LR-5215
[1-9]

RESEARCH ARTICLE **Legume Research- An International Journal**

Development and Validation of Soybean [*Glycine max* (L.) Merrill] Core Sets and Identification of Trait-specific Accessions from the Best Core Set

Kaveri Chawan[1], P. Ravishankar[1], S. Ramesh[1], T. Onkarappa[1], H.H. Sowmya[1]

10.18805/LR-5215

ABSTRACT

Background: Core collection of germplasm accelerates breeding objective of a crop. A core set of trait specific accessions reduces time, money and valuable man power in any crop breeding system and Standard stratified clustering approach is the most preferred approach to construct the coreset.

Methods: During summer of 2020-21, the genetic variability of 2000 soybean germplasm accessions were assessed at University of Agricultural Sciences, Bengaluru. The clustering approach was deployed to develop 8 core sets from the 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 of the base collection. During *kharif* 2021, 300 germplasm accessions (15% core size) were evaluated at two sites *viz*., GKVK, Bengaluru and KVK Doddaballapur, to identify trait-specific accessions from the best core set.

Result: 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 found promising for the combination of desirable traits from the best core set suggesting their preferential use in breeding programme.

Key words: Core collection, Germplasm, K-means clustering, Standard stratified clustering, Trait-specific.

INTRODUCTION

Soybean [*Glycine max* (L.) merrill] is an important pulse and oilseed crop rich in protein (40%) and oil (20%). Over 1,70,000 soybean germplasm collection is maintained at various gene banks throughout the world and forms the basis of the diversity available in soybean (Carter *et al.,* 2004). Increased use of available genetic resources is required for diversifying genetic base cover to changing consumer and end-user preference and address biotic and abiotic stress emerging to the effect of climate change. The large size of germplasm, possibility of duplicate/ redundant accessions and /or repeated sampling of the same accessions are obstacles to effective management and evaluation (Upadhyaya *et al*., 2001; Jain *et al*., 2017). The use of only a small portion of the available germplasm within any crop is attributed to lack of quality information on evaluation data (Brown 1989) emphasising the need for a core set. A good core collection includes cultivars, breeding lines, land races and wild species and should have distinguished characterstics like representativeness, low redundancy manageability, data completeness and usefulness (Van hintum *et al.,* 2000). Core collections developed in medicago (Diwan *et al.,* 1995), chickpea (Upadhyaya *et al.,* 2001), pigeonpea (Reddy *et al.,* 2005), cowpea (Mahalakshmi *et al.,* 2007), rice (Yan *et al.,* 2007) is effectively utilized in improvement of respective crops.

Several methods were developed and adopted to construct a core set like standard, stratified clustering (SSC) (Brown, 1989), M-Stat (Gousenard and Bataillon 2001), genetic distance sampling (Jansen and Van Hintum, 2007) and core hunter (Thachuk *et al.,* 2009). The SSC approach is preferred by most researchers to ensure selection of common alleles (Crossa *et al.,* 1995). Several workers (Minmin *et al*., 2020; Darai *et al*., 2020) have identified trait -specific accessions in soybean.

Current investigation was carried out to develop core sets from a large group of 2000 germplasm accessions using the SSC approach and the best set was selected using various univariable and multivariate validation statistics. Trait-specific accessions were finally listed as useful information for the scientific group.

---

[1]Department of Genetics and Plant Breeding, College of Agriculture, Gandhi Krishi Vigyana Kendra, University of Agricultural Sciences, Bengaluru-560 065, Karnataka, India.

Corresponding Author: H.H. Sowmya, Department of Genetics and Plant Breeding, College of Agriculture, Gandhi Krishi Vigyana Kendra, University of Agricultural Sciences, Bengaluru-560 065, Karnataka, India. Email: sowmyahh.cta@gmail.com

How to cite this article: Chawan, K., Ravishankar, P., Ramesh, S., Onkarappa, T. and Sowmya, H.H. (2024). Development and Validation of Soybean [*Glycine max* (L.) Merrill] Core Sets and Identification of Trait-specific Accessions from the Best Core Set. Legume Research. doi: 10.18805/LR-5215.

