**Original Research Article**

**Assessment of Genetic Divergence in Soybean (*Glycine max* [L.] Merrill) Using Mahalanobis D² Statistics and Principal Component Analysis**

**Abstract**

Soybean (*Glycine max* [L.] Merrill) is a globally significant leguminous crop, valued for its high protein and oil content, symbiotic nitrogen fixation ability, and adaptability to diverse agro-climatic conditions. Despite increased cultivation and advancements in agronomic practices, its productivity remains constrained due to a narrow genetic base and the complex polygenic nature of yield-related traits. The present investigation aimed to assess genetic divergence and identify key traits contributing to phenotypic variability in soybean employing Mahalanobis D² statistics and Principal Component Analysis (PCA). The experiment was conducted during the *Kharif* 2023 at the Research Farm, Department of Genetics and Plant Breeding, RVSKVV, Gwalior, Madhya Pradesh, India. Ninety-two soybean genotypes were evaluated using a Randomized Block Design with two replications for 10 quantitative traits *viz*., days to 50% flowering, days to maturity, plant height (cm), numbers of primary branches per plant, numbers of pods per plant, numbers of seeds per pod, 100- seed weight (g), biological yield (g), harvest index (%) and yield per plant (g). Mahalanobis D² analysis grouped the genotypes into eight clusters, revealing presence of substantial genetic divergence. Biological yield, plant height, and yield per plant were the major contributors to total divergence. Significant inter-cluster distances were investigated, particularly between Clusters IV and VIII, suggested the existence of highly divergent genotypes suitable for use in hybridization programme. Cluster VII and VIII were identified as potential sources for improving yield and biomass traits. PCA revealed that four principal components with eigenvalues greater than one accounted for 72.18% of the total variation. PC1 contributed the most (28.82%). The Scree plot confirmed the significance of the first four PCs, enabling dimensional reduction and efficient trait prioritization. This integrated approach demonstrated the effectiveness of multivariate analysis in exploring genetic variability and supports the strategic selection of parents for soybean improvement. The findings hold promise for enhancing productivity, adaptability, and sustainability in future breeding programmes targeting for diverse agro-ecological environments.

**Keywords:** Cluster analysis, Genetic divergence, Mahalanobis D2 statistic, Principal Component Analysis, Soybean, Yield components.

**1. Introduction**

*Glycine max* [L.] Merrill stands among the most economically vivacious and nutritionally enriched leguminous crops grown worldwide (Gautam et al., 2025a). It is particularly valued for its seeds, which comprise approximately 40% high-quality protein and around 20% oil, rendering its indispensable in human diets, livestock nutrition and numerous industrial applications (Sharma *et al.,* 2023; Mishra *et al.,* 2024a; Mishra *et al.,* 2024b; Jhariya *et al*., 2025a). Its multifaceted utility positions it is a key crop in addressing pressing challenges related to food and nutritional security, especially amidst rising population pressures and the progressive decline in cultivable land per capita (Gautam et al., 2025b). Apart from its nutritional and industrial contributions, it plays a fundamental role in promoting sustainable agriculture (Mishra *et al.,* 2021a; Sharma *et al.,* 2021). Its symbiotic relationship with *Bradyrhizobium* spp. facilitates biological nitrogen fixation, thereby minimizing the reliance on chemical nitrogen fertilizers and enhancing soil nutrient status. This ecological service not only contributes to environmentally sustainable farming systems but also improves the productivity of sequential crops in rotation-based agriculture (Gitonga *et al.,* 2021; Abd-Alla *et al.,* 2023). Besides, soybean’s short growth period, efficient resource use efficiency, and wide adaptability to varied agro-climatic conditions make it particularly suitable for cultivation in rainfed and low-input farming systems (Mishra *et al.,* 2020; Mishra *et al.,* 2021b; Mishra *et al*., 2021c; Mishra *et al.,* 2024c). In the Indian agricultural scenario, soybean has attained strategic importance, particularly in the rainfed agro-ecologies of central and western India including states such as Madhya Pradesh, Maharashtra, and Rajasthan (Agarwal *et al.,* 2013; Prashnani *et al.,* 2024). Its widespread cultivation is attributed to its tolerance to abiotic stress conditions and compatibility with existing cropping systems. The integration of soybean into traditional agricultural practices has been demonstrated to enhance overall system productivity and contribute to long-term soil health through its organic matter contributions and beneficial agro-ecological interactions (Mishra *et al.,* 2025a; Mishra *et al.,* 2025b).

Despite continuous expansion in the area under cultivation and advancements in agronomic management, its productivity has not reached its full potential. This stagnation is primarily attributed to a limited genetic base and the intricate polygenic control of yield and its associated traits (Tripathi *et al.,* 2022; Hamza *et al.,* 2024; Tahakik *et al.,* 2024). Consequently, a comprehensive understanding of genetic variability, trait associations, and the underlying genetic architecture is crucial for the formulation of effective breeding and selection strategies (Alemu *et al.,* 2024; Mishra *et al.,* 2024d).

Multivariate statistical methodologies, remarkably Mahalanobis D² analysis (Mahalanobis, 1928) and Principal Component Analysis (PCA) (Massey,1965; Jolliffe, 1986), have been widely employed for dissecting genetic diversity and pinpointing the traits that most significantly contribute to genotypic variation (Vanisri *et al.,* 2020; Jadhav *et al.,* 2021; Jhariya *et al*., 2025b). Mahalanobis D² statistics offer a quantitative measure of genetic divergence, allowing for the categorization of genotypes into discrete clusters. This facilitates the identification of genetically divergent parents, which is essential for maximizing heterosis in hybridization programmes (Raina *et al.,* 2015; Mark & Workman, 2018; Kanavi *et al.,* 2019). In contrast, PCA reduces the dimensionality of complex, multivariate datasets by transforming correlated variables into a smaller set of uncorrelated principal components. This enables the extraction of the primary sources of variation within the population and aids in trait prioritization (Wang, 2009; Roessner *et al.,* 2011). In this context, the present investigation was conducted to assess the extent of genetic divergence among a diverse collection of soybean genotypes using Mahalanobis D² analysis, and to determine the principal traits contributing to phenotypic variability through PCA. This integrated approach is intended to enhance the efficiency of parent selection and guide the development of superior, high-yielding cultivar (s) with broad adaptability to varying agro-ecological environments.

