INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
21–1010/2022/27–3–184–192
DOI: 10.17957/IJAB/15.1915
http://www.fspublishers.org

Full Length Article

Estimating the Breeding Potency of a Soybean Core Set

E. Vijayakumar[1], R. Sudhagar[2*], C. Vanniarajan[1], J. Ramalingam[4], V. Allan[3] and N. Senthil[4]
[1]*Department of Plant Breeding and Genetics, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Madurai, India*
[2]*Sugarcane Research Station, Tamil Nadu Agricultural University, Melalathur, India*
[3]*Department of Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore, India*
[4]*Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore, India*
[*]For correspondence: sudhagar.r@tnau.ac.in
Received 30 November 2021; Accepted 08 January 2022; Published 30 March 2022

Abstract

Multi-environment evaluation of a core set helps in the identification of trait-specific genetic stocks for further exploitation to sustain soybean productivity. TNAU soybean core set comprising of 50 soybean germplasm (thirty national and twenty international lines) was evaluated for variability pattern, trait association and adaptability using ten quantitative and twenty-one qualitative characters over environments, years and seasons to identify potential parents for utilization. Both quantitative and qualitative traits-based diversity analyses indicated the presence of adequate variability. The principal component analysis based on quantitative traits reduced the total variation into three major components (73.72%). Genotypes JS9305, JS(SH)99-02, IC16009 and TNAU20048 were tagged as highly divergent lines. Single plant yield (0.92) and number of pods per plant (0.92) contributed the maximum for the rotated PC1. The association analysis revealed that number of pods per plant (r = 0.88) and number of clusters per plant (r = 0.81) contributed significantly to the single plant yield. The qualitative cluster analysis divided the core set into eight clusters. A good variability was observed for plant type, pod color, seed coat color, and hilum color. In the validation experiments, a higher yield was witnessed in the hybridization involving distantly related accessions (14.1%) than the closely related parents (7.5%). It is concluded that the TNAU soybean core set has real breeding potency for future exploitation. The parents JS9305, JS(SH)99-02, IC16009 and TNAU20048 possessed a higher degree of divergence, heterosis with better adaptability and offers scope of breeding utility.

Keywords: Core collection; Clustering; Genetic divergence; PCA; Soybean; Validation

Introduction

Soybean [*Glycine max* (L.) Merr.] is the "Miracle legume" owing to its extraordinary oil, protein, vitamin and mineral contents with several health benefits (Vishwanath *et al.* 2021). It provides several vital raw materials for many industries thereby contributing significantly to the Indian economy. Agriculturally, its cultivation substantially enriches soil fertility by atmospheric nitrogen fixation (Jain *et al.* 2018). The genetic yield potential of soybean can be realized in the Indian sub-continent through (i) characterization of various germplasm collections maintained at different regions (ii) identification of trait-specific adaptable accessions and (iii) utilization of such identified genotypes in the targeted pre-breeding programs. The geographical situation makes a shift in the phenological pattern and yield in soybean (Bisen *et al.* 2015). Further, a narrow genetic base and stress susceptibility reduce yield in soybean (El-Harty *et al.* 2018) and therefore necessitate the identification of region-specific donors for yield improvement.

Pre-breeding by utilizing such genotypes gives an exceptional opportunity to broaden the genetic base and thereon evolution of potential segregants becomes possible (Sharma *et al.* 2013). A core set of germplasm represents the total variability of that respective collection and it facilitates precise pre-breeding activities (Upadhyaya *et al.* 2009). Muthamizhan *et al.* (2016) developed a core set representing the soybean germplasm pool of Tamil Nadu Agricultural University whose variability pattern, trait dependability, and adaptability are to be assessed. The utility of quantitative traits in core set characterization is well documented (Upadhyaya *et al.* 2009). The qualitative traits are helpful in the identification of morphological markers which eases the selection process in different plant breeding cycles and provide reliable opportunities for varietal characterization. The Principal Component Analysis (PCA) and association analysis are helpful in tagging desirable genotypes and traits with their contribution to the total divergence (Mohan *et al.* 2019). Yadav (2016)

established the utility of the morphological descriptors (NDUS – Novelty Distinctness Uniformity and Stability) for soybean cultivar identification. The usefulness of R language and UPGMA in germplasm characterization were well reported (Epskamp *et al.* 2019).

The production of soybean in India has been drastically reduced over the years due to poorly adapted genotypes and rainfed cultivation (Rao and Chaitanya 2020) where Tamil Nadu is not an exception. The TNAU soybean germplasm is constituted over a period of time (1975-2018) through inclusion of various trait specific exotic and indigenous genotypes with a view to develop location specific germplasm. A TNAU soybean core set was developed first of its kind in Tamil Nadu soybean improvement programs, whose validation is the need of the hour to improve soybean productivity in Tamil Nadu, India. Therefore, the present study aimed to evaluate the adaptability of a TNAU soybean core set, identification of desirable trait specific genotypes, and their validation through targeted hybridization programs.

Materials and Methods

Experimental Materials

The experimental material comprised of a soybean core collection of 43 accessions and seven checks developed from a global germplasm collection maintained at Dr. Ramaiah gene Bank, Tamil Nadu Agricultural University (TNAU), Coimbatore (Muthamizhan *et al.* 2016). The core set includes thirty National and twenty international genotypes whose geographical origin is depicted in Fig. 1.