Submitted: 27-07-2023 **Accepted:** 07-03-2024 **Online:** 14-05-2024

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 2020-21 following augmented design at University of Agricultural Sciences, Bangalore. Each block has consisted of 100 germplasm accessions and three checks (replicated twice). Observations were recorded on five randomly selected plants based on visual observation for 13 qualitative traits and 7 quantitative traits. 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, *i.e.*, 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) and 'class coverage' (Kim *et al.*, 2007) 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) 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 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) 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 higherexpression of the accessions for the other four traits (mean + 2 SD), 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

Table 1: Summary of validation statistics to identify representative and best core set (s) of soybean germplasm accessions.

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								
Significant Chi square (No. of traits)	9	2	8	2	9	2	12	2
Shannon-Weaver diversity index (H')	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
Class coverage	96.7	96.7	97.8	96.7	97.8	96.7	97.8	96.7
II Quantitative traits								
Significant 't' test (No. of traits)	2	3	2	3	2	4	0	3
Significant 'F' test(No. of traits)	3	2	3	2	2	4	0	3
SMD %	0.6	0.99	1.82	3.37	0.65	1.28	1.42	3.53
VD %	10.01	6.89	14.43	13.92	24.52	6.79	9.87	29.79
CR %	80.06	90.53	80.96	91.32	84.22	90.11	85.21	92.11
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.

for proportional and logarithmic sampling with random allocation of sizes 10 percent and 15 percent, 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 1).

All the core sets were comparable to those in the base collection for quantitative trait means ('t' test was significant for ≤4 traits). The SMD (%) of all eight core sets was less than 4, confirming their representativeness for quantitative trait means. Logarithmic sampling with preferred allocation strategy (of 15% size) retained higher VD (%), CR (%) and VR (%) than other approaches-based core sets (Table 1).

Comparison of SSC strategies

In the current study, a core size of 15% has retained more VD (%), CR (%) and VR (%), indicating greater representativeness of the core sets (Table 2). 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 (%) (Table 2). Among the two allocation strategies, preferred allocation was superior to random allocation, as evidenced by a lower SMD (%) and higher CR (%), VR (%) and VD (%) (Table 2) (Crossa *et al*., 1995).

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%), Purple flowers (53.8%), indeterminate leaf shape (50.47%) with light green (43.89%) leaves dominated the collection (Table 3a).

Pod and seed traits

Accessions bearing pubescent pods (97.36%) with tawny (64.69%) and light tawny (24.75%) colored pubescence, erect type pubescence (49.5%), dark brown pod accessions (38.61%) dominated the collection. Genotypes with yellow-

Table 2: 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

Table 3a: Variability for plant growth, leaf and floral traits and their frequency in core set of 300 soybean germplasm accessions.

Traits	Score	Classified as	Frequency	Percentage
Hypocotyl color	1	Green	73	24.09
	2	Purple	230	75.91
Leaf shape	1	Broad	69	22.77
	2	Indeterminate	162	53.47
	3	Narrow	72	23.76
Leaflet color	1	White	83	27.39
	2	Light green	133	43.89
	3	Green	47	15.51
	4	Dark green	40	13.2
Growth habit	1	Determinate	181	59.74
	2	Semi-determinate	64	21.12
	3	Indeterminate	58	19.14
Flower color	1	White	66	21.78
	2	Light purple	37	12.21
	3	Purple	163	53.8
	4	Dark purple	37	12.21

colored seed coats (60.07%) were found to be prominent (Table 3b) (Shruthi *et al*., 2022).

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 for pods plant^{-1} at KVK, Dodballapura (Table 4). 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 for pods plant^{-1} at K-Block GKVK, University of Agricultural Sciences, Bengaluru. These results indicated differential performance of accessions and checks at both locations.

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 5) (Darai *et al*., 2020; Shruthi *et al*., 2020).

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 (Table 6 and 7). 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.

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 8) (Minmin *et al*. 2020). The accessions listed in Table 9 were promising for combination of desirable traits.

