PLANT SCIENCE TODAY
Vol x(x): xx–xx
https://doi.org/10.14719/pst.6401

eISSN 2348-1900

RESEARCH ARTICLE

Unravelling soybean yield potential: Exploring trait synergy, impact pathways, multidimensional patterns and biochemical insights

Riya Mishra[1*], M K Shrivastava[2], M K Tripathi[1], Pawan K Amrate[2], Yogendra Singh[2], Ravindra Solanki[1], Sanjeev Sharma[1] & Sachin Shukla[1]

[1]Department of Genetics and Plant Breeding, Rajmata Vijayaraje Scindia Agriculture University, Gwalior 474 002, India
[2]Department of Genetics and Plant Breeding , Jawaharlal Nehru Agriculture University, Jabalpur 482 004, India

*Correspondence email - riyam2710@gmail.com

Received: 29 November 2024; Accepted: 24 March 2025; Available online: Version 1.0: 17 May 2025

Cite this article: Mishra R, Shrivastava MK, Tripathi MK, Amrate PK, Singh Y, Solanki R, Sharma S, Shukla S. Unravelling soybean yield potential: exploring trait synergy, impact pathways, multidimensional patterns and biochemical insights. Plant Science Today (Early Access). https:/doi.org/10.14719/pst.6401

Abstract

As the major oilseed *Kharif* crop, soybean (*Glycine max* L. Merill) requires continuous improvement to compete with changing climatic conditions. For this purpose, research was conducted during the *Kharif* 2022, at AICRP Soybean Seed Breeding Farm, JNKVV, Jabalpur, focusing on soybean genetics, yield related traits and their implications for enhancing global food security. A total of 118 genotypes, comprising 115 Recombinant Inbred Lines (RILs) with three checks were meticulously analyzed. The study encompassed a comprehensive evaluation, employing correlation coefficient analysis, path coefficient analysis and principal component analysis. Ten critical yield-related traits were systematically recorded, including flowering patterns, branching architecture, plant height and seed yield. Notably, strong positive correlations were found between seed yield and harvest index, 100-seed weight, number of seeds per plant and biological yield. Path analysis unveiled these traits' direct and indirect effects on seed yield, with harvest index and biological yield as key contributors. Principal Component Analysis (PCA) successfully condensed this data into seven principal components, explaining 95.93 % of the variance, with Principal Component 1 (PC1) bearing the highest impact. PC1 incorporated traits critical for soybean improvement, including seed yield and 100-seed weight. Furthermore, biochemical analysis of 31 RILs and 2 checks revealed moisture content ranging from 3 to 6 %, fat content 16-21 %, proteins 34-42 % and ash ranged from 4-6 %. This research offers valuable insights into soybean genetics and trait interactions, providing a foundation for future breeding programs aimed at enhancing global food security in an ever-changing world.

Keywords: biochemical analysis; correlation; path coefficient analysis; principal component analysis (PCA); protein content; Recombinant Inbred Lines (RILs)

Introduction

Oilseed crops remain important in international trade and have long been essential to agricultural economies. The soybean (*Glycine max* L. Merill) stands out among these due to its high oil and protein content, which has earned it the moniker "Golden Bean" (1). Native to East Asia, it has emerged as a crucial global crop, significantly enhancing food security and supporting diverse product applications. Its role in advancing scientific research is equally substantial, driven by its economic and nutritional value across various sectors (2). A mainstay in many cuisines and industrial goods, soybean seeds are prized for their high protein content (40 %), oil content (20 %) and health-promoting phytonutrients. Essential amino acids, vitamins and minerals are found in soybeans and are used to make a variety of soy-based products such as tofu, soymilk and tempeh (3).

