

Check for
updates

Research Article

Harnessing Soybean Diversity: Evaluating Introduced Genotypes for Sustainable Yield in Pakistan

Zulqar Nain¹, Zaheer Ahmed², Ahmed Raza³, Maqsood Ali⁴, Uzair Ahmad¹, Faiz Ullah Arif¹, Muhammad Nasir¹, Muhammad Qasim¹, Muhammad Rizwan Shafiq¹, Sidra Iqbal¹

¹Department of Plant Breeding and Genetics, University of Agriculture Faisalabad, Constituent College, Depalpur, Okara, Pakistan.

²Center for Advanced Studies in Agriculture and Food Security, University of Agriculture Faisalabad, Pakistan.

³Agriculture Testing Lab, Punjab Agriculture Food and Drug Authority, Punjab Science Enclave, Multan Road, Lahore-53700, Pakistan.

⁴Department of Plant Pathology, University of Agriculture Faisalabad, Constituent College, Depalpur Okara, Pakistan.



ABSTRACT

Food security has become a challenging issue in the altered climate pattern, especially for developing countries. Soybean (*Glycine max* L.), a crop having high protein and edible oil content, has great potential but remains underutilized in Pakistan due to the lack of diverse, locally adapted genotypes. This study aimed to assess the genetic potential of fifteen exotic soybean genotypes, including one check cultivar, Faisal Soybean, by investigating various yield-related traits under the agroclimatic conditions of Okara, Pakistan. This experiment was conducted under a Randomized Complete Block Design (RCBD) with three replications. Analysis of variance revealed highly significant ($p < 0.01$) genetic variation for plant height (MS = 93.58), number of pods per plant (MS = 152.46), and number of seeds per pod (MS = 1.407). Principal component analysis (PCA) showed that the first two components accounted for 73.02% of the cumulative variability in quantitative traits and 86.12% in qualitative traits, with plant height and pods per plant identified as key yield-related contributing traits. Positive correlations were observed among several traits, such as plant height, primary branches ($r = 0.64$), pods per plant ($r = 0.41$), and seeds per pod ($r = 0.62$). Cluster analysis grouped the genotypes into distinct clusters, with genotypes G1, G10, and G11 showing unique characteristic profiles, while G4, G5, G13, and G15 formed a closely related group. These findings highlight considerable genotypic variation in major traits, contributing to the development of high-yielding, climate-resilient varieties adapted to local conditions.

Keywords: Soybean; Genotypic Variation; Yield Traits; Principal Component Analysis; Cluster Analysis.

INTRODUCTION

Soybean (*Glycine max* L.) is an annual oilseed crop that belongs to the family Leguminosae. It is an important legume crop utilized for animal feed and human consumption (Singh et al. 2025). Its seed contains 40 to 42% protein, 20 to 22% edible oil, along with several other essential minerals and vitamins (Amol et al. 2021). Soybeans grows in well-drained loamy soils with moderate fertility, neutral pH, sufficient organic matter, adequate rainfall, optimum temperatures (20 to 30°C), and proper sunlight. These factors significantly enhance growth and yield (Suparwata and Jamin, 2024).

Globally, the soybean crop is cultivated on an area of 140.47 million hectares, and the seed production is about 394.97 million metric tons. The USA, Brazil, and Argentina are major producers of the soybean crop (USDA, 2025).



Correspondence

Sidra Iqbal

dr.sidraiqbal@uaf.edu.pk

Article History

Received: February 03, 2026

Accepted: March 17, 2026

Published: March 30, 2026



Copyright: © 2024 by the authors.

Licensee: Roots Press, Rawalpindi, Pakistan.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license:
<https://creativecommons.org/licenses/by/4.0>

Soybeans are native to East Asia (China, Korea, and Japan), where they have been cultivated for about 5,000 years. In the 18th century, it spread to Europe, the USA, and other countries. Researchers also consider Australia and East Africa as possible centers of diversity for the genus *Glycine* (Jeong et al. 2019).

In Pakistan, soybean was introduced as an oilseed crop in the early 1960s. During the 1970s, adaptability and yield trials were conducted across the country in regions such as Khyber Pakhtunkhwa, Punjab, and Sindh to identify suitable varieties for cultivation (Asad et al. 2020). The maximum cultivated area was recorded in 1994 at 6613 hectares (Khurshid et al. 2017). Despite several decades passing, it has not gained popularity among farmers nationwide. The main reasons for the soybean shortage in Pakistan include limited breeding programs, low genetic improvement, a lack of suitable cultivated varieties, and different cropping patterns (Abbas, 2023).

The growing population and changing dietary habits in Pakistan are increasing the demand for edible oil. Since local oilseed production is insufficient, the country relies on imports, which place a significant burden on the economy due to high import bills (Zaidi, 2014). According to the Pakistan Economic Survey 2023-24, the total supply of edible oil from all sources is estimated at 3.188 million tonnes, of which domestic production is only 0.471 million tonnes. To bridge the gap between domestic production and consumption, 2.717 million tonnes of edible oil were imported worth 2,079 billion (US\$ 2.809 billion) (Pakistan Economic Survey, 2023-24).

Considering the several benefits of soybeans, there is an urgent need for a good and stable policy regarding production awareness for farmers to enhance its cultivation (Akram and Ahmad, 2019). To maximize the potential of soybean crops, it is important to develop high-yielding and climate-adapted varieties that can be grown under specific agronomic conditions. Genetic variation plays a critical role in improving yield, and understanding the relationship between variation and yield-related traits is critical for successful breeding (Mahbub et al. 2016). Genetic diversity among genotypes is essential for breeding programs because the segregating nature of diverse parents produces better heterosis and high-yielding cultivars. Assessing genetic variability helps in selecting systematic parents to improve yield through correlation analysis (Mahbub and Shirazy, 2016).

