The Bioscan
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES
20(2): S2: 954-957, 2025 www.thebioscan.com

MULTIVARIATE ANALYSIS OF YIELD-CONTRIBUTING TRAITS IN SOYBEAN (*GLYCINE MAX* (L.) MERRILL.): INSIGHTS FROM CORRELATION AND PRINCIPAL COMPONENT APPROACHES

SONALI SRIVASTAVA[1], PIYUSHA SINGH[2], ANAMISH TYAGI[1], AMAN SRIVASTAVA[1]

[1]Ph.D research scholar, Department of Genetics and Plant Breeding, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya-224 229, Uttar Pradesh, India

[2]Assistant Professor, Department of Genetics and Plant Breeding, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya-224 229, Uttar Pradesh, India.

ORCIDs: Sonali Srivastava (0009-0002-2666-8327), Piyusha Singh (0009-0004-7095-4659), Anamish Tyagi (0009-0008-8424-4606), Aman Srivastava (0009-0006-2816-2716)

Corresponding Email:- piyusha_singh@yahoo.com

DOI: 10.63001/tbs.2025.v20.i02.S2.pp954-957

KEYWORDS
Agronomic traits, Correlation analysis, Principal Component Analysis, Seed yield, Soybean genotypes

Received on:

18-05-2025

Accepted on:

04-06-2025

Published on:

03-07-2025

ABSTRACT

The present study aimed to assess the interrelationship among eight key agronomic traits in 247 diverse soybean genotypes to identify yield-contributing factors for effective selection. Field evaluation data were analysed using correlation and Principal Component Analysis (PCA). Correlation results showed that seed yield per plant had strong positive correlations with number of pods per plant (0.719**), seeds per pod (0.643**), and seed yield efficiency (0.787**). Conversely, it showed significant negative correlations with plant height (-0.761**), primary branches (-0.682**), and number of nodes (-0.611**). Days to maturity was negatively correlated with seed yield (-0.619*) and plant height (-0.428*), but positively associated with primary branches (0.67*). PCA revealed that PC1 and PC2 accounted for 58.95% and 13.32% of total variance, respectively. PC1 showed positive loadings for seed yield (0.393) and yield efficiency (0.3409), and negative for pods per plant (-0.4257). Findings suggest compact, early-maturing genotypes with higher pod and seed counts are ideal for yield improvement.

INTRODUCTION

One of the most significant leguminous crops in the world, soybean (*Glycine max* (L.) Merrill) is primarily farmed for their high protein and oil content (Rani *et al.,* 2023). Globally, the importance of soybean has grown rapidly, and today it is cultivated in over 120 million hectares across countries like the USA, Brazil, Argentina, India, and China (FAOSTAT, 2023). Soybean has a great significance economically and it is also a highly lucrative crop (FAO,2023). Commonly referred to as the "Golden Bean" or "Miracle Crop", soybean is prized for its dual composition of protein and oil. With world production reaching over 380 million tons in 2021, soybean plays a key role in ensuring food security, supporting the livestock sector, and driving economic growth in many countries (FAOSTAT, 2023).

In agricultural and breeding research, Correlation Analysis and Principal Component Analysis (PCA) are widely used statistical tools to explore the relationships among multiple traits in soybean. Correlation analysis is used to determine whether two traits vary together. For example, positive correlations have been consistently observed between hundred seed weight, seed area, seed width, and seed perimeter, which means selecting for one of these traits often improves the others (Sun *et al.*, 2022). Similarly, a negative correlation between seed aspect ratio and seed circularity suggests that more elongated seeds tend to be less circular in shape (Sun *et al.*, 2022).

Correlation matrices also help in avoiding undesirable combinations. For instance, in seed sprout quality studies, positive correlations between phenolic content and hypocotyl diameter can be beneficial for selecting lines with better antioxidant activity and seedling vigour (Li *et al.*, 2023).