Experimental Season and Location

The preliminary core set development experiment was conducted at the Department of Pulses, Centre for Plant Breeding and Genetics, TNAU, Coimbatore, during *Kharif* season 2016. The adaptability of the core set was confirmed through second and third-year confirmative experiments during *Rabi* 2016 and summer 2017, respectively.

Experimental Design and Data Documentation

The experiments were conducted in a Randomized Block Design (RBD) with three replications. The spatial pattern was 30 cm × 10 cm. The genotypes were grown in four-meter rows. All the recommended packages of practice were followed to raise healthy crops as set forth in Crop Production Guide for Tamil Nadu (TNAU 2016). Both quantitative and qualitative traits were utilized to assess the potency of the soybean core set.

Quantitative Trait-based Experiments

For quantitative trait based PCA, in each replication ten plants were randomly selected per genotype and ten following quantitative characters plant height, days to fifty percent flowering, days to maturity, leaf area, number of primary branches, number of clusters per plant, number of pods per plant, number of seeds per pod, hundred seed weight, and single plant yield were recorded. Days to fifty percent flowering and maturity were recorded at appropriate stages. Leaf area was measured at the third fully opened leaf from meristem. The other traits were measured at harvest.

Qualitative Trait-based Experiments

For qualitative trait-based cluster analysis, in each replication ten plants were randomly selected per genotype and observations on twenty-one qualitative characters (Table 1) were documented as per the soybean descriptor of Bioversity International at appropriate growth stages (Ramteke and Murlidharan 2012) and used for clustering.

Validation of the Soybean Core Set

To test verify the potential of the TNAU soybean core set, a set of contrast and closely related parents were selected for effecting hybridizations during *Kharif*, 2017. The parents were selected based on the results of quantitative trait based PCA, qualitative trait-based similarity index, and yield performance. The genotypes that are scattered far apart in PCA biplot were tagged as diverse and the accessions located near the origin were considered as closely related parents. The genotypes with higher and lower similarity indices were treated as closely related and *vice versa.* Accordingly, the genotypes JS(SH)99-02, IC16009, TNAU20048 and JS9305 were selected as more diverse (group I) while the accessions NRC77, UGM75, EC30198 and Co (Soy)3 were earmarked as closely related (group II). Owing to yield potential, the genotypes Co (Soy)3 and IC16009 were utilized as females in the group I and II respectively while other genotypes were used as male parents. In each group, three crosses were made. The progenies of crosses were handled separately. The F_1 seeds (20–30 seeds per cross) were sown during *Rabi*, 2017. The true F_1s were identified based on the morphological characters used in the confirmative experiments and forwarded to F_2 generation (200–250 seeds per cross) during *Kharif*, 2018. The performances of F_2 segregants were quantified using the ten quantitative characters and compared.

Statistical Analysis

The quantitative data documented during the mentioned three seasons were pooled using Pbtools software (http://bbi.irri.org/). The PCA (Jackson and Edward 1991) for the pooled data was performed using 'factoextra' and 'FactomineR' packages of R studio version 1.0.136 (Lê *et al.* 2008). The correlation and path analysis were done as

Table 1: The list of qualitative characters used for characterization of the TNAU soybean core set

S. No	Qualitative traits	S. No	Qualitative traits
1	Leaflet shape	10	Plant height
2	Petiole presence	11	Hilum colour
3	Pubescence density	12	Hypocotyl anthocyanin pigmentation
4	Days to 50% flowering	13	Pod pubescence colour
5	Flower colour	14	Seed cotyledon colour
6	Growth type	15	Seed size
7	Pod colour	16	Seed lusture
8	Pod pubescence	17	Plant growth habit
9	Days to maturity	18	Seed coat colour
	Biochemical based tests		
19	Seed coat peroxidase activity		
20	Sodium hydroxide test		
21	Potassium hydroxide test		

Fig. 1: Geographical origin of the TNAU soybean core set

suggested by Pearson (1901) and Dewey and Lu (1959), respectively. The visualization of correlation and path analysis were obtained from 'corrplot' and 'semPlot' packages of R studio version 1.0.136 (Wei *et al.* 2017; Epskamp *et al.* 2019). The genetic associations between genotypes based on qualitative data were estimated by Jaccard's similarity coefficient. The similarity matrix was used to group the genotypes by Sequential Agglomerative Hierarchical Non-overlapping (SHAN) clustering technique utilizing the Unweighted Pair Group Method with Arithmetic Averages (UPGMA) method. The statistical analysis for qualitative data was carried out using NTSYSpc software (version 2.2).

Results

Both quantitative and qualitative traits were utilized in the current investigation to understand the breeding value of the soybean core set and the results were validated through a hybridization experiment.