To identify valuable characteristics of genotypes that are helpful in future breeding programs, various statistical techniques, including ANOVA, PCA, correlation, and cluster analyses are performed to identify elite parents that would be suitable for hybridization. While several previous Pakistani studies on soybean have focused on individual yield-related components (Ullah et al. 2021), this study aimed to evaluate genetic variation in fifteen exotic soybean genotypes under the agroclimatic conditions of Okara, Pakistan. The goal was to identify favorable parents for future breeding programs focused on developing high-yielding and climate-resilient varieties.

MATERIALS AND METHODS

Experimental site and material

The experiment was conducted from August 25, 2022, to April 2023 at the research area of the University of Agriculture Faisalabad Constituent College, Depalpur, Okara, Pakistan. Experimental material consisted of fifteen exotic soybean genotypes, including one locally cultivated variety (Faisal soybean), which was used as a check line (Table 1).

Experimental design and field layout

This experiment was conducted under a Randomized Complete Block Design (RCBD) with three replications. The total experimental plot area was 203 square meters, with each genotype sown in two rows. The row length was 3 meters, the row-to-row spacing was 0.75 meters, and the plant-to-plant distance was 3 inches. Sowing was done using a manual seed drill to ensure seed depth and maintain the desired plant population in a specific area. All intercultural operations, such as irrigation, weeding, and fertilizer application, were carried out according to standard recommendations as described by Mondal and Wahhab (2001).

Data collection

Five plants were randomly selected and tagged from each genotype at the stage of flower initiation, resulting in a total sample size of 15 plants per genotype. Eight morphological traits were recorded, including:

Quantitative traits:

Plant height (PH): The plant height was measured from the base to the tip of the main stem for all selected plants at maturity by using a measuring scale.

Primary branches (PB): The number of primary branches was recorded by counting the branches that were arising from the main stem of the plant.

Number of pods per plant (NPP): The number of total pods was counted from selected plants at maturity.

Number of seeds per pod (NSP): NSP were counted from 10 randomly selected pods per plant, and the average number of seeds of 10 pods was recorded.

Number of leaflets (NL): The number of leaflets was counted from the fully expanded leaves in the center of the main stem of the plant.

Qualitative traits:

Pubescence (Pub): Pubescence was observed based on the presence or absence of hairs on stems and pods.

Leaf pubescence (LP): Leaf pubescence was observed based on the presence or absence of hairiness on the leaf surface.

Leaf pubescence density (LPD): Leaf pubescence density was observed based on the density of hairs on the leaves of the whole plant.

Statistical analysis details

The data were analyzed using different statistical methods including Analysis of variance (ANOVA), with significance tested at $p < 0.05$ (*) and $p < 0.01$ (**) (Johnson et al. 1955), Principal component analysis (PCA) to assess the structure of the data and determine the main contributing traits (Jolliffe, 2002). Correlation analysis was employed to determine the correlations among traits, employing Pearson's correlation coefficient formula, with significance determined at $p < 0.05$ and $p < 0.01$ (Pearson, 1895). Cluster analysis was used to group genotypes into clusters using hierarchical clustering (Ward Jr, 1963). All the statistical analyses were performed in Statistix 8.1 software (Analytical Software, 2005), as well as R software (Team, 2024).

Table 1. List of the fifteen soybean genotypes used in this research.

Sr. No.	Genotype No.	Genotype Name
1	G1	UAFDO-SL-01
2	G2	UAFDO-SL-77
3	G3	UAFDO-SL-99
4	G4	UAFDO-SL-101
5	G5	UAFDO-SL-187
6	G6	UAM-SB-47
7	G7	UAM-SB-162
8	G8	UAM-SB-200
9	G9	FAISAL SOYBEAN (Check)
10	G10	ORI-Soy-25
11	G11	ORI-Soy-91
12	G12	NBG-M104
13	G13	NBG-403
14	G14	NBG-Soy-SG
15	G15	PGRA-9

RESULTS

Analysis of variance (ANOVA) was performed to measure the genetic variation within fifteen soybean genotypes, including one locally cultivated check variety (G9: Faisal Soybean) for several qualitative and quantitative traits. Analysis of variance revealed highly significant differences ($p < 0.01$) between plant height (MSS = 93.58, EMS = 20.83), number of pods per plant (MSS = 152.46, EMS = 41.896), and number of seeds per pod (MSS = 1.407, EMS = 0.167). Plant Height, number of pods per plant, and number of seeds per pod displayed highly significant results across all genotypes. The genotypic mean square was greater than the error mean square, which endorsed the presence of strong genetic effects. While other traits, such as primary branches (PB), number of leaflets (NL), leaf pubescence (LP), Leaf pubescence density (LPD), and pubescence (Pub), revealed insignificant differences among genotypes, these traits indicated limited genetic variation within the genotypes (Table 2).

Coefficient of variation (CV%) is a useful statistical measure to check the variability of the traits around their mean. Usually, the lower CV value suggests well experimental consistency and better control of environmental variations. For instance, CV values ranged from 6.67% for the number of leaflets (NL) to 78.94% for pubescence (LP). In this study, traits like NL (6.67%) and PH (13.00%) exposed low CV values, reflecting more accuracy and consistency in data. Conversely, a higher CV value was observed for the trait LP (78.94%), which may highlight a stronger environmental influence, which can reduce the reliability of conclusions drawn for those traits.

Table 2. ANOVA table for all traits of 15 soybean genotypes, values represent mean sum square (MSS).

SOV	DF	PH	NPP	NSP	PB	NL	LP	LPD	Pub
Replication	2	35.544	102.18	0.120	0.323	0.088	0.622	0.088	0.022
Genotype	14	93.580**	152.46**	1.407**	0.558	0.041	0.184	0.212	0.022
Error	28	20.834	41.896	0.167	0.477	0.041	0.241	0.279	0.022
CV	--	13.00	29.51	13.81	20.63	6.67	78.94	37.16	15.25

Note: ** indicates significance at $p < 0.01$, * indicates significance at $p < 0.05$.