Principal Component Analysis (PCA) is another powerful method used to reduce data complexity by identifying key factors (principal components) that explain the majority of variation among traits. In a study involving 589 soybean accessions, the first two principal components explained 87.6% of the total variability, with PC1 representing seed size-related traits (seed weight, area, perimeter) and PC2 representing shape traits (circularity and roundness) (Sun *et al.*, 2022). Similarly, another study that evaluated 101 soybean sprout lines found that three PCs captured nearly 78% of the total variation, with PC1 strongly related to traits like total phenolics, seed weight, and edible sprout rate (Li *et al.*, 2023). PCA plots also help identify clusters of superior genotypes, which can then be targeted for breeding and improvement programs. These tools have proven vital in screening for drought tolerance, high-yield potential, disease resistance, and seed longevity, among other traits (Sun *et al.*, 2022; Li *et al.*, 2023).

Additionally, soybean has a rich history of domestication and cultivation spanning over 5000 years, with its origin in ancient China and later expansion to many parts of the world (Sedivy *et al.*, 2017). Its economic value is derived from its exceptional protein and oil content, its environmental role in nitrogen fixation, and its wide industrial and agricultural uses. Correlation and PCA are essential tools that help researchers understand trait relationships and reduce complexity in data analysis, thereby accelerating the breeding of superior soybean varieties. In today's context of climate change and food security challenges, soybean continues to be a crop of immense relevance to global agriculture and sustainability (FAO, 2023).

MATERIALS AND METHODS

A field trial was conducted during 2023-24 Kharif season using an augmented design to evaluate 45 genetically diverse soybean (*Glycine max* (L.) Merrill) lines including three checks for key agronomic, physiological, and seed quality traits under natural field condition at the Genetics and Plant Breeding Farm of Acharya Narendra Deva University of Agriculture and Technology in Ayodhya, Uttar Pradesh, India, for optimal crop growth. Genotypes were laid in 3 blocks for the study of eight traits *viz*., days to maturity, plant height in cm, number of primary branches per plant, number of nodes per plant, number of pods per plant, number of seeds per pod, seed yield efficiency (%) and seed yield per plant (g). Collection of data was done by selecting five plant/plot randomly. Collection of data was done by selecting five plant/plot in each block.

RESULTS AND DISCUSSION

The correlation analysis (Table 1) shows how different traits in soybean are connected with each other. These relationships help in identifying which traits are more important and how they influence seed yield directly or indirectly.

Days to Maturity

The present investigation showed that days to maturity had a negative and significant relation with seed yield (-0.619*) and plant height (-0.428*), meaning that late-maturing plants gave lower yield and were shorter. However, it had a positive and significant relation with number of primary branches per plant (0.67*), showing that late-maturing plants had more branches.

Plant height

The present study showed that plant height had a positive and significant relation with number of pods per plant (0.697**), number of primary branches (0.64**) and number of nodes (0.621**), meaning taller plants had more pods, branches, and nodes. But it showed negative and significant correlation with seed yield per plant (-0.761**) and seed yield efficiency (-0.549**), showing that taller plants may not always give high yield.

Number of primary branches per plant

Number of nodes (0.923**), number of pods (0.878**), plant height (0.64**), and number of seeds per pod (0.518**) showed positive and significant relation with number of primary branches. But seed yield (-0.68**) and seed yield efficiency (-0.516**) had negative and significant relation, meaning more branches did not always lead to better yield.

Number of nodes per plant

This trait had positive and significant relation with number of branches (0.923**), pods (0.899**), plant height (0.621**) and seeds per pod (0.472**). But it was negatively and significantly related to seed yield (-0.611**) and yield efficiency (-0.459**), showing more nodes did not always give better performance.

Number of pods per plant

It had positive and significant relation with nodes (0.899**), branches (0.878**), plant height (0.697**), seeds per pod (0.44**), seed yield (0.719**), and yield efficiency (0.545**), indicating that more pods clearly helped in getting higher yield.

Number of seeds per pod

This trait showed positive and significant relation with branches (0.518**), nodes (0.472**), pods (0.44**), seed yield (0.643**), and yield efficiency (0.437**), meaning more seeds per pod directly helped in better productivity.

Seed yield efficiency (%)

It had a positive and significant relation with seed yield (0.787**), pods (0.545**) and seeds per pod (0.437**). But it was negatively related to plant height (-0.549**), branches (-0.516**), and nodes (-0.459**), showing that efficient plants were not always tall or heavily branched but gave better yield.

