International Journal of Agriculture and Plant Science
www.agriculturejournal.in
Online ISSN: 2664-7664
Received: 06-01-2026; Accepted: 05-02-2026; Published: 20-02-2026
Volume 8, Issue 1, 2026, Page No. 139-143

Multivariate analysis of genetic divergence in soybean genotypes

Sharmila Kumari[1], B L Meena[2], Karan Sachdeva[3]
[1] Genetics and Plant Breeding, College of Agriculture, Jodhpur, Agriculture University, Jodhpur, Rajasthan, India
[2] Associate Professor, Genetics and Plant Breeding, College of Agriculture, Hindoli, Bundi, Agriculture University, Kota, Rajasthan, India
[3] Genetics and Plant Breeding, SKN College of Agriculture, SKNAU, Jobner, Rajasthan, India

Abstract
Soybean is an important oilseed and protein-rich crop. Assessing genetic diversity is critical for choosing appropriate parents to enhance seed yield as well as better agronomic and quality parameters. In this context, the research was carried out during *Kharif* 2021 under rainfed conditions at the Agricultural Research Station, Ummedganj, Agriculture University, Kota, Rajasthan. Thirty-two soybean genotypes were evaluated in a Randomized Block Design with three replications to study genetic divergence using Mahalanobis D^2 statistics and principal component analysis. D^2 analysis grouped the genotypes into seven clusters, with Cluster II being the largest, followed by Cluster I. Cluster I had superior mean performance for yield and related traits, including number of pods per plant, biological yield and seed yield per plant. Cluster IV recorded the highest protein content, while Cluster V exhibited maximum oil content, indicating their importance for quality improvement. The greatest inter-cluster distances were found between Cluster II and Cluster VII, followed by Cluster II with Clusters V and VI, indicating significant genetic divergence and potential for heterosis via hybridization. The top three principle components accounted for 80.66% of the overall variation and plant height, days to 50% flowering, days to maturity, protein content, biological yield per plant, seed yield per plant, number of branches per plant and number of pods per plant were the most important contributors to genetic variability in the PCA analysis. Overall, the result revealed presence of significant level of genetic diversity among soybean genotypes. Parents chosen from extremely diversified clusters can be employed in breeding programs to produce superior, higher-yielding cultivars with desirable agronomic attributes.

Keywords: Soybean genetic diversity, mahalanobis d analysis, principal component analysis (pca)

Introduction

Soybean (*Glycine max* L. Merrill) is regarded as one of the most significant legume crops globally due to its high nutritional value and versatility in food, feed and commercial applications. It is a main source of edible oil and plant protein, making a considerable contribution to human diet and animal feed. The seed encompasses nearby 37-42% protein and 17-24% oil, as well as vital amino acids, unsaturated fatty acids and vitamins (Balasubramaniyan and Palaniappan, 2003) [2]. Soybean is commonly referred to as a "golden bean" because of its richness in nutritious aspect and contribution in boosting soil fertility by biological nitrogen fixation. (Kumar *et al*., 2015) [12]. In 2023-24, soybean was grown in India on about 13.25 million hectares, producing nearly 13.06 million tonnes with an average yield of 985 kg per hectare. In Rajasthan, the crop occupied around 11.26 million hectares, with a production of 11.70 million tonnes and an average productivity of 1039 kg per hectare (Directorate of Oilseeds Development, 2023-24) [4].

However, its productivity is frequently limited in India due to restricted genetic basis, selfpollinated nature and vulnerability to numerous biotic and abiotic challenges (Soni *et al*. 2025) [26]. These problems highlight the importance of exploring and utilizing the genetic heterogeneity present among soybean genotypes in order to develop superior and stable cultivars. Multivariate statistical techniques for assessing genetic diversity have proven to be a useful approach in crop development studies. Principal component analysis and cluster analysis are common statistical approaches for studying and quantifying genetic diversity. PCA is very beneficial for simplifying large datasets by finding essential characteristics that contribute the most to total variability. It reduces the dimensionality of data and highlights the major components responsible for variation among genotypes, thereby helping researchers focus on the most influential characters (Vianna *et al*. 2013) [28]. Cluster analysis groups genotypes into different clusters as per their similarities and differences, providing a clear understanding of genetic relationships and divergence. Genotypes placed in distant clusters are generally more diverse and can serve as valuable parents in breeding programmes, as their crossing may generate greater variability in the following generations. Together, these methods provide a clear understanding of genetic relationships and assist in identifying diverse genotypes that can serve as suitable parents in breeding programmes (Sivabharathi *et al*. 2023) [24]. Hence, the present study was conducted to assess the extent of genetic diversity among thirty-two soybean genotypes through multivariate analysis. The study aimed to estimate the extent of variability, identify key traits contributing to divergence and classify genotypes into groups to support effective parent selection for future soybean improvement.

