Genetic Diversity and Population Structure of Soybean (Glycine max L.) through Integrated Morphological and SSR Analyses

Sabeena Riasat
University of Agriculture
Farah Rasheed
University of Agriculture
Raheela Rehman
University of Agriculture
Rida Kanwal
University of Agriculture
Muhammad Hassaan Khan
University of Agriculture
Haris Moaz
University of Agriculture
Muhammad Atif
University of Agriculture
Usman Irfan
University of Bari Aldo Moro
Abdul Waheed
University of Agriculture
Ali Haider[2]
University of Agriculture
Muhammad Moaz Latif
Avignon Université - Jean-Henri Fabre Campus
Fatima Tul Kubra
University of Agriculture
Hameed Alsamadany
King Abdulaziz University
Hind A.S. Alzahrani
Imam Abdulrahman bin Faisal University
Zaheer Ahmed
zaheer.ahmed@uaf.edu.pk

University of Agriculture

Research Article

Keywords: Genetic diversity, Soybean (Glycine max), SSR markers, Morphological traits, Polymorphism Information Content (PIC), Germplasm characterization, Marker-assisted breeding

Posted Date: December 15th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7920345/v1

Additional Declarations: No competing interests reported.

Abstract

Background

Soybean (*Glycine max* L.Merr.) is an important crop for protein and oil production, and its genetic improvement relies on the availability of diverse germplasm. In this study, 20 soybean genotypes, including one locally adapted conventional cultivar (AARI-21) and nineteen exotic accessions from the USDA germplasm collection, were evaluated using morphological traits and validated 13 simple sequence repeat (SSR) markers to determine genetic diversity in oil composition and yield. This study was conducted to identify genetically diverse lines to be used in future soybean improvement programs.

Results

Significant phenotypic variation was observed, with PI-548512 and PI-556554 showing thicker stem diameters, while PI-556487 produced more branches and a distinct flowering pattern. Principal component analysis (PCA) indicated that the principal contributors to Dim1 were the stem diameter, cotyledon length, flowering days, seeds per pod, and the principal contributors to Dim2 were plant height, nodes, pods per plant, and 100-seed weight. On the molecular scale, SSR markers were highly polymorphic, with a mean PIC of 0.53. Markers like SR. SS001, SR.PR.1, and SR.OL4 were very informative (PIC > 0.5), and this indicates that they have high discriminatory ability. The genotypes were clustered according to the SSR banding patterns in a cluster analysis that identified three distinct clusters, indicating that there is a lot of genetic variation. Some of the most genetically variable identified lines were AARI-21 (local) and PI 232998 and PI 559393 (introductions).

Conclusions

The results reveal significant morphologgical and molecular diversity between the studied germplasm, which can be utilized in molecular-assisted breeding to enhance the adaptation, quality traits and yield potential of soybean.

INTRODUCTION

Soybean (*Glycine max* L.) is one of the most significant legume crops in the world. It is enriched with protein (35 to 42%) and oil (18 to 22%) [31]. Soybeans are used in food, feed, and biofuel processing due to their nutritional importance [9, 39, 55]. It is referred to as the golden bean, the queen of pulses, and the wonder crop [1, 27, 30], which possesses a fine source of key vitamins, including E, K, riboflavin, thiamine, niacin, and choline. Moreover, soybean consists of various bioactive compounds, such as isoflavones, chlorogenic acid, and ferulic acid [12, 36]. Currently, Brazil leads the world in soybean production as well as productivity [47]. Whereas other leading soybean-producing countries have recorded substantial growth in their yield over the past decades, Pakistan has not experienced growth,

which places it miles behind [37]. Soybean is cultivated on an area of over 120 million hectares all over the world [17] in comparison, the size of the land devoted to soybeans in Pakistan is pathetic [34]. Although many agro-ecological factors in the country allow its cultivation to be very successful, soybean has never been able to occupy a certain position in the state's cultivation pattern, mainly because of socioeconomic restrictions and insufficient policy support [4, 25).

Soybean was first cultivated in Northeast China about 1700–1100 B.C., and later expanded to Asia, Europe, and the United States [7, 8]. Currently, its main producers are Brazil, the United States of America, Argentina, China, and India, with the output as 169, 118.84, 49, 20.65, and 12.58 million metric tons, respectively [49]. In Pakistan, during the 1960s, the production of soybeans was initiated, but commercial attempts failed due to low return on investment, inadequate production practices, and unsuitable varieties. Lack of high-yielding, climate-resistant, and pest-resistant cultivars continues to be a key limitation [3, 24, 38].

It is estimated that the pressure of soybean diseases will only increase and become harder to control in the future, and such a problem is worsened by anthropogenic climate change. Over time, the world has experienced a warming rate of about 0.18°C per decade since the 1980s [32], and the high-emission scenarios project global warming of up to 6°C by the century's end [20]. Although some areas might have high temperatures and total precipitation, this will be typified by an increase in hydroclimatic variability. It involves increased rates of severe weather phenomena and more intense and extended droughts, which result in less predictable and more unpredictable growing conditions, contributing to the further evolution and propagation of pathogens [16].

Traditionally, genetic variation in soybeans has been investigated based on morphological characters, including flower color, leaf shape, and growth habit. These descriptors are, however, very much dependent on the environment, as they do not demonstrate actual genetic variation [46, 21]. To visualize actual variation, molecular markers are utilized nowadays. Diversity assessment in soybeans has been established using several different molecular markers, including random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP), inter-simple sequence repeats (ISSR), and single-nucleotide polymorphism (SNP), all of which come with disadvantages in terms of their low reproducibility and redundancy [2]. Simple sequence repeats (SSRs) are, however, more reliable for analysis because they are co-dominant, highly polymorphic, reproducible, and abundant throughout the genome [6, 43, 52]. Such properties make SSRs particularly useful in DNA fingerprinting, diversity surveys, and marker-based breeding in soybeans. DNA fingerprinting using SSRs offers accurate data on genetic diversity and helps to identify elite alleles in breeding programs [22, 26, 54]. However, the use of such molecular tools is relatively limited in Pakistan, leaving a gap in advancing enhanced soybean varieties.

The current research utilized a panel of validated 13 SSR markers that had been previously described as being linked to important seed quality measures, like the concentration of oil, protein content, and the abundance of specific fatty acids such as oleic, linolenic, and stearic acid. Such markers, being adjacent

to trait-associated genomic regions, provide an accurate means of identifying genotype-specific allelic variations. By targeting functional regions of the soybean genome, these SSRs enable both the assessment of genetic diversity and the identification of valuable lines with desirable nutritional profiles, thus enhancing the effectiveness of molecular breeding strategies. This study aims to evaluate the genetic diversity of 20 soybean genotypes using both morphological and SSR markers to identify genetically diverse lines that could be valuable in future soybean breeding programs.

MATERIALS AND METHODS

3.1. Plant Materials:

This study was carried out at the Farm at the University for Agriculture, Faisalabad (Geographic Coordinates: 31.4815 N, 73.0667 E) and in the Soybean Laboratory, Center for Advanced Studies in Agriculture and Food Security (CAS-AFS), University of Agriculture, Faisalabad. Twenty soybean genotypes (one local-adapted Pakistani cultivar (AARI-21) and nineteen introductions from the United States Department of Agriculture (USDA) soybean germplasm collection with a wide genetic representation were used in this research. The germplasm was imported from USDA in 2016 and followed all the quarantine measures. Seeds were sown in the Farm research area of University of Agriculture, Faisalabad following Randomized Complete Block design with three replications in spring 2024. Row to row and plant to plant distance was maintained at 18 and 3 inches and all the recommended agronomic practices were followed. One bag of SOP and half a bag of DAP were applied as basal dose while 20 Kg of urea was applied at the three leaf stage, flowering stage, and pod formation stage. Complete characterization information of all the accessions is provided in Supplemental Table 1. For DNA extraction, young and healthy leaf samples were collected from the field, instantly placed in liquid nitrogen and then taken to the Soybean Laboratory to extract genomic DNA.

3.2. Genomic DNA Extraction and SSR Analysis:

Genomic DNA from the young leaves of all genotypes was extracted using the GeneJET Plant Genomic DNA Purification Mini Kit (ThermoFisher Scientific, Cat. No. K0791), according to the manufacturer's instructions. The quality of the extracted DNA was validated by loading it onto a 1% TAE agarose gel, which was visualized under a UV transilluminator. Finally, the DNA was quantified with the help of a spectrophotometer, and the measurements were taken at 280 nm.

