RESEARCH ARTICLE Legume Research- An International Journal, Volume 48 Issue 6: 932-937 (June 2025)

Genetic Diversity Studies of Soybean [*Glycine max* (L.) Merrill] Germplasm Accessions using Cluster and Principal Component Analysis

R.C. Sivabharathi[1], A. Muthuswamy[2], K. Anandhi[1], L. Karthiba[1]

10.18805/LR-5071

ABSTRACT

Background: Soybean [*Glycine max* (L.) Merrill] is a self-pollinated diploid leguminous crop (2 n=40) originated from China. It is called as 'Cow of the field' or 'Gold from soil' because of high oil and protein content. The objective of the study is to determine genetic diversity using cluster and principal component analysis among 135 soybean germplasm.

Methods: The experiment was conducted at the Department of Pulses, Tamil Nadu Agricultural University, Coimbatore during *rabi,* 2021-22. A total of 135 germplasm was laid down in augmented block design II. Principal component analysis and cluster analysis were performed to determine the genetic diversity among 135 soybean germplasm accessions using GRAPES and Ggt 2.0 software respectively.

Result: Based on the cluster and principal components analysis, wide diversity were observed in MAUS 60, JS 98-61, MACS 1460, EC 18736 and PK 1038 genotypes and traits *viz.,* number of clusters per plant, number of pods per plant and single plant yield contributed to divergence. The above mentioned genotypes and traits could be used for selection and future breeding programme.

Key words: Cluster, Eigen value, Principal components, Soybean, Variation.

INTRODUCTION

Soybean is a miracle golden bean of the 21st century because of extraordinary high protein content (38-43%) combined with high amount of oil (17-19%). Soybean was originated from north-east Asia, particularly China (Hymowitz and Newell, 1981). Soybean [*Glycine max* (L.) Merrill] is well-adapted to different agro-climatic conditions of tropical, sub-tropical and temperate zones. Soybean ranks first among oilseeds in the world and has now found a prominent place in India. World soybean production was estimated as 385.527 million tonnes. Brazil ranks first in soybean production with 144 million tonnes followed by the United States, Argentina, China and India (Anonymous, 2021). Production in India accounts for 12.04 million tonnes from 11.45 million hectares with average productivity of 1051 kg/ha. Madhya Pradesh is the soybean bowl of India, contributing more than 89 per cent of the country's soybean production, followed by Maharashtra and Rajasthan (Anonymous, 2022).

The genetic diversity analysis helps in selecting appropriate parents for combining new alleles for the trait in crop improvement programmes (Shadakshari *et al.*, 2011). The principal component and cluster analysis have been successfully used to classify and measure the pattern of genetic diversity in soybean germplasm (Kayan and Adak, 2012). The PCA was carried out to determine the genetic relatedness of genotypes, the interdependence of different traits and the significance of traits in relation to total variance. The cluster analysis helps in grouping the genotypes based on the variation occurring among them. The genotypes selected from these two techniques could be used for selection.

[1]Department of Pulses, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore-641 003, Tamil Nadu, India.

[2]Agricultural College and Research Institute, Karur-639 001, Tamil Nadu, India.

Corresponding Author: K. Anandhi, Department of Pulses, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore-641 003, Tamil Nadu, India.
Email: anandhiagri@gmail.com

How to cite this article: Sivabharathi, R.C., Muthuswamy, A., Anandhi, K. and Karthiba, L. (2025). Genetic Diversity Studies of Soybean [*Glycine max* (L.) Merrill] Germplasm Accessions using Cluster and Principal Component Analysis. Legume Research. 48(6): 932-937. doi: 10.18805/LR-5071.

Submitted: 10-11-2022 **Accepted:** 15-02-2023 **Online:** 15-03-2023

MATERIALS AND METHODS

During the *rabi* season of 2021-22, the experiment was conducted at the Department of Pulses, Tamil Nadu Agricultural University, Coimbatore. The experimental material consisted of 135 soybean genotypes including five check varieties *viz.,* NRC 132, NRC 142, NRC 147, MACS 1460 and CO (Soy) 3 is given in (Table 1). The experiment was set up in an augmented block design II with row to row and plant to plant distance of 30 cm × 10 cm respectively. To determine the genetic divergence across the genotypes, Ggt 2.0 software was used to construct the tree diagram and the clusters were formed using Euclidian distances and UPGMA tree clustering method. The components were

extracted using the principal component analysis approach by using the GRAPES software (Kujane *et al*., 2019).