Materials and methods

A field experiment was conducted under rainfed conditions at the Research Farm of the Agricultural Research Station, Ummedganj, Agriculture University, Kota, Rajasthan. Thirty-two diverse soybean genotypes were evaluated in a

Randomized Block Design with three replications to study genetic diversity using Mahalanobis D² statistics and Principal Component Analysis (PCA). Each genotype was sown in three rows of 3 m length, with a spacing of 45 cm between rows and 10 cm between plants. All recommended agronomic practices were followed to ensure proper crop growth and development. Observations were recorded on days to 50% flowering, days to maturity, plant height, number of primary branches per plant, number of pods per plant, number of seeds per pod, 100-seed weight, biological yield per plant, seed yield per plant, harvest index, protein and oil content. Mean values of the recorded data were used for statistical analysis. Genetic divergence among the genotypes was estimated using Mahalanobis D² (Mahalanobis, 1936) [13] and based on the D² values, the genotypes were grouped into clusters following Tocher's method (Rao, 1952) [22]. Furthermore, Principal Component Analysis (Massey (1965) and Jolliffe (1986) [11, 16] was applied to identify the major traits contributing to genetic variability among the genotypes. The PCA and Mahalanobis D² analyses were performed using INDOSTAT software to assess the extent of genetic divergence.

Result and discussion

Mahalanobis D² analysis

Based on Mahalanobis D² distances, the thirty-two soybean genotypes were grouped into seven distinct clusters using Tocher's method and a dendrogram illustrating the cluster pattern was constructed (Table 1; Fig 1). The analysis showed considerable variation among the genotypes for the traits under study. Cluster II was the largest, comprising sixteen genotypes and accounting for 50.00% of the total genotypes, followed by Cluster I with eleven genotypes representing 34.38%. Clusters III to VII each contained one genotype and contributed 3.13% of the total genotypes (RVS 2011-10, AUKS 261, JS 95-60, JS 20-34 and AUKS 258), showing their high divergence. Comparable clustering patterns have been reported in earlier studies, where twenty-eight genotypes were grouped into five clusters (Mahbub *et al*., 2016) [14], eighty genotypes into six clusters (Al-Hadi *et al*., 2017) [1] and sixty genotypes into fourteen clusters (Gautam *et al*., 2025) [5]. Differences in the number of clusters may be attributed to variation in genetic background and the traits included in the analysis.

Table 1: Cluster-wise distribution of soybean genotypes into seven groups

Cluster	Total number of genotypes	Name of the genotypes
1	11	AUKS 255, AUKS 259, AUKS 260, AUKS 262, AUKS 263, AUKS 264, AUKS-207, AUKS-212, AUKS-213, AUKS-224 and AUKS-238
2	16	AUKS 245, AUKS 246, AUKS 247, AUKS 248, AUKS 249, AUKS 250, AUKS 251, AUKS 252, AUKS 253, AUKS 254, AUKS 256, AUKS 257, RKS 18, AUKS-229, RSC 10-52 and MACS 1710
3	1	RVS 2011-10
4	1	AUKS 261
5	1	JS 95-60
6	1	JS 20-34
7	1	AUKS 258

