

Revelation of Genetic Diversity in Advanced Breeding Lines of Soybean using Principal Component Analysis

SATISH KUMAR NAGAR, M. K. SHRIVASTAVA*, PAWAN KUMAR AMRATE, KUMAR JAI ANAND, PRATIK KUMAR AND AMIT KUMAR

Dept. of Genetics and Plant Breeding, Jawaharlal Nehru Krishi Vishwa Vidyalaya, Jabalpur, Madhya Pradesh (482004), India

*Email: shrivastava.manoj03@gmail.com

Abstract

The present investigation was carried out to examine the genetic variability and important yield-affecting characteristics of soybean (*Glycine max* (L.) Merrill) using Principal Component Analysis (PCA). The experiment was conducted at JNKVV, Jabalpur, during the Kharif season of 2022, with 30 advanced breeding lines and five national checks in a randomized block design with three replications. PCA helped us to simplify the data and identify the most essential traits contributing to yield. The analysis revealed that three principal components (PCs) with eigenvalues greater than 1.00 explained 86.3% of the total variability among the studied traits. PC1, which accounted for 67.7% of the variation, was linked to seed yield per plant, primary branches per plant, days to maturity, days to 50% flowering, number of seeds per plant, number of pods per plant, harvest index, and biological yield per plant. PC2 and PC3, contributing 10.0% and 8.6% of the variation, respectively, were associated with the number of clusters per plant, pods per plant, 100-seed weight, and days to 50% flowering. High PC scores in these components indicated genotypes with significant variability, notably JS 25-37, JS 25-31, JS 25-39 for PC1, JS 25-22, JS 25-47 for PC2, and JS 25-45, JS 25-33 for PC3. These findings underscore the importance of these genotypes and traits for targeted breeding programs to enhance soybean yield.

Keywords: soybean, Principal Component Analysis, randomized block design, harvest index

Introduction

Soybean (*Glycine max* (L.) Merrill) is a venerable and indispensable seed legume crop, highly valued in global agriculture for its high protein and oil content (Amrate, 2024, Amrate *et al.*, 2024). As a member of the family *Leguminosae*, soybeans are mainly self-pollinating plants (Dubey *et al.*, 2018b, Mehra *et al.*, 2019, Amrate *et al.*, 2020). The genus *Glycine*, belonging to the tribe *Phaseoleae* and subfamily *Fabaceae*, has a complex taxonomic background (Soni *et al.*, 2022, Banerjee, 2021, Banerjee *et al.*, 2022). Within this genus, the subgenera *Glycine* and *Soja* are particularly noteworthy, with the former encompassing 26 wild perennial species (Jawarkar, 2021, Jawarkar *et al.*, 2022, Pallavi *et al.*, 2020,). The cultivated soybean (*Glycine max*) and its wild annual counterpart (*Glycine soja*) fall under the subgenus *Soja*, sharing a chromosome number of $2n = 40$ (Jha, 2016, Satpute *et al.*, 2016, Shilpashree *et al.*, 2021, Barela *et al.*, 2022). The origins of soybean can be traced back to northern and central China, with landraces developing in regions such as China, Korea, and India. Soybean versatility extends to various industrial applications, including biodiesel production and the manufacture of a wide array of food products (Jha *et al.*, 2016, Mishra *et al.*, 2018, Sileshi, 2019).

This distinctive crop is particularly noteworthy due to its high protein level, which falls within the range of 37–42%. It also has an ash content of 5–6%, a carbohydrate content ranging from 25–29%, and an oil content varying from 17–24%, depending on the specific varieties and locations. (Verma *et al.*, 2015, Banerjee *et al.*, 2023, Uikay *et al.*, 2022). Given the intricate nature of yield as a trait, effective selection and breeding necessitate sophisticated analytical techniques (Pachori *et al.*, 2023, Upadhyay *et al.*, 2020). Yield is impacted by various interconnected factors, making it difficult to pinpoint and prioritize the traits that make the most significant contributions to genetic gain (Yadav *et al.*, 2017, Banerjee *et al.*, 2022). This is where Principal Component Analysis (PCA) becomes indispensable (Dubey *et al.*, 2018a, Singh *et al.*, 2019).