Together with big producers like Brazil, US and Argentina, India accounts for 3 % of the worlds' soybean production. Indias' soybean production increased dramatically, from 0.426 tons per hectare in 1970 to 1.158 t/ha in 2022. This expansion highlights Indias' growing significance in the world soybean market (2, 4, 5). However, the biotic and abiotic stressors pose significant obstacles to soybean production, affecting farmers worldwide. Heat stress, heavy metal contamination, salinity, drought and flooding are critical abiotic stressors, whereas disease, pest and weed are biotic stresses affecting soybean yield (6). Therefore, with notable advancements in gene manipulation to induce desirable traits, soybeans have drawn attention to genetic improvement over the past 2 decades (7). To meet consumer demands and address issues with global food security, biotechnological tools have also been applied to design soybeans with improved features, including herbicide, insects and disease resistance. Yet, conventional breeding is still used for soybean improvement (8).

Using the novel genetic variation in breeding could significantly increase soybean production. The development of improved soybean varieties depends on the efficient selection of parental genotypes based on genetic and phenotypic variation (9, 10). Inbred lines are created by repeatedly self-pollinating heterozygous populations or F2 progeny to produce homozygous lines in plant breeding; this process typically takes 6-10 generations. To take advantage of heterosis, a variety of breeding techniques can be employed for these inbred parents, such as backcrossing, pedigree method and doubled haploids (11). Utilizing these Recombinant Inbred Lines (RILs) population may improve our comprehension of soybeans' genetic regulation, evolution and molecular mechanism (12).

Deciphering the genetic foundation of yield-related traits in soybeans is challenging due to the interwoven nature of factors like flowering time, plant height, branching and seed characteristics. These complex interactions complicate the identification of specific traits contributing to yield, highlighting the need for sophisticated analytical methods beyond traditional breeding approaches (13). Correlation analysis, path coefficient analysis and principal component analysis (PCA) are powerful tools for addressing these complexities. Correlation analysis reveals trait associations, while path analysis provides a more profound understanding by quantifying both direct and indirect effects on yield, though it may face constraints from multicollinearity issues (14). PCA simplifies trait data by grouping key attributes into principal components, identifying those with the strongest links to yield. Biochemical parameters, such as proteins, oil content and secondary metabolites, are crucial for understanding yield-related traits, seed quality and soybean market value (15). Integrating biochemical analysis with morphological and agronomic studies helps breeders uncover metabolic pathways influencing yield. This combination of analytical methods enables breeders to pinpoint influential traits, promoting improvements in soybean yield, quality and resilience.

Materials and Methods

Characteristics of the experimental site

Jabalpur district, situated in the Mahakoshal region of Madhya Pradesh, spans 10160 km^2. The district lies between the Narmada and Son river basins, predominantly within the Narmada valley. The area is a narrow plain enclosed by highlands, with fertile alluvial, black cotton soil in the western and southern parts. It falls under agro-climatic subregion 10.1, classified as a subhumid dry eco-region and is part of the Rice-Wheat agroecological zone. Jabalpur experiences a subtropical, semi-arid climate with an average temperature of 25.3 °C, ranging from 5.7 °C to 43.9 °C. The average relative humidity is 59.4 %, varying between 10.5 % and 99.2 %. Jabalpurs' fertile soil, substantial rainfall and agroclimatic conditions make it suitable for intensive agriculture, necessitating tailored management strategies to optimize productivity and sustainability.

Experimental details

The research was conducted during the *kharif* season 2022 at the AICRP Soybean Seed Breeding Farm, Department of Genetics and Plant Breeding at College of Agriculture, JNKVV, Jabalpur, Madhya Pradesh. The experimental material for this study comprised a total of 118 genotypes, comprising 115 Recombinant Inbred Lines (RILs) and three checks (JS 97-52, NRC-37 and JS-335). The RILs were acquired from IISR, Indore and were planted using an Augmented block design. The genetic assessment encompassed an evaluation of the correlation coefficient, path coefficient and principal component analysis. By recording the observations of 3 plants from each RIL and, after that, taking average, observations were systematically recorded for ten yield-related traits, including the number of days to 50 % flowering, days to maturity, number of primary branches per plant, plant height, number of pods per plant, number of seeds per plant, biological yield per plant, harvest index per plant and seed yield per plant. The data analysis was done using the R-statistical package.