**2. Material & Methods**

**2.1 Experimental Site**

The present study was undertaken during the *Kharif*, 2023 at the experimental farm, Department of Genetics and Plant Breeding, College of Agriculture, Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya, Gwalior, Madhya Pradesh, India. The experimental site is situated in a region characterized by a hot and arid climate, with substantial seasonal temperature fluctuations. June typically records the highest temperatures, peaking at around 48°C, while winter months may experience a decline to 4°C. The annual temperature range at the location spans 2°C to 50°C. The area receives an average annual rainfall between 750 mm and 800 mm, primarily concentrated during the southwest monsoon period, which extends from late June to the end of September, accompanied occasionally by scattered winter showers. During the crop growth period: July to October 2023, a cumulative rainfall of 907.7 mm was recorded. However, precipitation during this period was inconsistent and irregularly distributed. Despite these fluctuations, the overall climatic conditions during the growing season were within normal ranges, with the mean maximum and minimum temperatures recorded 35.2°C and 24.5°C, respectively.

**2.2 Experimental Details**

The experimental trial was conducted employing a Randomized Block Design (RBD) with two replications to assess the performance of ninety-two soybean genotypes. Each genotype was cultivated in plots consisting of three rows, with an inter-row spacing of 30 cm and intra-row plant spacing of 10 cm. The dimensions of each plot were maintained 3.0 meters in length and 1.20 meters in width. For the evaluation of quantitative traits, observations were recorded on five aimlessly selected plants from each plot. Throughout the cropping period, standard agronomic practices were uniformly applied to ensure optimal plant growth and the accurate expression of genotypic potential.

**2.3 Statistical Analysis**

Mahalanobis’ D² statistic (Mahalanobis, 1936), in combination with Tocher’s clustering method (Rao,1952) was employed to form clusters based on the calculated D2 values. Moreover, Principal Component Analysis (PCA) as described by Massey (1965) and 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 employed to analyse PCA and OP-STAT for D2 analysis.

**3. Results & Discussion**

**3.1 Mahalanobis D² statistics**

Genetic divergence among ninety-two soybean genotypes was investigated using Mahalanobis D² statistics and Tocher’s clustering method (Rao, 1952), which grouped the genotypes into eight distinct clusters (Fig. 1). This analysis revealed existence of substantial variability among the evaluated genotypes, highlighting the presence of a broad genetic base, a prerequisite for planning of effective crop improvement strategies. The traits contributing most significantly to total genetic divergence were biological yield (26.42%) followed by plant height (21.69%) and yield per plant (17.94%), numbers of seeds per pod (11.61%) and 100-seed weight (9.87%) (Table 1). These findings emphasize the critical role of biomass-related and yield-attributing traits in distinguishing genotypic variation. Such observations are consistent with previous studies that also identified plant architecture and reproductive output as key contributors to phenotypic divergence in soybean (Raina & Khan, 2023; Xu *et al.,* 2024).

**Table 1: Contribution of different characters toward clustering in soybean genotypes**

|  |  |
| --- | --- |
| **Characters** | **Contribution %** |
| Days to 50% flowering | 1.08 % |
| Days to maturity | 0.88 % |
| Plant height (cm) | 21.69 % |
| Numbers of primary branches per plant | 2.27 % |
| Numbers of pods per plant | 2.96 % |
| Numbers of seeds per pod | 11.61 % |
| 100- seed weight (g) | 9.87 % |
| Biological yield (g) | 26.42 % |
| Harvest index (%) | 5.28 % |
| Yield per plant (g) | 17.94 % |
| **Total** | **100%** |

## Table 2: Inter and intra cluster D2 values for different clusters

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cluster** | **Cluster I** | **Cluster II** | **Cluster III** | **Cluster IV** | **Cluster V** | **Cluster VI** | **Cluster VII** | **Cluster VIII** |
| **Cluster I** | **9.25** | 14.22 | 15.97 | 11.89 | 14.50 | 11.89 | 13.30 | 25.71 |
| **Cluster II** |  | **10.79** | 17.85 | 13.38 | 21.19 | 15.18 | 15.67 | 21.68 |
| **Cluster III** |  |  | **9.83** | 22.23 | 16.83 | 18.73 | 21.09 | 17.12 |
| **Cluster IV** |  |  |  | **0.00** | 20.96 | 11.02 | 14.38 | 30.05 |
| **Cluster V** |  |  |  |  | **12.73** | 17.66 | 17.06 | 26.82 |
| **Cluster VI** |  |  |  |  |  | **0.00** | 16.10 | 26.92 |
| **Cluster VII** |  |  |  |  |  |  | **0.00** | 28.23 |
| **Cluster VIII** |  |  |  |  |  |  |  | **0.00** |

The intra-cluster distances (Table 2) arrayed between 0.00 to 12.73, with Cluster V displaying the highest intra-cluster divergence (D² = 12.73), suggesting presence of remarkable genotypic variability within this group. In contrast, Clusters IV, VI, VII and VIII exhibited zero intra-cluster distances due to the presence of only one genotype per cluster, indicating their unique genetic identities. The highest inter-cluster divergence was observed between Cluster IV and Cluster VIII (D² = 30.05), followed by Cluster IV and Cluster VII (D² = 28.23), and Cluster VI and Cluster VIII (D² = 26.92) (Table 2). These pronounced inter-cluster distances indicated the presence of highly divergent genotypes, which could be effectively utilized as parents in future hybridization programmes to generate heterotic combinations and transgressive segregants as earlier suggested by Mazur (2023) and Ragade *et al.* (2024).