3.3. PCR Analysis

In the present experiment, 12 SSR primer pairs were designed using Primer3 software, and one primer set from the previous study (soysat001) was used. The list of primers, along with their sequences and descriptions, is provided in Supplemental Table 2 (Table S2). Polymerase Chain Reaction (PCR) was carried out using a thermocycler in a 20 µl total reaction volume, consisting of 2 µl template DNA (50 ng/µl), 0.5 µl of forward and reverse primer each (5 pmol), 7 µl of 2X DreamTaq Green PCR Master Mix (Thermo Fisher Scientific, Cat. No. K1081), and 10 µl of nuclease-free water. The amplification profile

was set as an initial denaturation at 94°C for 3 minutes, followed by denaturation (94°C, 1 min), a primer-specific annealing temperature for 1 min, and extension at 72°C for 1 min. A final elongation step was carried out at 72°C for 8 minutes for 35 cycles. PCR amplicons were resolved on 2% agarose gels (Thermo Fisher Scientific, Cat. No. R0611), and a 1 kb DNA ladder (Thermo Fisher Scientific, Cat. No. 10416014) was used as a molecular weight standard, and bands were visualized under a ChemiDoc MP imaging system (Bio-Rad, USA).

3.4. Morphological Characterization and Yield Component Analysis

The various morphological traits were recorded at the two-week stage of the plant, i.e., seedling height (SI, in cm), hypocotyl length (HL, in mm), cotyledon length (CL, in mm), plant height (PH, in cm), number of branches (NoB), stem diameter at maturity (StD, in cm), number of leaflets (NoL), flowering days (FD, in days), number of pods per plant (NoPP), number of seeds per pod (NoSP), seed diameter (SeD, in mm), and 100 seed weight (SW, in grams).

3.5. Data Analysis:

The data from the experiment were carefully documented and then analyzed with a set of statistical methods chosen because of their suitability for the experimental design. In this detailed analysis, analysis of variance (ANOVA) (Steel et al., 1997), t-tests, and the principal component analysis (PCA) (Pearson, 1901) of different parameters were applied, and the polymorphism information content (PIC) was calculated. The data were further interpreted by visualization in heatmaps depicting the profile of markers in the genotypes and the results using cluster analysis. All the analytical methods were carried out using the principles of established statistics.

RESULTS

This study assessed the genetic variability of 20 soybean genotypes (one local Pakistani cultivar and 19 accessions from USDA soybean germplasm collection) by analyzing both phenotypic characteristics and SSR-based molecular markers. The outcomes are described separately for morphological and DNA-level data.

4.1. Phenotypic Diversity:

Analysis of Morphological and Yield Related Traits:

Graphical representation of the average data revealed significant inter-genotypic differences for most morphological traits. PI 543855 exhibited the highest seedling height, differing significantly from all other genotypes. Hypocotyl length was notably high in PI-556519, whereas genotypes such as PI-556737, PI-556554, PI-548530, PI-548512, and PI-538774 showed no significant variation. Cotyledon length remained statistically similar among the genotypes, while cotyledon width showed significant divergence, with the local Pakistan line PI-546480 displaying the maximum value (Figure 1). Figure H2.

Variation in seedling traits among 20 soybean genotypes, including seedling height, hypocotyl length, cotyledon length, and cotyledon width.

Plant height was highest in PI-548530, while the stem diameter at maturity showed significant variation in PI-548512 and PI-556554. Similarly, PI 548530 also demonstrated significant differences in seed diameter and number of nodes on the main stem. The locally adapted line AARI-21 showed the minimum node count on the main stem. The highest number of branches was observed in PI-556487, while leaflet number remained uniform across all genotypes (Figure 2).

Pod number per plant varied significantly among genotypes, including PI-548530, PI-639284, PI-559933, and PI-556784. In contrast, the number of seeds per pod showed no significant differences among genotypes. Flowering days were significantly higher in PI-556487. Notably, the indigenous line PI-546480 and the exotic line PI-556487 recorded the highest 100-seed weight, suggesting superior yield potential (Figure 3).

Principal Component Analysis of Morphological and Yield-Related Traits:

Principal component analysis (PCA) was used to assess the contribution of morphological and yield trait variability among soybean genotypes (Figure 4). The first two components, Dim1 and Dim2, accounted for 34.8% and 22.9% of the total variation in morphological traits, while for yield traits, Dim1 and Dim2 explained 37.8% and 33.3% of the total variation, respectively.

Among the morphological traits, stem diameter (StD), cotyledon length (CL), and flowering days (FD) were strongly aligned with Dim1, indicating that these traits were major contributors to the first principal component. Their longer vectors and horizontal orientation reflected their importance in genotype differentiation along this axis. The number of nodes on the main stem (NNMS) and plant height (PH), on the other hand, were more in line with Dim2, indicating that these characteristics were primarily responsible for the second component. Traits such as seedling height (SH), cotyledon width (CW), and hypocotyl length (HL) were located near the origin of the plot, indicating minimal influence on overall variability among genotypes (Figure 4a).

In yield traits, seed diameter (SD) and 100-seed weight (100SW) were strongly associated with Dim1, reflecting their higher discriminatory power along the horizontal axis. On the other hand, the number of pods per plant (NPP) and the number of seeds per pod (NSP) were aligned more with Dim2, suggesting that these traits were the major contributors to variability along the vertical axis (Figure 4b).

The spread and orientation of vectors in both biplots suggest a wide diversity in trait expression among the genotypes and identify specific traits with higher discriminatory power for future selection.

Correlation analysis of Morphological and Yield Characteristics:

The correlation analysis was also conducted to identify the relationships between morphological and yield-related characteristics of soybean genotypes. In the morphological characteristics (Figure 5a),

hypocotyl length (HL) showed considerable positive correlations with cotyledon length (CL) and cotyledon width (CW), which relate to the coordinated growth of the early seedling. The height of the plant (PH) exhibited a positive and significant correlation with the number of nodes on the main stem (NNMS), the number of branches on the main stem (NBMS), and the number of leaflets (NoL), indicating that increased vertical growth was accompanied by a more complex structure. Stem diameter at maturity (StD) was also positively related to PH and NNMS, as it contributes to the vigor and strength of the plant. Seedling height (SH), in contrast, showed weak relationships with the majority of traits, meaning that it has very few impacts on overall morphological performance in terms of its elements.

In yield-related characteristics (Figure 5b), one element of this correlation, seed diameter (SD), was shown to be positively and significantly correlated with 100-seed weight (100SW), validating their joint contribution to seed size and weight. There was a positive correlation between the number of pods per plant (NPP) and both the number of seeds per pod (NSP) and SD, implying that those genotypes that produced more pods also produced more seeds. On the other hand, NSP had a negative relationship with 100SW; an increase in the number of seeds per pod did not significantly lead to an increase in the weight of individual seeds because of the trade-off effect. In general, the correlation matrices revealed high levels of interdependence among the trait entities, where PH, NNMS, and StD could be described as important morphological measures of plant vigor, with SD and 100SW becoming key yield potential determinants. These are some of the key traits that need to be selected for in soybean improvement programs.

4.2. SSR Marker Polymorphism: PCR amplification using SSR markers successfully characterized soybean genotypes, producing a clear and reproducible banding pattern. The amplified fragments revealed polymorphic variation among the genotypes (Figure 6).

Marker Informativeness (PIC analysis):

The polymorphic information content (PIC) values of the 13 SSR markers ranged from 0.0975 (SR.OL1) to 0.9375 (SR.SS001), with an average of 0.53. Among them, 46% were highly informative (PIC > 0.50), 31% moderately informative (0.25–0.50), and 23% low informative (PIC < 0.25) (Table S3). The most informative markers (SR.SS001, SR.PR.1, SR.OL4) exhibited strong discriminatory potential, while markers such as SR.OL1 and SR.LA1 were the least informative. Since the mean PIC value exceeded 0.5, the marker set was considered highly informative, demonstrating its effectiveness for diversity analysis in soybean genotypes.