Biochemical analyses were conducted to estimate fat, ash and protein content in 31 RILs and 2 checks. These RILs were selected based on their superior performance across various traits and were subsequently subjected to biochemical analysis to support further crop improvement objectives. The Pelican SOCS Plus apparatus, operating on Soxhlets' extraction methodology, was employed to study fat content as prescribed by AOAC . A muffle furnace was used to heat samples to extract ash over 6 h. AOAC protein content was estimated using the Micro-Kjeldahl digestion and distillation process, with further titration performed thereafter.

Experimental design

Augmented block design (ABD) is a statistical experiment design used primarily in plant breeding trials where many test genotypes must be evaluated with limited resources. It is preferred over other designs when the availability of experimental material, land and labour is constrained, as it allows unreplicated test entries while including a few replicated check varieties for variance estimation. Replications are not used for all test entries due to resource limitations, but the repeated checks help estimate experimental error and adjust for environmental variation, ensuring reliable comparison among treatments.

Results and Discussion

Correlation coefficient analysis

This investigation aimed to assess the correlation between seed yield per plant and various yield-related traits in soybean (Table 1). Seed yield per plant showed a highly significant positive correlation with harvest index (0.844), followed closely by 100-seed weight (0.803), number of seeds per plant (0.633) and biological yield (0.477). These strong positive correlations indicate that genotypes with higher harvest indexes and larger seeds produce greater seed yields. The findings align with the idea that allocating more photosynthates to reproductive structures can boost seed yield (16-18). Additionally, aligning with the previous

Table 1. Correlation coefficient analysis for yield and its contributing traits in soybean RILs

	dff	**dtm**	**nbp**	**npopp**	**nspp**	**ph**	**hsw**	**by**	**hi**	**sy**
dff	1									
dtm	-0.083	1								
nbp	-0.012	0.013	1							
npopp	0.256**	-0.031	0.269**	1						
nspp	0.215*	0.022	0.361***	0.703***	1					
ph	-0.181*	-0.041	0.338***	0.252**	0.173*	1				
hsw	-0.07	0.065	0.392***	0.473***	0.709***	0.1	1			
by	-0.088	0.689***	0.15	0.03	0.189*	0.005	0.304***	1		
hi	-0.001	-0.16	0.370***	0.416***	0.609***	0.162	0.729***	-0.067	1	
sy	-0.042	0.231**	0.406***	0.388***	0.633***	0.147	0.803***	0.477***	0.844***	1

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1' '1. dff: Days to 50 % flowering; dtm: Days to maturity; nbp: No. of primary branches per plant; npopp: No. of pods per plant; nspp: No. of seeds per plant; ph: Plant height; hsw: 100 Seed weight; by: Biological yield; hi: Harvest Index

investigations the correlation analysis shed light on the interplay between traits and yield. Days to 50 % flowering exhibited a highly significant positive correlation with the number of pods per plant (0.256) and the number of seeds per plant (0.215) (19-22). The scatter plot further corroborates these findings, reinforcing the reliability and consistency of data (Fig. 1).This suggests that early flowering genotypes tend to produce more pods and seeds. Extending the reproductive phase can thus increase the number of potential seed-bearing structures, potentially leading to higher yields (23, 24). However, negative correlations between days to maturity and several other traits, including the number of pods per plant, plant height and harvest index, suggest that early maturing genotypes may sacrifice yield potential (25, 26). Striking a balance between early maturity and optimal yield is a critical consideration for soybean breeding programs. Moreover, the correlation between the number of primary branches per plant and seed yield highlights the significance of branching architecture in determining yield potential. This aligns with previous studies that have found a positive association between increased branching and higher pod and seed numbers (27, 28).

Furthermore, the positive correlation between the number of pods per plant and the number of seeds per plant supports the idea that genotypes with more pods tend to produce more seeds, ultimately contributing to higher seed yield (29). These findings provide valuable insights for soybean breeding programs that optimize yield by considering the interplay between flowering, branching and reproductive structures. Balancing these factors can be instrumental in developing soybean genotypes with enhanced yield potential.

