International Journal of Plant & Soil Science

Volume 37, Issue 6, Page 581-593, 2025; Article no.IJPSS.138281
ISSN: 2320-7035

Genetic Diversity Assessment of Soybean Genotypes Using D^2 and Principal Component Analysis for Breeding Advancements

Yamini Gautam [a], Goutam Mohbe [a], Lalita Bishnoi [b], Anurag Sharma [a], Riya Mishra [a], Sanjeev Sharma [a], Jagendra Singh [a] and M.K. Tripathi [a]*

[a] *Department of Genetics & Plant Breeding, Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya, Gwalior, Madhya Pradesh, India.*
[b] *Sardarkrushinagar Dantiwada Agricultural University, Dantiwada, Gujarat, India.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/ijpss/2025/v37i65537

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://pr.sdiarticle5.com/review-history/138281

Original Research Article

Received: 16/04/2025
Accepted: 18/06/2025
Published: 21/06/2025

ABSTRACT

Soybean (*Glycine max* L. Merrill), major leguminous crop rich in protein and oil, plays a critical role in human nutrition and sustainable agriculture. Genetic diversity analysis is vibrant to identify better parents for developing high-yielding cultivars with improved agronomic traits. The present investigation was undertaken during *Kharif* 2024 at the Zonal Agricultural Research Station,

---

**Corresponding author: E-mail: drmanojtripathi64@gmail.com;*

Cite as: *Gautam, Yamini, Goutam Mohbe, Lalita Bishnoi, Anurag Sharma, Riya Mishra, Sanjeev Sharma, Jagendra Singh, and M.K. Tripathi. 2025. "Genetic Diversity Assessment of Soybean Genotypes Using D² and Principal Component Analysis for Breeding Advancements". International Journal of Plant & Soil Science 37 (6):581-93. https://doi.org/10.9734/ijpss/2025/v37i65537.*

Morena, RVSKVV, Gwalior, M.P., India to evaluate genetic divergence among 60 elite soybean genotypes. The experiment was laid out in a Randomized Block Design with two replications, and observations were recorded on 13 quantitative traits. Mahalanobis' D^2 statistic and Tocher's clustering method were employed to estimate genetic divergence, while Principal Component Analysis (PCA) was used to identify main contributors of traits to create variability. Cluster analysis grouped the genotypes into 14 distinct clusters, with maximum intra-cluster distance observed in Cluster IV (139.66) while highest inter-cluster divergence ($D^2 = 456.59$) between Clusters XIII and VIII, suggesting potential for creating superior recombinants through hybridization. Cluster VIII displayed the highest mean seed yield per plant (19.56 g), along with higher values for numbers of pods per plant and seeds per plant, and harvest index. PCA revealed five principal components with eigenvalues >1.0, collectively accounting for 77.39% of the total variation. Genotypes like JS-21-17, Cat492A, NRC-142, and AUKS-21-5 were identified as promising based on high PC scores and desirable agronomic traits. The study emphasizes the utility of multivariate analysis in identifying genetically diverse and agronomically superior soybean genotypes. These genotypes hold significant promise as parental lines in future breeding programmes aimed to enhance yield potential, stress tolerance, and overall productivity in soybean.

Keywords: Cluster analysis; Mahalanobis D^2 analysis; genetic diversity; genetic variability; principal component analysis; soybean.

1. INTRODUCTION

Soybean (*Glycine max* L. Merrill) is a globally significant leguminous crop that serves as a cornerstone for both nutritional security and sustainable agriculture (Mishra et al., 2020; Mishra et al., 2024a). Renowned for its exceptional nutritional composition, soybean seeds contain approximately 40% high-quality protein and about 20% oil, rendering it a vital source of essential amino acids and dietary fats (Mishra et al., 2021a; Mishra et al., 2024b). These attributes make soybean a key ingredient in human nutrition, animal feed formulations, and a wide range of industrial products including biodiesel, bioplastics, and food-grade emulsifiers (Mishra et al., 2021b; Sharma et al., 2021; Dharshini et al., 2024; Mishra et al., 2024c). Beyond its nutritional and economic value, it plays a pivotal ecological role in agroecosystems. As a leguminous crop, it establishes a symbiotic association with *Bradyrhizobium japonicum* (formerly) or *Rhizobium japonicum*, facilitating biological nitrogen fixation (BNF) (Mishra et al., 2021c; Nakei et al., 2022; Mishra et al., 2024a; Mishra et al., 2025a). This natural process converts atmospheric nitrogen (N_2) into bioavailable forms, significantly reducing the need for synthetic nitrogen fertilizers. Consequently, the inclusion of soybean in crop rotation systems not only improves soil fertility and structure but also contributes to long-term soil health and productivity (Mahmud et al., 2020; Mishra et al., 2021d; Roy et al., 2024). This dual role-nutritional and ecological-positions makes soybean as a strategic crop for addressing global challenges related to food security, soil degradation and climate-resilient farming (Brivery, 2021; Mishra et al., 2021e; Majidian et al., 2024).

In recent decades, the escalating global demand for soybean driven by its multifaceted utilization in food, feed and industrial sectors has emphasized the urgent need to develop cultivars that are not only high-yielding but also resilient to climatic adversities and enriched with superior nutritional attributes (Upadhyay et al., 2020; Hamza et al., 2024; Mishra et al., 2025b). The intensification of breeding efforts to meet these demands is, however, challenged by the narrow genetic base that characterizes many commercial soybean varieties (Tripathi et al., 2023; Sharma et al., 2023; Mishra et al., 2024d). This genetic bottleneck significantly hampers the scope for improvement in complex quantitative traits such as grain yield, phenological maturity, biotic and abiotic stress tolerance and nutrient use efficiency. To overcome these limitations, it is imperative to systematically assess the extent of genetic diversity within and among elite soybean germplasm collections (Collins et al., 2008; Mishra et al., 2024e). A comprehensive understanding of genetic divergence serves as a critical foundation for identifying genetically distant and agronomically superior parental lines, which can be strategically utilized in hybridization programmes (Singer et al., 2021; Mishra et al., 2024e; Aziz & Masmoudi, 2025). Such diversity assessments not only facilitate the broadening of the genetic base but also enhance the efficiency of marker-assisted selection (MAS), genomic selection (GS) and other molecular breeding approaches intended to accelerate genetic gains (Sandhu et al., 2022; Mishra et al., 2024f; Sharma et al., 2025). Ultimately, integrating genetic diversity analysis into breeding pipelines is essential for ensuring the long-term sustainability and adaptability of soybean cultivation in the face of changing agro-climatic conditions and increasing global food and nutritional demands (Salgotra & Chauhan, 2023; Mishra et al., 2024e).