RESULTS AND DISCUSSION

Cluster analysis

The analysis of variance showed highly significant differences among the 135 soybean genotypes for all the ten quantitative traits studied. The 135 soybean genotypes were grouped into twelve clusters by using UPGMA tree clustering in Ggt 2.0 software as given in (Table 2). The cluster IV was the largest cluster with 30 genotypes followed by clusters VIII, II, V and III with 24, 22, 13 and 10 genotypes respectively. Cluster VI and VII were the solitary clusters with genotype *viz.,* MAUS 60 and JS 98-61 respectively. The genotypes within the same cluster showed less variation whereas between the clusters exhibited maximum variation. Therefore, the genotypes from different clusters could be selected for crop improvement.

Principal component analysis

Principal component analysis was performed using the mean data of ten quantitative traits using GRAPES software. The quantitative traits of 135 germplasm accessions were categorized into ten different principal components based on the total variation. The eigen value of first four principal components among ten PCs were more than one as given in Table 3 and these four PCs contribute the cumulative percentage of variation of 79.77 per cent. The contribution of each trait to total variation is presented in Table 4. Reddy *et al*. (2024) reported 68.61% of variance was contributed by first three principal components out of ten principal components formed with 24 french bean germplasms. Kumar *et al*. (2010) observed 95% of total variation was contributed by first ten PCs out of the fourteen PCs formed with 64 groundnut breeding lines.

The first principal component showing 42.17% variation was associated mainly with number of clusters per plant, number of pods per plant and number of branches per plant. The outcomes of Ghiday and Sentayehu (2015) and Dubey *et al*. (2018) were similar with the present study for number of branches per plant and number of pods per plant. Jain *et al*. (2023) reported 28.6% of total variation was contributed PC1 and is associated with number of pods per plant, days to flowering and plant height in 40 chickpea genotypes. The second principal component contributing 15.72% variation was mainly related to hundred seed weight and single plant yield. Similar outcomes were reported by Singh and Shrestha (2019) and Dubey *et al*. (2018) for hundred seed weight and single plant yield respectively. The third principal component conferring 11.24% variation was mainly connected with number of pods per plant and number of pods per cluster. Similar findings were observed by Ghiday and Sentayehu (2015) for number of pods per plant. The fourth principal component exhibiting 10.64% of variation and it was mainly coupled with hundred seed weight and Dubey *et al*. (2018) and Singh and Shrestha (2019) had similar findings on hundred seed weight.

The scree plot given in (Fig 1) clearly depicts that PC1 had highest variation, followed by PC2, PC3 and PC4. Based on PC1 and PC2, the genotypes were scattered along the biplot as shown in (Fig 2). The cos2 loading value ranges from 0.00 to 1.00 and these values were used to indicate the divergence among genotypes. MACS 1460, EC 18736

Table 1: List of soybean genotypes used in the study.