Principal component analysis

Principal Component Analysis (PCA) was performed for both quantitative and qualitative traits separately. PC1 to PC5 represented the variance summary for quantitative traits, and PC6 to PC8 represented the variance summary for qualitative traits (Table 3). In quantitative traits, the first two principal components (PC1 and PC2) accounted for 52.24% and 20.77% of the variance, respectively, cumulatively explaining 73.02% of the total variance. This indicates that most of the variation between soybean genotypes can be explained by these two components. PC1 solely accounted for 52.24%, highlighting its dominance in explaining variation within these traits. Whereas qualitative traits were best summarized by PC1 (57.23%) and PC2 (28.89%), which together explained 86.12% of the overall variability. This indicates that only the first two factors efficiently describe the majority of the variations in pubescence related characteristics.

Table 3. Summary of PCA results for quantitative and qualitative traits.

Component	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard Deviation	1.616	1.019	0.838	0.632	0.497	1.31	0.931	0.645
Proportion of Variance	0.522	0.208	0.14	0.08	0.049	0.572	0.289	0.139
Cumulative Proportion	0.522	0.73	0.871	0.951	1	0.572	0.861	1

PC1 to PC5 represent the quantitative traits summary, and PC6 to PC8 represent the qualitative traits summary.

The PCA biplots graphically exhibited the distribution and contribution of genotypes to the trait. Soybean genotypes grouped according to quantitative traits. NPP, PH, PB, NL, and NSP, which are positively correlated, comprise the first principal component (PC1). NL, however, was aligned with PC2.

This implies that genotypes along PC1 vary in terms of yield-related traits, whereas genotypes along PC2 differ with respect to leaf structure. Genotypes G1, G10, and G11 were located distantly, which suggests distinct trait profiles, while genotypes G4, G5, G13, and G15 grouped closely together, proposing similarity in trait expression (Figure 1).

Qualitative traits emphasized the impact of Pb and LPD along PC1, while LP contributed to PC2. Genotypes along PC1 were highly differentiated based on characteristics of pubescence, whereas genotypes along PC2 varied in terms of leaf pubescence. The biplot clearly differentiated genotypes with dense pubescence and those with sparse leaf cover, validating the role of qualitative parameters in classification. Genotype G1 was positioned alone, reflecting a distinct expression of traits, whereas genotypes G4, G11, and G13 formed a tight cluster, reflecting similarity in pubescence related characteristics (Figure 2).

Correlation analysis

Correlation coefficient analysis provides insights into the associations among traits. The Correlation heatmap graph showed the strength and direction of the relationships among the traits (Figure 3). The correlation matrix revealed several important associations between traits.

The numbers of primary branches (PB) indicated a strong correlation with both plant height (PH: $r = 0.64$) and number of seeds per pod (NSP: $r = 0.74$) among the quantitative traits. NSP exhibited a positive relationship with both the number of pods per plant (NPP: $r = 0.40$) and PH ($r = 0.62$), whereas the number of leaflets (NL) showed weak and mostly negative correlations with other quantitative traits, except for a positive relationship with Leaf pubescence density (LPD; $r = 0.37$).

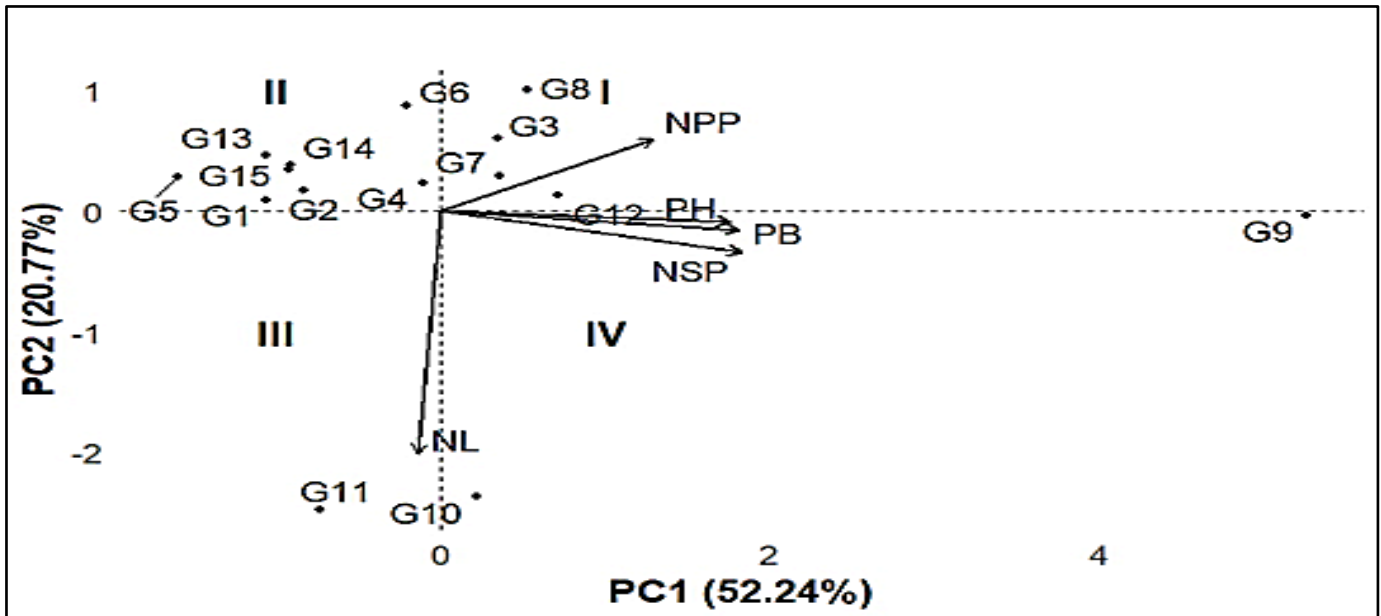


Figure 1. Principal component analysis (PCA) biplot for quantitative traits of 15 soybean genotypes. Vectors represent quantitative traits, PH = plant height, PB = primary branches, NPP = number of seeds per plant, NSP = number of seeds per grain, and NL = number of leaflets. Each point represents an individual genotype (G1-G15). This biplot explains 73.02% of the cumulative variation (PC1 = 52.24%, PC2 = 20.77%).