Genetic divergence analysis using multivariate statistical approaches offers a robust framework for dissecting the variability present in diverse genotypes. Mahalanobis' D^2 statistic (Mahalanobis, 1936), in conjunction with Tocher's clustering method (Rao,1952), is widely employed to quantify genetic distances and to categorize genotypes into homogeneous groups based on multiple traits. This helps in identifying divergent parents for hybridization, thereby maximizing heterosis and producing superior recombinants (Raina et al., 2015; Mark & Workman, 2018; Kanavi et al., 2019; Sharma et al., 2023; Yadav et al., 2023). Furthermore, Principal Component Analysis (PCA) serves as a powerful tool to reduce data dimensionality (Massey,1965; Jolliffe, 1986) and to highlight the most influential traits contributing to genetic variability (Wang, 2009; Roessner et al., 2011). The integration of these techniques facilitates precise genotype selection and accelerates breeding efforts. With the objective of assessing the extent of genetic divergence using D^2 analysis and PCA, thirteen quantitative traits encompassing phenological, morphological and yield-attributing characteristics were investigated. The aim was to identify genetically diverse and agronomically superior genotypes that could serve as valuable parental lines in future breeding programmes.

2. MATERIALS AND METHODS

2.1 Experimental Site

The field experiment was conducted during the *Kharif*, 2024 at the Research Farm, Zonal Agricultural Research Station, Morena, Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya, Gwalior, Madhya Pradesh, India. The experimental location is situated at 26.5°N latitude and 78.0°E longitude, with an elevation of approximately 177 meters above mean sea level. The soil of the experimental field was medium-black, well-drained, and topographically uniform. The site experienced typical monsoonal climatic conditions, with temperatures ranging between 25°C to 30°C, which are considered ideal for soybean growth. Rainfall during the crop period was within the low to moderate range, which aligns with the agro-climatic requirements of soybean; however, the region is also prone to yield fluctuations due to intermittent drought, continuous rainfall, or waterlogging.

2.2 Experimental Details

The experiment was laid out using Randomized Block Design (RBD) with two replications. A total of 60 genotypes of soybean [*Glycine max* (L.) Merrill] were evaluated. The seeds were collected from multiple sources, viz., College of Agriculture, JNKVV, Jabalpur, RAK College of Agriculture, Sehore, and RVSKVV, Gwalior, to ensure a diverse genetic base. Each plot consisted of rows spaced 30 cm apart, with a plant-to-plant distance of 10 cm. The length of each row was 5-meters-long. All recommended agronomic practices including land preparation, sowing, fertilization, and pest management were uniformly followed throughout the crop period to ensure optimum growth and reliable data collection. To ensure comprehensive observations, a random selection of five plants was made from each replication of every genotype, serving as the foundation for data collection across all measured traits.

2.3 Statistical Analysis

Mahalanobis' D^2 statistic (Mahalanobis, 1936), in conjunction with Tocher's clustering method (Rao,1952) was used to form clusters based on the calculated D^2 values. Furthermore, Principal Component Analysis (PCA) (Massey,1965; Jolliffe, 1986) was employed to highlight the most influential traits contributing to genetic variability. To determine the genetic divergence among the genotypes, Agri Analyzer software was used to analyse PCA and OP-STAT for D^2 analysis.

3. RESULTS AND DISCUSSION

3.1 Mahalanobis D^2 Analysis

The genetic divergence among sixty soybean genotypes was estimated using Mahalanobis's D^2 statistic and subsequent clustering was carried out through Tocher's method (Fig.1). This multivariate analysis revealed presence of considerable genetic variability among the

Table 1. Percent contribution of characters towards divergence in 60 soybean genotypes

Source	Characters	Contribution %
1	Days to 50% flowering	1.13 %
2	Days to podding	3.62 %
3	Days to seed filling	3.90 %
4	Days to maturity	12.55 %
5	Plant height (cm)	1.81 %
6	Numbers of primary branches	0.78 %
7	Numbers of pods per plant	28.93 %
8	Numbers of seeds per pod	1.91 %
9	Numbers of seeds per plant	30.79 %
10	Hundred- seed weight (g)	2.08 %
11	Harvest index (%)	3.62 %
12	Biological yield (g/plant)	2.37 %
13	Seed yield per plant (g/plant)	5.89 %

Table 2. Distribution of soybean genotypes in different clusters

Cluster No.	Name of genotypes	Total number of genotypes
I	JS-20-116, JS-26, JS-20-79, JS-22-12, JS-24-26, RVS-23-26, RVS-23-15, RVS-23-23, RVS-2001-4, RVS-23-5, RVS-23-12, RVS-23-20, RSC-10-46, AMS-264, AMS-2021-3, TS-208, Rajsoya-24, DS1510, Himso1695, NRC-201, NRC-166, NRC-138, NRC-192, NRCSL-7, KDS-1203, KBSL-23-36, Pusa Sipani BS 8, PS 1569, MAUS-787, MAUS-791, SL-1315	31
II	VLS- 104, KDS-1201, NRC-142, Cat 87, CAUMS-3, AMS-100-39, ASB-85, ASB-93, RVSM-2012-4, JS-20-94, JS-22-01	11
III	JS-25-03	1
IV	JS-21-07, RSC-10-52, AS-26, NRC-255, NRC-152, KDSIS-1394, BAUS(M)-6	7
V	SL-311	1
VI	DLSB-40	1
VII	Cat492A	1
VIII	JS-21-17	1
IX	MACS-824	1
X	KSS-213	1
XI	NRCSL-4	1
XII	RVS-23-10	1
XIII	JS-23-05	1
XIV	AUKS-21-5	1

genotypes for the investigated quantitative traits (Table 1). The contribution of individual traits to the overall divergence indicated that the numbers of seeds per plant donated the highest (30.79%), closely followed by the numbers of pods per plant (28.93%), days to maturity (12.55%), and seed yield per plant (5.89%). Traits like numbers of primary branches per plant (0.78%), days to 50% flowering (1.13%), plant height (1.81%) and numbers of seeds per pod (1.91%) contributed the least, indicating their relatively lower role in genetic divergence.