Path coefficient analysis

The investigation aimed to assess the correlation between seed yield per plant and various yield-related traits in soybean and employed a path coefficient analysis to understand these traits' direct and indirect effects on seed yield. This analysis partitions the observed correlation between the dependent and independent variables into direct and indirect impacts, shedding light on the underlying relationships between traits. The findings and numerical values from the analysis offer valuable insights into the factors influencing soybean seed yield (Table 2).

The results suggest that specific traits exert significant direct effects on seed yield. Harvest index and biological yield stand out with considerably high positive direct effects (0.88772 and 0.844, respectively). These findings align with established investigations, indicating that genotypes with higher harvest index values, reflecting efficient resource allocation to reproductive structures and greater biological yield, representing total dry matter production, tend to produce higher seed yields (30-33). Additionally, other traits such as days to 50 % flowering (0.01477), number of primary

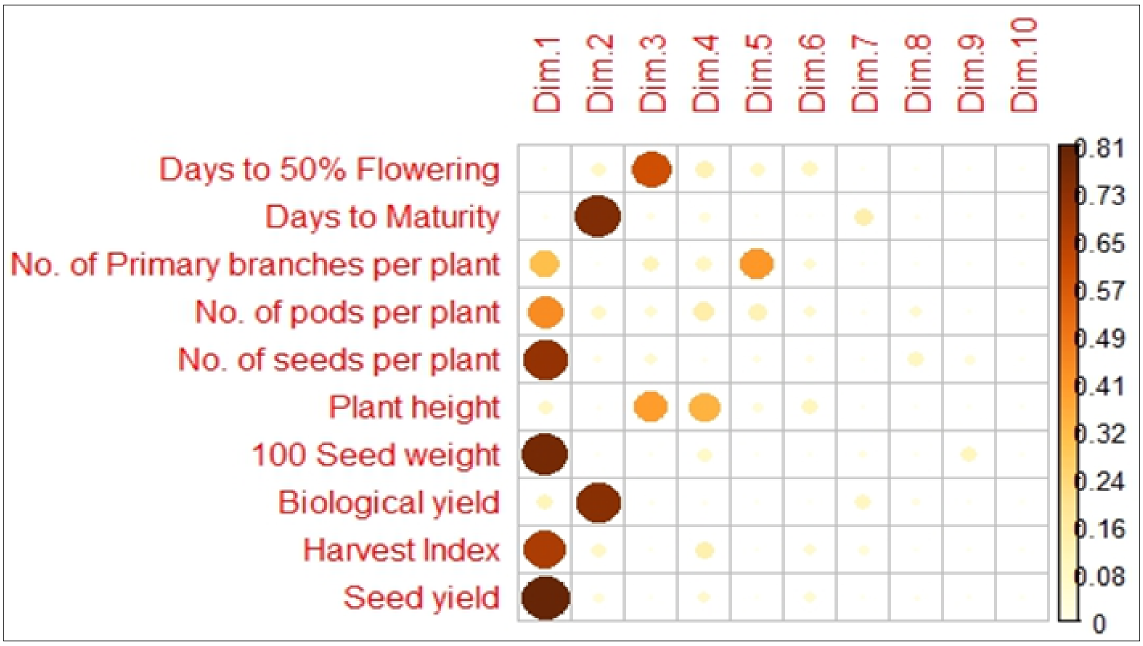

Fig. 1. Scatter plot diagram for different quantitative traits.

Table 2. Path coefficient analysis for yield and its component traits in soybean RILs

Characters	dff	dtm	nbp	npopp	nspp	ph	hsw	by	hi
dff	**0.01477**	0.05870	-0.00002	0.00451	-0.00775	-0.00201	-0.00024	-0.04983	0.00000
dtm	-0.00118	**-0.00734**	0.00002	-0.00052	-0.00074	-0.00045	0.00021	0.38203	-0.14204
nbp	-0.00015	-0.00007	**0.00218**	0.00468	-0.01329	0.00380	0.00134	0.08305	0.32846
npopp	0.00384	0.00022	0.00059	**0.01733**	-0.02584	0.00280	0.00161	0.01661	0.37284
nspp	0.00310	-0.00015	0.00079	0.01213	**-0.03692**	0.00190	0.00244	0.10520	0.54151
ph	-0.00266	0.00029	0.00074	0.00433	-0.00628	**0.01119**	0.00034	0.00000	0.14204
hsw	-0.00103	-0.00044	0.00085	0.00815	-0.02621	0.00112	**0.00343**	0.16610	0.64803
by	-0.00133	-0.00506	0.00033	0.00052	-0.00701	0.00000	0.00103	**0.55367**	-0.06214
hi	0.00000	0.00117	0.00081	0.00728	-0.02252	0.00179	0.00251	-0.03876	**0.88772**
Residual Effect^2 =	0.002011886								