SSR Marker Principal Component Analysis:

The contribution of the variability of SSR markers of the soybean genotypes was assessed by principal component analysis (PCA) (Figure 7). Dim1 and Dim2 had 30.3 and 15.3 percent of the overall variation, respectively. The SSR markers, including SR.LA1, SR.LA2, SR.OA1, SR.SA1, and SR.OA1, had a strong level of correlation with Dim1, which implies that they were significant contributors to the first major factor. They had evolved to have longer vectors and lay horizontally in the semi conservation of

genotypic diversity along this axis. Conversely, SR.SS001 and SR.PR1 had greater associations with Dim2, postulating that these markers were the major contributors to the second component. Their values of contribution and vertical orientation accentuated their contribution positions in the diversity pattern along Dim2. Markers including SR.OA2 and SR.OL3 were found to be close to the origin of the plot and so would also have limited effects on the overall variability between genotypes. The geometry of distributions and patterns of vectors in the biplot imply an extensive diversity pattern on a molecular basis and determines certain SSR markers with greater discriminatory capability in future diversity and association studies.

Correlation analysis of SSR markers:

The heatmap of the SSR data showed different patterns of clustering of the samples under investigation (Figure 8). The heatmap showed that some groups have a strong positive correlation (red gradient), especially SR.LA3, SR.SA1, and SR.LA2, and these groups were clustered together, showing a high level of similarity. On the same note, SR.OA1 and SR.OA3 had a significant correlation and clustered in the same cluster. There was a moderate correlation between SR.OL4, SR.OL1, and SR.LA1, indicating similar but weakly correlated profiles. SR.PR1 and SR.SS001, on the other hand, had less significant correlations with the majority of other samples (blue gradient), which puts them further apart in the cluster dendrogram. In general, the correlation heatmap was a very thorough visualization of the genetic similarity, as it not only demonstrated associations with high correlation between the groups but also revealed the dissimilar samples that are distinct and dissimilar from other samples, and thus, the genetic variability of the samples in the entries that are being studied.

Dendrogram-Based Grouping of Soybean Genotypes:

The dendrogram represents the hierarchical clustering of genotypes derived from the binary scoring of gel electrophoresis banding patterns (Figure 19). Each genotype was assessed for the presence (1) or absence (0) of specific bands, and the resulting binary matrix was used to compute genetic similarities.

The clustering reveals the existence of three major groups among the tested genotypes. The first cluster, located on the left side of the dendrogram, includes introduced genotypes such as PI 556914, PI 556520, and PI 564812, which exhibit a high degree of similarity in their banding profiles. The second group contains moderately related genotypes like PI 564261, PI 564287, and PI 556737, suggesting a shared but slightly divergent banding pattern. The third cluster encompasses the most genetically diverse genotypes, including PI 559393, AAR-21, and PI 232998, which may possess unique or less common bands.

The distinct grouping of genotypes within the dendrogram reflects a high degree of polymorphism among the studied samples. This genetic variability underscores the effectiveness of the applied markers in differentiating the genotypes, confirming their applicability for genetic diversity assessment and utilization in crop improvement strategies. The gel electrophoresis data used for clustering are provided in supplemental figure 1 (Figure S1).

DISCUSSION

The phenotypic evaluation revealed statistically significant differences among the analyzed soybean genotypes, as supported by Student's t-test results. Morphological traits proved useful in distinguishing and classifying soybean genotypes. Comparable findings were reported in earlier studies, where researchers evaluated a set of 70 genotypes based on 17 qualitative traits and another collection of 100 genotypes using 17 phenotypic traits [14, 33]. Notably, seedling height in genotype PI 543855 varied considerably compared to the hypocotyl length observed in PI 556519, indicating genotypic influence on early plant development. The cotyledon width in genotype PI-546480 was distinct from all other genotypes, although most displayed comparable cotyledon lengths. Among all, PI 548530 attained the maximum plant height, along with significant differences in seed diameter and node number, suggesting its potential for high-yielding traits. Genotypes PI-548512 and PI-556554 were distinguished by their relatively thicker stem diameters, whereas PI-556487 exhibited the highest branch number along with a distinct flowering pattern, suggesting better adaptability and growth vigor.

Furthermore, the genotypes PI 548530, PI 639284, PI 559933, and PI 556784 recorded the highest number of pods per plant. This finding correlates with a study by Baig and his colleagues [5]. Their research indicates that the number of pods varies significantly between soybean genotypes. The degree of variability in the number of seeds per pod was relatively small among the genotypes. It is important to note that both the local line PI-546480 and the exotic line PI 556487 had significantly higher seed weights; thus, both lines would be useful in a breeding program that aims to improve yield. These findings indicate that certain genotypes have desirable agronomic qualities and are good candidates for soybean breeding and hybrid development programs.

The PCA findings of the morphological and yield-based traits indicated that the soybean genotypes had a clear variability, with a separate contribution of the traits along the first two components. The first principal component was highly related to the stem diameter, cotyledon length, and flowering days, which means that these traits are very significant to the early morphological development and genotype differentiation. On the contrary, the height of the plants and the main stem nodes matched more with the second component, and it showed their contribution to the architectural diversity. On the same note, seed diameter and 100-seed weight played a significant role as independent variables, in relation to the first axis, whereas pod number and seeds per pod played a significant role as independent variables, in relation to the second axis. These findings are consistent with previous research studies by [13] and [15]. Conversely, some of the traits closer to the origin of the plot had minimal influence, implying that they were less variable among the genotypes. The biplot structure shows the overall high phenotypic diversity as well as the important traits, which could be used in breeding selection in the future.

The correlation analysis of morphological and yield parameters demonstrated that the parameters of soybean growth and productivity have a significant interdependence. The characteristics of plant height, stem diameter, and the number of nodes of the plant showed a strong positive correlation, which implied that these features are combined in the process of promoting morphological vigor. [53] observed strong

positive relationships between yield and plant height, number of branches, number of pods on a plant, and number of seeds per pod.

Likewise, seed diameter was significantly positively correlated with the weight of 100 seeds, which proves the joint effect of the two on yield potential. Conversely, the observed negative correlation between the number of seeds per pod and 100-seed weight was indicative of a trade-off effect, which is also apparent in the results of [5], who found that seed number and seed size were both negatively correlated in soybean. [10] also reported that 100-seed weight, the number of pods per plant, and protein content all positively correlate with seed yield per plant, but branches per plant and seeds per pod have negative correlations. This highlights the significance of trait combinations in breeding.

Understanding genetic diversity in a population is very essential for the characterization of the germplasm and provides important insight into conservation, evolution, and breeding strategies, and effective utilization [28]. This average PIC value (> 0.5) shows that the markers employed are very informative and that their use in estimating the degree of genetic differences is significant. The existence of highly informative markers, together with moderately informative ones, provides adequate discriminative power between soybean genotypes. Similar results were achieved by Zatybekov and colleagues, who showed a significantly high use of SSR markers in the analysis of genetic diversity [51]. The clustering by dendrogram has shown that the soybean genotypes have undergone genetic divergence, giving rise to three groups based on their SSR banding patterns. Genotypes belonging to the same cluster have similar allele profiles, whereas those belonging to different clusters are distinguished by significant polymorphism. The second cluster is entirely exotic, as previously reported by [44]. The high diversity of lines, especially in the third cluster, demonstrates the ability of SSR to capture genetic variation. Khatab and coworkers reported that distinct genetic groupings were found through clear SSR-based differentiation and clustering of 24 soybean genotypes based on agro-morphological traits and SSR markers [23]. This structured organization demonstrates the potential of these markers to define genetically diverse lines, which can be utilized in broadening the genetic base in soybean breeding endeavors.

CONCLUSION

This research was conducted to evaluate the genetic diversity and population structure of 20 soybean genotypes (including local Pakistani and exotic USDA lines) using a combined method of phenotypic measurements and SSR molecular markers to estimate genetic diversity based on oil composition and yield. There was a high level of phenotypic variation in some of the important agronomic traits, with certain genotypes, such as PI 548530 and PI 556487, showing better traits in terms of plant architecture and yield potential, respectively. The quality of the 13 trait-linked SSR markers was proven by the high average Polymorphism Information Content (PIC) value of 0.53, which indicates a great amount of revealed molecular genetic diversity. The overall outcomes of principal component analysis, cluster analysis, and genotypic profiling made a clear distinction between the three genetic groups in genotype, which showed a distinct distinction between the local and exotic germplasm. This organized diversity,

along with the discovery of specific allelic differences in lines like AARI-21 and PI 232998, is an important source of genetic resources. To summarize, the results show that the combination of both morphological and SSR marker analysis is an effective method for the characterization of soybean germplasm. The selected genetically superior and agronomically superior genotypes have enormous potential to be used in the breeding programs of Pakistan with the help of a marker-assisted breeding program. Their integration can add value to the production of high-yielding, resistant soybean varieties with improved nutritional composition, thus increasing local production and minimizing reliance on imports.