PCA is a critical dimension reduction technique that condenses a large set of potentially correlated variables into a smaller set of uncorrelated variables, known as principal components (Massay, 1965, Jolliffe, 1986). The first principal component explains the most variation in the data, while each subsequent component elucidates the maximum remaining variation. By narrowing down the number of traits to those that produce the most variability, PCA simplifies the analysis and enhances the efficiency

of genotype selection (El-Hashash, 2016, Singh and Shrestha, 2019). In soybean breeding, PCA helps identify key traits that impact yield and other essential agronomic characteristics. This technique assists in ranking genotypes based on principal component scores, enabling the selection of superior breeding lines (Dubey *et al.*, 2018b). By focusing on the most influential traits, breeders can attain more significant genetic gain and enhance soybean crops' overall productivity and quality. This study seeks to apply PCA to advanced soybean breeding lines to identify the most significant traits contributing to yield and rank the genotypes accordingly. The results will offer valuable insights for soybean breeding programs, ultimately aiding in developing high-yielding, resilient soybean varieties.

MATERIAL AND METHODS

The experiment took place at the AICRP Soybean Seed Breeding Farm, Department of Plant Breeding and Genetics, College of Agriculture, JNKVV, Jabalpur (M.P.), during the *Kharif* season of 2022. The experimental site had uniform topography and medium

black soil with a pH of 7.5, and was free from waterlogged conditions. A total of 30 advanced soybean breeding lines were planted, along with five national checks provided by JNKVV Jabalpur (Table 1). The experiment was laid out in a Randomised complete Block Design (RCBD) with three replications to minimize experimental error and field variation. Each genotype was sown in two rows of 3 meters in length. The row-to-row distance was maintained at 40 cm, and the plant-to-plant distance was kept between 7 and 8 cm. All recommended agronomic practices and packages were employed to ensure the healthy growth of the crop. Observations were recorded from five competitive plants randomly selected from each advanced breeding line in each replication. Data were collected to evaluate yield and its contributing attributes. Principal Component Analysis (PCA) was performed following the methods described by Massay (1965) and Jolliffe (1986). PCA is a well-known dimension reduction technique that reduces a large set of variables into a smaller set of principal components. This technique was used to identify and analyze the key factors contributing to yield and its attributes.

Table-1. List of Genotypes

S. No	Genotypes	Pedigree
1	JS 25-20	JS 20-116 x AMSMB 5-18
2	JS 25-21	JS 20-116 x Cat 489A
3	JS 25-22	PS1092 x JS21-08
4	JS 25-23	SL 738 x JS 95-60
5	JS 25-24	JSM 283 x 21-08
6	JS 25-25	NRC-148 x JS 335
7	JS 25-26	JS 20-94 x JS 95-60
8	JS 25-27	JS 22-04 x Pb -1
9	JS 25-28	JS 22-05 x JS 20-34
10	JS 25-29	Pb-1 x 20-34
11	JS 25-30	JS 22-05 x Pb-1
12	JS 25-31	JS 20-116 x JS 95-60
13	JS 25-32	JS 20-98 x AMS 475
14	JS 25-33	JS -22-04 x JS 93-05
15	JS 25-34	(JS 20-98 x JS 20 -24) x (JS 95-60)
16	JS 25-35	JS 20-98 x JS 95-60
17	JS 25-36	DS 31-05 x JS 20-69
18	JS 25-37	MAUS -740 x JS 20-98
19	JS 25-38	JS 20-98 x Pb -1
20	JS 25-39	JS- 22-04 x JS -93-05
21	JS 25-40	JS 20-116 x AMSMB5-18
22	JS 25-41	JS 20-116 x Cat 489 A
23	JS 25-42	JS 20-98 x AMS 475