Table 3b: Variability for pod and seed traits and their frequency in core set of 300 soybean germplasm accessions.

Traits	Score	Classified as	Frequency	Percentage
Plant pubescence	0	Absent	8	2.64
	1	Present	295	97.36
Plant pubescence color	0	Absent	8	2.64
	1	Grey	24	7.92
	2	Light tawny	75	24.75
	3	Tawny	196	64.69
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
Pod color	1	Light brown	90	29.7
	2	Brown	96	31.68
	3	Dark brown	117	38.61
	4	Black	0	0
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
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

Table 4: 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

Table 4: contd...

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

Table 5: Descriptive statistics for seven quantitative traits in a core set of 300 soybean germplasm accessions.

Traits	Mean±Std. Error	Range		Standardized range	Coefficient of variability		Broad-sense h² (%)	GAM (%)
		Min	Max		GCV%	PCV%		
Days to 50% flowering	64.23±0.44	50	80	30	10.970	11.480	91.34	21.61
Plant height (cm)	44.78±0.48	33	69	36	17.370	17.860	94.60	34.8
Secondary branches plant^{-1}	3.97±0.04	3	6	3	17.730	18.170	95.21	35.63
Pods plant^{-1}	40.67±0.64	14	70	56	25.450	26.360	93.25	50.63
Days to 80% maturity	119.16±0.54	97	138	41	6.400	7.750	68.16	10.89
100 seed weight (g)	9.59±0.12	3.9	15	11.1	19.690	20.120	95.72	39.68
Seed yield plant^{-1} (g)	70.5±1.54	21.6	131.2	109.6	36.590	36.980	97.90	74.58

Table 6: Estimates of quantitative traits means of a core set of soybean germplasm accessions belonging to different clusters.

Size of the cluster	Mean of clusters										F Statistic	Probability
	C_1 25	C_2 38	C_3 23	C_4 30	C_5 36	C_6 53	C_7 21	C_8 19	C_9 38	C_{10} 20		
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}
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}
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
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}
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}
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}
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 7: Estimates of quantitative traits variances among a core set of soybean germplasm accessions belonging to different clusters.

Size of the cluster	Variance of clusters										F Statistic	Probability
	C_1 25	C_2 38	C_3 23	C_4 30	C_5 36	C_6 53	C_7 21	C_8 19	C_9 38	C_{10} 20		
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}
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}
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
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}
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}
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}
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 8: Promising trait-specific accessions in a core set of 300 soybean germplasm.

Traits	Selection criteria	Range	Germplasm accessions
Days to 50% flowering	Earliness (Mean-1SD) ≤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-1SD) ≤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+2SD) ≥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+2SD) ≥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-1SD) ≤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+2SD) ≥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+2SD) ≥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-

Table 8: Continue...

Table 8: Continue...

	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 9: 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}

CONCLUSION

The soybean core collection developed 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 and addressing the climate change challenges. 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 programme to generate variability for developing farmer-acceptable varieties with consumer/end-user-preferred traits.

Conflict of interest

All authors declared that there is no conflict of interest.

REFERENCES

Brown, A.H.D. (1989). Core collections: A practical approach to genetic resources management. Genome. 31: 818-824.

Carter, T.E., Nelson, R.L., Sneller, C.H. and Cui, Z. (2004). Genetic Diversity in Soybean. In: Soybeans: Improvement, Production and Uses. [Boerma, H., Specht, J. (eds)], American Society of Agronomy, Madison, (Soybean monograph). p. 303-450.

Crossa, J., Delacy, I.H. and Taba, S. (1995). The use of Multivariate Methods in Developing a Core Collection. In: Core Collections of Plant Genetic Resources [Hodgkin, T., Brown, A.H.D., Van Hintum, TH.J.L. and Morales, E.A.V. (eds.)]. John Wiley and Sons, UK. Pp. 77-92.

Darai, R., Dhakal, K.H. and Sah, R.P. (2020). Genetic variability of soybean accessions for yield and yield attributing traits through using multivariate analysis. Int. J. Plant Res. 4(3): 10-16.