## Table 3: Cluster mean for yield and its attributing traits of soybean genotypes

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cluster** | **Days to 50% flowering** | **Days to maturity** | **Plant height (cm)** | **Numbers of primary branches per plant** | **Numbers of pods per plant** | **Numbers of seeds per pod** | **100-seed weight (g)** | **Biological yield (g)** | **Harvest index (%)** | **Yield per plant (g)** |
| **Cluster I** | 42.12 | 105.81 | 45.51 | **8.45** | 23.02 | 2.45 | 10.91 | 57.76 | 25.72 | 13.43 |
| **Cluster II** | 41.72 | 109.77 | 65.42 | 10.00 | 22.64 | 2.72 | 10.56 | 88.12 | 15.81 | 13.10 |
| **Cluster III** | **38.83** | 105.07 | 44.47 | 9.25 | **20.58** | 2.17 | 10.05 | 101.88 | 12.66 | 11.88 |
| **Cluster IV** | **45.50** | **115.94** | 53.95 | 9.00 | 26.50 | **3.00** | **9.54** | 45.00 | 24.98 | **11.24** |
| **Cluster V** | 42.00 | **93.64** | **34.53** | 8.25 | 23.58 | 2.50 | 10.96 | 60.77 | 28.64 | 14.60 |
| **Cluster VI** | 42.50 | 105.51 | 50.10 | 9.00 | 22.00 | **2.00** | 10.39 | **22.70** | **58.99** | 13.39 |
| **Cluster VII** | 42.50 | 115.78 | 49.00 | **10.50** | 25.00 | **3.00** | 12.13 | 86.00 | 20.58 | **17.70** |
| **Cluster VIII** | 40.00 | 95.96 | **69.00** | 10.00 | **29.00** | **3.00** | **12.09** | **169.50** | **9.33** | 15.81 |

**Table 4: Distribution of 92 soybean genotypes in 8 different clusters employing Tocher Method**

|  |  |  |
| --- | --- | --- |
| **Cluster No.** | **Numbers of genotypes** | **Name of the genotypes** |
| I | 56 | RVS23-1, RVS23-2, RVS23-3, RVS23-10, RVS23-11, RVS23-12, RVS 23-13, RVS23-14, RVS23-15, RVS23-16, RVS23-17, RVS23-18, RVS23-19, RVS23-20, RVS23-21, RVS23-22, RVS23-23, RVS23-24, RVS23-25, RVS23-26, RVSM35, JS93-05, JS335, JS20-69, Raj Soya 24, Raj Soya 18, JS20-29, NRCSL5, JS24-26, NRCSL7, SKAUS3, KDS1203, NRC253, MACS1756, Lok Soya 2, Himso1695, NRCSL-8, JS24-34, RSC10-52, KSS213, NRC254, DS1529, MACS1745, Asb93, VLS105, NRCSL4, NRC 257, MAUS814, Asb85, RSC1165, BAUS124, DLSB 40, CAUMS3, RVSM 12-21, NRC259, AS34, RVSM 2011-35, TS-208 |
| II | 18 | RVS23-4, RVS23-5, RVS23-6, RVS76, NRC-260, NRC-196, RVS23-7, RVS23-9, NRC256, RVS23-8, AMS 2021-4, SL1311, RVS2001-4, Himso 1696, KDS 1188, AMS 2021-3, PS 1693, AUKS 212 |
| III | 7 | VLS104, NRC255, Pusa Sipani, SPS-433, JS20-98, DS1510, AS55 |
| IV | 1 | RSC1172 |
| V | 7 | JS 20-34, MAUS 824, NRC258, JS95-60, RVS12-8, Pusa Sipani, BS-9 |
| VI | 1 | TS-156 |
| VII | 1 | PS-1696 |
| VIII | 1 | JS 20-116 |

**Fig. 1: Distribution of 92 Soybean genotypes in 8 different clusters employing Tocher Method**

Cluster mean analysis revealed wide variation across all investigated traits (Table 3). Cluster IV contained the highest mean for days to 50% flowering (45.50 days) and days to maturity (115.94 days), indicating its potential utility in breeding for longer duration genotypes as previously recommended by Sawarkar *et al.* (2025) and Korke *et al.* (2025). Cluster VIII exhibited the highest plant height (69.00 cm), numbers of pods per plant (29.00), and biological yield (169.50 g), making it a promising source for biomass-related traits. While Cluster VII had the highest 100-seed weight (12.13 g) and yield per plant (17.70 g), signifying its potential for improving seed size and productivity. Interestingly, Cluster VI displayed the maximum harvest index (58.99%), despite its low biological yield, highlighting efficient partitioning of assimilates to economical yield, a valuable trait for resource-limited environments (Swar *et al.,* 2021; Ghughe *et al.,* 2023). The distribution of genotypes among clusters further supports the existence of both homogenous and heterogenous groups (Table 4). Cluster I was the largest, encompassing 56 genotypes, potentially representing a shared genetic background or similar phenotypic traits. In contrast, the presence of single-genotype clusters including clusters IV, VI, VII, and VIII emphasizes the uniqueness of certain genotypes, which may possess rare or valuable traits warranting targeted utilization in breeding programmes as advocated by Mounika *et al.* (2022), Ghughe *et al.*  (2023), Nichal *et al.* (2023), Kumawat *et al.* (2024) and Paikra *et al.* (2025). Overall, the study not only demonstrated the effectiveness of Mahalanobis D² statistics and Tocher’s method in elucidating genetic divergence but also provides strategic insights for parent selection. The diverse clusters identified offer valuable resources for recombination breeding aimed to enhance complex traits such as yield potential, maturity duration, biomass production, and stress adaptability in soybean (Kanavi *et al.,* 2019; Sharma & Lal, 2020; Zafar *et al.,* 2023; Patel *et al.,* 2024).

**3.2 Principal Component Analysis**

Principal Component Analysis (PCA) was conducted to evaluate the multivariate structure and identify the major sources of variation among the soybean genotypes based on the measured quantitative traits. A total of ten principal components were extracted, out of which four components had eigenvalues greater than 1, accounting for 72.18% of the total variability observed among the traits (Table 5). The first principal component (PC1) alone explained the highest proportion of variability (28.82%) with an eigenvalue of 1.70, followed by PC2 (20.72%), PC3 (11.50%), and PC4 (11.14%).

The Scree plot (Fig. 2) illustrated the distribution of eigenvalues across principal components, where a clear inflection point was observed after the fourth component, indicating that the majority of the variance was captured by the first few components. The curve followed a typical elbow-shaped pattern, signifying that components beyond PC4 contributed minimal information. This confirms the suitability of employing the first four PCs for meaningful interpretation and genotype differentiation.