Sl. no.	Genotypes	Sl. no	Genotypes	Sl. no	Genotypes
1.	CLARK	46.	MACS 1148	91.	NRC 43
2.	CO 1	47.	MACS 1188	92.	VLS 53
3.	CO 2	48.	MACS 1238	93.	VLS 69
4.	CSB 0804	49.	MACS 1254	94.	PK 1158
5.	CSB 0806	50.	MACS 1259	95.	PK 1223
6.	CSB 0808	51.	MACS 1281	96.	PK 1243
7.	CSB 0809	52.	MACS 145	97.	MAUS 59
8.	CSB 0810	53.	MACS 565	98.	MAUS 60
9.	CSB 0811	54.	MACS 610	99.	MAUS 61
10.	EC 18678	55.	MACS 629	100.	NRC 44
11.	EC 18736	56.	MACS 693	101.	NRC 45
12.	JS 20-01	57.	MACS 694	102.	NRC 46
13.	JS 20-09	58.	MACS 715	103.	NRC 76
14.	JS 76119	59.	MACS 798	104.	NRC 78
15.	JS 76-1194	60.	MACS 94-2	105.	NRC 79
16.	JS 87-12	61.	MACS 985	106.	NRC 80-1
17.	JS 89-24	62.	MAUS 109	107.	NRC 82
18.	JS 90-21	63.	MAUS 144	108.	NRC 84
19.	JS 90-29	64.	MAUS 17	109.	NRC 95-06-03
20.	JS 92-22	65.	MAUS 2	110.	VLS 70
21.	JS 95-60	66.	MAUS 20	111.	VLS 75
22.	JS 95-98	67.	MAUS 311	112.	WC 37
23.	JS 97-52	68.	MAUS 34	113.	WC 67
24.	JS 98-21	69.	MAUS 39	114.	PK 1000
25.	JS 98-61	70.	MAUS 414	115.	PK 1303
26.	JS 98-63	71.	MAUS 417	116.	PK 25
27.	JS 98-68	72.	MAUS 52-1	117.	PK 257
28.	JS 99-12	73.	MAUS 55	118.	PK 258
29.	JS 99-128	74.	JS(SH) 2001-04	119.	PK 727
30.	JS 99-72	75.	JS(SH) 2002-14	120.	PK 768
31.	JS 99-76	76.	JS(SH) 8554	121.	PK 1011
32.	JS 99-77	77.	JS(SH) 89-2	122.	PK 1014
33.	JS 99-83	78.	MAUS 65	123.	PK 1024
34.	JS(SH)18608	79.	MAUS 68	124.	PK 1028
35.	JS(SH)89-49	80.	MAUS 71-07	125.	PK 1038
36.	JS(SH)90-91	81.	MAUS 81	126.	PK 1125
37.	JS(SH)91-93	82.	NRC 2006-M-6	127.	PK 1146
38.	JS(SH)92-46	83.	NRC 2007-G-1-13	128.	PK 1225
39.	JS(SH)93-37	84.	NRC 2007-I-3	129.	PK 701
40.	JS(SH)93-44	85.	NRC 2007-K-7-2	130.	PK 7247
41.	JS(SH)99-14	86.	NRC 21	131.	NRC 132
42.	MACS 1039	87.	NRC 25	132.	NRC 142
43.	MACS 1126	88.	NRC 29	133.	NRC 147
44.	MACS 1139	89.	NRC 34	134.	MACS 1460
45.	MACS 1140	90.	NRC 42	135.	CO (Soy) 3

Table 2: Clustering of 135 soybean germplasm accessions based on ten quantitative traits.

Cluster	Number of accessions	Genotype number	Genotypes name
I	15	1, 16, 99, 43, 117, 47, 50, 10, 132, 133, 71, 120, 83, 87, 131	CLARK, JS 87-12, MAUS 61, MACS 1126, PK 257, MACS 1188, MACS 1259, EC 18678, NRC 142, NRC 147, MAUS 417, PK 768, NRC 2007-G-1-13, NRC 27, NRC 132
II	224	24, 65, 69, 66, 17, 40, 68, 101, 6, 61, 33, 60, 67, 32, 29, 31, 54, 62, 63, 7, 30	CSB 0804, JS 98-21, MAUS 2, MAUS 39, MAUS 20, JS 89- 24, JS (SH) 93-44, MAUS 34, NRC 45, CSB 0808, MACS 985, JS 99-83, MACS 94-2, MAUS 311, JS 99-77, JS 99-128, JS 99-76, MACS 610, MACS 109, MAUS 144, CSB 0809, JS 99-72
III	10	23, 76, 81, 109, 51, 42, 107, 108, 111, 45	JS 97-52, JS (SH) 8554, MAUS 81, NRC 95-06-03, MACS 1281, MACS 1039, NRC 82, NRC 84, VLS 75, MACS 1140
IV	30	2, 52, 37, 34, 53, 56, 28, 36, 73, 77, 103, 96, 91, 35, 64, 14, 130, 102, 121, 126, 59, 116, 129, 100, 135, 113, 112, 55, 128, 58	CO 1, MACS 145, JS (SH) 91-93, JS (SH) 18608, MACS 565, MACS 693, JS 99-12, JS (SH) 90-91, MAUS 55, JS (SH) 89-2, NRC 76, PK 1243, NRC 43, JS (SH) 89-49, MAUS 17, JS 76119, PK 7247, NRC 46, PK 1011, PK 1125, MACS 798, PK 25, PK 701, NRC 44, CO (Soy) 3, WC 67, WC 37, MACS 629, PK 1225, MACS 715
V	13	123, 78, 118, 90, 94, 18, 70, 38, 84, 15, 80, 19, 89	PK 1024, MAUS 65, PK 258, NRC 42, PK 1158, JS 90-21, MAUS 414, JS (SH) 92-46, NRC 2007-I-3, JS 76-1194, MAUS 71-07, JS 90-29, NRC 34
VI	1	98	MAUS 60
VII	1	25	JS 98-61
VIII	24	3, 5, 104, 82, 119, 105, 115, 41, 48, 106, 12, 124, 72, 13, 79, 85, 26, 27, 44, 49, 46, 74, 75, 134	CO 2, CSB 0806, NRC 78, NRC 2006-M-6, PK 727, NRC 79, PK 1303, JS (SH) 99-14, MACS 1238, NRC 80-1, JS 20-01, PK 1028, MAUS 52-1, JS 20-09, MAUS 68, NRC 2007- K-7-2, JS 98-63, JS 98-68, MACS 1139, MACS 1254, MACS 1148, JS (SH) 2001-04, JS (SH) 2002-14, MACS 1460
IX	6	8, 9, 11, 57, 22, 39	CSB 0810, CSB 0811, EC 18736, MACS 694, JS 95-98, JS (SH) 93-37
X	4	97, 95, 125, 114	MAUS 59, PK 1223, PK 1038, PK 1000
XI	7	93, 88, 122, 92, 110, 127, 86	VLS 69, NRC 29, PK 1014, VLS 53, VLS 70, PK 1146, NRC 21
XII	2	20, 21	JS 92-22, JS 95-60