Quantitative Trait-based Experiments

Principal Component Analysis (PCA)

The PCA was performed to know the genetic relatedness of the genotypes, interdependence of various traits and importance of traits with respect to total variation. The quantitative trait based PCA divided the total variation into ten Principal Components (PCs). The first three PCs contributed 47.63, 15.66 and 10.43 percent to the total variation, respectively and therefore considered as major principal components (Table 2). The genotypes were scattered along the biplot based on the first two PCs (Fig. 2). The genotypes *viz.*, JS9305, JS(SH)99-02, IC16009 and TNAU20048 were located far apart, while the genotypes NRC77, UGM75, EC30198 and MAUS61 were placed closer to the origin. The interrelationship and contribution of quantitative characters to the total variation are represented in Fig. 3. The characters single plant yield, number of pods per plant and plant height were away from the origin that

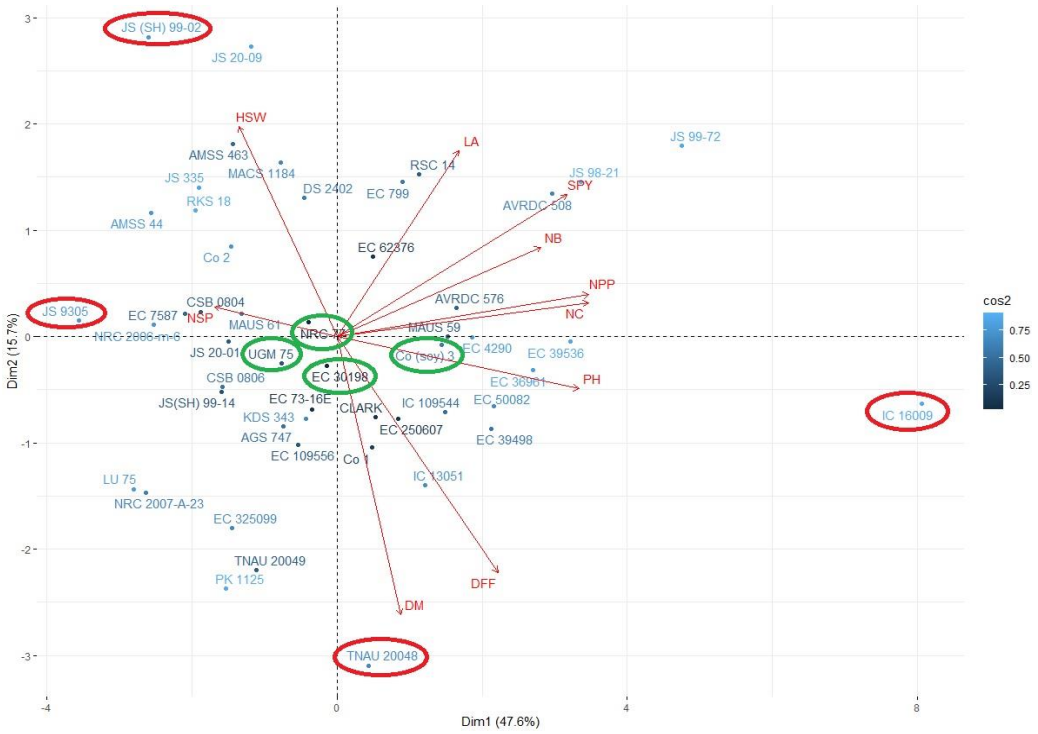

Fig. 2: Genetic divergence of TNAU soybean core set in biplot with cos^2 loadings
*circled in red color are diverse parents and in green are closely related parents used for hybridization

Table 2: Contribution of ten principal components to the total divergence with Eigen values

Principal components (PC)	Eigen value	Percentage of variance	Cumulative percentage of variance
PC1	4.76	47.63	47.63
PC2	1.57	15.66	63.29
PC3	1.04	10.43	73.72
PC4	0.83	8.34	82.06
PC5	0.74	7.44	89.50
PC6	0.42	4.23	93.73
PC7	0.29	2.95	96.68
PC8	0.20	1.97	98.65
PC9	0.10	1.02	99.67
PC10	0.03	0.33	100.00

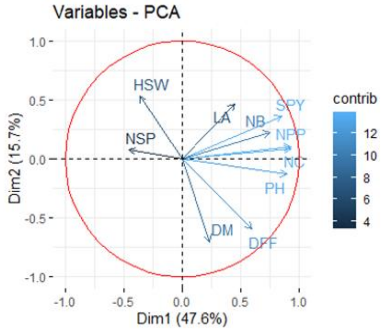

Fig. 3: Variables plot with contribution of quantitative characters to the total divergence

contributed the maximum to the divergence. On the contrary, leaf area and number of seeds per plant were closer to the origin and contributed the minimum. Further, the contributions of various quantitative characters to various PC's were analyzed to understand their importance and are depicted in a rotated component matrix (Fig. 4). Single plant yield (0.92) and number of pods per plant (0.92) contributed the maximum for the rotated PC1. Similarly, days to maturity (0.97) and number of seeds per pod (0.96) contributed the maximum for the rotated PC2 and PC3, respectively.