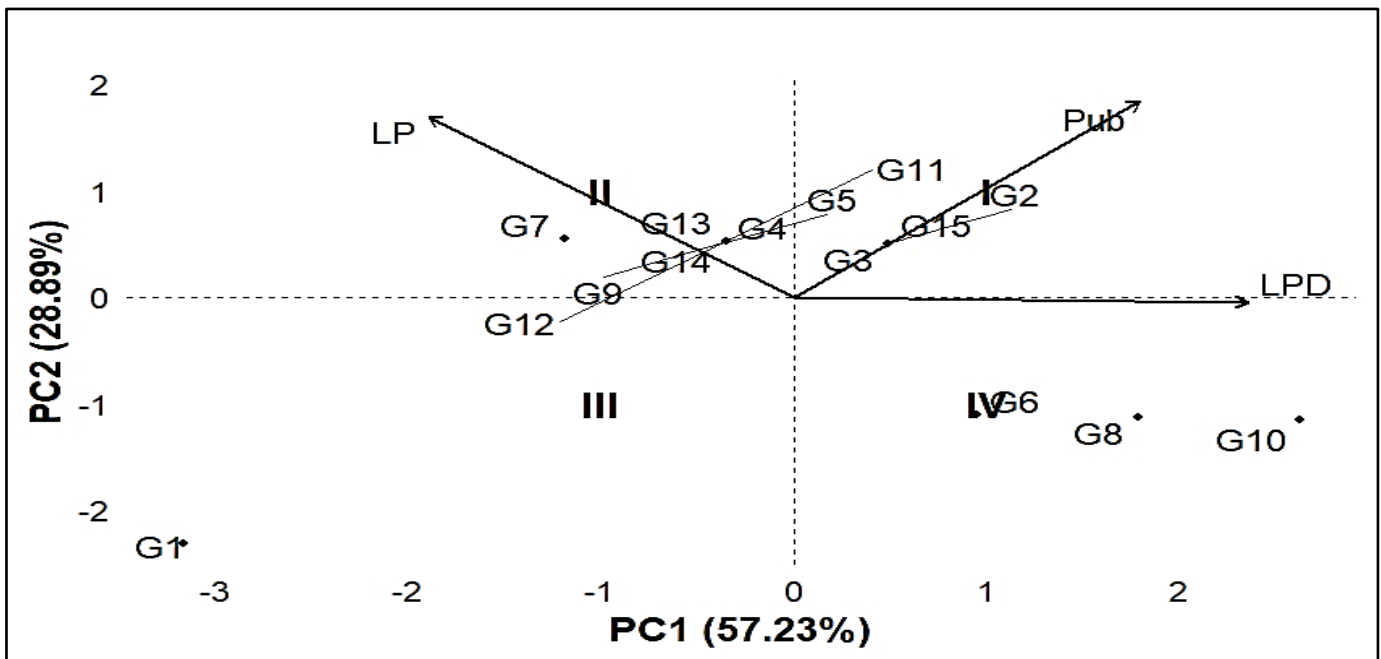


Figure 2. Principal component analysis (PCA) biplot for qualitative traits of 15 soybean genotypes. Vectors represent qualitative traits, Pub = pubescence, LP = leaf pubescence, and LPD = leaf pubescence density. Each point represents an individual genotype (G1-G15). This biplot explains 86.12% of the cumulative variation (PC1 = 57.23%, PC2 = 28.89%).

An exchange between the two traits may exist, as suggested by the negative correlation between NPP and leaf pubescence (LP; $r = -0.64$). Pubescence (Pb), identified as one of the qualitative traits, exhibited a weak relation with PB ($r = 0.14$) and a significant relation with LPD ($r = 0.44$). While LP and LPD ($r = -0.48$) showed a marginally negative association with each other, indicating that the lower leaf pubescence is correlated with denser pubescence (Table 4). The Correlation Coefficient analysis described the associations among traits based on degree and direction as direct or indirect effects on yield. It helps to identify the superior parents for the future breeding programs.