Based on the D^2 values, the 60 soybean genotypes were grouped into 14 clusters (Table 2). Cluster I was the largest, comprising 31 genotypes, tracked by Cluster II (11 genotypes) and Cluster IV (7 genotypes). The remaining 11 clusters were mono-genotypic, each containing a single genotype, suggesting the presence of unique and genetically diverse lines. The intra-cluster distances arrayed between 115.59 (Cluster I) to 139.66 (Cluster IV), while mono-genotypic clusters displayed an intra-cluster distance of 0.00. The maximum inter-cluster divergence (D^2=456.59) was observed between Cluster XIII and Cluster VIII, signifying substantial genetic distance and potential for use in hybridization programmes. Conversely, the minimum inter-cluster divergence (D^2 = 89.66) was recorded between Cluster X and Cluster V.

Cluster means for the quantitative traits exposed significant variability across clusters (Table 3). Cluster XII exhibited the highest mean for days to 50% flowering (43.00), while Cluster VII recorded the lowest (36.00). Days to podding ranged from 47.50 in Cluster VII to 60.00 in Cluster VI, and days to seed filling varied between 58.00 (Cluster XIV) and 75.00 (Cluster XI). The numbers of pods per plant was found highest in Cluster VIII (95.17) and lowest in Cluster IV (42.69). Cluster VIII also showed superior performance for several traits, including numbers of seeds per plant (222.66), 100-seed weight (12.43 g), harvest index (63.03%), and seed yield per plant (19.56 g), signifying its potential utility for high-yielding trait improvement. In contrast, Cluster XIII demonstrated the lowest seed yield (5.61 g) and harvest index (23.40%). The wide inter- and intra-cluster variability highlights the opportunity for selection and exploitation of diverse genotypes to improve yield and other agronomic traits in soybean breeding programmes. Similar studies have also been conducted by Kacchadia et al. (2014), Ghughe et al. (2023) and Paikra et al. (2025). for different traits of soybean.

The Mahalanobis D^2 and Tocher's clustering analyses revealed substantial genetic divergence among the sixty soybean genotypes, highlighting a rich genetic reservoir for breeding. Traits like numbers of seeds and pods per plant subsidized the most to overall divergence, underlining their importance as key selection criteria for yield improvement. In contrast, traits such as plant height, numbers of primary branches, and days to 50% flowering donated minimally, demonstrating uniformity or limited discriminative power (Sirisa et al., 2020; Paikra et al., 2025).

The clustering pattern presented a concentration of genotypes in major clusters (*e.g.*, Cluster I and II) alongside several mono-genotypic clusters, reflecting both closely related and highly distinct genotypes. Mono-genotypic clusters likely harbour unique alleles or rare trait combinations, making them valuable for enhancing genetic diversity through hybridization (Singh et al., 2020; Gupta et al., 2021; Nalajala et al., 2023).

The highest inter-cluster distance between clusters XIII and VIII suggests potential for generating heterotic progenies, while the lowest distance between Clusters X and V designates redundancy. Cluster VIII exhibited superior mean values for main yield attributing traits, making it ideal for direct selection or use as donor parents. Conversely, low-yielding clusters such as XIII and III may still offer value for traits like early maturity or stress resilience, supporting their targeted utilization in specific breeding objectives (Abdel-Monaem et al., 2022; Somu et al., 2024; Mulugeta et al., 2024).

3.2 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was performed to assess the underlying structure of the multivariate data set and to identify the most influential traits contributing to genetic variability among the 60 soybean genotypes (Table 4). The analysis generated five principal components (PCs) with eigenvalues greater than 1.0, cumulatively accounting for 77.39% of the total variation observed across the measured traits, in accordance with the Kaiser criterion. The first principal component (PC1) explained the maximum variance (22.775%), tracked by PC2 (19.563%), PC3 (16.347%), PC4 (10.353%), and PC5 (8.347%) (Table 4). These results indicated that the first five components captured the majority of the total variability and were thus, considered for further interpretation and selection strategies (Tepavcevic et al., 2021; Souza et al., 2023).

The scree plot supported the selection of these five PCs, showing a sharp decline in eigenvalues after PC5, reflecting diminishing returns in explained variance beyond this point (Fig. 2; Fig. 3). The trait loadings and genotype scores associated with these components provided insights into trait combinations and genotype performance (Kahlon et al., 2018; Dunna et al., 2023).

PC1 was primarily associated with yield-related traits such as numbers of pods per plant, numbers of seeds per plant, seed yield per plant, 100-seed weight, and harvest index. Genotypes viz., JS-21-17, JS-22-01, Cat492A, NRC-142, and ASB-85 exhibited high positive scores along with PC1, demonstrating superior performance for these key traits and signifying their potential utility in high-yield breeding programmes. PC2 highlighted variation driven by traits including harvest index, numbers of seeds per pod, 100-seed weight, and plant height. Genotypes *namely*: AS-26, AUKS-21-5, NRC-255, and RVSM-2012-4 had the highest scores, denoting their strength in biomass partitioning and architectural traits. PC3 emphasized

Table 3. Cluster means value for thirteen characters in soybean

Cluster	Days to 50% flowering	Days to podding	Days to seed filling	Days to maturity	Plant height (cm)	Numbers of primary branches	Numbers of pods per plant	Numbers of seeds per pod	Numbers of seeds per plant	100-seed weight	Biological yield	Harvest index	Seed yield per plant
Cluster I	40.37	54.55	65.32	95.23	37.82	4.03	61.42	2.69	106.07	8.69	25.32	39.08	9.41
Cluster II	39.64	51.00	62.64	94.00	38.47	4.82	88.83	2.77	153.03	8.86	30.61	43.15	12.55
Cluster III	40.00	54.00	63.00	88.00	38.50	3.67	45.83	3.16	58.00	8.89	15.92	35.23	5.61
Cluster IV	41.07	51.57	61.71	97.29	37.07	3.40	42.69	2.71	79.40	9.31	20.46	45.38	8.77
Cluster V	42.00	48.00	63.00	109.00	29.33	5.66	73.17	2.84	105.50	11.07	26.48	43.67	11.56
Cluster VI	41.00	60.00	72.00	104.00	36.50	6.33	63.17	2.84	144.34	10.77	25.56	45.35	11.60
Cluster VII	36.00	47.50	66.50	90.00	43.17	5.33	91.17	3.33	185.66	12.35	11.94	62.52	7.45
Cluster VIII	40.00	51.00	62.00	96.00	38.17	4.16	95.17	2.84	222.66	12.43	31.03	63.03	19.56
Cluster IX	40.50	52.00	71.00	106.00	31.83	5.00	51.83	2.00	129.16	11.56	29.39	49.27	14.48
Cluster X	40.00	52.00	65.00	103.00	37.00	7.00	84.83	2.33	98.33	10.85	39.03	26.75	10.44
Cluster XI	41.00	53.00	75.00	107.50	34.50	6.33	91.50	2.16	71.50	10.69	26.57	28.56	7.59
Cluster XII	43.00	59.00	66.00	108.00	31.83	4.67	68.83	3.00	202.34	11.35	26.33	61.00	16.07
Cluster XIII	42.00	51.00	66.00	102.00	33.33	4.33	61.33	2.50	41.83	10.06	28.42	23.40	6.65
Cluster XIV	37.00	50.00	58.00	63.00	40.17	3.33	72.83	3.16	127.33	8.51	27.5	41.78	11.49