Declarations

Not Applicable

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding:

This research was supported by Punjab Agricultural Research Board (PARB) project No. 20-350 awarded to Dr. Zaheer Ahmed.

Author's contribution:

SR conducted the research experiments and drafted the manuscript. FR assisted in data analysis and manuscript write-up. RR conceived the main idea of the study and edited the manuscript. HM, MA, UI, AW, and AH contributed to the research work and provided technical support. FTK assisted in data collection and preliminary laboratory work. MML, MHK, and RK critically reviewed and improved the manuscript. HA and HASA contributed to manuscript improvement and critical suggestions. ZA supervised and finalized the overall research work. All authors read and approved the final manuscript for submission.

Acknowledgments:

The authors are thankful to the University of Agriculture, Faisalabad (UAF), for providing institutional support throughout this research. Special thanks are also extended to the Center for Advanced Studies in Agriculture and Food Security (CAS-AFS), Soybean Laboratory, where the experimental work was conducted, and technical assistance was provided.

Clinical trial registration:

This study does not involve human participants, human data, or clinical trials. Therefore, trial registration is not applicable.

References

1. Akram Z, Ahmad Q. Future prospects of soybean in Pakistan. Biomed J Sci Tech Res. 2019; 15: 11562-63.
2. Amom T, Nongdam P. The use of molecular marker methods in plants: a review. IJCRR. 2017; 9:1-7.
3. Asad SA, Wahid MA, Farina S, Ali R, Muhammad F. Soybean production in Pakistan: experiences, challenges and prospects. IJAB. 2020; 24: 995-1005.
4. Asad SA, Shaheen F, Ahmad I, Mahmood I, Farooq M. Thermo-temporal Changes in Soybean Productivity under Rainfed and Irrigated Conditions in Pakistan. Gesunde Pflanzen. 2023; 75: 2437-47.
5. Baig D, Khurshid H, Arshad M, Jan SA, Khan MA, Nawaz N. 1. Evaluation of Soybean genotypes for yield and other economically important traits under rainfed condition. PAB. 2018; 7:1-7.
6. Bezerra AR, Sediyama T, da Silva FL, Borém A, da Silva AF, dos Santos Silva FC. Agronomical aspects of the development of cultivars. In: Soybean breeding. Cham: Springer International Publishing; 2017.p.395-411
7. Campobenedetto C, Mannino G, Agliassa C, Acquadro A, Contartese V, Garabello C, Bertea CM. Transcriptome analyses and antioxidant activity profiling reveal the role of a lignin-derived biostimulant seed treatment in enhancing heat stress tolerance in soybean. Plants. 2020; 9:1308.
8. Carciochi WD, Rosso LH, Secchi MA, Torres AR, Naeve S, Casteel SN, Kovács P, Davidson D, Purcell LC, Archontoulis S, Ciampitti IA. Soybean yield, biological N_2 fixation and seed composition responses to additional inoculation in the United States. Sci rep. 2019; 9:19908.
9. Cardoso CC, Mendes BM, Pasa VM. Production and characterization of cold-flow quality biofuel from soybean oil using different alky and benzyl alcohols. J Environ Chem Eng. 2018; 6: 2241-7.
10. Chavan BH, Dahat DV, Rajput HJ, Deshmukh MP, Diwane SL. Correlation and path analysis in soybean. IRJMS. 2016; 2:1-5.
11. Chen L, Nan H, Kong L, Yue L, Yang H, Zhao Q, Fang C, Li H, Cheng Q, Lu S, Kong F. Soybean AP1 homologs control flowering time and plant height. J Integ Plant Biol. 2020; 62:1868-79.

12. Devi KN, Singh LN, Devi TS, Devi HN, Singh TB, Singh KK, Singh WM. Response of soybean [*Glycine max* (L.) Merrill] to sources and levels of phosphorus. J Agric Sci. 2012; 4:44.
13. Deviona D, Roslim DI, Zul D, Suhartina S, Syafitri AR. Keragaman Genetik dan Heritabilitas Berbagai Karakter Tanaman Kedelai (*Glycine max* (L.) Merril). Agrosains: J Penelitian Agronomi. 2024; 2638-42.
14. Dhaliwal I, Khosla G, Singh TP, Gill BS, Kaushik P. DUS characterization of some released varieties and advanced breeding lines of soybean (*Glycine max* L.) under Punjab agroclimatic conditions. agriRxiv. 2020; 20210198131. https://doi.org/10.31220/agrirxiv.2020.00015
15. Dunna DS, Heisnam ND, Thokchom RD, Sinha B, Singh O. Principal component analysis among vegetable soybean genotypes (*Glycine max* L. Merrill). Environ Conser J. 2023; 24: 79-86.
16. Fones HN, Bebber DP, Chaloner TM, Kay WT, Steinberg G, Gurr SJ. Threats to global food security from emerging fungal and oomycete crop pathogens. Nature Food. 2020; 1: 332-42. https://doi.org/10.1038/s41558-020-0871-4
17. Food and Agriculture Organization of the United Nations. *World food and agriculture – Statistical yearbook 2022*. Rome. 2020. https://doi.org/10.4060/cc2211en
18. Hipparagi Y, Singh R, Choudhury DR, Gupta V. Genetic diversity and population structure analysis of Kala bhat (*Glycine max* (L.) Merrill) genotypes using SSR markers. Hereditas. 2017; 154: 9.
19. Hwang TY, Gwak BS, Sung J, Kim HS. Genetic diversity patterns and discrimination of 172 korean soybean (*Glycine max* (L.) merrill) varieties based on SSR analysis. Agriculture. 2020; 10: 77.
20. IPCC. *Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. 2021.
21. Kachare S, Tiwari S, Tripathi N, Thakur VV. Assessment of genetic diversity of soybean (*Glycine max*) genotypes using qualitative traits and microsatellite markers. Agric Res. 2020; 9: 23-34.
22. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. Microsatellite markers: an overview of the recent progress in plants. Euphytica. 2011; 177: 309-34.
23. Khatab IA, Morsy AR, Fares WM, Kumamaru T. Genetic Diversity of Soybean Genotypes revealed by Agro–morphological and SSR markers. J Fac Agric Kyushu Uni. 2016; 61: 79-87.
24. Khurshid H, Baig D, Jan SA, Arshad M, Khan MA. Miracle crop: the present and future of soybean production in Pakistan. MOJ Biol Med. 2017; 2: 189-91.
25. Khurshid H, Baig D, Nawaz N, Rabbani MA, Shinwari ZK. Intraspecific variation in soybean (*Glycine max*) revealed by multivariate approaches to morphometric traits and soybean mosaic virus resistance. 2020
26. Kumar A, Ramesh KV, Singh C, Sripathy KV, Agarwal DK, Pal G, Kuchlan MK, Singh RK, Prabha R, Kumar SJ. Bioprospecting nutraceuticals from soybean (*Glycine max*) seed coats and cotyledons. Indian J Agric Sci. 2019; 89: 2064-8.