S. No	Genotypes	Pedigree
24	JS 25-43	JS 22-05 x JS 20-34
25	JS 25-44	JS 22-04 x Pb-1
26	JS 25-45	JS 20-69 x Pb -1
27	JS 25-46	Cat 87 x JS 20-94
28	JS 25-47	JS 20-116 X JS 95-60
29	JS 25-48	Pb-1 x JS-20-34
30	JS 25-49	JS 22-05 x Pb-1
Checks		
1.	JS 20-34	JS 98-63 x PK 768
2.	JS 20-69	JS 97-52 X SL 710
3.	JS 20-116	JS 97-52 x JSM 120 A
4.	JS 20-94	JS 97-52 x JS 20-02
5.	JS 20-98	JS 97-52 x SL 710

Table-2. Eigen values, percentage of total variation and cumulative percentage for twelve traits in soybean

Characters	Principal components (PC)	Eigenvalue	Variability %	Cumulative Variability %
Days to 50% flowering	PC1	8.128	67.7	67.7
Days to maturity	PC2	1.201	10.0	77.7
Plant height (cm)	PC3	1.032	8.6	86.3
No of primary branches per plant	PC4	0.575	4.8	91.1
No of cluster per plant	PC5	0.428	3.6	94.7
No. of pods per plant	PC6	0.264	2.2	96.9
No. of seeds per pod	PC7	0.193	1.6	98.5
No. of seeds per plant	PC8	0.078	0.7	99.2
100 seed weight (g)	PC9	0.045	0.3	99.5
Biological yield per plant (g)	PC10	0.026	0.3	99.8
Harvest index	PC11	0.017	0.1	99.9
Seed yield per plant (g)	PC12	0.012	0.01	100



Fig.1 -Scree Plot laid between Eigen Value and Cumulative Variability of 12 PCs

Table 3. Principal components for twelve yield contributing traits of soybean genotypes

Traits	Principal components		
	PC1	PC2	PC3
Days to 50% flowering	0.318	0.025	0.101
Days to maturity	0.327	-0.045	0.17
Plant height (cm)	0.112	0.036	-0.866
No. of primary branches per plant	0.332	-0.148	0.132
No. of cluster per plant	0.071	0.81	0.239
No. of pods per plant	0.258	0.426	-0.152
No. of seeds per pod	0.316	-0.246	0.066
No. of seeds per plant	0.318	0.057	-0.212
100 seed weight (g)	0.304	-0.057	0.193
Biological yield per plant (g)	0.321	-0.203	0.099
Harvest index	0.293	0.158	-0.113
Seed yield per plant (g)	0.341	-0.059	-0.067

Table 4. PC scores of soybean genotypes

Genotype	PC1	PC2	PC3
JS 25-20	-0.711	0.493	0.528
JS 25-21	0.212	1.121	-0.565
JS 25-22	0.629	1.731	0.028
JS 25-23	-0.238	0.794	-0.733
JS 25-24	0.229	1.622	-1.523
JS 25-25	0.031	0.269	0.157
JS 25-26	0.519	-0.398	0.034
JS 25-27	-1.328	-0.26	0.556
JS 25-28	-0.288	0.282	-0.349
JS 25-29	-0.495	-0.419	0.021
JS 25-30	0.09	1.437	-0.037
JS 25-31	1.329	-0.222	-2.191
JS 25-32	0.81	1.3	-1.285
JS 25-33	0.119	0.537	1.478
JS 25-34	0.091	1.412	1.228
JS 25-35	1.082	0.618	-0.012
JS 25-36	-0.16	-0.383	0.493
JS 25-37	2.383	-1.576	-1.693
JS 25-38	1.414	-1.648	-1.05
JS 25-39	0.923	-1.176	1.355
JS 25-40	0.747	-0.464	0.458
JS 25-41	0.617	-0.453	0.318
JS 25-42	-0.767	-0.669	0.832
JS 25-43	-1.252	-0.233	-0.881
JS 25-44	-0.455	0.47	0.559
JS 25-45	1.35	-1.134	1.678
JS 25-46	-0.654	0.879	1.222
JS 25-47	0.851	1.728	1.11
JS 25-48	-0.248	-0.362	0.387
JS 25-49	0.651	-1.501	1.277
JS 20-34	-1.677	-0.303	-1.269
JS-20-69	-1.412	-1.015	-0.07
JS 20-116	-1.32	-1.528	0.294
JS-20-94	-1.105	-0.394	-1.225
JS-20-98	-1.967	-0.555	-1.13