branches per plant (0.002180), number of pods per plant (0.01733), plant height (0.01119) and 100-seed weight (0.00343) exhibit positive but relatively more minor direct effects on seed yield. These results align with existing studies, suggesting early flowering can enhance seed yield potential by synchronizing pod sets. In contrast, increased branching, pod number and plant height contribute to higher seed production (18, 33). Larger seed size, represented by 100-seed weight, is also associated with increased yield (22, 23).

Conversely, days to maturity (-0.00734) and number of seeds per plant (-0.03692) have negligible adverse effects on seed yield. This implies that genotypes with delayed maturity and more seeds per plant may experience reduced seed yields. More extended maturity periods can lead to a shorter reproductive phase, while a more significant number of seeds per plant can result in resource competition and smaller individual seed sizes. This finding concurs with prior studies emphasizing the role of indirect effects of traits on seed yield through other variables. For instance, days to maturity have a high indirect impact on seed yield through biological yield (0.382), indicating that delayed maturity can indirectly enhance yield by increasing overall biomass production (34, 35). Similarly, the number of primary branches per plant (0.32846) positively affects seed yield through the harvest index, suggesting that increased branching can contribute to greater resource allocation to reproductive organs (36). These findings underscore the complexity of trait interactions in influencing seed yield. They highlight the importance of considering direct and indirect effects when breeding for improved soybean seed yields.

Principal component analysis

Principal Component Analysis (PCA) is a powerful statistical method used for dimension reduction, aiming to condense a large set of variables into smaller ones while retaining the essential information (37). It mathematically transforms correlated variables into principal components that capture most of the datas' variability. In this study, PCA was applied to evaluate multiple soybean traits and the outcomes, including eigenvalues, percentage of variance, PCA scores and cumulative eigenvalues, are presented in Table 3-5 & Fig. 2.

Out of 10 (Principal Components) PCs generated, only seven exhibited eigenvalues greater than 0.5, collectively accounting for 95.93 % of the total variability among the studied traits. These seven PCs were prioritized for further analysis. PC1 showed the highest variability at 39.38 %, followed by PC2 (18.3 %), PC3 (12.75 %), PC4 (10.42 %), PC5 (6.913 %), PC6 (4.81 %) and PC7 (3.357 %), depicting the highest variability in PC1 (38, 39, 40). The scree plot, a graphical representation of eigenvalues and the percentage of cumulative variance associated with each PC confirmed the significance of these seven PCs (Fig. 3). The steep decline in variance after the seventh PC indicated that the first PC captured the majority of the variation. Aligning with the previous investigations, this underscores the importance of PC1 in explaining the variability among the studied traits and suggests that focusing on characteristics associated with PC1 may be most beneficial (41, 42). Further analysis revealed the composition of each principal component. PC1 encompassed traits such as seed yield, 100-seed weight, harvest index, number of pods per plant and number of seeds per plant. PC2 primarily comprised days to maturity and biological yield, while PC3 was associated with plant height. PC4 included days to 50% flowering and PC5 was linked to the number of primary branches per plant.