27. Kumar S, Sharma NK. Level of attitude towards soybean cultivation practices by the farmers. Ind Res J Extens Educ. 2018;18(2):41-5.
28. Li G, Ra WH, Park JW, Kwon SW, Lee JH, Park CB, Park YJ. Developing EST-SSR markers to study molecular diversity in Liriope and Ophiopogon. Biochem System Ecol. 2011; 39: 241-52. doi:10.1016/j.bse.2011.08.012
29. Lin F, Chhapekar SS, Vieira CC, Da Silva MP, Rojas A, Lee D, Liu N, Pardo EM, Lee YC, Dong Z, Pinheiro JB. Breeding for disease resistance in soybean: a global perspective. Theor Appl Genet. 2022; 135: 3773-872.
30. Maish DJ, Bajpai DA, Kaur DA. Indian soybean: the wonder crop. IJASR. 2022; 12: 97-108.
31. Mishra R, Tripathi MK, Sikarwar RS, Singh Y, Tripathi N. Soybean (*Glycine max* L. Merrill): A multipurpose legume shaping our world. Plant Cell Biotechnology and Molecular Biology. 2024; 25: 17-37.
32. NOAA. *Climate change: Global temperature*. National Oceanic and Atmospheric Administration. 2021. Retrieved from https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature
33. Pachori S, Tiwari S, Shrivastava MK, Amrate PK, Thakur S, Gupta S. Unveiling genetic diversity in soybean germplasm: A comprehensive snalysis through DUS characteristics. Electron J Plant Breed. 2023;14: 1541-8.
34. Pakistan, (Various Issues). Pakistan Economic Survey. Finance Division, Economic Advisor's Wing, Islamabad. 2021.
35. Pearson K. On lines and planes of closest fit to systems of points in space. London, Edinburgh Dublin Philos. Mag J Sci. 1901; 2: 559-572.
36. Rizzo G. The antioxidant role of soy and soy foods in human health. Antioxidants. 2020; 9: 635.
37. Sedibe MM, Mofokeng AM, Masvodza DR. Soybean production, constraints, and future prospects in poorer countries: a review. Production and utilization of legumes-progress and prospects. 2023.
38. Shahid MQ, Saleem MF, Khan HZ, Anjum SA. Performance of soybean (*Glycine max* L.) under different phosphorus levels and inoculation. Pak J Agric Sci. 2009; 46: 237-41.
39. Sharma A, Sharma S, Kumar N, Rana RS, Sharma P, Kumar P, Rani M. Morpho-molecular genetic diversity and population structure analysis in garden pea (*Pisum sativum* L.) genotypes using simple sequence repeat markers. PLoS One. 2022;17
40. Singh PK, Shrestha J. Evaluation of soybean [*Glycine max* (L.) Merrill] genotypes for agro-morphological traits using multivariate analysis. NepJAS. 2019; 18: 100-7.
41. Singh R, Arya M, Verma VC, Bisht IS. Genetic diversity and food value of black soybean (*Glycine max* (L.) Merr.) in Uttarakhand Hills, North-Western India. Proceedings of the National Academy of Sciences, India Section B: Biol Sci. 2023; 93: 743-53.
42. Steel RGD, Torrie JH, Dickey DA. Principles and procedures of statistics: A biometrical approach. 3rd ed. McGraw Hill Book Co. Inc., New York, USA. 1997; 352-358.

43. Sulistyo A, Indriani FC, Mejaya MJ, Sugiharto AN, Agranoff J. Genetic diversity of Indonesian soybean (*Glycine max* L. Merrill) germplasm based on morphological and microsatellite markers. In: IOP Conference Series: Earth and Environmental Science. IOP Publishing; 2019.
44. Thakur R, Kumari V, Gupta S, Chahota RK, Isobe S, Shirasawa K, Vangala R, Bisht A, Kumar N. Elucidating Morpho-Molecular Diversity in Soybean Germplasm Using SSR Markers. Plant Mol BiolRep. 2025:1-5.
45. Tiwari S, Tripathi N, Tsuji K, Tantwai K. Genetic diversity and population structure of Indian soybean (*Glycine max* (L.) Merr.) as revealed by microsatellite markers. Physiol Mol Biol Plants. 2019; 25: 953-64.
46. Ullah A, Akram Z, Malik SI, Khan KS. Assessment of phenotypic and molecular diversity in soybean [*Glycine max* (L.) Merr.] germplasm using morpho-biochemical attributes and SSR markers. Genet Res Crop Evol. 2021; 68: 2827-47.
47. United States Department of Agriculture. 2023. *World agricultural production* (Circular Series WAP 11-23). https://apps.fas.usda.gov/psdonline/circulars/production.pdf
48. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3—new capabilities and interfaces. Nucl acids res. 2012; 40
49. USDA/Foreign Agricultural Service. 2024-2025. Global market analysis. Available online with updates at https://www.fas.usda.gov/data/production/commodity/2222000
50. Wang J, Hu B, Jing Y, Hu X, Guo Y, Chen J, Liu Y, Hao J, Li WX, Ning H. Detecting QTL and candidate genes for plant height in soybean via linkage analysis and GWAS. Front plant sci. 2022; 12: 803820.
51. Zatybekov A, Yermagambetova M, Genievskaya Y, Didorenko S, Abugalieva S. Genetic diversity analysis of soybean collection using simple sequence repeat markers. Plants. 2023; 12: 3445.
52. Zhang C, Zha B, Yuan R, Zhao K, Sun J, Liu X, Wang X, Zhang F, Zhang B, Lamlom SF, Ren H. Identification of Quantitative Trait Loci for Node Number, Pod Number, and Seed Number in Soybean. Int J Mol Sci. 2025; 26: 2300.
53. Zhang Z, Gong W, Yang W, Ma L. Correlation between agronomic characters and yield in relay-planting soybeans. Soybean Sci. 2008; 26: 680.
54. Zhenbin H, Zhang D, Zhang G, Kan G, Hong D, Yu D. Association mapping of yield–related traits and SSR markers in wild soybean [*Glycine soja Sieb*. and Zucc.]. Breeding Sci. 2014; 63: 441–449
55. Zheng Z, Lei T, Wang J, Wei Y, Liu X, Yu F, Ji J. Catalytic cracking of soybean oil for biofuel over γ-Al2O3/CaO composite catalyst. J Braz Chem Soci. 2019; 30: 359-70.

Figures

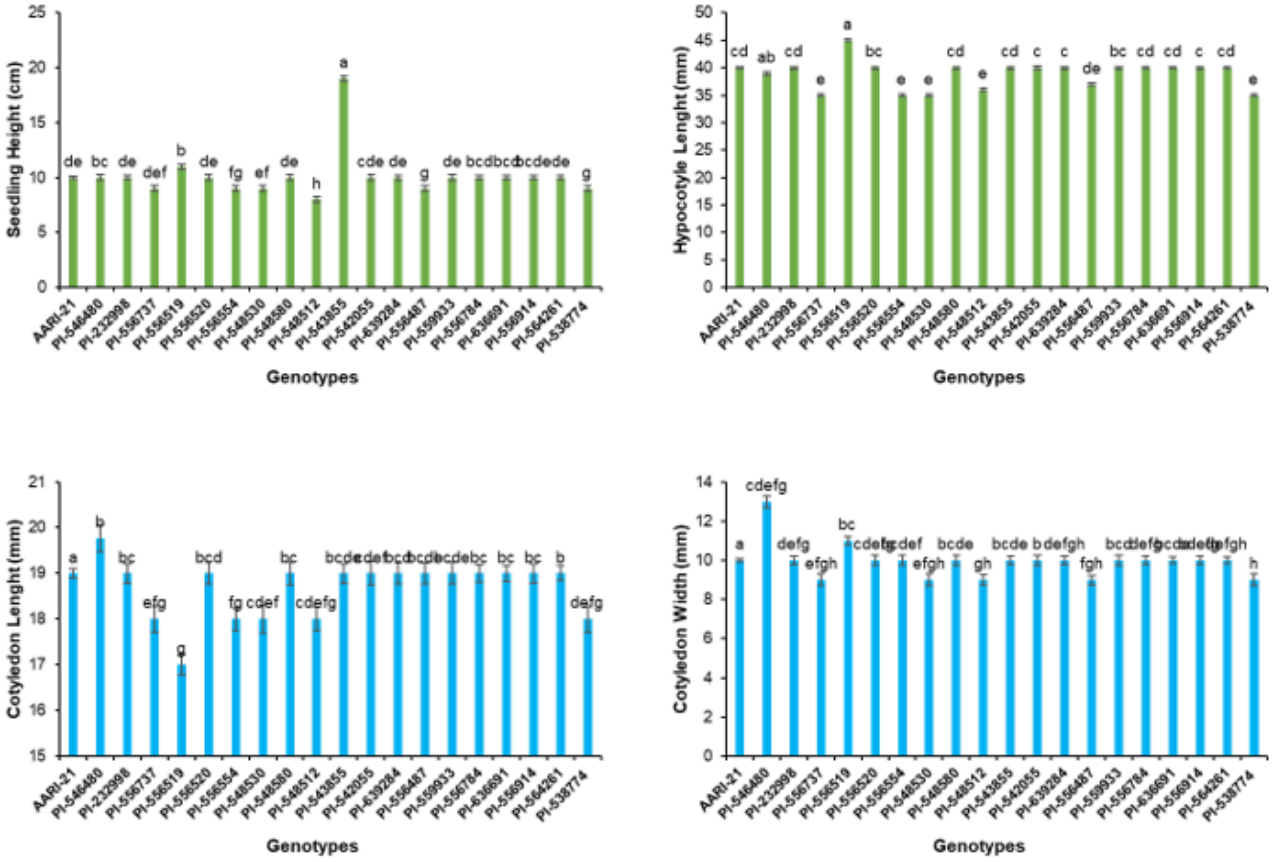

Figure 1

Variation in seedling traits among 20 soybean genotypes, including seedling height, hypocotyl length, cotyledon length, and cotyledon width