Table-5. List of genotypes selected in each PC

S.NO	PC 1	PC 2	PC 3
1.	JS 25-31	JS 25-21	JS 25-33
2.	JS 25-35	JS 25-22	JS 25-34
3.	JS 25-37	JS 25-24	JS 25-39
4.	JS 25-38	JS 25-30	JS 25-45
5.	JS 25-39	JS 25-34	JS 25-46
6.		JS 25-47	JS 25-47
7.			JS 25-49

Results and Discussion

The Principal Components Analysis (PCA) conducted in this investigation provided valuable insights into the variability of soybean traits. Of the twelve principal components (PCs) examined, three exhibited eigenvalues exceeding 1.00, representing 86.3% of the total variability. These three PCs were prioritized for further analysis. PC1 contributed the most to the variation at 67.7%, suggesting that selecting characteristics within this component would be advantageous. PC2 and PC3 contributed 10.0% and 8.6% to the overall variation, respectively. The rotated component matrix indicated that each principal component was associated with various yield and yield-attributing traits. PC1 was mainly associated with traits such as seed yield per plant, primary branches per plant, days to maturity, days to 50% flowering, number of seeds per plant, number of pods per plant, harvest index, and biological yield per plant. PC2, the second most variable component, was linked to traits like the number of clusters per plant and the number of pods per plant. PC3 included traits such as the number of clusters per plant, primary branches per plant, 100-seed weight, and days to 50% flowering. These findings are consistent with those reported by Dubey *et al.* (2018), Jha *et al.* (2016) and El-Hashash (2016), who also observed similar associations between these traits and principal components.

Principal component scores were computed for all genotypes across the three principal components. These scores can be utilized to predict precise selection indices based on the variability represented by each principal component. A high PC score for a particular genotype indicates in table 5.

Significant variability was observed in that genotype for the traits being considered. When looking at PC1, JS 25-37 showed the highest PC score, followed by JS 25-31, JS 25-39, JS 25-40, JS 25-26, JS 25-22, and JS 25-41. These genotypes displayed high values for traits such as the number of pods per plant, the number of seeds per plant, the number of primary branches per plant, the biological yield per plant, and the seed yield per plant. Moving on to PC2, JS 25-22

obtained the highest score, followed by JS 25-47, JS 25-24, JS 25-30, JS 25-34, JS 25-21, JS 25-20, JS 25-28, and JS 25-25, indicating high values for traits like days to 50% flowering and days to maturity. Lastly, for PC3, JS 25-45 achieved the highest score, followed by JS 25-33, JS 25-39, JS 25-34, JS 25-49, JS 25-46, JS 25-47, JS 25-42, JS 25-27, JS 25-20, JS 25-36, and JS 25-40. These genotypes exhibited high values for traits such as days to 50% flowering.

The studies conducted by various researchers such as Jha *et al.* (2014), Berhanu *et al.* (2019), Sileshi *et al.* (2019), Baria (2019), Singh *et al.* (2020), and Manav *et al.* (2017) all revealed consistent findings regarding different traits of soybean plants. Additionally, the PCA analysis identified JS 25-37, JS 25-39, JS 25-34 and JS 25-47 as promising genotypes with specific characteristics deemed significant across multiple principal components. The most important traits included the number of pods and seeds per plant, primary branches per plant, biological and seed yield per plant, days to flowering and maturity, and the harvest index. These traits should be given greater attention in future breeding programs to enhance soybean genotypes.

Conclusion

The analysis identified significant traits crucial for improving soybean productivity, such as number of pods and seeds per plant, primary branches per plant, biological and seed yield per plant, days to flowering and maturity, and the harvest index. Genotypes with high PCA scores, such as JS 25-37 and JS 25-22, JS 25-39, JS 25-34 and JS 25-47 demonstrate their potential for future breeding programs. Incorporating PCA into breeding strategies provides a focused approach to maximize genetic gain and effectively meet agricultural demands.