**Table 5: Principal components of yield and its related traits of soybean**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Traits** | **Principal component** | **Standard deviation** | **Proportion of variance** | **Cumulative proportion** |
| **Days to 50% flowering** | PC1 | 1.70 | 28.82 | 28.82 |
| **Days to maturity** | PC2 | 1.44 | 20.72 | 49.54 |
| **Plant height (cm)** | PC3 | 1.07 | 11.50 | 61.04 |
| **Numbers of primary branches per plant** | PC4 | 1.06 | 11.14 | 72.18 |
| **Numbers of pods per plant** | PC5 | 0.95 | 9.04 | 81.22 |
| **Numbers of seeds per pod** | PC6 | 0.84 | 7.10 | 88.32 |
| **100- seed weight (g)** | PC7 | 0.78 | 6.05 | 94.37 |
| **Biological yield (g)** | PC8 | 0.65 | 4.19 | 98.56 |
| **Harvest index (%)** | PC9 | 0.32 | 1.01 | 99.56 |
| **Yield per plant (g)** | PC10 | 0.21 | 0.44 | 100.00 |

**Table 6: List of PCA scores of the soybean in each PCs**

| **S. No.** | **Observations** | **PC1** | **PC2** | **PC3** | **PC4** |
| --- | --- | --- | --- | --- | --- |
| 1 | RVS 23-1 | -2.2143 | 0.3521 | -0.6417 | 0.1663 |
| 2 | RVS 23-2 | -0.2711 | -0.467 | 1.2828 | -0.7072 |
| 3 | RVS 23-3 | 1.7565 | -0.496 | -0.6271 | -0.7975 |
| 4 | RVS 23-4 | -0.574 | -1.3062 | -1.3486 | 0.7751 |
| 5 | RVS 23-5 | 2.4233 | 0.5297 | 1.5668 | -0.7005 |
| 6 | RVS 23-6 | 1.5812 | 0.5161 | 1.4872 | -1.4701 |
| 7 | RVS 23-7 | 0.0255 | 1.627 | 0.1092 | -1.3036 |
| 8 | RVS 23-8 | 0.5303 | 2.5781 | -0.3146 | 0.8997 |
| 9 | RVS 23-9 | 1.3097 | -0.7658 | 0.6226 | 1.3585 |
| 10 | RVS 23-10 | 0.2598 | 0.1154 | 1.1785 | 1.0612 |
| 11 | RVS 23-11 | 0.024 | 0.7675 | 1.0088 | 0.7429 |
| 12 | RVS 23-12 | 0.557 | 1.726 | 0.3165 | -0.1931 |
| 13 | RVS 23-13 | -0.122 | 2.1014 | -0.0597 | -1.4882 |
| 14 | RVS 23-14 | 0.9813 | -1.0913 | 0.7395 | 0.5273 |
| 15 | RVS 23-15 | -1.1102 | 0.9733 | 0.1262 | 0.5264 |
| 16 | RVS 23-16 | -1.1301 | -1.4912 | -1.2028 | -0.9813 |
| 17 | RVS 23-17 | -2.0641 | 0.1721 | 0.1035 | -0.0299 |
| 18 | RVS 23-18 | -2.6374 | 2.2744 | -0.8774 | -0.2338 |
| 19 | RVS 23-19 | -2.6655 | 0.2671 | -1.5003 | 0.8694 |
| 20 | RVS 23-20 | -1.1833 | -1.3921 | -0.207 | -0.3228 |
| 21 | RVS 23-21 | -1.964 | -2.0143 | -0.7171 | 0.0821 |
| 22 | RVS 23-22 | -1.2201 | 1.4169 | 1.4544 | 0.3593 |
| 23 | RVS 23-23 | -3.9136 | 2.5456 | -0.0411 | -1.5313 |
| 24 | RVS 23-24 | -0.7777 | -2.0017 | -1.7295 | 0.0222 |
| 25 | RVS 23-25 | 1.0523 | -1.1043 | 1.6568 | -0.4854 |
| 26 | RVS 23-26 | -1.3279 | -1.497 | 0.3865 | 0.4059 |
| 27 | RVSM 35 | 1.9182 | -0.6734 | -0.727 | -0.3454 |
| 28 | JS 20-34 | -3.1871 | -1.3022 | 1.1477 | -2.2231 |
| 29 | JS 93-05 | -1.7771 | -2.1308 | -0.3915 | 0.9554 |
| 30 | JS 95-60 | -0.8534 | -0.9 | 2.4753 | -1.6852 |
| 31 | JS 335 | 0.8871 | -0.2881 | -0.5517 | 0.2564 |
| 32 | JS 20-116 | 3.2452 | 2.4987 | -0.7755 | -1.4346 |
| 33 | JS 20-69 | 3.0061 | -0.5245 | 1.8113 | -0.8513 |
| 34 | JS 20-98 | 2.2569 | -0.4496 | -1.4473 | -1.3504 |
| 35 | RVS76 | 1.684 | 1.44 | 0.4539 | 0.4824 |
| 36 | RVS 2001-4 | 3.2973 | -0.3551 | -1.1709 | 0.4616 |
| 37 | Raj Soya 24 | 0.9776 | 1.7019 | -1.0689 | -2.1448 |
| 38 | Raj Soya 18 | 1.1463 | -0.712 | 1.2072 | -2.5272 |
| 39 | JS 20-29 | 1.4693 | 2.1377 | 0.429 | -1.535 |
| 40 | VLS 104 | 0.5998 | -1.2616 | -1.5743 | -1.491 |
| 41 | NRCSL 5 | -0.3308 | 0.3908 | -0.8672 | -1.8416 |
| 42 | JS 24-26 | 0.1631 | -0.2453 | 0.3556 | -0.2411 |
| 43 | NRCSL 7 | -0.118 | -2.3305 | -1.4036 | -1.0665 |
| 44 | SKAUS 3 | -0.3729 | -1.3053 | 1.3135 | 1.2258 |
| 45 | RVS 12-8 | -1.4774 | -2.1905 | -0.2542 | -2.2803 |
| 46 | KDS 1203 | -1.082 | -1.5434 | -0.5029 | 0.3469 |
| 47 | NRC 253 | -0.3725 | -1.434 | -0.1507 | 0.0731 |
| 48 | MACS 1756 | -0.9902 | 2.2393 | 0.1514 | 0.2062 |
| 49 | Lok Soya 2 | -0.0008 | -0.9497 | 1.5258 | 0.1164 |
| 50 | AMS 2021-3 | 2.247 | -1.167 | -0.6402 | 0.3854 |
| 51 | Himso 1695 | -1.4914 | 1.4851 | -0.7403 | -0.4135 |
| 52 | TS-156 | -2.4225 | -0.0576 | 1.0324 | -0.1263 |
| 53 | NRCSL-8 | -1.1355 | -1.4393 | -1.4994 | -0.1287 |
| 54 | JS 24-34 | -2.5079 | 2.6317 | -1.1783 | 0.2162 |
| 55 | RSC 10-52 | -2.8307 | 1.4692 | -0.2563 | -1.079 |
| 56 | DS 1510 | 1.6364 | -0.4758 | 0.1327 | -1.0076 |
| 57 | KSS 213 | -1.7678 | 1.3456 | 0.1674 | 1.6779 |
| 58 | MAUS 824 | -0.8322 | 2.5504 | -1.7879 | 0.5942 |
| 59 | NRC 254 | 0.9449 | -1.0149 | 1.1429 | 2.1277 |
| 60 | AMS 2021-4 | 2.7468 | -1.1554 | -0.8817 | 0.0573 |
| 61 | Himso 1696 | 3.3344 | -0.1267 | -0.1546 | 0.4583 |
| 62 | DS 1529 | 0.1595 | 1.3683 | 0.5938 | -0.2129 |
| 63 | KDS 1188 | 2.9306 | -0.373 | -0.1601 | 2.1791 |
| 64 | MACS 1745 | -0.9252 | -2.4255 | -0.3025 | 0.7869 |
| 65 | NRC 255 | -2.2191 | -1.7099 | -1.4606 | -0.3126 |
| 66 | Asb 93 | -2.5009 | 1.2344 | 1.3785 | 0.8274 |
| 67 | VLS 105 | -2.1317 | 1.8746 | 0.8688 | 2.1053 |
| 68 | NRCSL 4 | -1.0647 | -1.7996 | -0.4128 | 0.4246 |
| 69 | NRC 257 | 0.2074 | -1.991 | 0.8827 | 0.9379 |
| 70 | MAUS 814 | -0.9066 | 1.0646 | -0.2275 | -0.3564 |
| 71 | SL 1311 | 1.7806 | 1.7106 | -0.0832 | 1.0481 |
| 72 | Asb 85 | -1.2629 | -0.7144 | 1.379 | 0.8031 |
| 73 | PS 1693 | 2.3575 | 1.1078 | -1.6507 | -0.6931 |
| 74 | NRC 256 | 2.3735 | -0.2797 | 0.962 | -0.719 |
| 75 | RSC 1165 | -0.8618 | -1.4655 | 2.4462 | 0.8869 |
| 76 | BAUS 124 | -0.4632 | -2.0405 | -0.0489 | 0.2743 |
| 77 | DLSB 40 | 0.0823 | 0.1068 | -2.4346 | 2.2202 |
| 78 | NRC 258 | 1.1119 | 1.7342 | 0.7982 | -0.9334 |
| 79 | Pusa Sipani BS-9 | -3.0483 | 1.029 | 0.0674 | 0.6025 |
| 80 | PS-1696 | 0.8252 | 2.0904 | 0.0263 | 1.0805 |
| 81 | CAUMS 3 | 1.4336 | 1.1501 | -1.5876 | 1.6775 |
| 82 | AUKS 212 | 1.8354 | -0.5505 | 0.3708 | 0.2513 |
| 83 | RVSM 12-21 | -1.4724 | -0.6513 | -0.1412 | -1.0639 |
| 84 | NRC 259 | 1.0406 | -0.5858 | 1.0295 | 0.1469 |
| 85 | AS 34 | -0.1693 | -0.9479 | 1.1287 | 0.2765 |
| 86 | RVSM 2011-35 | 0.0935 | 1.7833 | 1.0759 | 0.3532 |
| 87 | RSC1172 | 0.7794 | -0.7297 | 1.4336 | 1.1336 |
| 88 | AS 55 | 2.2525 | 0.6241 | -1.8538 | 0.2178 |
| 89 | TS-208 | 0.2397 | -1.8971 | -0.8019 | -0.4881 |
| 90 | NRC-260 | -0.3612 | 1.1326 | 0.4249 | 1.1334 |
| 91 | NRC-196 | 2.1231 | 1.7153 | -0.1881 | 1.2226 |
| 92 | Pusa Sipani SPS-433 | 0.4242 | -1.758 | -1.607 | 0.7696 |