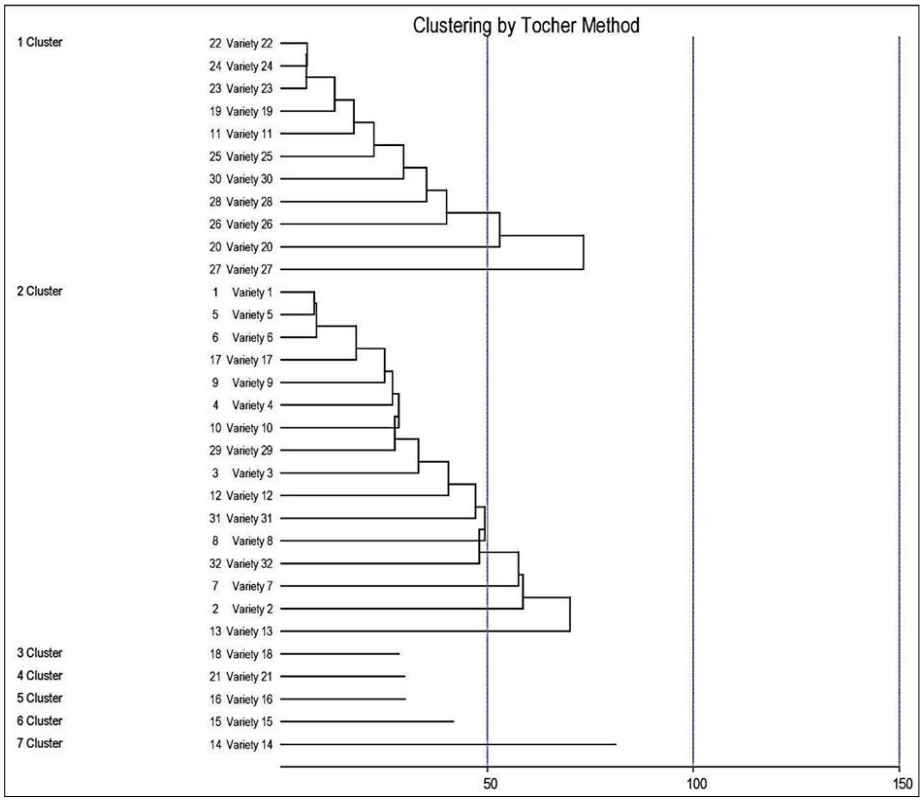

Fig 1: Distribution pattern of soybean genotypes into different clusters based on Tocher's method

The relative contribution of thirteen traits to total genetic divergence is displayed in Table 2. Plant height (27.82%) had the maximum contribution, followed by number of pods per plant (22.18%) and harvest index (20.77%). Moderate contributions were recorded from 100-seed weight and biological yield per plant, while the remaining traits had comparatively low influence. Similar findings were also reported by Mahbub *et al*. (2016) and Gautam *et al*. (2025) [5, 14].

Table 2: The individual and cumulative percentage contribution of different traits towards genetic divergence

S. No.	Traits	% Contribution	Cumulative % of variation
1	Days to 50% flowering	0.40	0.40
2	Days to maturity	2.42	2.82
3	Plant height (cm)	27.82	30.64
4	Pod length (cm)	3.83	34.47
5	No. of branches per plant	0.01	34.48
6	No. of pods per plant	22.18	56.66
7	No. of seeds per pod	0.40	57.06
8	100-seed weight (g)	9.88	66.94
9	Biological yield per plant (g)	6.35	73.29
10	Harvest index (%)	20.77	94.06
11	Protein content (%)	1.81	95.87
12	Oil content (%)	4.03	99.90
13	Seed yield per plant (g)	0.10	100.00

The intra-cluster mean values of all the studied traits, are displayed in Table 3, revealed substantial variation among the seven clusters across all the characters, indicating a broad range of diversity among the genotypes. Cluster II recorded the maximum mean values for days to flowering (44.08), followed by days to maturity (101.77) and plant height (35.89), indicating that the genotypes in this cluster were comparatively taller and took more time to mature. Cluster VII showed superior performance for number of branches per plant (3.51), 100-seed weight (17.9), pod length (4.71) and seed yield per plant (8.52), highlighting its importance for improving yield potential. On the other hand, Cluster I showed the highest values for number of pods per plant (59.94), followed by biological yield per plant (19.47) and seed yield per plant (9.56), making this group particularly promising for enhancing overall productivity. Cluster IV had the highest protein content (41.12), while Cluster V showed the highest oil content (18.2), suggesting suitability for protein and oil-related traits, respectively. Cluster III exhibited the maximum values for number of seeds per pod (3.0) with moderate values for other yield components, whereas Cluster VI showed the highest harvest index (50.07) along with average performance for most of the yield attributes. Overall, Cluster I and Cluster VII performed better for seed yield and its major contributing traits, while Clusters IV and V were notable for quality parameters. Therefore, genotypes from these clusters may serve as suitable parents in hybridization programmes aimed at combining high yield with improved quality traits.