Seed yield per plant

Seed yield had positive and significant relation with yield efficiency (0.787**), number of pods (0.719**) and seeds per pod (0.643**). But it showed negative and significant relation with plant height (-0.761**), branches (-0.682**) and nodes (-0.611**), showing that compact and efficient plants gave better results.

These results are consistent with previous studies. For instance, Sun *et al.,* (2022) found a substantial relationship between soybean output and features associated to pods. Li *et al.,* (2023) also emphasized that characteristics such as the number of seeds per pod and the quantity of pods were important factors in determining the final yield. These findings support the idea that yield is a complicated characteristic impacted by a number of interconnected factors, and that enhancing one may have a positive or negative impact on others.

Table 1 SIMPLE CORRELATION VALUES OF EIGHT KEY TRAITS.

Traits	Days to Maturity	Plant Height	Number of Primary Branches Per Plant	Number of Nodes Per Plant	Number of Pods Per Plant	Number of Seeds Per Pod	Seed Yield Efficiency	Seed Yield Per Plant
Days to Maturity	**1.0**	-0.428*	0.67*	-0.164 NS	-0.128 NS	0.07 NS	0.128 NS	-0.619*
Plant Height		**1.0**	0.64 **	0.621 **	0.697 **	0.241 NS	-0.549 **	-0.761 **
Number of Primary Branches Per Plant			**1.0**	0.923 **	0.878 **	0.518 **	-0.516 **	-0.682 **
Number of Nodes Per Plant				**1.0**	0.899 **	0.472 **	-0.459 **	-0.611 **
Number of Pods Per Plant					**1.0**	0.44 **	0.545 **	0.719 **
Number of Seeds Per Pod						**1.0**	0.437 **	0.643**

Seed Yield Efficiency							**1.0**	0.787 **
Seed Yield Per Plant								**1.0**

Note: * = Significant at 0.05; ** = Significant at 0.01; NS = Not Significant

PCA Analysis

Variance Distribution and Trait Correlation

Principal Component Analysis (PCA) was performed using data from eight important agronomic traits in soybean. The analysis was carried out using Singular Value Decomposition (SVD). A total of eight principal components (PCs) were generated.

Inference from Variance Distribution **(Table 2)**

As per Kaiser's criterion, only the components with an eigenvalue greater than 1 are considered meaningful for further interpretation. In this study, PC1 and PC2 fulfilled this condition. Principal Component 1 (PC1) contributed the maximum variation, accounting for 58.95% of the total variability among the traits, while Principal Component 2 (PC2) explained an additional 13.32%. Together, PC1 and PC2 captured a substantial 72.27% of the total variation, indicating that these two components effectively summarize the majority of the data's variability.

This implies that most of the variability present in the data could be effectively represented by the first two components alone.

Table 2 EIGENVALUE, PROPORTION OF VARIANCE AND CUMULATIVE PROPORTION OF EIGHT KEY COMPONENTS

Principal Component	Eigenvalue	Variance (%)	Cumulative Variance (%)
PC1	4.7	58.946	58.946
PC2	1.1	13.325	72.271
PC3	0.9	11.015	83.286
PC4	0.7	9.325	92.612
PC5	0.3	3.813	96.425
PC6	0.1	1.509	97.934
PC7	0.1	1.384	99.317
PC8	0.1	0.683	100

Table 3 showed inference from Trait Correlations with Principal Components **:** The principal component PC1 had positive correlation with variables seed yield per plant (0.393), seed yield efficiency (0.3409) and days to maturity (0.063), while negative correlation with variables number of seeds per pod (-0.2497), plant height (-0.3686), number of nodes per plant (-0.4105), number of primary branches per plant (-0.4213) and number of pods per plant (-0.4257).

The principal component PC2 had positive correlation with variables plant height (0.0645), while negative correlation with variables number of pods per plant (-0.0022), number of nodes per plant (-0.0237), seed yield efficiency (-0.0861), number of primary branches per plant (-0.108), seed yield per plant (-0.1766), number of seed plant per plant (-0.4791) and days to maturity (-0.8458).