Table 3: Eigen value, percentage of variance and cumulative proportion of the principal component.

Principal components	Eigen values	Percentage of variance	Cumulative percentage of variance
PC1	4.22	42.17	42.17
PC2	1.57	15.72	57.90
PC3	1.12	11.24	69.13
PC4	1.06	10.64	79.77
PC5	0.90	8.98	88.76
PC6	0.37	3.73	92.49
PC7	0.30	2.99	95.48
PC8	0.24	2.43	97.91
PC9	0.13	1.27	99.19
PC10	0.08	0.82	100.00

and PK 1038 were the genotypes far apart, while JS 89-24, NRC 25, NRC 2007-G-1-13, NRC 43 and PK 7247 were the genotypes closer to the origin. The genotypes away from origin with high cos2 loading value exhibited maximum divergence whereas genotypes closer to origin with less cos2 loading value exhibited minimum divergence. The quantitative trait contribution to total variation and its interrelationship is shown in Fig 3. The traits *viz.,* number of clusters per plant, number of pods per plant and single plant yield were far away and contributes maximum variation while the traits *viz.,* number of seeds per pod, number of pods per cluster and hundred seed weight were closer to the origin and showed minimum contribution to total variation. Similar finding was reported by Vijayakumar *et al*. (2022).

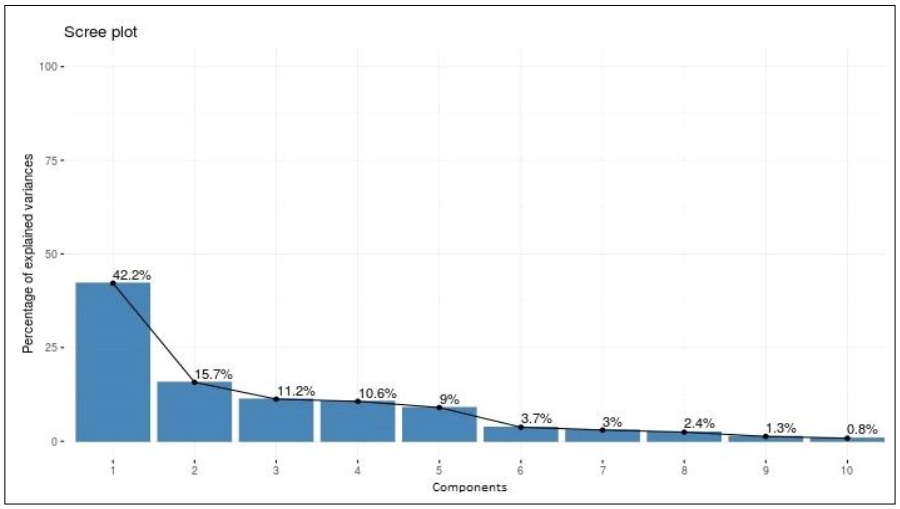

Fig 1: Scree plot of principal component analysis of soybean germplam accessions between percentage of variance and principal components.