Table 3: Cluster-wise mean values for thirteen traits in soybean genotypes

	Days to 50% flowering	Days to maturity	Plant height (cm)	Pod length (cm)	No. of branches per plant	No. of pods per plant	No. of seeds per pod	100-seed weight (g)	Biological yield per plant (g)	Harvest index (%)	Protein content (%)	Oil content (%)	Seed yield per plant (g)
Cluster I	38.85	96.06	27.47	3.78	2.97	59.94	2.50	12.24	19.47	49.10	39.54	17.80	9.56
Cluster II	44.08	101.77	35.89	3.30	2.80	57.51	2.46	9.78	15.42	36.00	39.18	16.73	5.35
Cluster III	38.00	93.00	28.98	4.12	2.13	28.60	3.00	12.57	13.05	49.45	39.82	16.90	6.46
Cluster IV	37.67	89.33	30.08	3.65	2.94	23.54	2.50	12.63	11.51	49.92	41.12	14.73	5.68
Cluster V	34.33	87.00	30.15	4.15	2.76	30.54	2.50	13.83	8.96	45.20	39.87	18.20	4.08
Cluster VI	34.33	87.00	30.15	4.15	2.76	30.54	2.50	11.67	13.01	50.07	38.13	17.90	6.52
Cluster VII	35.00	99.00	24.33	4.71	3.51	40.47	2.50	17.90	18.61	46.05	40.48	16.63	8.52

Table 4: Estimates of inter- and intra-cluster distances based on D^2 values

	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V	Cluster VI	Cluster VII
Cluster I	45.21						
Cluster II	175.02	50.52					
Cluster III	91.55	236.69	0				
Cluster IV	128.71	231.44	29.92	0			
Cluster V	118.69	266.86	30.07	39.25	0		
Cluster VI	93.65	319.06	52.45	75.65	41.7	0	
Cluster VII	144.48	424.83	124.84	189.69	132.77	144.05	0

The inter- and intra-cluster distances among the genotypes are presented in Table 4. The intracluster distance was highest in Cluster II (50.52), followed by Cluster I (45.21), indicating greater variability among the genotypes within these groups. In contrast, the monogenotypic clusters showed zero intra-cluster distance, as each contained only one genotype. Such clusters may possess rare traits and can be useful in hybridization programmes to enhance genetic diversity. Similar results have also been reported by Singh *et al*. (2020), Mounika *et al*. (2022), Ghuge *et al*. (2023),

Nalajala *et al*. (2023) and Paikra *et al*. (2025) [6, 18, 19, 20, 23]. Inter-cluster distances were generally higher than intra-cluster distances, reflecting a considerable level of genetic divergence among the clusters. The greatest inter-cluster distance was recorded between Cluster II and Cluster VII (424.83), followed by Cluster II and Cluster VI (319.06) and Cluster II and Cluster V (266.86). This highlight that genotypes in these clusters are highly divergent, suitable for hybridization to exploits heterosis, as also reported by Mazur (2023), Ragade *et al*. (2024) and Jhariya *et al*. (2025) [10, 17, 21]. In contrast, the smallest inter-cluster distance was recorded between Cluster III and Cluster IV (29.92), followed by Cluster III and Cluster V (30.07), indicating close genetic similarity among these groups. Overall, Cluster II showed comparatively high divergence than remaining clusters, while Clusters III, IV and V were comparatively closer to each other.

Principal component analysis (PCA)

Principal component analysis was performed to assess the contribution of different traits to the total genetic variability (Table 5). The results showed that the first three principal components together accounted for 80.66% of the total variation, suggesting that most of the diversity present in the material was effectively represented by these components. Among them, PC1 contributed the largest portion of variability (53.69%), followed by PC2 with 19.58% and PC3 with 7.40%. Comparable results have been documented in earlier studies. Shrivastava *et al*. (2025) and Mahbub *et al*. (2016) [14, 26] reported that the first three principal components explained 83.28% and 83.23% of the total variation, respectively. Sivabharathi *et al*. (2025) [25] observed 79.77% cumulative variation through four components, while Iqbal *et al*. (2008) [8] reported 69.77% variation explained by the first three components. These reports support the present findings, indicating that a limited number of principal components can effectively capture most of the genetic variability in soybean.