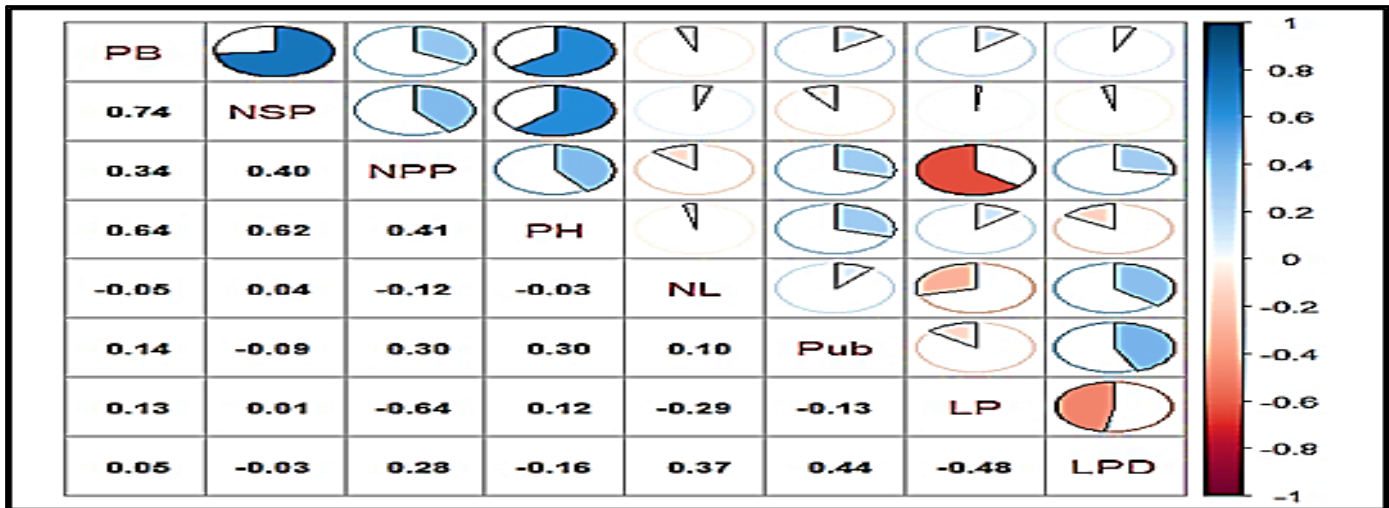


Figure 3. Correlation heatmap showing Pearson correlation coefficients among quantitative and qualitative traits in 15 soybean genotypes. Blue pie circles indicate positive correlations, and red pie circles indicate negative correlations. The size and color intensity of the pie circle indicate the strength of the associations. Abbreviations: PH = plant height, PB = primary branches, NPP = number of pods per plant, NSP = number of seeds per pod, NL = number of leaflets, Pub = pubescence, LP = leaf pubescence, LPD = leaf pubescence density.

Table 4. Pearson correlation coefficients among quantitative and qualitative traits in 15 soybean genotypes.

Trait	PB	NSP	NPP	PH	NL	Pub	LP	LPD
PB	1.00	0.74**	0.34*	0.64**	-0.05	0.14	0.13	0.05
NSP	0.74**	1.00	0.40*	0.62**	0.04	-0.09	0.01	-0.03
NPP	0.34*	0.40*	1.00	0.41*	-0.12	0.30*	-0.64**	0.28*
PH	0.64**	0.62**	0.41*	1.00	-0.03	0.30*	0.12	-0.16
NL	-0.05	0.04	-0.12	-0.03	1.00	0.10	-0.29	0.37*
Pub	0.14	-0.09	0.30*	0.30*	0.10	1.00	-0.13	0.44*
LP	0.13	0.01	-0.64**	0.12	-0.29	-0.13	1.00	-0.48**
LPD	0.05	-0.03	0.28*	-0.16	0.37*	0.44*	-0.48**	1.00

Cluster Analysis.

The quantitative trait hierarchical clustering (Number of Primary Branches, Number of Seeds per Pod, Number of Pods per Plant, Plant Height, and Number of Leaflets) provided an in depth distinction among genotypes based on numerical variations in traits (Figure 4). The check cultivar, G9 (Faisal Soybean), was most divergent, forming an independent group, and thus, a noticeable difference in their quantitative trait expression compared to other genotypes.

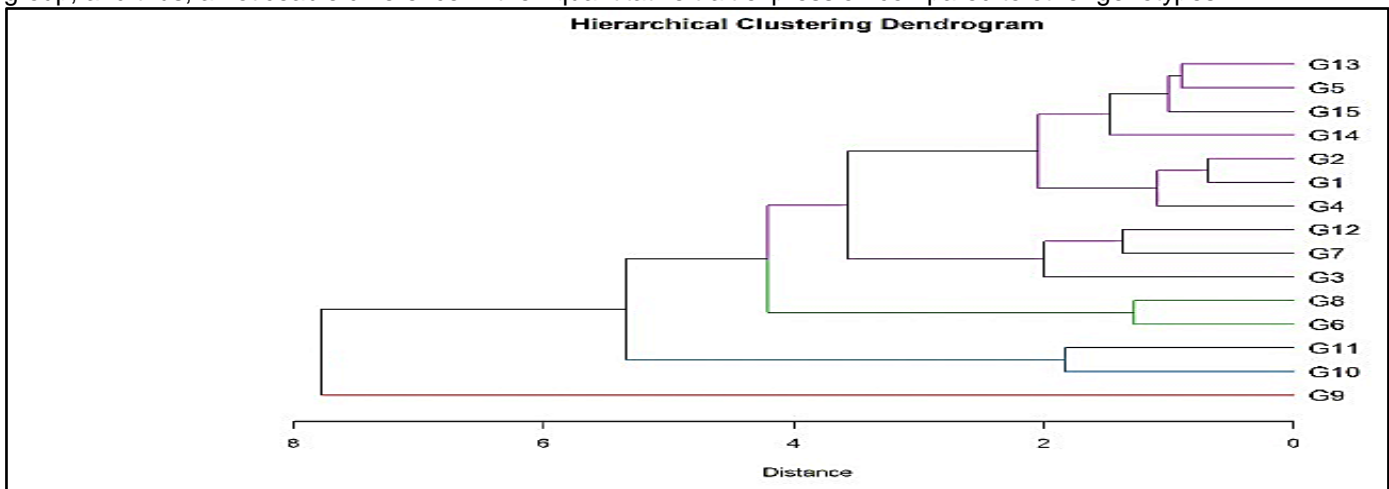


Figure 4. Hierarchical clustering dendrograms of 15 soybean genotypes based on quantitative traits. This dendrogram was built using Ward's method with Euclidean distance.

The genotypes, G10 and G11, were also highly distinguishable from each other, showing separate expression of

traits that set them apart from the general population. A large cluster formed by G2, G4, G12, G3, G7, and others revealed a greater association, thereby showing relatively similar quantitative distributions of traits. The clear separation of G6 and G8 into a different cluster also indicated high genetic diversity among the investigated soybean genotypes. The closest and farthest distances between clusters for quantitative traits are shown as the genetic differences among genotype clusters (Table 5).

Table 5. Nearest and Farthest Cluster Distances for Quantitative Trait Clustering of soybean genotypes.

