging to different clusters

Sl. No.	Size of the cluster	Variance of clusters										F Statistic	Probability
		C$_1$	C$_2$	C$_3$	C$_4$	C$_5$	C$_6$	C$_7$	C$_8$	C$_9$	C$_{10}$		
		25	38	23	30	36	53	21	19	38	20		
1	Days to 50% flowering	17.04	10.1	28.99	13.26	20.33	20.13	30.72	43.61	34.99	18.51	36.22	1×10^{-4}
2	Plant height (cm)	65.6	23.53	23.66	62.19	47.99	17.78	57.98	30.2	20.07	37.44	19.7	1×10^{-4}
3	Secondary branches plant^{-1}	0.64	0.4	0.44	0.48	0.49	0.41	0.59	0.69	0.55	0.31	0.54	0.62
4	Pods plant^{-1}	6.1	4.6	8.88	4.71	3.78	12.46	11.36	73.04	6.55	5.62	262.31	1×10^{-4}
5	Days to 80% maturity	31.13	45.76	40.84	20.99	53.56	65.4	19.45	42.45	54.09	57.01	19.41	1×10^{-4}
6	100 seed weight (g)	2.13	3.14	4.57	4.71	2.32	4.41	2.08	3.99	6.23	2.6	1.62	2×10^{-3}
7	Seed yield plant^{-1} (g)	41.3	13.88	50.05	19.66	20.07	48.69	34.63	254.07	16.87	34.19	349.32	1×10^{-4}

Table 10: Promising trait-specific accessions in a core set of 300 soybean germplasm