Genetic Association Studies

Even though, PCA hints the relationship between the quantitative characters, the magnitude of the relationship can be arrived from correlation and path analysis. The correlation and path analyses were performed to decipher the influence of various quantitative traits on single plant yield and the results are depicted in Fig. 5 and 6. The association analysis revealed that the number of pods per plant (r = 0.88) contributed the maximum for single plant yield. It was followed by number of clusters per plant (r = 0.81), plant height (r = 0.70) and number of primary branches (r = 0.64). The number of seeds per pod (r = -0.18) showed a negative association with single plant yield. The path analysis also revealed that the number of pods per plant (0.94) contributed the highest positive direct effect to single plant yield followed by the hundred seed weight (0.32). The

Table 3: Qualitative trait-based clustering of TNAU soybean core set

Clusters	No. of Genotypes	Genotypes
I	19	KDS343, JS20-01, JS99-72, JS(SH)99-02, JS(SH)99-14, NRC2007-A-23, NRC77, MACS1184, AMSS44, AMSS463, LU75, CLARK, CSB0804, CSB0806, Co2, RKS18, MAUS61, JS335 & EC250607
II	9	NRC2006-m-6, DS2402, TNAU20049, EC325099, UGM75, AGS747, EC30198, EC73-16E & JS9305
III	17	RSC14, IC16009, JS98-21, PK1125, IC13051, EC36961, EC39498, EC62376, AVRDC508, AVRDC576, EC50082, EC799, EC39536, EC4290, IC109544, Co1 & Co (Soy)3
IV	1	MAUS59
V	2	JS20-09 & EC7587
VI	1	TNAU20048
VII	1	EC109556

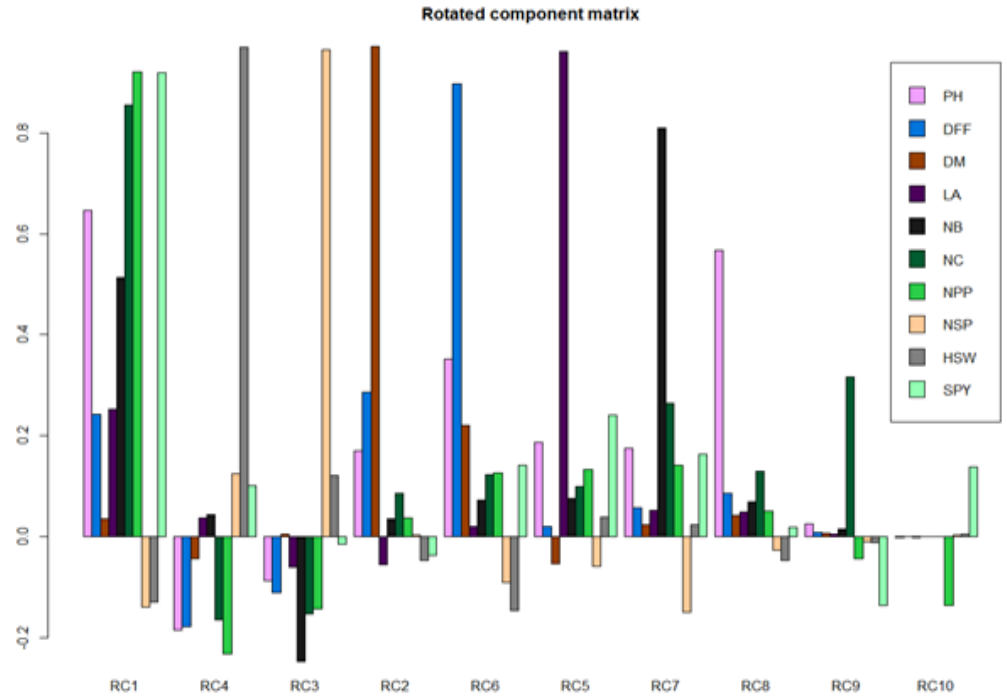

Fig. 4: Rotated component matrix for the ten principal components

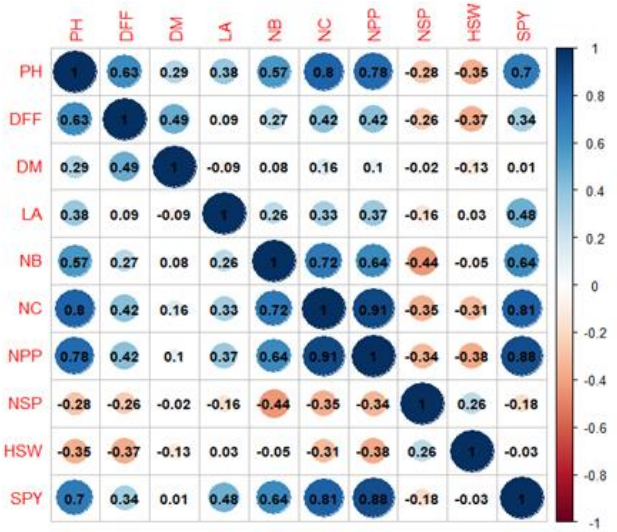

Fig. 5: Correlation between ten quantitative characters in the TNAU soybean core set