References

- Amrate, P. K. (2024). Survey and Present Status of Soybean Diseases in Central India. *International Journal of Bio-resource and Stress Management*. 15(5): 01-10. doi: <https://doi.org/10.23910/1.2024.5299>

- Amrate, P. K., Pancheshwar, D.K., and Shrivastava, M.K. (2018). Evaluation of soybean germplasm against charcoal rot, aerial blight and yellow mosaic virus disease in Madhya Pradesh. *Plant Disease Research*, 33(2):185-190.
- Amrate, P. K., Shrivastava, M.K., Pancheshwar, D.K., and Sharma, S. (2020). Charcoal Rot and Yellow Mosaic Virus Diseases of Soybean under Hot Spot Condition: Symptoms, Incidence and Resistance Characterization. *International Journal of Bio-resource and Stress Management*, 11(3): 268-273.
- Amrate, P. K., Chaukikar, K., Kharte, S., Pancheshwar, D.K., Marabi, R.S., Shrivastava, M.K. and Bhale, M.S. (2024). Distribution of Charcoal Rot of Soybean, its Influencing Factors and Pathogenic Variabilities in Different Regions of Madhya Pradesh. *Legume Research*. doi: 10.18805/LR-5262.
- Amrate, P. K., Shrivastava, M.K., Pancheshwar, D.K., Stuti, S. (2020). Charcoal rot and yellow mosaic virus disease of soybean under hot spot condition: symptoms, incidence and resistance characterization. *International Journal of Bio-resource and Stress Management*, 11(3): 268–273. doi: <https://doi.org/10.23910/1.2020.2104>.
- Banerjee J, Shrivastava M. K, Amrate P.K, Singh Y, Upadhyay A and Soni M. (2022). Genetic variability and association of yield contributing traits in advanced breeding lines of soybean. *Electronic Journal of Plant Breeding*. Vol 13(2): 597 – 607.
- Banerjee, J. (2021). Genetic Assessment and Morpho-Biochemical analysis in Advanced Breeding lines of Soybean. Msc. (Ag.) Thesis, Department of Plant Breeding and Genetics, JNKVV, Jabalpur (M.P.) JNKVV, Jabalpur: 156.
- Banerjee, J., Shrivastava, M. K., Amrate, P. K., Singh, Y., Upadhyay, A. and Soni, M. (2022). Genetic variability and association of yield contributing traits in advanced breeding lines of soybean. *Electronic Journal of Plant Breeding* 13(2): 597-607.
- Banerjee, J., Shrivastava, M.K., Singh, Y. Amrate, P.K. (2023). Estimation of genetic divergence and proximate composition in advanced breeding lines of soybean (*Glycine max* (L.) Merrill. *Environment and Ecology* 41(3C): 1960–1968. <https://doi.org/10.60151/envec/vywe5744>.
- Barela, A., Shrivastava, M.K., Mohare, S., Rahangdale, S., Jawarkar, S., Amrate, P.K., Singh, Y. (2022). Morphological characterization and recognition of new traits of soybean [*Glycine max* (L.) Merrill]. *International Journal of Environment and Climate Change* 12(12): 1497-1504.
- Dubey, N., Avinash, A.H. and Shrivastava, A.N. (2018a). Evaluation of genetic diversity among soybean [*Glycine max* (L.) Merrill] genotypes using multivariate analysis. *Plant Archives*, 18(1):908-912.
- Dubey, N., Avinash, H.A., and Shrivastava, A.N. (2018b). Principal component analysis in advanced genotypes of soybean [*Glycine max* (L.) Merrill] over seasons. *Plant Archives*, 18(1):501-506.
- El-Hashash, E.F. (2016). Genetic Diversity of Soybean Yield Based on Cluster and Principal Component Analyses. *Journal of Advances in Biology & Biotechnology*, 10 (3):1-9.
- Jawarkar, S. (2021). Genetic studies for yield associated trait in RILs of Soybean [*Glycine max* (L.) Merrill]". MSc Ag. College of Agriculture JNKVV, 160p.
- Jawarkar, S., Shrivastava, M.K., Satpute, G.K., Amrate, P.K., Barela, A. and Nagar, S.K. (2023). Morphological Characterization of Recombinant Inbred Lines of Soybean [*Glycine max* (L.) Merrill]. *International Journal of Environment and Climate Change* 13(9): 521-527.
- Jha, A, Shrivastava, A.N., and Mishra, S. (2016). Principal Component Analysis in Advanced Genotypes of Soybean (*Glycine max* (L.) Merrill)" during *Kharif*– 2014. *Advances in Life Sciences*, 5(9):3508-3513.
- Mehra, S., Shrivastava, M.K., Amrate, P.K. and Yadav, R.B. (2020). Studies on variability, correlation coefficient and path analysis for yield associated traits in soybean [*Glycine max* (L.) Merrill]. *Journal of Oilseeds Research*, 37(1): 56-59.
- Mishra, S., Jha, A., Pancheshwar, D.K. and Shrivastava, A.N. (2018). Study of Genetic Divergence in Advance Breeding Lines of Soybean [*Glycine max* (L.) Merrill] for Yield Attributing Traits. *International Journal of Bio-resource and Stress Management*, 9 (1):103-107.
- Pachori, S., Tiwari, S., Shrivastava, M.K., Amrate, P.K., Thakur, S. and Gupta, S. (2023). Unveiling genetic diversity in soybean germplasm: A comprehensive analysis through DUS characteristics. *Electronic Journal of Plant Breeding* 14(4): 1541-1547.
- Pallavi, K.S., Jiban, S, and Ujjawal, K.S.K. (2020). Multivariate analysis of soybean genotypes. *Journal of Agriculture and Natural Resources*, 3(1):69-76.
- Satpute, G. K., Gireesh, C., Shivakumar, M., Arya, G.K., Patel, R.K., Gupta, R. and Husain. S. (2016). Genetic variability and association studies in new soybean Germplasm accessions. *Soybean research*, 14(2):77-83.