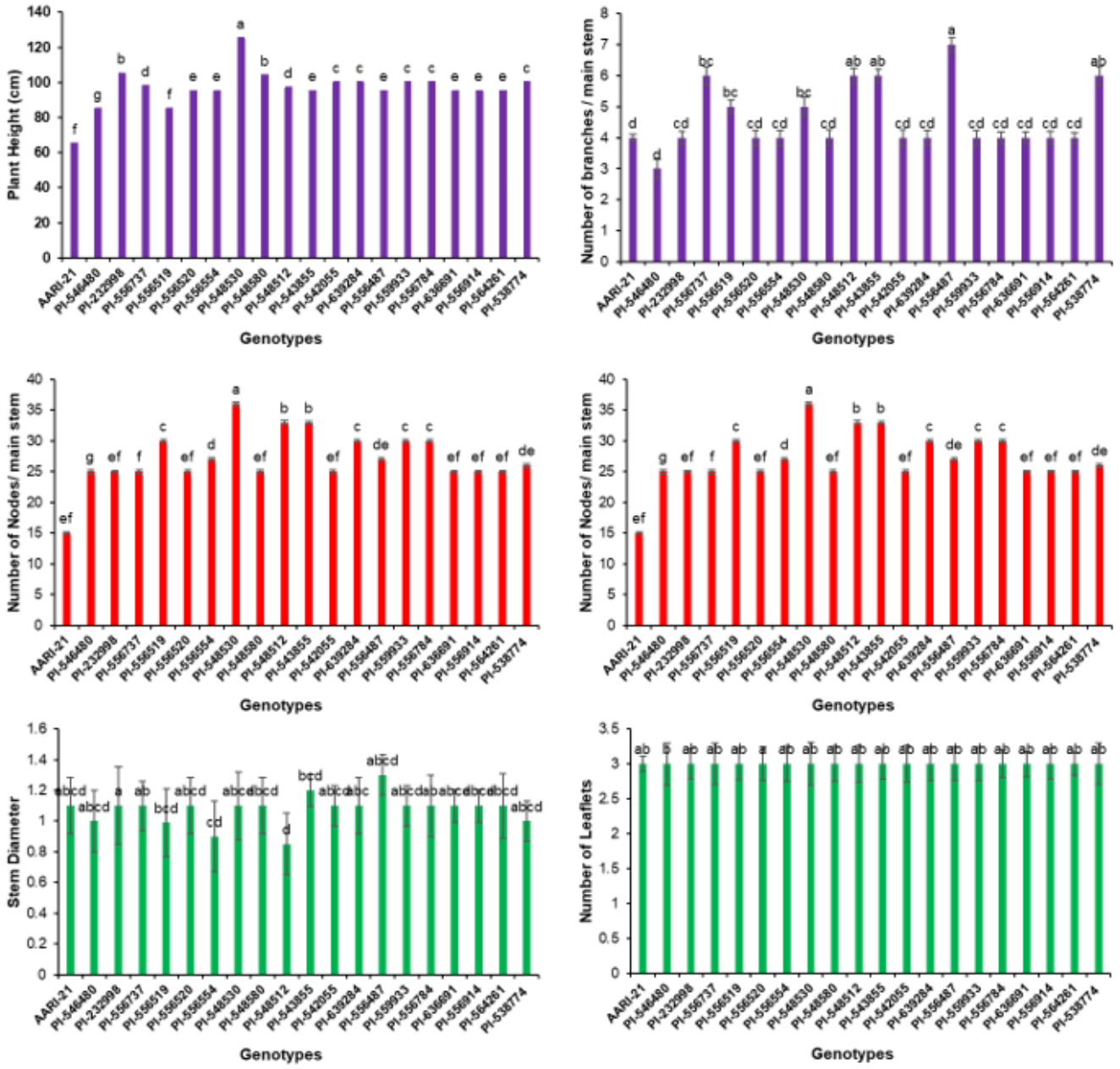

Figure 2

Variation in morphological traits among 20 soybean genotypes, including plant height (cm), number of branches per stem, number of nodes per main stem, stem diameter (cm), and number of leaflets

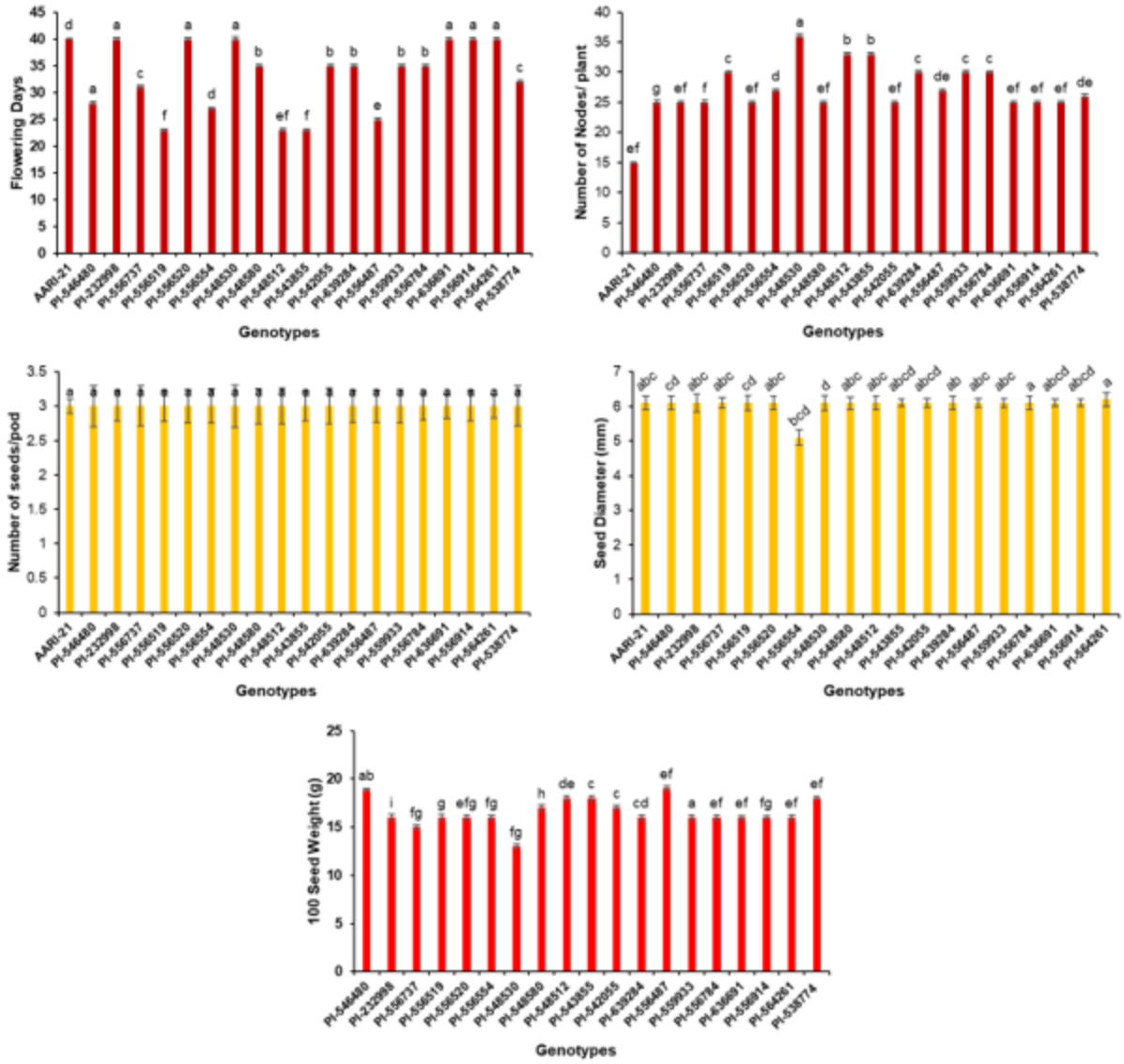

Figure 3

Variation in yield-related traits among 20 soybean genotypes, including flowering days, number of pods per plant, number of seeds per pod, seed diameter, and 100-seed weight.