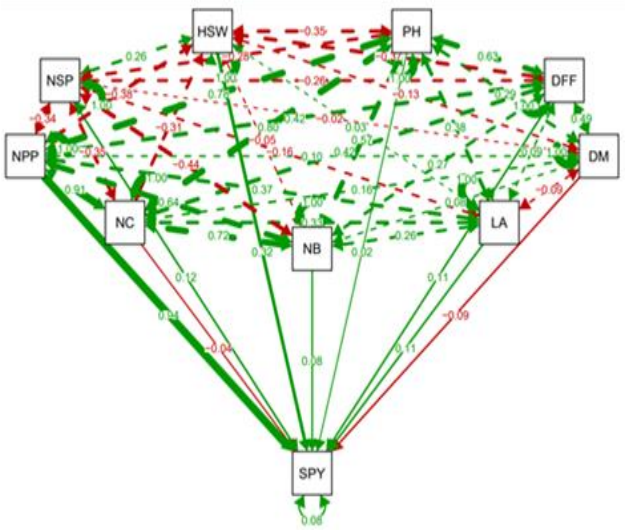

Fig. 6: Path analysis for ten quantitative characters in the TNAU soybean core set

Table 4: Similarity indices of selected parents for hybridization based on qualitative traits

Diverse parents					Closely related parents				
Parents	JS(SH)99-02	IC16009	TNAU20048	JS9305	Parents	NRC77	UGM75	EC30198	Co(Soy)3
JS(SH)99-02	1				NRC77	1			
IC16009	0.34	1			UGM75	0.80	1		
TNAU20048	0.13	0.25	1		EC30198	0.75	0.87	1	
JS9305	0.31	0.40	0.29	1	Co(Soy)3	0.7	0.76	0.89	1

Table 5: Performance of F_2 segregants involving distantly related soybean accessions

Characters	F_2 population															Parents			
	F_2(IC16009 × JS(SH)99-02)					F_2(IC16009 × TNAU 20048)					F_2(IC16009 × JS 9305)					JS(SH) 99-02	IC 16009	TNAU 20048	JS 9305
	Mean	Mini.	Maxi.	CD	CV	Mean	Mini.	Maxi.	CD	CV	Mean	Mini.	Maxi.	CD	CV	Mean	Mean	Mean	Mean
PH	65.8	32.6	85.9	8.9	7.4	75.5	54.6	94.3	11.5	9.1	80.2	61.3	91.6	12.3	7.9	15.9	87.9	46.0	26.2
DFF	39.2	28.0	47.0	4.2	7.9	43.7	39.0	50.0	8.7	6.9	38.5	35.0	42.0	3.4	6.3	31.0	41.7	45.7	35.0
DM	87.4	50.0	95.0	10.5	8.0	94.0	86.0	105.0	8.9	7.8	88.6	82.0	94.0	5.1	4.2	77.7	91.7	95.0	85.0
NPB	6.8	4.0	8.0	1.1	7.5	7.8	6.0	10.0	1.3	6.9	6.4	6.0	8.0	0.6	5.3	4.0	9.3	6.3	3.3
NCP	90.9	59.0	128.0	15.6	15.1	91.3	66.0	91.0	9.6	12.4	52.3	46.0	89.0	8.9	8.2	25.7	85.3	16.0	13.0
NPP	267.2	123.0	298.0	23.1	18.2	288	156.0	291.0	16.3	16.2	198.6	160.0	267.0	16.5	12.6	88.3	245.7	84.3	40.3
NSP	2.6	2.0	3.0	0.3	2.1	2.6	2.0	3.0	0.2	3.0	3.0	2.6	3.0	0.1	3.0	3.0	2.3	2.3	3.0
HSW	9.2	7.6	9.9	0.8	2.8	8.9	6.9	10.2	0.7	3.1	11.2	9.9	12.4	0.7	4.1	12.1	7.1	9.7	12.3
SPY	45.78	38.7	50.35	6.4	4.6	44.6	29.4	48.6	8.4	7.2	43.8	22.1	46.9	9.1	7.9	22.9	39.2	17.4	15.6

*CD @ 5%

Table 6: Performance of F_2 segregants involving closely related soybean core accessions

Characters	F_2 population															Parents			
	F_2(Co (Soy)3 × NRC 77)					F_2(Co (Soy)3 × UGM 75)					F_2(Co (Soy)3 × EC 30198)					NRC 77	UGM 75	EC 30198	Co (Soy)3
	Mean	Mini.	Maxi.	CD	CV	Mean	Mini.	Maxi.	CD	CV	Mean	Mini.	Maxi.	CD	CV	Mean	Mean	Mean	Mean
PH	40.6	27.3	66.8	8.6	9.6	40.6	32.7	59.6	6.7	10.8	56.2	50.6	62.1	2.8	7.6	22.3	22.4	53.6	55.6
DFF	37.2	32.0	44.0	3.2	7.2	42.6	40.0	44.0	1.8	5.0	40.5	38.0	44.0	1.9	5.8	35.3	41.3	39.0	40.0
DM	86.4	82.0	92.0	3.5	7.6	88.6	84.0	92.0	3.1	7.6	88.0	83.0	92.0	2.0	8.6	84.3	86.0	84.3	90.0
NPB	6.4	5.0	7.0	0.3	5.6	6.1	5.0	7.0	0.4	5.1	6.0	5.0	7.0	0.6	7.6	6.0	5.3	5.0	6.3
NCP	36.8	32.0	42.3	2.6	10.2	41.3	36.0	48.0	3.9	9.4	43.0	30.0	45.0	3.1	9.8	35.3	26.0	35.3	40.0
NPP	120.5	96.0	140.2	8.6	12.3	118.9	102.0	123.0	4.9	8.6	120.4	70.0	130.0	9.9	10.2	112.3	75.0	72.0	111.7
NSP	2.7	2.3	3.0	0.2	4.3	2.7	2.7	3.0	0.2	3.5	2.7	2.7	3.0	0.2	4.9	2.7	2.7	2.7	2.7
HSW	9.9	8.5	10.6	0.9	5.6	10.2	10.0	10.6	0.1	2.6	10.2	9.9	10.6	0.3	8.6	8.4	10.2	9.6	10.4
SPY	25.1	20.1	28.3	2.5	10.2	24.2	18.9	26.4	1.6	6.6	24.6	18.5	27.5	2.7	5.9	22.3	18.3	15.8	22.9