The current results are consistent with previous studies that showed that traits such as pods per plant and seed weight can be efficiently captured by PCA for genotype selection (Gul *et al.*, 2021). Patil *et al.*, (2020) also reported that the initial principal components in soybeans typically represent yield and maturity traits, thereby bolstering their role in strategic breeding programs. The usefulness of PCA in locating important inter-trait connections and streamlining intricate datasets in soybeans through dimensionality reduction was also highlighted by Li *et al.*, (2023).

Table 3 Correlation Between Principal Component and Eight Key Traits.

Traits	PC1	PC2
Days to Maturity	0.063	-0.8458
Plant Height	-0.3686	0.0645
Number of Primary Branches Per Plant	-0.4213	-0.108
Number of Nodes Per Plant	-0.4105	-0.0237
Number of Pods Per Plant	-0.4257	-0.0022
Number of Seeds Per Pod	-0.2497	-0.4791
Seed Yield Efficiency	0.3409	-0.0861
Seed Yield Per Plant	0.393	-0.1766

CONCLUSION

The present study aimed to understand the relationship among eight key agronomic traits in soybean and their influence on seed yield. Correlation analysis revealed that yield is a complex trait, strongly affected by several other characteristics. Seed yield per plant showed a strong positive association with number of pods per plant, seeds per pod, and seed yield efficiency, suggesting that increasing these traits can directly enhance yield. In contrast, plant height, number of nodes, and number of branches, though important for plant structure, showed negative correlations with yield, indicating that compact and efficient plants may perform better. Principal Component Analysis (PCA) supported these findings, with PC1 and PC2 explaining 72.27% of total variation. PC1 was positively linked to yield traits and negatively to vegetative ones. PC2 showed a strong negative loading for days to maturity, favoring early-maturing types. These findings are supported by Gul et al. (2021), Patil et al. (2020), and Li et al. (2023).

REFERENCES

- FAO,2023. The State of Food and Agriculture 2023: Revealing the true cost of food to transform agrifood systems. FAO, Rome.
- FAOSTAT,2023. Food and Agriculture Organization of the United Nations.
- Graham, P.H. and Vance, C.P. 2000. Nitrogen fixation in perspective: An overview of research and extension needs. *Field Crops Res.* 65(2–3): 93–106.
- Gul, I., Khan, N.U., Ahmad, H., Khan, A.A., Bibi, F. and Saeed, A. 2021. Multivariate analysis and trait association in soybean. *Pak. J. Bot.* 53(2): 789–796.
- Li, X., Zhang, J., Chen, H. and Zhang, Y. 2023. Diversity analysis and evaluation of 101 soybean germplasms based on sprout quality. *Plant Diversity* (in press).
- Li, Y., Zhou, Q., Wang, L. and Liu, X. 2023. Correlation and PCA analysis of phenolic traits and sprout quality in soybean. *Front. Plant Sci.* 14: Article 1229495.
- Patil, S.S., Gudadinni, R.M., Biradar, B.D. and Ninganur, B.T. 2020. Principal component analysis in soybean [*Glycine max* (L.) Merrill] for yield attributing traits. *Legume Res.* 43(3): 385–390.
- Rani, R., Raza, G., Ashfaq, H., Rizwan, M., Razzaq, M.K., Waheed, M.Q., Shimelis, H., Babar, A.D. and Arif, M. 2023. Genome-wide association study of soybean (*Glycine max* (L.) Merrill) germplasm for dissecting the quantitative trait nucleotides and candidate genes underlying yield-related traits. *Front. Plant Sci.* 14: Article 1229495.
- Sedivy, E.J., Akpertey, A. and Miklas, P.N. 2017. Genetic studies and breeding strategies for soybean improvement. In: Janick, J. (Ed.). *Plant Breeding Rev.* Vol. 41: 45–96.
- Sun, S., Li, Y., Zhang, H. and Zhao, L. 2022. Genetic dissection of seed size and shape traits using GWAS and PCA in soybean. *Front. Plant Sci.* 13: Article 886300.
- Sun, X., Wang, T. and Yu, D. 2022. Image-based phenotypic analysis and classification of soybean seed traits. *Plants* 11(15): 1989.