Shilpashree, N., Devi, S.N., Manjunathagowda, D.C., Muddappa, A., Abdelmohsen, S.A., Tamam, N., Elansary, H.O., El-Abedin, T.K.Z., Abdelbacki, A.M., Janhavi, V. (2021). Morphological characterization, variability and diversity among vegetable soybean (*Glycine max* L.) genotypes. *Plants* 10(4): 671.

Singh, P.K., 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

Soni M, Shrivastava M.K, Singh Y, Sharma S, Amrate P.K and Banerjee J. (2022). Evaluation the Genetic Variability and Association Analysis of Soybean (*Glycine max* (L.) Merrill for Some Physiological Parameters. *International Journal of Plant & Soil Science* 34(10): 17-26.

Uikey, S., Sharma, S., Amrate, P.K., Shrivastava, M.K. (2022). Identification of rich oil-protein and disease resistance genotypes in soybean [*Glycine max* (L.) Merrill]. *International Journal of Bio-resource and Stress Management*. 13(5): 497-506. <https://doi.org/10.23910/1.2022.2478>.

Upadhyay, P., Shrivastava, M.K., Amrate, P.K. and Yadav, R.B. (2020). Yield determining traits, genetic variability and character association in exotic lines of soybean [*Glycine max* (L.) Merrill]. *Soybean Research* 18(2): 95-102.

Verma, K., Talukdar, A., Shivakumar, M., Kumar, B., Lal, S.K., Srivastava, N., Sapra, R.L., Harish, G.D., Girmilla, V. (2015). Biochemical screening for trypsin inhibitor factors and morpho-molecular characterization of soybean (*Glycine max* L. Merr.). *Indian Journal of Genetics* 75(4): 490-496.

Yadav, U.S., Mishra, S. and Shrivastava, A.N. (2017). Genetic diversity analysis based on seedling and physiological traits in advance generations of soybean [*Glycine max* (L.) Merrill]. *Frontiers in Crop Improvement* 5: 307-310.