*CD @ 5%

PH–Plant Height (cm): DM- Days to Maturity: DFF- Days to Fifty percent Flowering: NPB- Number of Primary Branches: NCP- Number of Clusters per Plant: NPP- Number of Pods per Plant: NSP- Number of Seeds per Pod: HSW- Hundred Seed Weight: SPY- Single Plant Yield (g)

negative direct effect on single plant yield was exerted by days to maturity (-0.09), and number of clusters per plant (-0.04). The highest positive indirect effect to single plant yield was observed with number of pods per plant and number of clusters per plant (0.91) followed by number of clusters per plant and plant height (0.80). The residual effect was only eight percent.

Qualitative Trait based Experiments

The variations for qualitative traits categorized the genotypes into eight clusters (Table 3), the cluster I was the biggest (19 genotypes) followed by cluster III (17) and cluster II (9). The genotypes MAUS59 and TNAU20048 formed a solitary cluster individually. Further, it also revealed a significant variation for growth habit, leaf type, hilum colour and pod colour (Fig. 7). A relationship between hypocotyl and flower colour was established. The genotypes with purple hypocotyl produced purple flowers, while the green hypocotyl genotypes produced white flowers. A total of 38 genotypes were earmarked as determinate type, 10 and two genotypes were categorized as semi-determinate and indeterminate respectively. The biochemical based KOH and NaOH tests also grouped the genotype EC109556 as a solitary cluster. The seed coat peroxidase test equally distributes the genotypes into two groups based on the presence or absence of enzyme activity.

The results of quantitative (PCA biplot) (Fig. 2) and qualitative trait (similarity index) (Table 4) based experiments along with single plant yield (Table 5 and 6) were utilized to tag potential genotypes for targeted hybridizations. Accordingly, two groups of parents were formed. The genotypes located far apart (encircled as red) with less similarity index and average yield potential (single plant yield above the national check JS335 *i.e.,* > 15.2 g) were categorized as group I (IC16009, JS(SH)99-02, TNAU20048 and JS9305). While group II comprised of closely related parents (encircled as green) with higher similarity index and good yield (Co(soy)3, NRC77, UGM75 and EC30198). Based on single plant the genotype IC