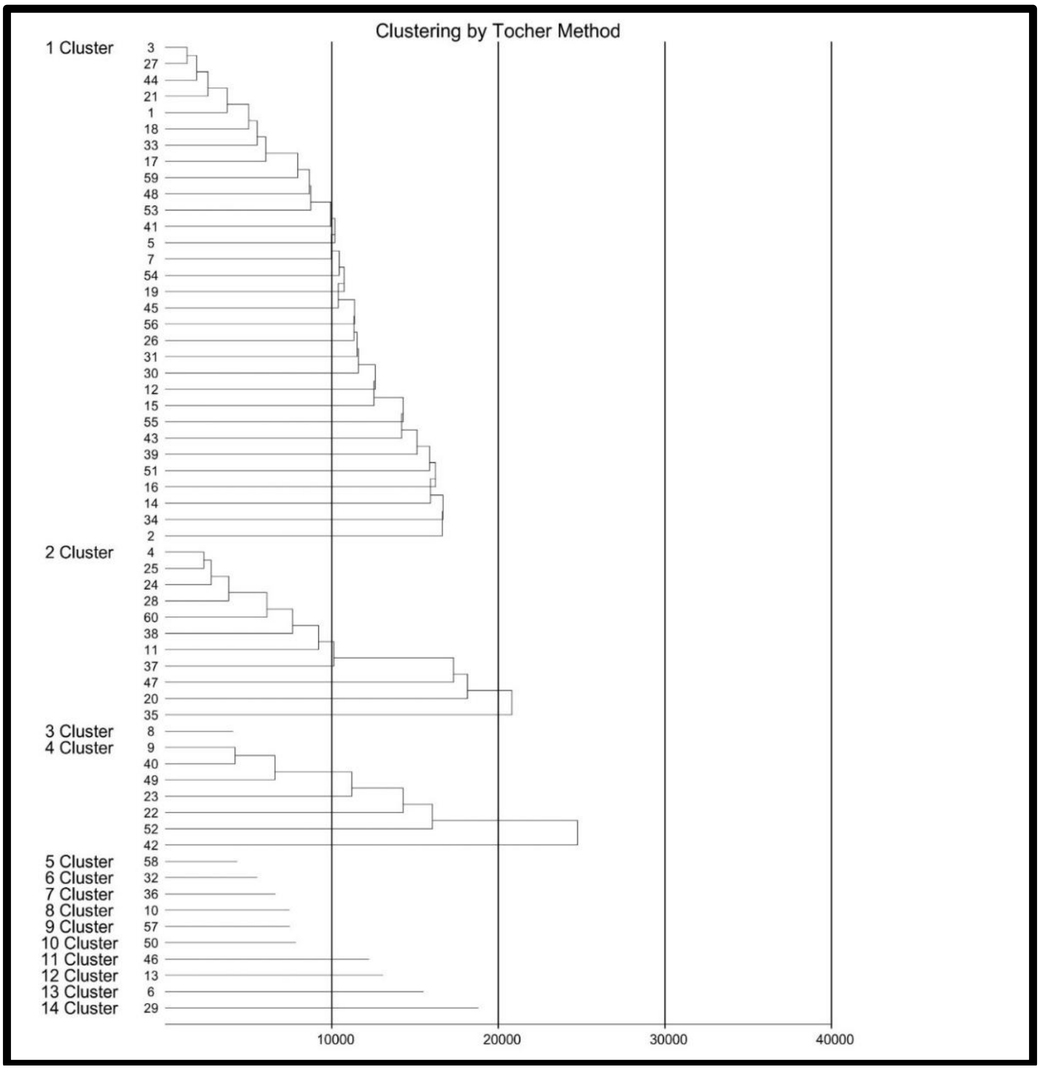

Fig. 1. Distribution of soybean genotypes in different clusters employing Tocher's method

Table 4. Principal components for yield contributing characters of soybean genotypes

Traits	Principal Components				
	PC1	**PC2**	**PC3**	**PC4**	**PC5**
Days to 50% flowering	-0.232	-0.383	-0.204	0.107	-0.263
Days to podding	-0.319	-0.311	-0.172	0.212	-0.227
Days to seed filling	-0.231	-0.425	-0.119	0.068	0.157
Days to maturity	-0.192	-0.419	-0.181	-0.254	0.173
Plant height (cm)	0.040	0.005	0.242	0.632	0.268
Numbers of primary branches	0.179	-0.315	-0.055	0.052	0.590
Numbers of pods per plant	0.431	-0.213	0.181	0.034	0.024
Numbers of seeds per pod	0.088	0.095	-0.335	0.575	0.074
Numbers of seeds per plant	0.451	-0.227	-0.129	0.185	-0.209
Hundred- seed weight (g)	0.224	0.034	-0.350	-0.303	0.449
Harvest index (%)	0.218	0.147	-0.558	0.006	-0.157
Biological yield (g/plant)	0.183	-0.348	0.465	-0.067	-0.094
Seed yield per plant (g/plant)	0.439	-0.228	-0.091	-0.102	-0.346

Table 5. Eigen values, percentage of total variation (%) and cumulative percentage (%) for different traits in soybean