**Table 7: Promising genotypes on the basis of scores of principal components of soybean**

|  |  |  |  |
| --- | --- | --- | --- |
| **PC 1** | **PC 2** | **PC 3** | **PC 4** |
| Himso1696 | JS24-34 | JS95-60 | DLSB 40 |
| RVS2001-4 | RVS23-8 | RSC1165 | KDS1188 |
| JS20-116 | MAUS824 | JS 20-69 | NRC254 |
| JS20-69 | RVS23-23 | RVS23-25 | VLS105 |
| KDS1188 | JS 20-116 | RVS23-5 | KSS213 |
| AMS2021-4 | RVS23-18 | Lok Soya 2 | CAUMS 3 |
| RVS23-5 | MACS1756 | RVS23-6 | RVS23-9 |
| NRC256 | JS 20-29 | RVS23-22 | SKAUS 3 |
| PS1693 | RVS23-13 | RSC1172 | NRC-196 |
| JS20-98 | PS-1696 | Asb 85 | RSC1172 |
| AS55 | VLS105 | Asb 93 | NRC-260 |
| AMS2021-3 | RVSM2011-35 | SKAUS 3 | PS-1696 |
| NRC-196 | NRC258 | RVS23-2 | RVS23-10 |
| RVSM35 | RVS23-12 | Raj Soya18 | SL1311 |
| AUKS212 | NRC-196 | RVS23-10 |  |
| SL1311 | Raj Soya 24 | JS20-34 |  |
| RVS23-3 | RVS 23-7 | NRC254 |  |
| RVS76 | Himso 1695 | AS34 |  |
| DS1510 | RSC10-52 | RVSM 2011-35 |  |
| RVS23-6 | RVS 76 | TS-156 |  |
| JS20-29 | RVS 23-22 | NRC259 |  |
| CAUMS3 | DS 1529 | RVS23-11 |  |
| RVS23-9 | KSS 213 |  |  |
| Raj Soya18 | Asb 93 |  |  |
| NRC258 | CAUMS 3 |  |  |
| RVS23-25 | NRC-260 |  |  |
| NRC259 | PS1693 |  |  |
|  | MAUS 814 |  |  |
|  | Pusa Sipani BS-9 |  |  |