Diwan, N., Mcintosh, M.S. and Bauchan, G.R. (1995). Methods of developing a core collection of annual Medicago species. Theor. Appl. Genet. 90: 755-761.

Gousenard, B. and Bataillon, T.M. (2001). MSTART: An algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. J. Hered. 92(1): 93-94.

Hu, J., Zhu, J. and Xu, H.M. (2000). Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. Theor. Appl. Genet. 101: 264-268.

Jain, R.K., Joshi, A., Chaudhary, H.R., Dashora, A. and Khatik, C.L. (2017). Study on genetic variability, heritability and genetic advance in soybean [*Glycine max* (L.) Merrill]. Legume Res. 41(4): 532-536. doi: 10.18805/LR-3874.

Jansen, J. and Van Hintum, J.L. (2007). Genetic distance sampling: A novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. Theor. Appl. Genet. 114(3): 421-428.

Kim, K.W., Chung, H.K., Cho, G.T., Ma, K.H., Chandrabalan, D., Gwag, J.G., Kim, T.S., Cho, E. G. and Park, Y.J. (2007). Power core: A program applying the advanced M strategy with a heuristic search for establishing core sets. Bioinformatics. 23: 2155-2162.

MacQueen, J.B. (1967). Some Methods for Classification and Analysis of Multivariate Observations, Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press. 1: 281-297.

Mahalakshmi, V., NG, Q., Lawson, M. and Ortiz, R. (2007). Cowpea [*Vigna unguiculata* (L.) Walp.] core collection defined by geographical, agronomical and botanical discripters. Plant Genetic Resources: Characterization and Utilization. 5(3): 113-119.

Minmin, L., Ying, L., Chunsheng, W., Xue, Y., Dongmei, L., Xiaoming, Z., Chongjing, X., Yan, Z., Wenbin and Lin, Z. (2020). Idnetification of trait contributing to high and stable yield in different soybean varieties across three chinese lattitudes. Front. Plant. Sci. 10: 1642. https://doi.org/10.3389/fpls.2019.01642.

Reddy, L.J., Upadhyaya, H.D., Gowda, C.L.L. and Singh, S., (2005), Development of core collection in chickpea [*Cajanas cajan* (L.) Millspaugh] using geographic and quaalitative morphological discripters. Genet. Resour. Crop Evol. 52: 1049-1056.

Shannon, C.E. and Weaver, W. (1949). The Mathematical Theory of Communication. University Illinois Press, Urbana, USA.

Shruthi K., Siddaraju R., Naveena K., Ramanappa T.M., Vishwanath K. (2020). Assessment of variability based on morphometric characteristics in the core set of soybean germplasm accessions. Legume Resarch. 44(4): 375-381. doi: 10.18805/LR-4286.

Shruthi K., Siddaraju R., Naveena K., Ramanappa T.M., Gireesh C., Vishwanath K., Nagaraju K.S. (2022). Trait based modelling approach for selection of elite germplasm accessions in soybean [*Glycine max* (L). Merrill]. Legume Res. 45(7): 822-827. doi: 10.18805/LR-4567.

Thachuk, C., Crossa, J., Franco, J., Dreisigacker, S., Warburton, M. and Davenport, G.F. (2009). Core hunter: An algorithm for sampling genetic resources based on multiple genetic measures. BMC Bioinformatics. 10: 243. doi: 10.1186/1471-2105-10-243.

Upadhyaya, H.D., Bramel, P.J. and Singh, S. (2001). Development of a chickpea core subset using geographic distribution and quantitative traits. Crop Sci. 41: 206-210.

Van Hintum, T.H.J.L., Brown, A.H.D., Spillane, C. and Hodgkin, T. (2000). Core collections of plant genetic resources. IPGRI Technical Bulletein 3. International Plant Genetic Resources Institute, Rome, Italy.

Yan, W., Rutger, J.N., Bryant, R.J., Bockelman, H.E., Fjellstrom, R.G., Chen, M.H., Tai, T.H. and Mcclung, A.M. (2007). Development and evaluation of a core subset of the USDA rice germplasm collection. Crop Sci. 47: 869-878.