Cluster	Nearest Cluster	Distance	Farthest Cluster	Distance
G9	G10	2.1	G13	7.8
G10	G11	1.9	G6	6.5
G11	G8	2.0	G15	6.3
G2	G4	1.7	G9	6.9
G6	G8	1.5	G9	7.2

Hierarchical clustering among the qualitative characteristics such as Pubescence, Leaf Pubescence, and Leaf Pubescence Density assigned the 15 soybean genotypes into different groups based on their common traits (Figure 5). The genotypes were segregated by similarity, with the check variety G9 grouped with G5 and G11 at a moderate distance. The pattern of clustering indicated that these genotypes have similar expressions of qualitative traits. One of the largest clusters was composed of genotypes G4, G12, G13, G14, and G7, which showed close similarity in their qualitative traits. On the other hand, G6, G8, and G10 belonged to a different cluster exhibiting their unique set of traits. The hierarchical dendrogram portrays the genetic diversity among the genotypes, which is vital for breeding schemes aimed at improving qualitative traits. The farthest and nearest cluster distances for qualitative traits provided information on the levels of similarity and diversity among genotypes (Table 6).

Table 6. Nearest and Farthest Cluster Distances for Qualitative Trait Clustering of soybean genotypes.

Cluster	Nearest Cluster	Distance	Farthest Cluster	Distance
G9	G5	1.8	G6	5.7
G5	G11	1.6	G10	5.2
G11	G9	1.8	G8	5.3
G4	G12	1.4	G6	5.9
G12	G13	1.2	G9	4.8

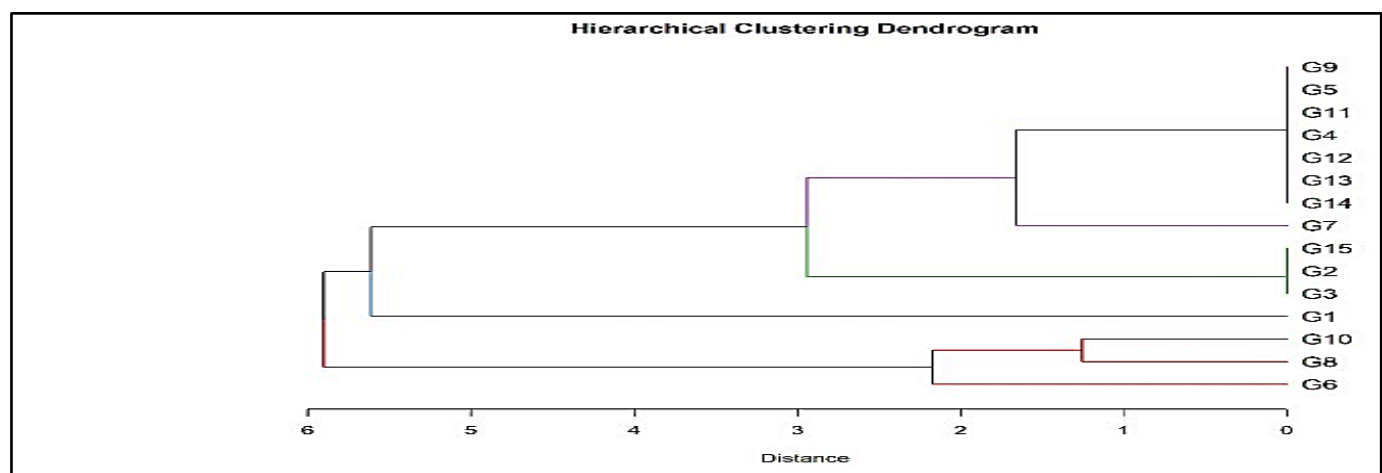


Figure 5. Hierarchical clustering dendrograms of 15 soybean genotypes based on qualitative traits. These dendrograms were built using Ward's method with Euclidean distance.

Both qualitative and quantitative traits showed clustering groups that provided valuable information about genotype selection for breeding plans. To support the trait enhancement in soybean breeding programs, the genetic distance between clusters identified possible parental lines for hybridization. Additionally, the ten

greatest and smallest inter-genotypic distances highlighted important genotypic relationships that will be essential for future breeding programs (Table 7).

Table 7. Ten highest and ten lowest inter-genotypic distances among the 15 genotypes of soybean.

Genotype Pair	Distance
G9 – G13	7.8
G6 – G9	7.2
G10 – G6	6.5
G9 – G2	6.9
G11 – G15	6.3
G9 – G5	5.7
G10 – G8	5.3
G5 – G11	1.6
G6 – G8	1.5
G4 – G12	1.4

Identification of superior genotypes.

Based on statistical analyses, potential high-yielding genotypes including UAFDO-SL-01, ORI-Soy-25, and ORI-Soy-91 were identified, which surpassed in plant height, number of pods, number of seeds per pod, and distinct trait profiles. These genotypes are recommended as potential parent genotypes for improved yield components.

Genotypes UAFDO-SL-101, UAFDO-SL-187, NBG-403, and PGRA-9 exhibited balanced trait expression and were found to be suitable for varietal development. The genotype UAFDO-SL-01 with dense pubescence was recommended for pest resistance.

Locally cultivated variety Faisal soybean was used as a check variety for comparison. It showed moderate performance but was genetically different from most of the genotypes. This suggests that crossing Faisal soybean with test genotypes, especially UAFDO-SL-01, ORI-Soy-25, or ORI-Soy-91, can provide sufficient genetic diversity for selection.

DISCUSSION

Soybean is an important legume crop known for its high protein and edible oil content, properties that make them essential for food and animal feed. It grows optimally in well-drained soils, warm climates, and short day length. In Pakistan, although the growing conditions are favorable, soybeans have not become very popular because of insufficient breeding programs, lack of superior varieties, and inappropriate cropping patterns. Pakistan relies heavily on imported edible oil, straining the economy. Boosting domestic production of oilseed crops can contribute in fulfilling growing demand and reducing import expenses. To achieve this, farmer awareness needs to be enhanced, and high-yielding and climate-resilient varieties should be developed. This research aimed at assessing the genetic diversity of desirable exotic soybean genotypes to determine the most suitable lines for future breeding programs.