The large amount of variability captured by PC1 indicates that it is associated with traits that contribute significantly to overall genetic diversity, possibly including yield-related traits and morphological characteristics as suggested by Jain *et al.* (2020), Gupta *et al.* (2021) and Sharma *et al*. (2023). The gradual decline in variability across subsequent components reflects weaker trait associations and reduced discriminatory power, which is consistent with the findings of previous studies employing PCA in leguminous crops (de Souza *et al.,* 2023; Yadav *et al*., 2023; Yadav *et al.,* 2024). The PCA also helped in identifying genotypes with high principal component scores, which were distinctly separated in the multidimensional trait space (Table 6; Table 7). These genotypes represent potential candidates for selection and parental inclusion in breeding programmes (Table7), particularly those targeting trait combinations associated with yield, biomass, seed weight, and plant architecture (Dunna *et al.,* 2023; Cui *et al.,* 2024; Mishra *et al.,* 2025b).

**Fig. 2: Scree plot for different principal components**

Overall, the PCA effectively reduced data dimensionality while retaining most of the information, allowing for a clearer understanding of trait interrelationships and genotype differentiation. The findings emphasized the importance of multivariate statistical approaches in characterizing genetic resources and identifying superior genotypes for crop improvement programmes. The insights gained from this analysis may aid in prioritizing genotypes and traits for targeted breeding strategies in soybean (El-Hashash, 2016; Kahlon *et al.,* 2018; Leite *et al.,* 2018; Singh *et al.,* 2020; Uikey *et al.,* 2021; Tepavcevic *et al.,* 2021).

**Conclusion**

The present study underscores the critical role of multivariate statistical tools-Mahalanobis D² statistics and Principal Component Analysis (PCA) in dissecting the genetic diversity and trait interrelationships among 92 soybean genotypes. The significant genetic divergence revealed through D² analysis and Tocher’s clustering not only reflects a wide genetic base but also provides a solid foundation for strategic parent selection to be used in further breeding programmes. Remarkably, traits such as biological yield, plant height, and seed yield per plant emerged as pivotal drivers of genotypic differentiation, making them main selection indices for yield enhancement. The identification of distinct and unique single-genotype clusters highlights the presence of novel alleles or rare combinations, representing untapped genetic reservoirs that can be harnessed for trait introgression and broadening the genetic pool. The extraction of four principal components accounting for over 72% of total variation reinforces the efficiency of PCA in dimensionality reduction and trait prioritization. The dominance of PC1, heavily loaded with yield-contributing traits, offers a streamlined focus for breeders aiming at productivity improvements. Collectively, the integration of D² statistics and PCA provides a comprehensive framework for exploring genetic diversity, guiding parent selection, and rationalizing breeding strategies. The promising genotypes identified in this study serve not only as immediate candidates for crossing but also as foundational germplasm for long-term genetic enhancement. These insights may significantly aid in the development of high-yielding, stress-resilient, and agro-climatically adaptable soybean cultivars, thereby contributing to food security, climate resilience, and sustainable agricultural intensification in legume-based systems.

**Disclaimer (Artificial intelligence)**

Author(s) hereby declare that no generative AI technologies such as Large Language Models (ChatGPT, manuscript.