In PC1, plant height, days to flowering, days to maturity, protein content and seed yield per plant showed high positive loadings, indicating their strong role in total variability. PC2 was mainly influenced by biological yield per plant, number of branches per plant, number of pods per plant and seed yield per plant. PC3 was largely dominated by oil content with a very high positive loading, suggesting its importance in differentiating genotypes. These results indicate that both yield-related and quality traits were important in explaining the observed diversity and should be considered during genotype selection for crop improvement. The high variability explained by PC1 suggests its strong association with key morphological and yield traits, whereas the gradual decline in variation across subsequent components reflects their lower contribution to total diversity, as also reported in earlier studies Jain *et al*. (2020), Gupta *et al*. (2021), de Souza *et al*., (2023) and Yadav *et al*. (2023) [3, 7, 9, 29].

Table 5: Principal components showing eigenvalues, % variance and factor loadings of yield contributing traits

	PC 1	**PC 2**	**PC 3**
Eigen values	6.98	2.55	0.96
Percent variation (%)	53.69	19.58	7.40
Cumulative proportion (%)	53.69	73.27	80.66
Characters	PC 1	PC 2	PC 3
Days to 50% flowering	0.35	0.10	0.08
Days to maturity	0.30	0.24	-0.30
Plant height (cm)	0.35	-0.06	0.02
Pod length (cm)	-0.34	0.10	-0.12
No. of branches per plant	-0.07	0.45	-0.11
No. of pods per plant	0.22	0.39	0.13
No. of seeds per pod	-0.29	0.19	0.04
100-seed weight (g)	-0.33	0.19	-0.12
Biological yield per plant (g)	-0.15	0.52	-0.16
Harvest index (%)	-0.33	0.01	0.24
Protein content (%)	0.33	-0.07	-0.08
Oil content (%)	0.08	0.26	0.85
Seed yield per plant (g)	0.25	0.38	-0.14

Conclusion

The evaluation of thirty-two soybean genotypes under rainfed conditions showed a high level of genetic variability for thirteen agronomic and quality traits, as revealed through D^2 and PCA analyses. The genotypes were grouped into seven distinct clusters and wide inter-cluster distances indicated a broad genetic base and considerable divergence. Differences in cluster mean values for yield and quality parameters further emphasized the scope for selecting suitable parents. Genotypes from clusters exhibiting better performance in seed yield, protein and oil content may be effectively used in hybridization programmes to develop superior recombinants. This strategy can contribute to the development of improved soybean varieties with enhanced productivity, quality and better adaptation to rainfed conditions.

References

1. Al-Hadi G, Islam RM, Karim AM, Islam TM. Morpho-physiological characterization of soybean genotypes under subtropical environment. Genetika,2017:49(1):297–311.
2. Balasubramaniyan P, Palaniappan SP. Principles and practices of agronomy. Agrbios, 2003, 45–46.
3. De Souza RR, Cargnelutti Filho A, Toebe M, Bittencourt KC. Sample size and genetic divergence: A principal component analysis for soybean traits. European Journal of Agronomy, 2023, 149.
4. Directorate of Oilseeds Development. Soybean area, production and productivity statistics. Ministry of Agriculture Farmers Welfare, Government of India, 2023–24.
5. Gautam Y, Mohbe G, Bishnoi L, Sharma A, Mishra R, Sharma S, *et al*. Genetic diversity assessment of