The results of this research presented a pivotal analysis of genetic diversity and variation of traits among soybean genotypes, focusing on important factors that control their phenotypic expression. Plant height, number of pods per plant, and number of seeds per pod indicated significant variation among genotypes. Similar findings have been reported by Malek et al. (2014), who observed significant genetic variation in yield-related traits in soybean germplasm. Stability in qualitative traits (Pub, LP, LPD) indicates no genetic differences among whole genotypes, consistent with Ullah et al. (2024), who reported stability in qualitative traits across soybean genotypes. These genetic variations provide raw material for selection and hybridization for further use in breeding programs.

PCA revealed that a large amount of the variability was contributed by the first two principal components, of which PC1 covers the bulk of the variation between qualitative and quantitative traits. Previous studies have also described the plant height, seeds per pod, and number of primary branches as the significant contributors to soybean genetic variation. These results are consistent with Wang et al. (2020), who identified plant height and pod size as important differentiating factors in soybean genetic resources. Similarly, Kumar et al. (2015) reported significant PCA loading values of plant height and seed traits, the importance of pubescence and density of leaf pubescence in PC1 in qualitative traits emphasizing their role in genotypic differentiation, supporting previous research on the contribution of pubescence to genetic diversity in legumes by Iqbal et al. (2008). The clear positional relationships between

genotypes G1, G10, and G11 in the PCA biplot support the unique characteristics of these genotypes and demonstrate their importance in introducing new genetic variations into Pakistan's breeding program.

The correlation analysis revealed significant positive associations of primary branches, plant height, and seed-related traits. The strong correlation between plant height and seeds per pod indicated that taller plants have greater reproductive potential, a trend also reported in high-yielding soybean varieties. Dubey et al. (2015) reported similar positive correlations between yield components of high-yielding soybean varieties, supporting the reliability of this association. Whereas the negative association between pods per plant and leaf pubescence suggested an anticipated possible trade-off between yield potential and adaptive trait. Moe and Girdthai (2013) also showed that pubescence traits can affect physiological processes such as water retention and pest resistance, which can affect yield. Weak correlations were observed between leaflet number and other traits, indicating that leaflet number may not be a major determinant of the yield potential, consistent with Xu et al. (2024) in other legume crops.

The hierarchical clustering analysis also confirmed the genetic diversity between soybean genotypes. The clustering of Faisal Soybean (G9) with G5 and G11 for qualitative characters suggested genetic similarities among them, whereas its clustering in quantitative trait clusters signified that it possesses unique agronomic traits. The identification of big clusters containing multiple genotypes implied that some groups have similar genetic backgrounds, which could be advantageous for breeding programs aimed at improving certain traits. Consistent with Al-Hadi et al. (2017), the genetic distances that exist between clusters give important information to use in parent selection during hybridization programs, as hinted in earlier research on the genetic diversity of soybean.

Identifying genetically distinct genotypes with complementary traits offers practical opportunities for soybean improvement in Pakistan. Genotypes UAFDO-SL-01, ORI-Soy-25, and ORI-Soy-91 can serve as parental lines for crossbreeding, while closely related clusters such as UAFDO-SL-101, UAFDO-SL-187, NBG-403, and PGRA-9 support trait stabilization and varietal development. As Nasir et al. (2025) demonstrated, these strategies address the challenge of limited systematic breeding material in Pakistan.

Overall, the findings of this research align with earlier studies on soybean genetic variation and offer important insights for breeding strategies. The strong differentiation observed in traits like plant height and pubescence indicates that these traits should be emphasized in screening programs for yield improvement and adaptability. The observed correlations among traits related to yield also indicated the possibility of using trait-based selection to improve productivity. Future research should investigate the molecular basis of these variations and further elucidate the genetic basis of these traits, which will enable the development of high-yielding, climate-resistant soybean varieties for Pakistan.

CONCLUSION

The study highlighted the significant genetic diversity among fifteen soybean genotypes assessed under agroclimatic conditions in Pakistan. Key morphological traits, such as PH, NPP, and NSP exhibited significant variation, suggesting their potential importance for future breeding programs. Promising candidates for the production of climate resilient and high yielding varieties were identified by the PCA and cluster analysis, which clearly differentiated genotypes based on the performance of qualitative and quantitative traits. The positive correlations among PH, PB, NPP, and NSP indicated the potential for improving overall yield. In addition to providing a solid foundation for soybean cultivar development initiatives. Three genotypes, UAFDO-SL-01, ORI-Soy-25, and ORI-Soy-91, were identified as superior genotypes and recommended for future breeding programs.

ACKNOWLEDGMENT

The authors gratefully acknowledge Prof. Dr. Zaheer Ahmed at Center for Advanced Studies in Agriculture and Food Security, University of Agriculture, Pakistan for providing the seed material used in this study. The authors are also thankful to the University of Agriculture Faisalabad, Constituent College Depalpur (Okara), for providing the necessary research facilities and technical support to successfully conduct the experiment.

AUTHOR CONTRIBUTIONS

Zulqar Nain: Data collection, statistical analysis and manuscript drafting. Zaheer Ahmed: Provided seed material. Ahmed Raza: Technical support, sowing planning, field management. Muhammad Qasim, Maqsood Ali, Uzair Ahmad, Faiz Ullah Arif, Muhammad Nasir: Field management and data collection. Muhammad Rizwan Shafiq: Data

verification, manuscript review. Sidra Iqbal: Conceptualization, study supervision, research design, critical revision of manuscript, final approval of version for submission. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

COMPETING OF INTEREST

The authors declare no competing interests.