Characters	Principal Components (PCs)	Eigen value	Variability (%)	Cumulative Proportion (%)
Days to 50% flowering	PC1	3.046	22.775	22.775
Days to podding	PC2	2.521	19.563	42.338
Days to seed filling	PC3	2.102	16.347	58.685
Days to maturity	PC4	1.384	10.353	69.038
Plant height (cm)	PC5	1.125	8.347	77.385
Numbers of primary branches	PC6	0.689	4.689	82.074
Numbers of pods per plant	PC7	0.667	4.442	86.516
Numbers of seeds per pod	PC8	0.528	3.962	90.478
Numbers of seeds per plant	PC9	0.494	3.131	93.609
Hundred- seed weight (g)	PC10	0.473	2.910	96.520
Harvest index (%)	PC11	0.362	1.999	98.519
Biological yield (g/plant)	PC12	0.283	1.305	99.825
Seed yield per plant (g/plant)	PC13	0.015	0.175	100.000

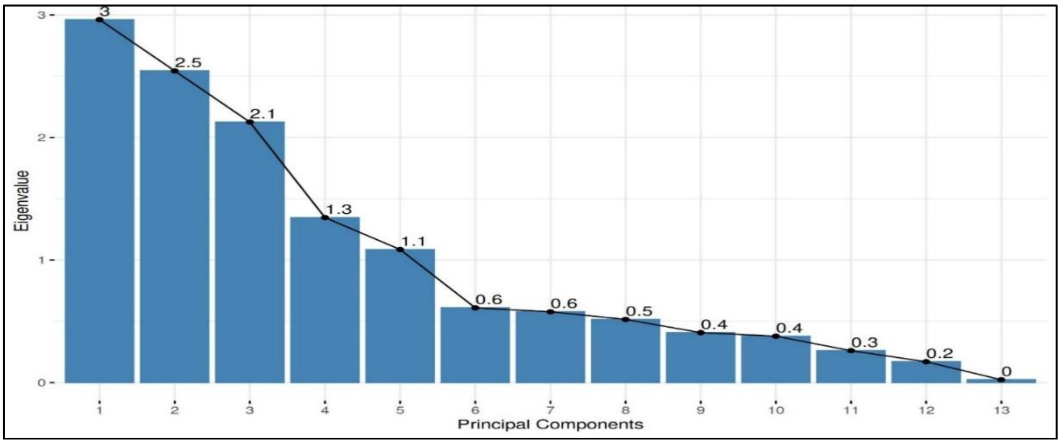

Fig. 2. Scree plot laid between eigenvalue and principal components

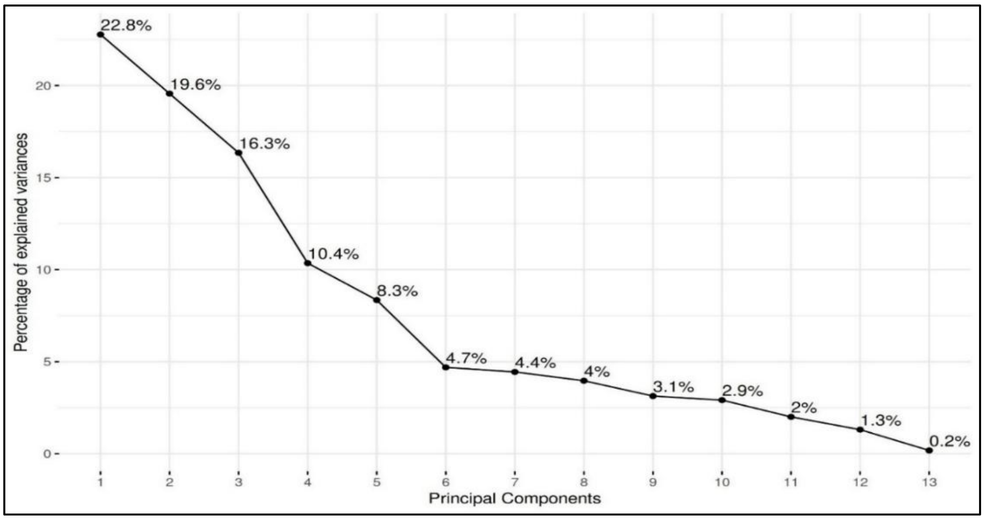

Fig. 3. Scree plot between variability (%) and principal components

contributions from biological yield, plant height, and numbers of pods per plant, with Cat 87, RVS-23-15, and NRC-152 genotypes showing superior scores. These genotypes may be preferred for biomass-focused breeding approaches. PC4 captured variation in traits for instance plant height, numbers of seeds per pod, days to podding, numbers of seeds per plant, and days to 50% flowering. Genotypes *i.e*., NRC-192, KBSL-23-36, BAUS(M)-6, and AMS-264 had the highest PC scores, indicating their adaptability and earliness in flowering and pod development. PC5 explained variation related to numbers of primary branches, 100-seed weight, plant height, days to maturity, and days to seed filling. Genotypes *viz*., NRCSL-4, Cat492A, KSS-213, and RVS-2001-4 were prominent in this component, making them suitable candidates for improving maturity and branching characters. Comparable investigations have also been conducted by Mannan et al. (2010), Uikey et al. (2018), Dunna et al. (2023) and Mishra et al. (2025b).

The PCA results facilitated the identification of superior genotypes based on multi-trait performance and highlighted specific trait combinations that can be prioritized in soybean breeding programmes to enhance yield, plant architecture, and maturity patterns (El-Hashash, 2016; Leite et al., 2018; Singh et al., 2020).