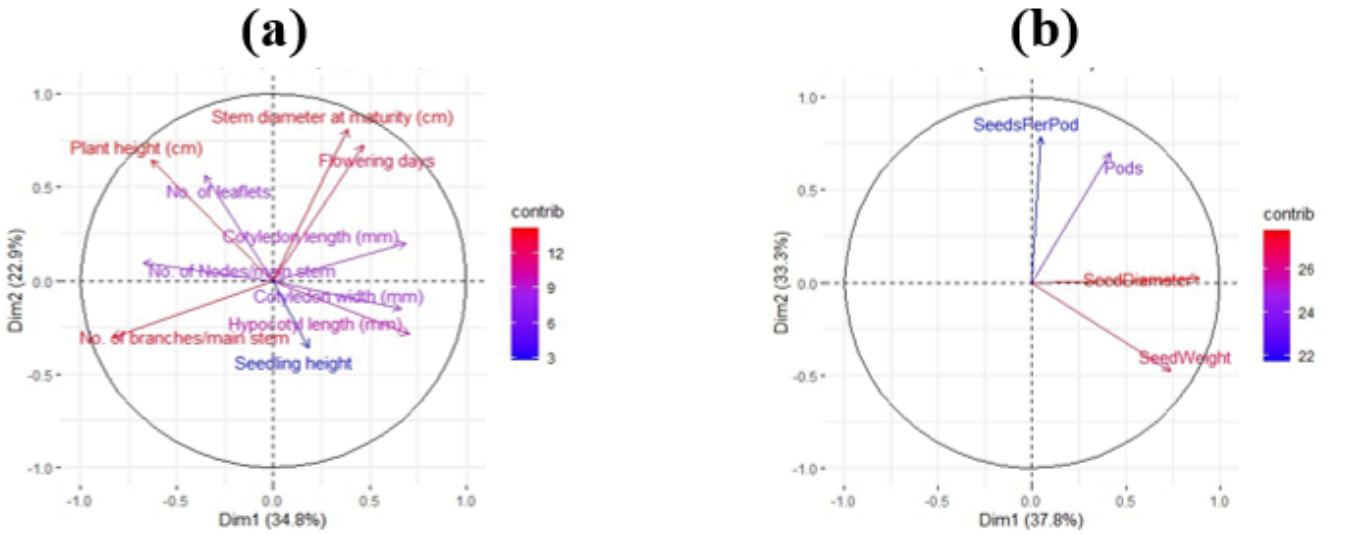

Figure 4

Principal Component Analysis of (a) morphological and (b) yield-related traits

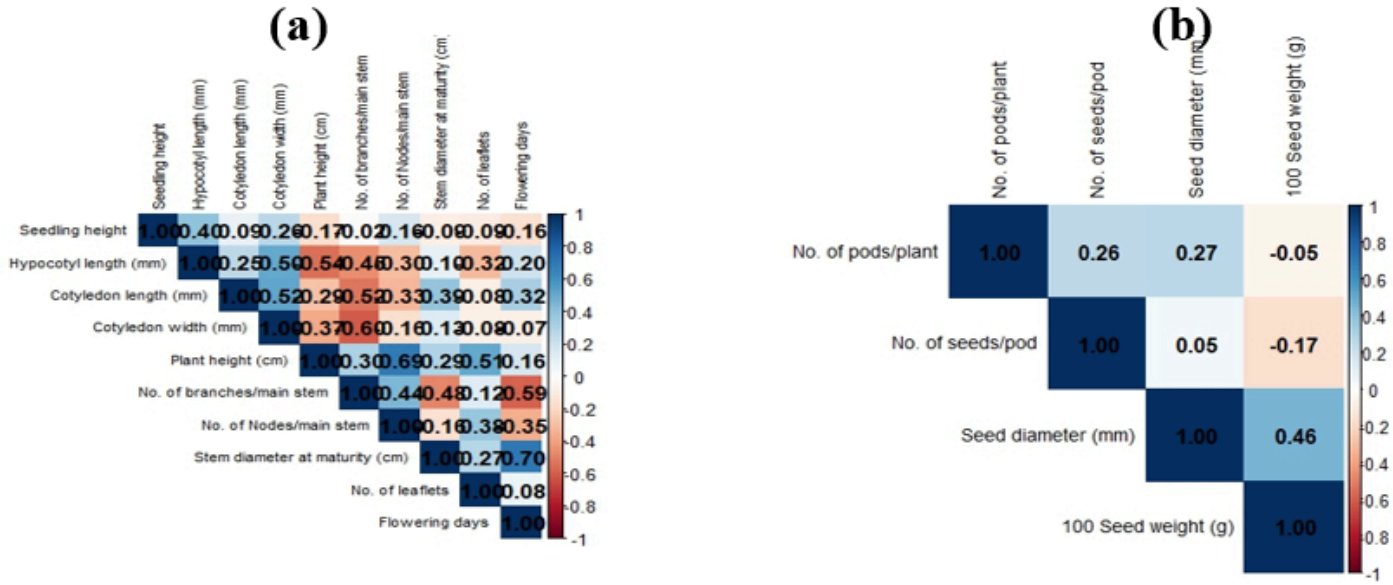

Figure 5

Correlation analysis of **(a)** morphological traits **(b)** yield-related traits