soybean genotypes using D² and principal component analysis for breeding advancements. International Journal of Plant & Soil Science,2025:37(6):581–593.
6. Ghuge VR, Surnar DV, Jadhav RS, Randhave BS, Thakur NR. Genetic diversity analysis in soybean (*Glycine max* L.). Journal of Oilseeds Research, 2023, 40(Special Issue).
7. Gupta D, Muralia S, Gupta NK, Gupta S, Jakhar ML, Sandhu JS. Genetic diversity and principal component analysis in mungbean [*Vigna radiata* (L.) Wilczek] under rainfed condition. Legume Research, 2021.
8. Iqbal Z, Arshad M, Ashraf M, Mahmood T, Waheed A. Evaluation of soybean [*Glycine max* (L.) Merrill] germplasm for some important morphological traits using multivariate analysis. Pakistan Journal of Botany,2008:40(6):2323–2328.
9. Jain SK, Sharma LD, Gupta KC, Kumar V, Sharma RS. Principal component and genetic diversity analysis for seed yield and its related components in chickpea (*Cicer arietinum* L.) genotypes. Legume Research, 2020.
10. Jhariya R, Tripathi MK, Mishra R, Sharma S, Solanki R, Singh J. Assessment of genetic divergence in soybean (*Glycine max* [L.] Merrill) using Mahalanobis D² statistics and principal component analysis. Journal of Advances in Biology Biotechnology,2025:28(8):346–360.
11. Jolliffe IT. Principal component analysis. Springer Verlag, 1986.
12. Kumar A, Pandey A, Aochen C, Pattanayak A. Evaluation of genetic diversity and interrelationships of agro-morphological characters in soybean (*Glycine max*) genotypes. Proceedings of the National Academy of Sciences, India Section B: Biological Sciences,2015:85(2):397–405.
13. Mahalanobis PC. On the generalized distance in statistics. Proceedings of the National Institute of Sciences of India,1936:2:49–55.
14. Mahbub MM, Rahman MM, Hossain MS, Nahar L, Shirazy BJ. Morphophysiological variation in soybean (*Glycine max* (L.) Merrill). American-Eurasian Journal of Agricultural Environmental Sciences,2016:16(2):234–238.
15. Malik MFA, Ashraf M, Qureshi AS, Khan MR. Investigation and comparison of some morphological traits of the soybean populations using cluster analysis. Pakistan Journal of Botany,2011:43(2):1249–1255.
16. Massey WF. Principal components regression in exploratory statistical research. Journal of the American Statistical Association,1965:60:234–246.
17. Mazur O. Genetic determination of elements of the soybean yield structure and combining ability of hybridization components. Acta Fytotechnica et Zootechnica,2023:26(2):163–178.
18. Mounika KLS, Devi TR, Devi HN, Ramaiah KS, Kumar MS, Sinha B, *et al*. Genetic diversity study in soybean [*Glycine max* (L.) Merrill] based on agro-morphological characters. Electronic Journal of Plant Breeding,2022:13(3):1005–1011.
19. Nalajala S, Singh NB, Jeberson MS, Yumnam S, Sinha B. Genetic divergence studies for yield and its component traits in mung bean (*Vigna radiata* L. Wilczek). Environment Conservation Journal,2023:24(3):14–20.
20. Paikra N, Karishma, Nag S. Assessment of genetic variation in soybean (*Glycine max* (L.) Merrill) using Mahalanobis D² analysis. International Journal of Advanced Biochemistry Research,2025:9(1S):981–984.
21. Ragade O, Deshmukh M, Kale S, Kumbhar S, Mahajan S, Patil A. Evaluation of 37 soybean genotypes (*Glycine max* (L.) Merrill) for genetic diversity. International Journal of Advanced Biochemistry Research,2024:8(11):517–522.
22. Rao CR. Advanced statistical methods in biometrical research. John Wiley and Sons, New York, 1952.
23. Singh PK, Shrestha J, Kushwaha UKS. Multivariate analysis of soybean genotypes. Journal of Agriculture and Natural Resources,2020:3(1):69–76.
24. Sivabharathi RC, Muthuswamy A, Anandhi K, Karthiba L. Genetic diversity studies of soybean [*Glycine max* (L.) Merrill] germplasm accessions using cluster and principal component analysis. Legume Research,2023:1(6):1–5.
25. Sivabharathi RC, Muthuswamy A, Anandhi K, Karthiba L. Genetic diversity studies of soybean [*Glycine max* (L.) Merrill] germplasm accessions using cluster and principal component analysis. Legume Research – An International Journal,2025:48(6):932–937.
26. Soni M, Shrivastava MK, Khare V, Amrate PK, Singh Y. Multivariate analysis of soybean genotypes: Uncovering agro-morphological insights. Plant Science Today, 2025, 12.
27. Srivastava S, Singh P, Tyagi A, Srivastava A. Multivariate analysis of yield-contributing traits in soybean (*Glycine max* (L.) Merrill): Insights from correlation and principal component approaches. The Bioscan,2025:20(2 (S2):954–957.
28. Vianna VF, Unêda-Trevisoli SH, Desidério JA, De Santiago S, Charnai K, Júnior JAF, *et al*. The multivariate approach and influence of characters in selecting superior soybean genotypes. African Journal of Agricultural Research,2013:8(30):4162–4169.
29. Yadav RK, Tripathi MK, Tiwari S, Asati R, Chauhan S, Sikarwar RS, *et al*. Evaluation of genetic diversity through D² statistics in chickpea (*Cicer arietinum* L.). International Journal of Environmental and Climate Change,2023:13(10):1598–1611.