4. CONCLUSION

The present investigation exposed substantial genetic variability among the 60 soybean genotypes for all 13 quantitative traits investigated, as confirmed by highly significant differences in the analysis of variance. Cluster analysis grouped the genotypes into 14 distinct clusters, highlighting both inter- and intra-cluster genetic divergence. The presence of mono-genotypic clusters and wide inter-cluster D^2 distances signifies the existence of considerable genetic diversity, which can be strategically utilized in soybean improvement programmes. Clusters VIII and X were notable for exhibiting high cluster means for economically important traits such as numbers of pods per plant, seed yield, biological yield, and harvest index, indicating their potential as promising sources for selecting high-yielding genotypes. Principal Component Analysis further supported these findings by capturing approximately 77.39% of the total variation in the first five components. Genotypes *viz*., JS-21-17, Cat492A, NRC-142, and AUKS-21-5 consistently displayed high principal component scores across multiple PCs, signifying their contribution to traits of agronomic importance. These genotypes can serve as ideal parents in hybridization programmes aimed to combine desirable traits like early maturity, increased seed weight, and enhanced yield components. Overall, the integrated approach of multivariate analysis and PCA not only elucidated the genetic architecture of the studied genotypes but also provided a robust framework for selection and breeding. The findings emphasize the significance of exploiting diverse genetic backgrounds to enhance genetic gains and ensure sustainable soybean improvement under varying agro-climatic circumstances.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

Abdel-Monaem, M. A., Abido, W. A. E., Hadházy, Á., Ghoneima, M. H., EL-Mansy, Y. M., & EL-Shazly, M. W. (2022). Genetic divergence among Egyptian cotton genotypes under water deficit conditions. *Acta Ecologica Sinica*, *42*(2), 11–18. https://doi.org/10.1016/j.chnaes.2020.11.007

Abdul Aziz, M., & Masmoudi, K. (2025). Molecular breakthroughs in modern plant breeding techniques. *Horticultural Plant Journal*, *11*(1), 15–41. https://doi.org/10.1016/j.hpj.2024.01.004

Brivery, S. (2021). The significance of soybean production in the face of changing climates in Africa. *Cogent Food & Agriculture*, *7*(1). https://doi.org/10.1080/23311932.2021.1933745

Collins, N. C., Tardieu, F., & Tuberosa, R. (2008). Quantitative trait loci and crop performance under abiotic stress: Where do we stand? *Plant Physiology*, *147*(2), 469–486. https://doi.org/10.1104/pp.108.118117

Dharshini, M. S., Macwana, S. S., Parmar D. J., & Patil, K. (2024). Multi environment analysis of soybean genotypes to delineate stability and adaptability for yield and quality parameters. *Journal of Scientific Research and Reports, 30(5),* 259–275. https://doi.org/10.9734/jsrr/2024/v30i51941

Dunna, D. S., Heisnam, N. D., Thokchom, R. D., Sinha, B., & Singh, O. (2023). Principal component analysis among vegetable soybean genotypes (*Glycine max* L. Merrill). *Environment Conservation Journal*, *24*(3), 79–86. https://doi.org/10.36953/ECJ.14532442

El-Hashash, E. (2016). Genetic diversity of soybean yield based on cluster and principal component analyses. *Journal of Advances in Biology & Biotechnology*, *10*(3), 1–9. https://doi.org/10.9734/JABB/2016/29127

Ghuge, V R., Surnar, D V., Jadhav, R S., Randhave, B S., & Thakur N R. (2023). Genetic diversity analysis in soybean (*Glycine max* L.). *Journal of Oilseeds Research*, *40* (Special issue). https://doi.org/10.56739/jor.v40iSpecialissue.145778

Gupta, D., Muralia, S., Khandelwal, V., & Nehra, A. (2021). Assessing diversity of sesame genotypes using cluster analysis and principal component analysis. *International Journal of Current Microbiology and Applied Sciences*, *10*(01), 304–312. https://doi.org/10.20546/ijcmas.2021.1001.038

Hamza, M., Basit, A. W., Shehzadi, I., Tufail, U., Hassan, A., Hussain, T., Siddique, M. U., & Hayat, H. M. (2024). Global impact of soybean production: A review. *Asian Journal of Biochemistry, Genetics and Molecular Biology*, *16*(2), 12–20. https://doi.org/10.9734/ajbgmb/2024/v16i2357

Jolliffe, I. T. (1986). Principal component analysis. (2nd ed.). Springer Verlag.

Kachhadia, V.H., Baraskar, V. V., Vachhani, J. H., Patel, M. B. & Barad, H. R. (2014). Genetic divergence in soybean [*Glycine max* L. Merrill.]. *Electronic Journal of Plant Breeding*, 5(3), 563-566. Retrieved from https://www.ejplantbreeding.org/index.php/EJPB/article/view/53

Kahlon, C. S., Li, B., Board, J., Dia, M., Sharma, P., & Jat, P. (2018). Cluster and principal component analysis of soybean grown at various row spacings, planting dates and plant populations. *Open Agriculture*, *3*(1), 110–121. https://doi.org/10.1515/opag-2018-0011

Kanavi, M. S. P., Rangaiah, S., & Anusha, C. R. (2019). Genetic diversity analysis through D^2 statistic for quantitative traits in germplasm lines of green gram [*Vigna radiata* (L.) Wilczek]. *International Journal of Current Microbiology and Applied Sciences*, *8*(06), 847–855. https://doi.org/10.20546/ijcmas.2019.806.102

Leite, W. de S., Unêda-Trevisoli, S. H., Silva, F. M. da, Silva, A. J. da, & Mauro, A. O. di. (2018). Identification of superior genotypes and soybean traits by multivariate analysis and selection index. *Revista Ciência Agronômica*, *49*(3). https://doi.org/10.5935/1806-6690.20180056

Mahmud, K., Makaju, S., Ibrahim, R., & Missaoui, A. (2020). Current progress in nitrogen fixing plants and microbiome research. *Plants*, *9*(1), 97. https://doi.org/10.3390/plants9010097

Mahalanobis, P.C. (1936). On the generalized distance in statistics. *Proc Natl Inst Sci India*. *2(1)*, 49-55.

Majidian, P., Ghorbani, H. R., & Farajpour, M. (2024). Achieving agricultural sustainability through soybean production in Iran: Potential and challenges. *Heliyon*, *10*(4), e26389. https://doi.org/10.1016/j.heliyon.2024.e26389

Mannan, M. A., Karim, M. A., Khaliq, Q. A., Haque, M. M., Mian, M. A. K., & Ahmed, J. U. (2010). Assessment of genetic divergence in salt tolerance of soybean (*Glycine max* L.) genotypes. *Journal of Crop Science and Biotechnology*, *13*(1), 33–37. https://doi.org/10.1007/s12892-009-0091-y

Mark, H., & Workman, J. (2018). The statistics of spectral searches. In *Chemometrics in Spectroscopy* (pp. 507–511). Elsevier. https://doi.org/10.1016/B978-0-12-805309-6.00075-1

Massey, W. F. (1965). Principal components regression in exploratory statistical research. *Journal of the American Statistical Association, 60*, 234–246.