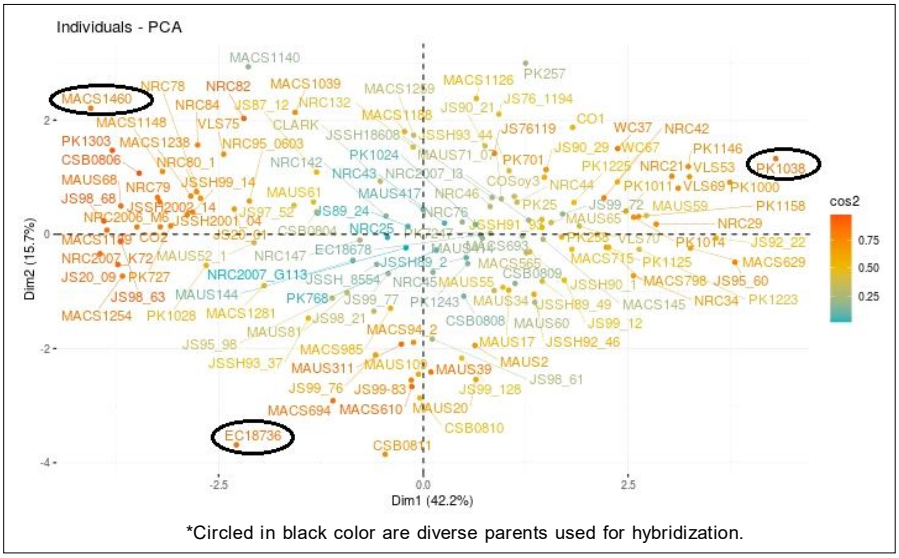

Fig 2: Genetic divergence of 135 soybean germplasm accessions in biplot with cos2 loadings.

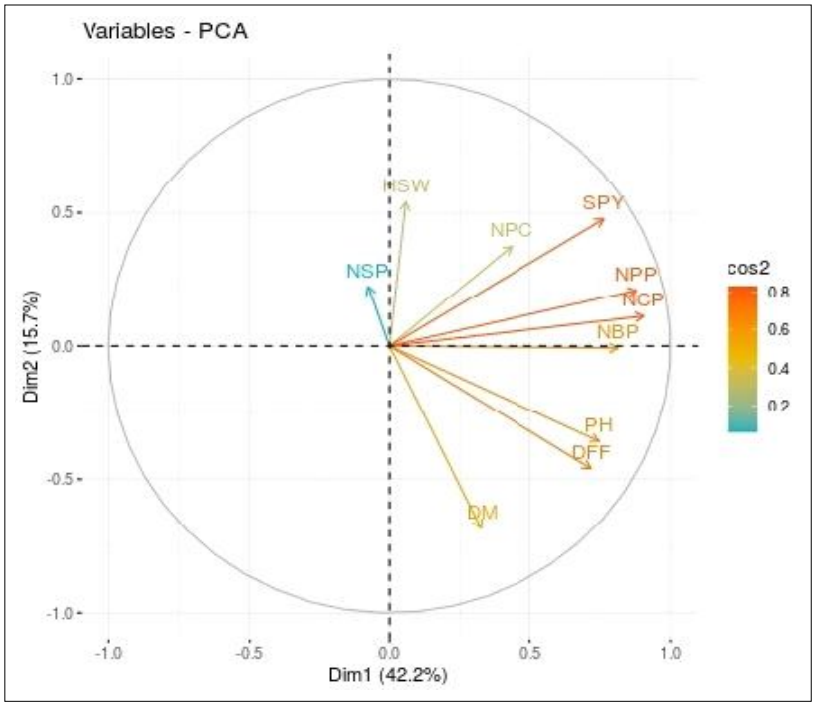

Fig 3: Variables plot with contribution of quantitative characters to the total divergence.

Table 4: Component matrix representing Eigen vectors and scores of principal components for the quantitative traits.

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
DFF	0.35	-0.37	-0.19	0.21	-0.06	0.28	-0.54	0.52	-0.15	0.02
DM	0.16	-0.55	-0.30	0.41	-0.27	0.04	0.26	-0.48	0.21	-0.07
PH	0.36	-0.29	0.09	-0.16	0.06	-0.85	0.03	0.11	-0.13	0.01
NBP	0.40	-0.01	-0.16	-0.12	0.33	0.30	0.68	0.21	-0.30	-0.04
NCP	0.44	0.09	0.08	-0.17	0.22	0.10	-0.06	-0.03	0.73	0.41
NPC	0.21	0.30	0.27	0.23	-0.74	-0.03	0.28	0.29	0.01	0.20
NPP	0.43	0.17	0.28	-0.12	-0.09	0.12	-0.16	-0.21	0.08	-0.77
NSP	-0.04	0.18	-0.74	-0.45	-0.33	-0.11	0.01	0.16	0.20	-0.18
HSW	0.03	0.43	-0.27	0.67	0.33	-0.28	0.02	0.17	0.18	-0.21
SPY	0.37	0.38	-0.24	0.02	-0.04	0.00	-0.27	-0.51	-0.46	0.34