**References**

1. Abd-Alla, M. H., Al-Amri, S. M., & El-Enany, A.W.E. (2023). Enhancing rhizobium–legume symbiosis and reducing nitrogen fertilizer use are potential options for mitigating climate change. *Agriculture*, 13(11): 2092. <https://doi.org/10.3390/agriculture13112092>
2. Agarwal, D. K., Billore, S. D., Sharma, A. N., Dupare, B. U., & Srivastava, S. K. (2013). Soybean: introduction, improvement, and utilization in India-problems and prospects. *Agricultural Research*, 2(4): 293–300. <https://doi.org/10.1007/s40003-013-0088-0>
3. Alemu, A., Åstrand, J., Montesinos-López, O. A., Isidro y Sánchez, J., Fernández-Gónzalez, J., Tadesse, W., Vetukuri, R. R., Carlsson, A. S., Ceplitis, A., Crossa, J., Ortiz, R., & Chawade, A. (2024). Genomic selection in plant breeding: Key factors shaping two decades of progress. *Molecular Plant*, 17(4): 552–578. <https://doi.org/10.1016/j.molp.2024.03.007>
4. Cui, B., Liu, J., Zhang, M., Wan, H., Wei, G., Wei, Z., & Liu, F. (2024). CO2 elevation modulates the growth and physiological responses of soybean (*Glycine max* L. Merr.) to progressive soil drying. *Plant Growth Regulation*, 103(1): 139–150. https://doi.org/10.1007/s10725-023-01092-z
5. 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>
6. 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>
7. Gautam, Y., Tripathi, M.K., Mishra, R., Mohbe, G., Sharma, A., Sharma, S., Singh, J. and Bishnoi L. (2025a). “Genetic evaluation and trait association analysis for yield enhancement in soybean (*Glycine Max* L. Merrill)”. *Archives of Current Research International* *25* (6):589-603. <https://doi.org/10.9734/acri/2025/v25i61304>.
8. Gautam, Y., Mohbe, G., Bishnoi, L., Sharma, A., Mishra, R., Sharma, S., Singh, J., and Tripathi, M.K. (2025b). 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.
9. 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 (Specialissue). <https://doi.org/10.56739/jor.v40iSpecialissue.145778>
10. Gitonga, N. M., Njeru, E. M., Cheruiyot, R., & Maingi, J. M. (2021). *Bradyrhizobium* inoculation has a greater effect on soybean growth, production and yield quality in organic than conventional farming systems. *Cogent Food & Agriculture*, 7(1). <https://doi.org/10.1080/23311932.2021.1935529>
11. Gupta, D., Muralia, S., Gupta, N. K., Gupta, S., Jakhar, M. L., & Sandhu, J. S. (2021). Genetic diversity and principal component analysis in mungbean [*Vigna radiata* (L.) Wilczek] under rainfed condition. *Legume Research,* <https://doi.org/10.18805/LR-4568>
12. 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>
13. Jadhav, R. A., Mehtre, S. P., Patil, D. K., & Gite, V. K. (2021). Multivariate analysis using d2 and principal component analysis in mung bean [*Vigna radiata* (L.) Wilczek] for study of genetic diversity. *Legume Research*, <https://doi.org/10.18805/LR-4508>
14. Jain, S. K., Sharma, L. D., Gupta, K. C., Kumar, V., & Sharma, R. S. (2021). Principal component and genetic diversity analysis for seed yield and its related components in the genotypes of chickpea (*Cicer arietinum* L.). *Legume Research,* <https://doi.org/10.18805/LR-4489>
15. Jhariya, R., Tripathi, M.K. Mishra, R., Solanki, R. and Sharma S. (2025a). Assessment of soybean (*Glycine max* L.) genotypes on the basis of biochemical parameters. *Plant Cell Biotechnology & Molecular Biology,* 26 (7-8):212-23. <https://doi.org/10.56557/pcbmb/2025/v26i7-89433>.
16. Jhariya, R., Tripathi, M.K. Mishra, R., Solanki, R. and Sharma S. (2025b). Field based disease indexing of soybean genotypes against yellow mosaic virus. *Journal of Scientific Research and Reports* 31 (6):719-28. https://doi.org/10.9734/jsrr/2025/v31i63167.
17. Jolliffe, I. T. (1986). Principal component analysis (2nd ed.). Springer Verlag.
18. 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>
19. Kanavi, M. S. P., Rangaiah, S., & Anusha, C. R. (2019). Genetic diversity analysis through d2 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>
20. Korke, S., Patil, D., & Pawar, S. B. (2025). Study on genetic diversity in mungbean. *International Journal of Advanced Biochemistry Research*, 9(5S): 125–131. <https://doi.org/10.33545/26174693.2025.v9.i5Sb.4351>
21. Kumawat, A., Mishra, S., Sen, R., & Gour, L. (2024). Genetic divergence analysis of soybean (*Glycine Max* L.) genotypes using Mahalanobis multivariate analysis. *Annals of Plant and Soil Research*, 26(1): 172–174. <https://doi.org/10.47815/apsr.2024.10348>
22. 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>
23. Mahalanobis, P. C. (1928). On the generalized distance in statistics. Proceedings of the National Academy of Science, 19: 201–208.
24. 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>
25. Massey, W. F. (1965). Principal components regression in exploratory statistical research. Journal of the American Statistical Association, 60: 234–246.
26. Mazur, O. (2023). Genetic determination of elements of the soybean yield structure and combining ability of hybridization components. *Acta Fytotechnica et Zootechnica*, 26(2): 163–178. <https://doi.org/10.15414/afz.2023.26.02.163-178>
27. Mishra, N., Tripathi, M.K., Tiwari, S., Tripathi, N., &Trivedi, H.K. (2020). Morphological and molecular screening of soybean genotypes against yellow mosaic virus disease. *Legume Research*. <https://doi.org/10.18805/LR-4240>
28. Mishra, N., Tripathi, M.K., Tiwari, S., Tripathi, N., Sapre, S. & Ahuja, A. (2021a). Cell suspension culture and *in vitro* screening for drought tolerance in soybean using poly-ethylene glycol. *Plants*, 10(3):517. <https://doi.org/10.3390/plants10030517>
29. Mishra, N., Tripathi, M.K., Tiwari, S., Tripathi, N., Gupta, N., & Sharma, A. (2021b). Morphological and physiological performance of Indian soybean [*Glycine max* (L.) Merrill] genotypes in respect to drought. *Legume Research.* LR-4550.
30. Mishra, N., Tripathi, M.K., Tripathi, N., Tiwari, S., Gupta, N.& Sharma A. (2021c). Changes in biochemical and antioxidant enzymes activities play significant role in drought tolerance in soybean. *Int. J Agric. Technol.* 17,1425-46. 41.
31. 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>
32. Mishra, R., Tripathi, M.K., 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>
33. Mishra, R., Tripathi, M.K., Tripathi, N., Singh, J., Yadav, P.K. & Sikarwar, R.S. (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](https://doi.org/https%3A//doi.org/10.37446/volbook032024/22-68)