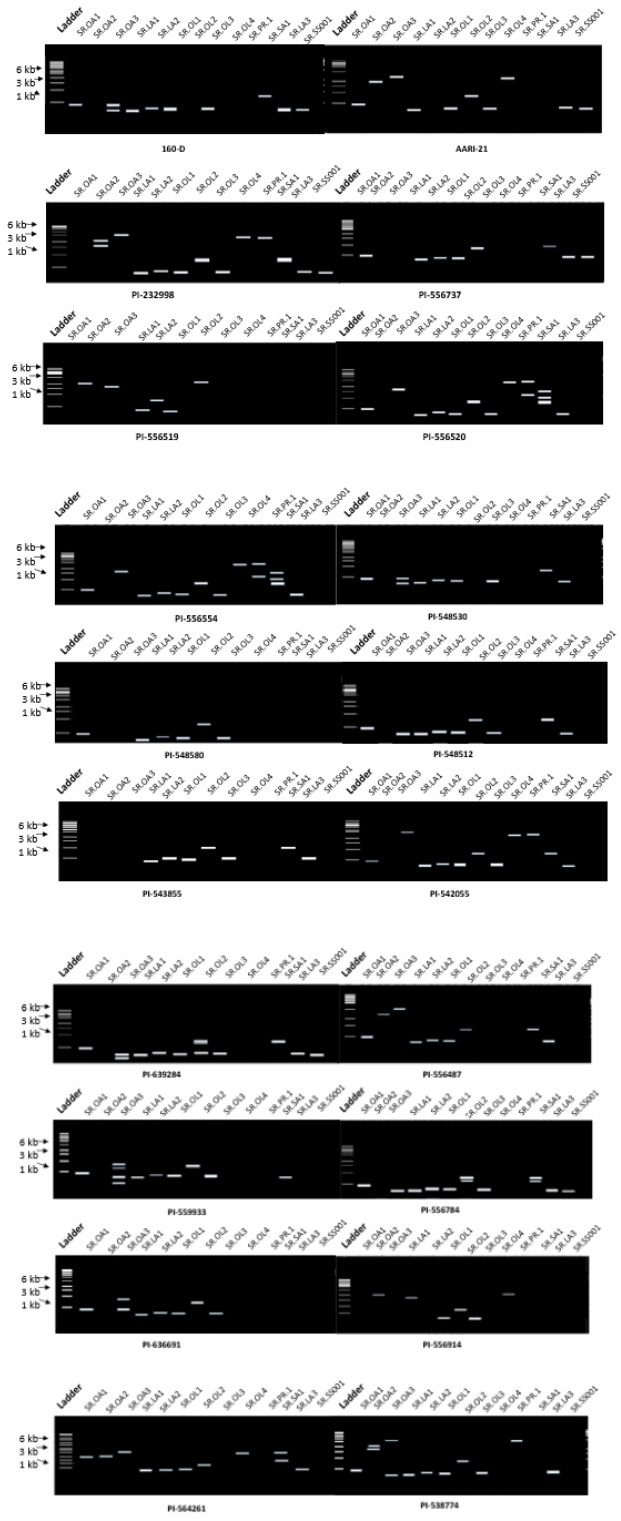

Figure 6

PCR based molecular characterization of soybean genotypes through SSR markers

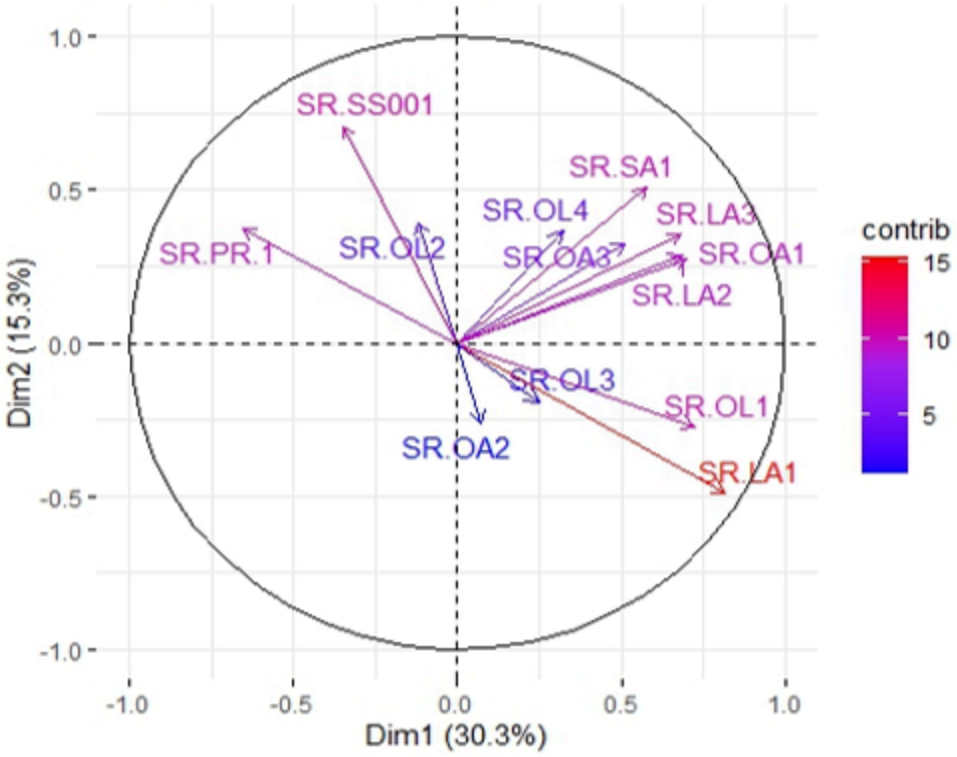

Figure 7

Principle Component Analysis of SSR markers

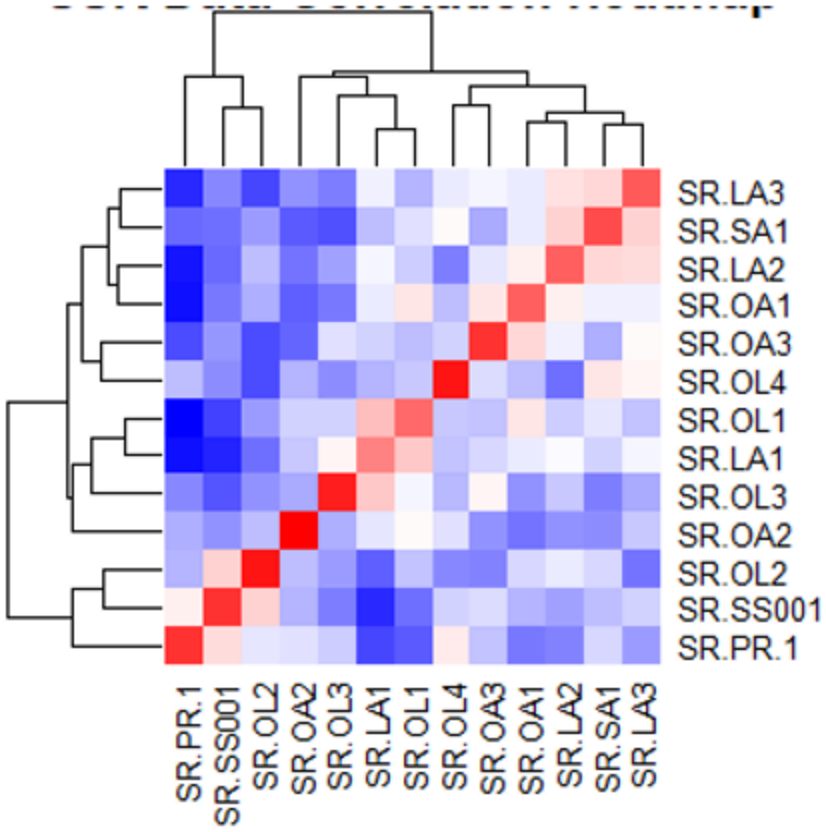

Figure 8

Correlation analysis of 13 SSR markers

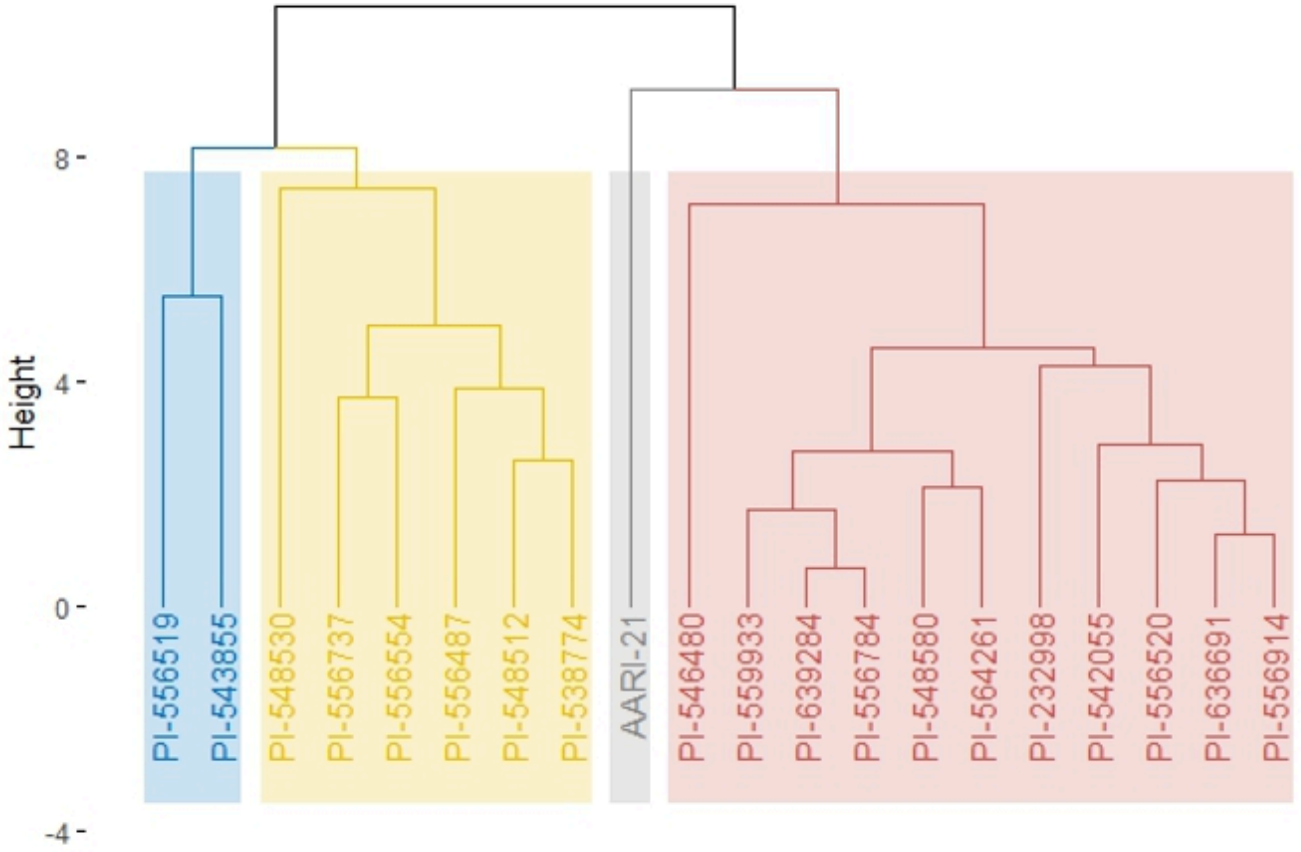

Figure 9

Dendrogram showing genetic relationships among 20 soybean genotypes based on binary scoring of agarose gel banding patterns.