Traits	Selection criteria	Range	Germplasm accessions
Days to 50% flowering	Earliness (Mean-1 SD) $\leq$ 56.58	50-56.58	TGX 1895-33 F, SQL 150, SPEL 56 A, SQL 40, SQL 18, JS 20-82, IC 625899, JS 20-60, EC 333899, EC 333905, EC 333912, EC 329157, EC 251531, MAUS 47, JS 20-61, AGS 2L, AGS 112, AGS 97, NB 172, PP 4, LOCAL 1, AMS 243-7-3, PS 1477A, CAT 1003, PP 47, JS 20-93, SQL 95, SQL 6, EC 572020, EC 572126, EC 457179, EC 457164, EC 481360, EC 389157, EC 457367
Plant height (cm)	Dwarf (Mean-1 SD) $\leq$ 36.45	33-36.45	EC 329157, TGX 996-4F6, EC 39049, VLS 62, JS 20-51, EC 357998, TGX 709-7E, EC 39376, JS-20-51, NKG 43, EC 251871
Secondary branches plant^{-1}	High (Mean + 2 SD) $\geq$ 4.73	4.73-6	EC 100031, EC 39049, UPSV 45, JS 98-97, VLS 62, PS 1480, TGX 802-100 F, TGX 868-26 F, PK 726, TGX 1895-33 F, TNAU 5-55, SOJA SAVANNA 6, SL 47, MAUS 60, MAUS 81, NRC 59, GP 562, JS 20-49, JS (HS) 91-93, NRC 21, TGX 825-17D, TGX 824 32D, PK 640, PK 1133, P 1120, TGX 814-148F, TGX 849-47F, TGX 1073-54, P 540, TGX 814-54E, PLSO 92, PLSO 978, SQL 95, EC 572020, EC 274755, EC 457475, EC 457266, EC 389164, EC 389157, EC 389392, EC 396057, EC 389165, EC 394859, EC 241827, EC 114573, EC 241669, EC 241778, NRC 133, IC 279791, IC 393206, CAT 2070B, IC 501844
Pods plant^{-1}	High (Mean + 2 SD) $\geq$ 51.73	51.73-70	EC 100031, EC 39045, EC 39049, EC 2581, V 61, V 31, VLS 11, UGM 73, JS 99-76, K 53, UPSM 181, UPSM 1024, TGX 802-100 D, TK 5, SEHORE 7, TGX 1895-33 F, PKS 45, SOJA SAVANNA 6, SQL 18, MACS 263, MAUS 60, MAUS 81, JS 75-19, EC 287468, EC 343311, EC 274684, EC 250577, GP-38, TGX 293-41E, GP-434, PK-1038, AGS 74, UPSM-534, DN 290, AMS 26 A, DSB 9, AMS 243-7-3, LEE 46, JS 79-307, JS 20-93, HIMSO 1597, ICAL 138, TGX 825-3D, TGX 816-3E, TGX 824 32D, P 1120, TGX 814-148F, PK 731, SIZTA 194, SL 142, EC 589400, EC 389160B, EC 389173, EC 394859, IC 25000
Days to 80% maturity	Earliness (Mean-1 SD) $\leq$ 109.71	97-109.71	EC 77210, VLS 62, KARUNE, EC 333905, GC-84040-7-1, RICUM, GP-434, AGS 33, AMS MB 51-18, AGS 97, FASSY COOK, AK 887, LOCAL 1, AMS 243-7-3, MAUS 5286, J 732, CAT 1003, JS 20-75, P 540, TGX 814-54E, SQL 95, SQL 6, EC 572020, EC 274755, EC 528639, EC 538840, EC 538815, EC 572126, EC 457475, EC 457164, EC 457266, EC 383165, EC 389157, EC 457367, EC 396057, EC 396058B, EC 241309, EC 241709, IC 345625, IC 279791, IC 341347, IC 393206, CAT 2070B
100 seed weight (g)	High (Mean + 2 SD) $\geq$ 11.6	11.6-15	VP 1177 B, VP 1147 A, PS 1506, VLS 62, PKS 15, KB 17, TGX 854-42D-1, SQL 91, SQL 40, SL 96, JS 20-82, JS 20-51, MACS 1259, JS 20-30, JS 20-60, EC 287468, EC 274747, EC 281462, EC 291451, EC 251446, EC 309512, EC 309512, EC 329157, EC 343311, G-2258, AGS 74, DS 76-1-139-2, LOCAL 1, PS 1477A, MAUS 5286, TGX 85B-48, TGX 825-3D, TS 98-1, TGX 709-7E, P 540, PK 644, SQL 6, PRAB 1, EC 589400, EC 457179, EC 457266, EC 456610, EC 456610, EC 393223, EC 114573, IC 501760
Seed yield plant^{-1} (g)	High (Mean + 2 SD) $\geq$ 97.38	97.38-131.2	EC 100031, EC 39045, EC 39049, EC 2581, V 61, V 31, VLS 11, UGM 73, JS 99-76, K 53, UPSM 181, UPSM 1024, TGX 802-100 D, TK 5, SEHORE 7, TGX 1895-33 F, PKS 45, SOJA SAVANNA 6, SQL 18, MACS 263, MAUS 60, MAUS 81, JS 75-19, EC 287468, EC 343311, EC 274684, EC 250577, GP-38, TGX 293-41E, GP-434, PK-1038, AGS 74, UPSM-534, DN 290, AMS 26 A, DSB 9, AMS 243-7-3, LEE 46, JS 79-307, JS 20-93, HIMSO 1597, ICAL 138, ICAL 138, TGX 825-3D, TGX 816-3E, TGX 824 32D, P 1120, TGX 814-148F, PK 731, SIZTA 194, SIZTA 194, SL 142, EC 589400, EC 481498, EC 389160B, EC 389173, EC 394859, EC 333880, IC 25000

Table 11: Promising accessions identified for multiple traits in a core set of 300 soybean germplasm

Accessions	Traits
TGX 1895-33 F	Days to 50% flowering, Pods plant^{-1}, Secondary branches plant^{-1}, Seed yield plant^{-1}
AMS 243-7-3	Days to 50% flowering, Pods plant^{-1}, Secondary branches plant^{-1}, Days to 80% maturity, Seed yield plant^{-1}
EC 39049	Plant height, Pods plant^{-1}, Secondary branches plant^{-1}, Seed yield plant^{-1}
VLS 62	Plant height, Days to 80% maturity, Secondary branches plant^{-1}, 100 seed weight
LOCAL 1, SQL 6	Days to 50% flowering, Days to 80% maturity, 100 seed weight
SQL 95, EC 572020, EC 389157	Days to 50% flowering, Days to 80% maturity, Secondary branches plant^{-1}
SQL 18, JS 20-93	Days to 50% flowering, Pods plant^{-1}, Seed yield plant^{-1}
EC 329157	Days to 50% flowering, Plant height, 100 seed weight
P 1120, EC 100031, TGX 824 32D, TGX 814-148F, SOJA SAVANNA 6, MAUS 60, MAUS 81	Secondary branches plant^{-1}, Pods plant^{-1}, Seed yield plant^{-1}
EC 287468, EC 343311, AGS 74, TGX 825-3D, EC 589400	100 seed weight, Pods plant^{-1}, Seed yield plant^{-1}
GP-434	Days to 80% maturity, Pods plant^{-1}, Seed yield plant^{-1}