34. Mishra, R., Tripathi, M.K., Shrivastava, M.K., & Amrate, P.K. (2024d). 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](https://doi.org/https%3A//doi.org/10.37446/volbook032024/1-21)
35. 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*, 26(7–8):32–52. <https://doi.org/10.56557/pcbmb/2025/v26i7-89345>
36. 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>
37. Mounika, K. L. S., Devi, Th. R., Devi, H. N., Ramaiah, K. S., Kumar, M. S., Sinha, B., & Singh, N. G. (2022). Genetic diversity study in soybean [*Glycine max* L. Merrill] based on agro-morphological characters. *Electronic Journal of Plant Breeding*, 13(3): 1005–1011. https://doi.org/10.37992/2022.1303.130
38. Nichal, S. S., Patil, P. V, Dandge, M. S., & Chandankar, G. D. (2023). Genetic diversity studies in soybean (*Glycine max* L. Merrill) germplasm. *Journal of Oilseeds Research*, 40(Specialissue). <https://doi.org/10.56739/jor.v40iSpecialissue.145617>
39. Paikra, N., Karishma, & Nag, S. (2025). Assessment of genetic variation in soybean (*Glycine max* (L.) Merrill) using Mahalanobis D2 analysis. *International Journal of Advanced Biochemistry Research*, 9(1S): 981–984. <https://doi.org/10.33545/26174693.2025.v9.i1Sm.3652>
40. Patel, D., Vaja, M. B., Patel, R., & Solanki, P. (2024). Phenotypic diversity using D2 statistics in finger millet [*Eleusine coracana* (L.) Gaertn.]. *International Journal of Advanced Biochemistry Research*, 8(3): 947–951. <https://doi.org/10.33545/26174693.2024.v8.i3k.881>
41. Prashnani, M., Dupare, B., Vadrevu, K. P., & Justice, C. (2024). Towards food security: Exploring the spatio-temporal dynamics of soybean in India. *PLOS ONE*, 19(5): e0292005. <https://doi.org/10.1371/journal.pone.0292005>
42. Ragade, O., Deshmukh, M., Kale, S., Kumbhar, S., Mahajan, S., & Patil, A. (2024). Evaluation of 37 soybean genotypes (*Glycine max* (L.) Merrill) for genetic diversity. *International Journal of Advanced Biochemistry Research*, 8(11): 517–522. <https://doi.org/10.33545/26174693.2024.v8.i11g.2890>
43. Raina, A., & Khan, S. (2023). Field assessment of yield and its contributing traits in cowpea treated with lower, intermediate, and higher doses of gamma rays and sodium azide. *Frontiers in Plant Science*, 14. <https://doi.org/10.3389/fpls.2023.1188077>
44. 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>
45. Rao, C. R. (1952). Advanced statistical methods in biometric research. John Willey and Sons. pp-390.
46. 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>
47. Sawarkar, A., Das, S., Saha, S., Datta, S., Das, A. K., & Dasgupta, T. (2025). Genetic cause and effect interrelationships for yield and yield components of mungbean [*Vigna radiata* (L.) Wilczek]. *Legume Research*, <https://doi.org/10.18805/LR-5338>
48. 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.)](https://www.researchgate.net/profile/Manoj-Tripathi-10/publication/370419935_Discriminant_function_analysis_for_yield_improvement_in_bread_wheat_Triticum_aestivum_L/links/644fa7b697449a0e1a6d07d3/Discriminant-function-analysis-for-yield-improvement-in-bread-wheat-Triticum-aestivum-L.pdf). *The Pharma Innovation Journal*, *12* (5): 224-232.
49. 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,* <https://doi.org/10.18805/LR-4678>
50. Sharma, R., & Lal, G. M. (2020). Study of genetic divergence in soyabean (*Glycine max*. (L.) Merrill) germplasm. *International Journal of Current Microbiology and Applied Sciences*, 9(9): 2532–2537. <https://doi.org/10.20546/ijcmas.2020.909.317>
51. 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>
52. de Souza, R. R., 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>
53. Swar, B. R., Swarnalatha, V., Reddy, M. R., & Vanisree, S. (2021). Genetic diversity studies in MAGIC population of soybean (*Glycine max* (L.) Merrill) based on Mahalanobis D2 distance. *International Journal of Plant & Soil Science*, 18–25. <https://doi.org/10.9734/ijpss/2021/v33i530433>
54. Tahakik, R. R., Deshmukh, A. G., Moharil, M. P., Jadhav, P. V., Kogade, V. T., More, K. D., & Shinde, V. P. (2024). Transitioning from the green revolution to the gene revolution: strengthening nutritional security using climate resilient traditional crops. *Bulletin of the National Research Centre*, 48(1): 123. <https://doi.org/10.1186/s42269-024-01281-4>
55. 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>
56. Tripathi, N., Tripathi, M. K., Tiwari, S., & Payasi, D. K. (2022). Molecular breeding to overcome biotic stresses in soybean: update. Plants, 11(15), 1967. <https://doi.org/10.3390/plants11151967>
57. 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>
58. Vanisri, S., Charlse, W., Laxmi, I., Rahul, V. S. et al. (2020). Mahalanobis D2 and principal component based genetic divergence, selection criteria and genetic variability studies in cold tolerant rice (*Oryza sativa*) genotypes. *International Journal of Chemical Studies*, 8(4): 1754–1761. https://doi.org/10.22271/chemi.2020.v8.i4r.9862
59. 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>
60. Xu, L. W., Pang, J., Turner, N. C., Lambers, H., & He, J. (2024). Genotypic differences in soybean (*Glycine max* L.) in yield response to phosphorus fertilizer are associated with difference in biomass and phosphorus content. *Journal of Agriculture and Food Research*, 18, 101267. <https://doi.org/10.1016/j.jafr.2024.101267>
61. Yadav, R. K., Tripathi, M. K., Tiwari, S., Asati, R., Chauhan, S., Sikarwar, R.S., & Yasin, M. (2023). [Evaluation of genetic diversity through D2 Statistic in chickpea (*Cicer arietinum* L.)](https://www.researchgate.net/profile/Manoj-Tripathi-10/publication/373492816_Evaluation_of_Genetic_Diversity_through_D_2_Statistic_in_Chickpea_Cicer_arietinum_L/links/64ef24590f7ab20a86654528/Evaluation-of-Genetic-Diversity-through-D-2-Statistic-in-Chickpea-Cicer-arietinum-L.pdf). *Int. J. Environ. Clim. Change*, 13 (10), 1598-1611.
62. Yadav, S. K., Singh, N., & Yadav, H. K. (2024). Genetic variability and trait association analysis in linseed (*Linum usitatissimum* L.) for yield and related traits. *Oil Crop Science*, 9(3): 151–159. <https://doi.org/10.1016/j.ocsci.2024.06.003>
63. Zafar, S. A., Aslam, M., Zaman Khan, H., Sarwar, S., Saad Rehman, R., Hassan, M., Masood Ahmad, R., A. Gill, R., Ali, B., Al-Ashkar, I., Ibrahim, A., Atikur Rahman, M., & el Sabagh, A. (2023). Estimation of genetic divergence and character association studies in local and exotic diversity panels of soybean (*Glycine max* L.) genotypes. *Phyton*, 92(6), 1887–1906. <https://doi.org/10.32604/phyton.2023.027679>