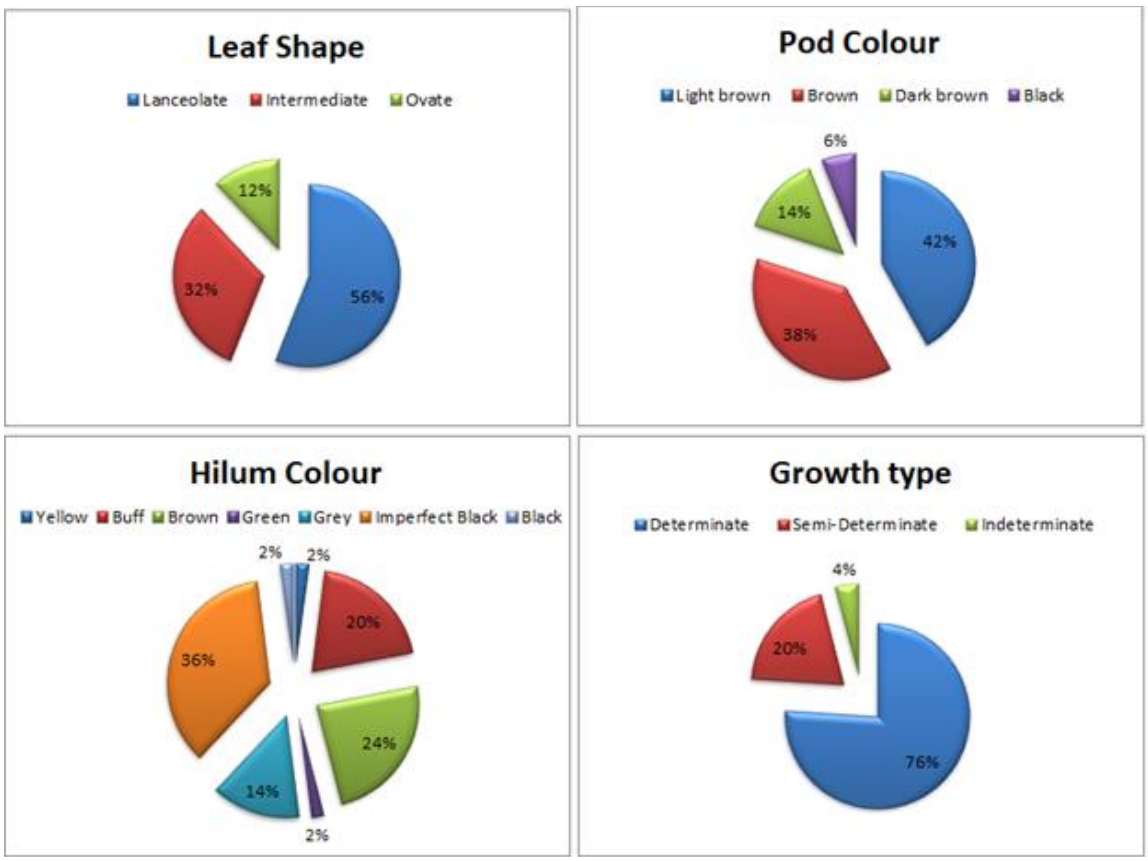

Fig. 7: Variations observed for qualitative traits in the TNAU soybean core set

16009 (39.2 g) and Co (soy) 3 (22.9 g) were designated as female parents in group I and II respectively while the other three parents were treated as pollen parents. In each group three crosses were made and the progenies were evaluated separately. The true F_1s were forwarded to F_2 and the agronomic performances were evaluated along with respective parents (Table 5 and 6).

Validation of Breeding Value of Soybean Core Set

The above listed distantly and closely related genotypes utilized for hybridization to validate the results of quantitative and qualitative trait-based experiments. The genotypes were hybridized, the F1s were grown in an ideal condition, the true F1s were forwarded to F_2 generation, and the yield performances were analyzed. The F_2 segregants of closely related accessions showed the average single plant yield of 25.1 g [F_2 (Co(soy)3 × NRC77)], 24.2 g [F_2 (Co(soy)3 × UGM75)] and 24.6 g [F_2 (Co(soy)3 × EC30198)] compared to the better parent Co(soy)3 (22.9 g). Similarly, the diverse accessions possessed the average single plant yield of 45.78 g [F_2 (IC16009 × JS(SH)99-02)], 44.6 g [F_2 (IC16009 × TNAU20048)] and 43.2 g [F_2 (IC16009 × JS9305)] compared to the better parent IC16009 with 39.2 g (Table 5 and 6).

Discussion

Soybean is one of the major legume crops that grown in different seasons and environments in India for multi-purposes. The potential yield is not realized due to the incidence of various stresses and geographical origin (Malik *et al*. 2011). Therefore, identification of adaptable, region and trait-specific genotypes and their subsequent utilization in the pre-breeding activities are necessitated. A core set reduces the complexity in the *modus operandi* of germplasm. The PCA helps to reduce the complexity of multidimensional data into fewer principal components (PC's) without losing major information (Jolliffe and Cadima 2016). In the present study, the first three PC's contributed about 73.72% of total variation and were considered as significant as their Eigenvalues were more than one. The Eigenvalues of more than one contributes the maximum to the total variation while less than one contributes less (Gerrano *et al*. 2019). The genotypes scattered along the biplot based on the first two PC's represented the variability among the soybean core set. The genotypes JS9305, JS(SH)99-02, IC16009 and TNAU20048 were located far apart and considered as highly divergent lines. The characters *viz.*, single plant yield, number of pods per plant, and plant height were away from an origin that

contributed the maximum for the divergence. The contributions of various quantitative characters to various PC's revealed that the single plant yield and number of pods per plant, days to maturity and number of seeds per pod contributed the maximum for the rotated PC1, PC2 and PC3, respectively. Earlier, a similar finding was reported by (Lazaridi *et al.* 2017). The ultimate aim of any breeding program is to improve the yield which is a complex trait and influenced by contributing traits. An association study on these traits helps in ideotype breeding (Jain *et al.* 2018). The results of correlation and path analyses revealed that number of pods per plant and number of clusters per plant contributed significantly to the single plant yield that was parallel with the findings of Jain *et al.* (2018).

Qualitative traits categorized the genotypes into eight clusters. The genotypes within a cluster are considered as less divergent while the genotypes grouped in different clusters are more divergent. The qualitative trait-based clustering also confirms the grouping pattern of genotypes as per quantitative traits (Ramteke and Murlidharan 2012). The determinate growth type accessions found in the study might be preferred under rainfed conditions as it conserves moisture (Malik *et al.* 2011). The accessions that possessed a lanceolate leaflet shape might be linked to drought tolerance. Similar findings were reported by Malik *et al.* (2011). Variations in hilum colour and pod colour are due to light intensity, temperature, drought, disease injury, and other environmental factors (Yadav and Sharma 2001). The KOH, NaOH and seed coat peroxidase tests helped in effective grouping in the present study. Earlier, the utility of biochemical tests in clustering was reported (Agrawal and Sharma 1989).

To validate the above results, four diverse and closely related parents were selected based on quantitative and qualitative clustering patterns and hybridized. The F_2 generation of closely related and diverse accessions showed an average yield increase of 7.5 and 14.1 percent, respectively than their better parent. These results indicated that the TNAU soybean core set is more diverse and can be effectively used for pre-breeding activities.