Conclusions

The soybean core collection created in this study will provide valuable genetic resources for soybean breeders and researchers for screening soybean germplasm and identifying desirable genotypes for economically important traits. Core collection development will also aid in addressing climate change challenges because core collections represent the genetic variability of the entire collection and desirable genotypes can be easily identified. The soybean core collection created in the current study can also be used in

association mapping studies to identify the genes and QTLs linked to numerous economically significant features. The trait-specific accessions from the best core set are suggested for preferential use in crossing programmes to generate variability for developing farmer-acceptable varieties with consumer/end-user-preferred traits.

References

1. Allen LH, Boote KJ. Crop ecosystem responses to climatic change: Soybean. In Climate Change and Global Crop Productivity (K.R. Reddy and H.F Hodges, Eds.). CABI Publishers, New York; c2000. p. 133-160.
2. Carter TE, Nelson RL, Sneller CH, Cui Z. Genetic diversity in soybean. In Boerma H, Specht J (eds) Soybeans: Improvement, production, and uses. American Society of Agronomy, Madison, (Soybean monograph); c2004. p. 303-450.
3. Carver BF, Burton JW, Carter Jr TE, Wilson RF. Response to environmental variation of soybean lines selected for altered unsaturated fatty acid composition. Crop Science. 1986;26:1176-1180.
4. Cure JD, Acock B. Crop responses to carbon dioxide doubling: a literature survey. Agricultural and Forest Meteorology. 1986;38:127-145.
5. Curry RB, Jones JW, Boote KJ, Peart RML, Allen H, Pickering NB. Response of soybean to predicted climate change in the USA. In Climate Change and Agriculture: Analysis of Potential International Impacts (C. Rosenzweig, Ed.). ASA Special Publication. 1995;59:163-182.
6. Debnath D, Babu S. Prospects for sustainable intensification of soybean production in Sub-Saharan Africa. African Journal of Agricultural and Resource Economics. 2020;15(4):365-371.
7. Desclaux D, Roumet P. Impact of drought stress on the phenology of two soybean (*Glycine max* L. Merr) cultivars. Field Crops Research. 1996;46:61-70.
8. Dornbos DL, Mullen RE. Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. Journal of the American Oil Chemists Society. 1992;69:228-231.
9. Dutton B. Top 10 Soybean Producing Countries; c2019. Available from: www.cropprophet.com.
10. FAOSTAT. Databases: Soybean Production and Trade Data; c2019. Available from: http://www.fao.org/faostat/en/#data.
11. FAOSTAT. Databases: Soybean Production and Trade Data; c2020. Available from: http://www.fao.org/faostat/en/#data.
12. Gibson LR, Mullen RE. Soybean seed composition under high day and night growth temperatures. Journal of American Oil Chemists Society. 1996;73:733-737.
13. Global Soybean Market Outlook; c2022. Available from: https://www.expertmarketresearch.com/reports/soybean-market.
14. Hu J, Zhu J, Xu HM. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. Theor. Appl. Genet. 2000;101:264-268.
15. Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwag JG, *et al*. Power Core: A program applying the advanced M strategy with a heuristic search for establishing core sets. Bioinformatics. 2007;23:2155-2162.
16. MacQueen JB. Some Methods for classification and analysis of multivariate observations, Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press. 1967;1:281-297.
17. Mall RK, Lal M, Bhatia VS, Rathore LS, Singh R. Mitigating climate change impact on soybean productivity in India: A simulation study. Agricultural and Forest Meteorology. 2004;121:113-125.
18. Piper EL, Boote KJ. Temperature and cultivar effects on soybean seed oil and protein concentrations. Journal of American Oil Chemists Society. 1999;76:1233-1241.
19. Ramesh K, Ashok K, Patra, Ashis KB. Best Management Practices for Soybean under Soybean-Wheat System to Minimize the Impact of Climate Change Article. Indian Journal of Fertilizers. 2017;13(2):42-55.
20. Shannon CE, Weaver W. The mathematical theory of communication. University Illinois Press, Urbana, USA; c1949.
21. Thomas JMG, Boote KJ, Allen LH, Gallo-Meagher M, Davis JM. Elevated temperature and carbon dioxide effects on soybean seed composition and transcript abundance. Crop Science. 2003;43:1548-1557.