DFF: Days to fifty per cent flowering; NPC: Number of pods per cluster; DM: Days to maturity; NPP: Number of pods per plant, PH: Plant height (cm); NSP: Number of seeds per pod; NBP: Number of branches per plant; HSW: Hundred seed weight (g); NCP: Number of clusters per plant; SPY: Single plant yield (g).

CONCLUSION

The traits number of clusters per plant, number of pods per plant and single plant yield and genotypes MACS 1460, EC 18736 and PK 1038 could be used for future crop improvement programme. In addition, the genotypes MAUS 60 and JS 98-61 were grouped in solitary clusters and therefore these genotypes could also be used for breeding work.

ACKNOWLEDGEMENT

Authors are grateful to all professors of Department of Pulses, Tamil Nadu Agricultural University, Coimbatore for providing facilities and their immense support to complete the research work successfully.

Conflict of interest

All authors declared that there is no conflict of interest.

REFERENCES

Anonymous, (2021). Oilseeds - World Markets and Trade, a USDA Publications.

Anonymous, (2022). Oilseeds - World Markets and Trade, a USDA Publications.

Dubey, N., Avinashe, H.A. and Shrivastava, A. (2018). Principal component analysis in advanced genotypes of soybean [*Glycine max* (L.) Merrill] over seasons. Plant Archives. 18(1): 501-506.

Ghiday, T. and Sentayehu, A. (2015). Genetic divergence analysis on some soybean [*Glycine max* (L.) Merrill] genotypes grown in Pawe, Ethiopia. American-Eurasian Journal of Agriculture and Environmental Science. 15(10): 1927-1933.

Hymowitz, T. and Newell, C. (1981). Taxonomy of the genus glycine, domestication and uses of soybeans. Economic Botany. 35(3): 272-288.

Jain, S.K., Sharma, L.D., Gupta, K.C., Kumar, V. and Sharma, R.S. (2023). Principal component and genetic diversity analysis for seed yield and its related compnents in the genotypes of chickpea (*Cicer arietinum* L.). Legume Research-An International Journal. 46(9): 1174-1178. doi: 10.18805/LR-4489.

Kayan, N. and Adak, M.S. (2012). Associations of some characters with grain yield in chickpea (*Cicer arietinum* L.). Pak. J. Bot. 44(1): 267-272.

Kujane, K., Sedibe, M.M. and Mofokeng, A. (2019). Genetic diversity analysis of soybean [*Glycine max* (L.) Merr.] genotypes making use of SSR markers. Australian Journal of Crop Science. 13(7): 1113-1119.

Kumar, S., Marappa, N. and Govindaraj, M. (2010). Classification of new germplasm and advanced breeding lines of groundnut (*Arachis hypogea* L.) through principal component analysis. Legume Research-An International Journal. 33(4): 242-248.

Reddy, R., Pandey, M., Singh, J., Singh, P.M. and Rai, N. (2024). Principal component analysis and stability of genotypes in French bean (*Phaseolus vulgaris* L.). Legume Research-An International Journal. 47(5): 745-750. doi: 10.18805/LR-4569.

Shadakshari, T., Kalaimagal, T., Senthil, N., Boranayaka, M., Kambe, G.R. and Rajesha, G. (2011). Genetic diversity studies in soybean [*Glycine max* (L.) Merrill] based on morphological characters. Asian Journal of Bio Science. 6(1): 7-11.

Singh, P. and Shrestha, J. (2019). Evaluation of soybean [*Glycine max* (L.) Merrill] genotypes for agro-morphological traits using multivariate analysis. Nepalese Journal of Agricultural Sciences. 18: 100-107.

Vijayakumar, E., Sudhagar, R., Vanniarajan, C., Ramalingam, J., Allan, V. and Senthil, N. (2022). Estimating the breeding potency of a soybean core set. Intl. J. Agric Biol. 27: 184-192.