Conclusion

The multivariate analysis for various quantitative and qualitative traits revealed the existence of significant variation in the TNAU soybean core set. The crosses between identified diverse accessions have a net worthy yield advantage over the closely related genotypes. It is concluded that the variation present in the TNAU soybean core set has breeding potency and hence could be used in the pre-breeding programs.

Acknowledgements

The authors are grateful to Dr. D. Packiaraj, Dr. J.R. Kanna Bapu, Dr. K. Ganesamurthy, Dr. V. Thiruvengadam, Dr. S. Ganesh Ram, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University for their scientific support.

Author Contributions

RS and CV planned the experiments. EV and VA performed the experiments, statistically analyzed the data and made illustrations. EV and RS wrote the draft manuscript. JR, CV and NS interpreted the results and made the final Manuscript.

Conflicts of Interest

All authors declare no conflicts of interest.

Data Availability

The datasets generated during the study are all included in the manuscript. Further inquiries can be directed to the corresponding author

Ethics Approvals

Ethics approval was not required for this study

References

Agrawal RL, BL Sharma (1989). Identification of mungbean varieties on the basis of seedling growth response to GA_3 and DDT treatment. *Seed Res* 17:84–37

Bisen A, D Khare, P Nair, N Tripathi (2015). SSR analysis of 38 genotypes of soybean (*Glycine max* (L.) Merr.) genetic diversity in India. *Physiol Mol Biol Plants* 21:109–115

Dewey DR, K Lu (1959). A correlation and path-coefficient analysis of components of crested wheatgrass seed production 1. *Agron J* 51:515–518

El-Harty EH, SS Alghamdi, MA Khan, HM Migdadi, M Farooq (2018). Adaptability and stability analysis of different soybean genotypes using biplot model. *Intl J Agric Biol* 20:2196–2202

Epskamp S, MS Epskam, S Mplus Automation (2019). Package 'semPlot'. R Package Version

Gerrano AS, WSR Jansen, FR Kutu (2019). Agronomic evaluation and identification of potential cowpea (*Vigna unguiculata* L. Walp) genotypes in South Africa. *Acta Agric Sci B Soil Plant Sci* 69:295–303

Jackson JE, A Edward (1991). *User's guide to principal components*. John Willey Sons. Inc., New York, USA

Jain RK, A Joshi, HR Chaudhary, A Dashora, CL Khatik (2018). Study on genetic variability, heritability and genetic advance in soybean [*Glycine max* (L.) Merrill]. *Leg Res* 41:532–536

Jolliffe IT, J Cadima (2016). Principal Component Analysis: A review and recent developments. *Phil Trans Roy Soc A* 374:1–16

Lazaridi E, G Ntatsi, D Savvas, PJ Bebeli (2017). Diversity in cowpea (*Vigna unguiculata* (L.) Walp.) local populations from Greece. *Genet Resour Crop Evol* 64:1529–1551

Lê S, J Josse, F Husson (2008). FactoMineR: An R package for multivariate analysis. *J Stat Softw* 25:1–18

Malik FA, M Ashraf, AS Qureshi, MR Khan (2011). Investigation and comparison of some morphological traits of the soybean populations using cluster analysis. *Pak J Bot* 43:1249–1255

Mohan S, A Sheeba, T Kalaimagal (2019). Genetic diversity and association studies in Greengram [*Vigna radiata* (L.) Wilczek]. *Leg Res* 44:779–784

Muthamizhan G, R Sudhagar, V Thiruvengadam, S Ganeshram, JR KannanBapu (2016). Formulation of core set in base collections of soybean (*Glycine max* (L.) Merrill) germplasm accessions. *Adv Life Sci* 5:7909–7917

Pearson K (1901). I. Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Phil Trans Roy Soc A* 195:262–273

Ramteke R, P Murlidharan (2012). Characterization of soybean (*Glycine max*) varieties as per DUS guidelines. *Ind J Agric Sci* 82:572–577

Rao DE, KV Chaitanya (2020). Changes in the antioxidant intensities of seven different soybean (*Glycine max* (L.) Merr.) cultivars during drought. *J Food Biochem* 44:e13118

Sharma S, HD Upadhyaya, RK Varshney, CLL Gowda (2013). Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. *Front Plant Sci* 4:309-322

TNAU – Tamil Nadu Agricultural University (2016). *Crop Production Guide (CPG)*, pp:173–178. Tamil Nadu Agricultural University and Department of Agriculture, Government of Tamil Nadu, India

Upadhyaya HD, CLL Gowda, KN Reddy, S Singh (2009). Augmenting the pearl millet core collection for enhancing germplasm utilization in crop improvement. *Crop Sci* 49:573–580

Vishwanath, R Hunje, P Shinde (2021). Influence of Seed Quality Upgrading Processing Machines in Soybean var. DSb-21. *Leg Res* 1:1–5

Wei T, V Simko, M Levy, Y Xie, Y Jin, J Zemla (2017). Package 'corrplot'. *Statistician* 56:1–18

Yadav S (2016). Soybean (*Glycine max* L.) germplasm evaluation through morphological and quality characterization. *Intl J Appl Pure Sci Agric* 12:141–145

Yadav SP, SP Sharma (2001). Variation for hilum colour and its stability during four crop seasons in soybean (*Glycine max*). *Ind J Agric Sci* 110:45–51