REFERENCES

- Analytical Software. (2005). Statistix 8.1 user's manual. Analytical Software.
- Al-Hadi, G., R.M. Islam, A.M. Karim and T.M. Islam. 2017. Morpho-physiological characterization of soybean genotypes under subtropical environment. *Genetika* 49:297-311.
- Akram, Z. and Q. Ahmad. 2019. Future prospects of soybean in Pakistan. *Biomed J Sci Tech Res* 15:11562-11563.
- Asad, S.A., M.A. Wahid, S. Farina, R. Ali and F. Muhammad. 2020. Soybean production in Pakistan: experiences, challenges and prospects. *International Journal of Agriculture and Biology* 24:995-1005.
- Amol, V., Bhati, K. R., & Bhati, K. R. (2021). Nutritive benefits of soybean (*Glycine max*). *Indian J Nutr Diet*, 522-33.
- Abbas, G. (2023). Effect of Soybean Unavailability Situations and COVID-19 on the Poultry Industry of Pakistan: A Comprehensive Analysis Problems Faced and Its Solution for Sustainable Animal Production. *Pakistan Journal of Science*, 75(2).
- Dubey, N., A. Shrivasthava, H. Avinash and S. Jaiwar. 2015. Genetic variability, correlation and path analysis for yield and yield contributing characters in soybean (*Glycine max* L.). *Electronic Journal of Plant Breeding* 6:318-325.
- Iqbal, Z., M. Arshad, M. Ashraf, T. Mahmood and A. Waheed. 2008. Evaluation of soybean [*Glycine max* (L.) Merrill] germplasm for some important morphological traits using multivariate analysis. *Pakistan Journal of Botany* 40:2323-2328.
- Johnson, H.W., H. Robinson and R. Comstock. 1955. Estimates of genetic and environmental variability in soybeans.
- Jolliffe, I.T. 2002. *Principal component analysis for special types of data*. Springer.
- Jeong, S.-C., J.-K. Moon, S.-K. Park, M.-S. Kim, K. Lee, S.R. Lee, N. Jeong, M.S. Choi, N. Kim and S.-T. Kang. 2019. Genetic diversity patterns and domestication origin of soybean. *Theoretical and Applied Genetics* 132:1179-1193.
- Kumar, A., A. Pandey, C. Aochen and A. Pattanayak. 2015. 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* 85:397-405.
- Khurshid, H., D. Baig, S. Jan, M. Arshad and M. Khan. 2017. Miracle crop: the present and future of soybean production in Pakistan. *MOJ Biology and Medicine* 2:189-191.
- Mondal, M. and M. Wahhab. 2001. *Production Technology of Crops*. Oil seed Research Center, Bangladesh Agril. Res. Inst. Joydebpur, Gazipur:1-10.
- Moe, S. and T. Girdthai. 2013. Relationships of soybean [*Glycine max* (L.) Merrill] accessions based on physiological and agro-morphological traits. *International Journal of Chemical, Environmental & Biological Sciences (IJCEBS)* 1:375-380.
- Malek, M., M.Y. Rafii, M. Shahida Sharmin Afroz, U.K. Nath and M.M.A. Mondal. 2014. Morphological characterization and assessment of genetic variability, character association, and divergence in soybean mutants. *The Scientific World Journal* 2014:968796.
- Mahbub, M., M.M. Rahman, M. Hossain, L. Nahar and B. Shirazy. 2016. Morphophysiological variation in soybean (*Glycine max* (L.) Merrill). *American-Eurasian Journal of Agricultural & Environmental Sciences* 16:234-238.
- Mahbub, M.M. and B.J. Shirazy. 2016. Evaluation of genetic diversity in different genotypes of soybean (*Glycine max* (L.) Merrill). *American Journal of Plant Biology* 1:24-29.
- Nasir, B., Rahman, S. U., Ali, A., Shafique, E., Zia, N., Ahmad, N., ... & Bukhari, R. (2025). Resilient soybeans for a changing climate: Analyzing traditional and emerging new plant breeding technologies to combat abiotic stresses. *Acta Physiologiae Plantarum*, 47(11), 102.
- Pearson, K. 1895. *Contributions to the Mathematical Theory of Evolution*. III. Regression, Heredity, and Panmixia. *Proceedings of the Royal Society of London* 59:69-71.
- Pakistan Economy. 2023. *Agriculture*. *Crops* 3:3-006.
- Suparwata, D. and F. Jamin. 2024. Analysis of Organic Fertilizer Use in Improving Soil Quality and Agricultural Yields in Indonesia. *West Science Agro*, 2 (01), 17–27.
- Singh, K. H., Gupta, S., Shivakumar, M., & Nataraj, V. (2025). Soybean [*Glycine max* (L.) Merr.] Breeding. In *Fundamentals of Legume Breeding: A Text for Students and Practitioners* (pp. 283-298). Singapore: Springer Nature Singapore.
- Team, R.C. 2024. *RA language and environment for statistical computing*, R Foundation for Statistical Computing.
- Ullah, A., Akram, Z., Malik, S. I., & Khan, K. S. U. (2021). Assessment of phenotypic and molecular diversity in

- soybean [*Glycine max* (L.) Merr.] germplasm using morpho-biochemical attributes and SSR markers. *Genetic Resources and Crop Evolution*, 68(7), 2827-2847.
- Ullah, A., Z. Akram, G. Rasool, M. Waris and H. Khurshid. 2024. Agro-morphological characterization and genetic variability assessment of soybean [*Glycine max* (L.) Merr.] germplasm for yield and quality traits. *Euphytica* 220:67.
- USDA. 2025 World Agricultural Production. Circular Series WAP. Foreign Agricultural Service, USDA, USA, p. 1-30.
- Ward Jr, J.H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* 58:236-244.
- Wang, G., Q. Zhou, M. He, X. Zhong and G. Tang. 2020. Wilting index and root morphological characteristics used as drought-tolerance variety selection at the seedling stage in soybean (*Glycine max* L.). *Plant Growth Regulation* 92:29-42.
- Zaidi, S.M.H. 2014. Edible oil imports in Pakistan. *South Asian Journal of Management* 8:1-8.