Mishra N, Tripathi MK, Tiwari S, Tripathi N, Trivedi HK. (2020). Morphological and molecular screening of soybean genotypes against yellow mosaic virus disease. *Legume Research*. https://doi.org/10.18805/LR-4240

Mishra, N., Tripathi, M.K., Tiwari, S., Tripathi, N., Sapre, S. & Ahuja, A. *et al*. (2021a). Cell suspension culture and *in vitro* screening for drought tolerance in soybean using poly-ethylene glycol. *Plants*.; 10;10(3):517. https://doi.org/10.3390/plants10030517

Mishra, N., Tripathi, M.K., Tiwari, S., Tripathi, N., Gupta, N., Sharma, A., & Solanki1 R.S. (2021b). Evaluation of diversity among soybean genotypes *via* yield attributing traits and SSR molecular markers. *Current Journal of Applied Science and Technology*, *40*, 21: 9-24.

Mishra, N., Tripathi, M.K., Tripathi, N., Tiwari, S., Gupta, N., & Sharma, A. (2021c). Validation of drought tolerance gene-linked microsatellite markers and their efficiency for diversity assessment in a set of soybean genotypes. *Current Journal of Applied Science and Technology*. *30*,48–57. https://doi.org/10.9734/cjast/2021/v40i2531515

Mishra N, Tripathi MK, Tripathi N, Tiwari S, Gupta N, Sharma A, *et al*. (2021d). Changes in biochemical and antioxidant enzymes activities play significant role in drought tolerance in soybean. *Int. J Agric. Technol. 17*,1425-46. 41.

Mishra, N., Tripathi, M.K., Tiwari, S., Tripathi, N., Gupta, N., & Sharma, A. (2021e). Morphological and physiological performance of Indian soybean [*Glycine max* (L.) Merrill] genotypes in respect to drought. *Legume Res Int. J*. LR-4550.

Mishra, R., Karada, M. S., Agnihotri, D., & Yadav, N. K. (2024f). Advances in molecular genetics: Techniques and applications. In A. Tigga, M. Shahani, K. Modunshim, Longkho K, & Kumari S (Eds.), *Advances in Genetics and Plant Breeding* (pp. 44–65). Stella International Publication.

Mishra, R., Shrivastava, M.K., Amrate, P.K., Sharma, S., Singh, Y., Tripathi, & M.K. (2025a). Phenotypic diversity and trait analysis of soybean recombinant inbred lines. *Plant Cell Biotechnology and Molecular Biology*. May 17;26(7–8):32–52. https://doi.org/10.56557/pcbmb/2025/v26i7-89345

Mishra, R., Shrivastava, M.K., Tripathi, M.K., Amrate, P.K., Singh, Y. & Solanki R. *et al*. (2025b). Unravelling soybean yield potential: Exploring trait synergy, impact pathways, multidimensional patterns and biochemical insights. *Plant Science Today*. https://doi.org/10.14719/pst.6401

Mishra, R., Tripathi, M. K., Sikarwar, R. S., Singh, Y., & Tripathi, N. (2024a). Soybean (*Glycine max* L. Merrill): A multipurpose legume shaping our world. *Plant Cell Biotechnology and Molecular Biology*, *25*(3–4), 17–37. https://doi.org/10.56557/pcbmb/2024/v25i3-48643

Mishra, R., Tripathi, M., Tripathi, N., Singh, J. & Tiwari S. (2024b). Nutritional and anti-nutritional factors in soybean. *Acta Scientific Agriculture*. 3: 8(11):46–63. https://doi.org/10.31080/ASAG.2024.08.1432

Mishra, R., Tripathi, M.K., Tripathi, N., Singh, J., Yadav, P.K. & Sikarwar, R.S. *et al.* (2024c). Breeding for major genes against drought stress in soybean. In: Tripathi MK, Tripathi N, editors. *Advances in Plant Biotechnology*. Cornous Publications LLP, Puducherry, India. p. 22–68. https://doi.org/https://doi.org/10.37446/volbook032024/22-68

Mishra, R., Tripathi, M. K., Shrivastava, M. K., Amrate, P. K., Singh, J., & Singh, Y. (2024d). From conventional to modern plant breeding: How far have we come? In M. K. Tripathi & R. Mishra (Eds.), *Recent Advances in Plant Breeding* (Vol. 1, pp. 1–20). Cornous Publications LLP. https://doi.org/10.37446/volbook102024/1-20

Mishra, R., Tripathi, M.K., Shrivastava, M.K., & Amrate, P.K. (2024e). Genetic diversity in crop improvement: A cornerstone for sustainable agriculture and global food security. In: Tripathi MK, Tripathi N, editors. Advances in Plant Biotechnology. Cornous Publications LLP; p. 1–21. https://doi.org/https://doi.org/10.37446/volbook032024/1-21

Mulugeta, T., Abate, A., Tadesse, W., Bezabih Woldeyohannes, A., Tefera, N., Shiferaw, W., & Tiruneh, A. (2024). Multivariate analysis of phenotypic diversity elite bread wheat (*Triticum aestivum* L.) genotypes from ICARDA in Ethiopia. *Heliyon*, *10*(16), e36062. https://doi.org/10.1016/j.heliyon.2024.e36062

Nakei, M. D., Venkataramana, P. B., & Ndakidemi, P. A. (2022). Soybean-nodulating rhizobia: Ecology, characterization, diversity, and growth promoting functions. *Frontiers in Sustainable Food Systems*, *6*. https://doi.org/10.3389/fsufs.2022.824444

Nalajala, S., Singh, N. B., Jeberson, M. S., Yumnam, S., & Sinha, B. (2023). Genetic divergence studies for yield and its component traits in Mung bean (*Vigna radiata* L. Wilczek). *Environment Conservation Journal*, *24*(3), 14–20. https://doi.org/10.36953/ECJ.14422436

Paikra, N., Karishma, & Nag, S. (2025). Assessment of genetic variation in soybean (*Glycine max* (L.) Merrill) using Mahalanobis D^2 analysis. *International Journal of Advanced Biochemistry Research*, *9*(1S), 981–984. https://doi.org/10.33545/26174693.2025.v9.i1Sm.3652

Raina, D., Dhillon, W. S., Gill, P. P. S., & Singh, N. P. (2015). Assessment of genetic divergence using Mahalanobis D^2 and principal component analysis of qualitative and quantitative characters in pomegranate genotypes under sub-tropics. *Indian Journal of Horticulture*, *72*(4), 451. https://doi.org/10.5958/0974-0112.2016.00001.3

Rao, C.R. (1952). Advanced statistical methods in biometrical research. *New York: Wiley & Sons*.

Roessner, U., Nahid, A., Chapman, B., Hunter, A., & Bellgard, M. (2011). Metabolomics – The combination of analytical biochemistry, biology, and informatics. In *Comprehensive Biotechnology* (pp. 435–447). Elsevier. https://doi.org/10.1016/B978-0-444-64046-8.00027-6

Roy, K., Haque, K. N., Samanta, K., Acharya, R., Kanthal, S., Kundu, S., Sarkar, T., & Sengupta, S. (2024). Biological nitrogen fixation: Reducing the N footprints of the environment. *International Journal of Advanced Biochemistry Research*, *8*(4S), 133–137. https://doi.org/10.33545/26174693.2024.v8.i4Sb.936

Salgotra, R. K., & Chauhan, B. S. (2023). Genetic Diversity, Conservation, and utilization of plant genetic resources. *Genes*, *14*(1), 174. https://doi.org/10.3390/genes14010174

Sandhu, K. S., Shiv, A., Kaur, G., Meena, M. R., Raja, A. K., Vengavasi, K., Mall, A. K., Kumar, S., Singh, P. K., Singh, J., Hemaprabha, G., Pathak, A. D., Krishnappa, G., & Kumar, S. (2022). Integrated approach in genomic selection to accelerate genetic gain in sugarcane. *Plants*, *11*(16), 2139. https://doi.org/10.3390/plants11162139

Sharma, S., Tripathi, M.K., Tiwari, S., Solanki, R.S., Chauhan, S., Tripathi, N., Dwivedi, N. & Tiwari P. N. (2023). Discriminant function analysis for yield improvement in bread wheat (Triticum aestivum L.). *The Pharma Innovation Journal*, 12 (5),224-232.

Sharma, A., Tripathi, M. K., Tiwari, S., Gupta, N., Tripathi, N., & Mishra, N. (2021). Evaluation of soybean (*Glycine max* L.) genotypes on the basis of biochemical contents and anti-oxidant enzyme activities. *Legume Research - An International Journal*. https://doi.org/10.18805/LR-4678

Sharma, A., Mishra, N., Tripathi, N., Nehra, S., Singh, J., & Tiwari S. *et al*. (2023). Qualitative trait based variability among soybean genotypes. *Acta Scientific Agriculture*. 1;02–13. https://doi.org/10.31080/ASAG.2023.07.1212

Sharma, S., Shukla, S., Mishra, R., Sanu, K., Mishra, G., Paliwal, S., Chauhan, S., & Mishra, N. (2025). Marker assisted selection - a crucial technique to facilitate modern plant breeding. In M. K. Tripathi & N. Tripathi (Eds.), *Advances in Plant Biotechnology* (Vol. 1, pp. 188–207). Cornous Publications LLP. https://doi.org/https://doi.org/10.37446/volbook032024/188-207

Singer, S. D., Laurie, J. D., Bilichak, A., Kumar, S., & Singh, J. (2021). Genetic variation and unintended risk in the context of old and new breeding techniques. *Critical Reviews in Plant Sciences*, *40*(1), 68–108. https://doi.org/10.1080/07352689.2021.1883826

Singh, P. K, Shrestha, J., & Kushwaha, UKS. (2020). Multivariate analysis of soybean genotypes. *Journal of Agriculture and Natural Resources*, 3(1), 69-76. https://doi.org/10.3126/janr.v3i1.27092

Sirisha, A. B. M., Haseena Banu, S. K., & Saritha, R. (2020). Genetic Divergence Studies using Mahalanobis D square analysis in sesame (*Sesamum indicum* L.) germplasm. *International Journal of Current Microbiology and Applied Sciences*, *9*(8), 2224–2229. https://doi.org/10.20546/ijcmas.2020.908.255

Somu, G., Meena, N., & Badigannavar, A. (2024). Exploring genetic variability for morphological and yield contributing traits in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Southern India. *Genetic Resources and Crop Evolution*.

https://doi.org/10.1007/s10722-024-02251-5

Souza, R. R. de, Cargnelutti Filho, A., Toebe, M., & Bittencourt, K. C. (2023). Sample size and genetic divergence: a principal component analysis for soybean traits. *European Journal of Agronomy*, *149*, 126903. https://doi.org/10.1016/j.eja.2023.126903

Tepavčević, V., Cvejić, J., Poša, M., Bjelica, A., Miladinović, J., Rizou, M., Aldawoud, T. M. S., & Galanakis, C. M. (2021). Classification and discrimination of soybean (*Glycine max* (L.) Merr.) genotypes based on their isoflavone content. *Journal of Food Composition and Analysis*, *95*, 103670. https://doi.org/10.1016/j.jfca.2020.103670

Tripathi, M.K., Tripathi, N., Tiwari, S., Mishra, N., Sharma, A., & Tiwari, S, *et al*. (2023). Identification of Indian soybean (*Glycine max* [L.] Merr.) genotypes for drought tolerance and genetic diversity analysis using SSR markers. *Scientist*, *3(3)*, 31-46.

Uikey, S., Sharma, S., Shrivastava, M. K., & Amrate, P. K. (2021). Study of principal component analyses for pod traits in soybean. *International Journal of Agricultural Sciences*, *17*(2), 341–349. https://doi.org/10.15740/HAS/IJAS/17.2/341-349

Upadhyay, S., Singh, A.K., Tripathi, M.K., Tiwari, S., & Tripathi, N. (2020). Validation of simple sequence repeats markers for charcoal rot and Rhizoctonia root rot resistance in soybean genotypes. *I.J.A.B.R. 10(2),*137-144.

Wang, F. (2009). Factor Analysis and Principal-Components Analysis. In *International Encyclopedia of Human Geography* (pp. 1–7). Elsevier. https://doi.org/10.1016/B978-008044910-4.00434-X

Yadav, R. K., Tripathi, M. K., Tiwari, S., Asati, R., Chauhan, S., Sikarwar, R.S., & Yasin, M. (2023). Evaluation of genetic diversity through D^2 Statistic in chickpea (Cicer arietinum L.). *Int. J. Environ. Clim. Change*, 13 (10), 1598-1611.

Peer-review history:
The peer review history for this paper can be accessed here:
https://pr.sdiarticle5.com/review-history/138281