The interpretation of these findings concurs with the previous studies suggesting that the traits contributing to PC1 have the most substantial impact on the overall variability observed in the soybean RILs (43-45). Therefore, breeders and researchers may prioritize these traits when designing soybean breeding programs. The results for PC2, PC3, PC4 and PC5 also offer valuable insights into the relationships and contributions of days to maturity, biological yield, plant height, days to 50 % flowering and number of primary branches per plant, respectively. The application of PCA in this study successfully reduced a complex set of soybean traits into seven principal components, collectively explaining

Table 3. Eigen value, percentage of total variation and cumulative percentage for corresponding traits in soybean RILs

Characters	Principal component (PC)	Eigen value	Cumulative %	Variability %
Days to 50 % Flowering	PC1	1.9845	39.38	39.38
Days to Maturity	PC2	1.353	57.69	18.3
No. of Primary branches per plant	PC3	1.1291	70.44	12.75
No. of pods per plant	PC4	1.0206	80.85	10.42
No. of seeds per plant	PC5	0.83142	87.765	6.913
Plant height	PC6	0.6935	92.57	4.81
100 Seed weight	PC7	0.57943	95.932	3.357
Biological yield	PC8	0.47761	98.213	2.281
Harvest Index	PC9	0.4219	99.99	1.78
Seed yield	PC10	0.02584	100	0.007

Table 4. PC scores of soybean RILs having positive values >0.5 in each PCs

	PC1	PC2	PC3	PC4	PC5
RIL 115-1	-1.113	0.433	-0.310	0.856	0.844
RIL 115-4	1.252	0.541	-0.914	0.841	0.167
RIL 115-36	-0.334	-0.746	-0.667	1.985	-0.684
RIL 115-41	-2.400	0.027	-0.799	1.416	0.320
RIL 115-42	0.881	-0.290	-0.601	1.266	0.144
RIL 115-43	-1.355	0.813	-0.632	0.671	0.368
RIL 115-45	3.001	-0.429	-0.657	0.650	0.102
RIL 115-51	2.097	-0.537	0.151	-0.026	0.990
RIL 115-57	0.140	0.735	-0.577	0.260	0.425
RIL 115-62	-0.579	-0.922	-2.137	1.003	0.075
RIL 115-64	-0.231	0.771	-0.515	1.721	-0.111
RIL 115-65	3.769	1.176	-0.988	0.088	0.921
RIL 115-66	0.098	0.704	-0.466	0.230	1.024
RIL 115-67	0.784	-0.528	-0.870	0.516	-0.287
RIL 115-69	-1.720	-0.048	-2.689	1.004	0.206
RIL 115-71	-1.746	-0.108	-1.395	-0.261	0.221
RIL 115-74	3.331	0.420	-0.419	-0.659	0.075
RIL 115-76	4.059	-0.275	-1.705	-0.277	0.961
RIL 115-89	3.423	1.242	-0.873	-0.362	0.423
RIL 115-91	-1.234	0.154	-0.768	0.357	-0.867
RIL 115-92	0.590	-0.872	-0.528	0.269	-0.064
RIL 115-93	1.159	0.531	-0.150	-0.784	1.288
RIL 115-96	2.822	-1.321	0.213	-0.507	0.494
RIL 115-99	2.786	0.027	-0.580	0.544	0.419
RIL 115-108	2.769	-0.347	0.868	-1.049	2.049
RIL 115-118	-1.111	-0.691	-0.302	-1.137	-0.923
RIL 115-132	-0.422	-0.598	0.373	0.084	1.183
RIL 115-140	-1.575	0.139	-0.778	0.751	0.262
RIL 115-141	-3.143	0.778	-1.058	0.161	0.307
RIL 115-144	1.782	-1.637	-0.959	1.707	-0.133
RIL 115-150	1.674	-0.385	-1.828	-0.649	0.475
RIL 115-153	-1.803	-0.872	-1.945	1.123	0.630
RIL 115-154	-2.835	-2.395	0.504	-2.339	-1.148
RIL 115-155	-1.256	-0.805	-1.146	0.348	1.476
RIL 115-156	1.172	-0.009	0.132	-0.309	-0.511
RIL 115-159	-2.148	-1.561	-2.225	-1.260	-0.085
RIL 115-160	-4.514	-1.201	-0.368	-0.781	-0.724
RIL 115-161	-2.722	0.971	-1.009	-1.784	0.506
RIL 115-162	-2.139	-2.484	-1.365	-0.967	0.040
RIL 115-163	0.361	1.234	-0.683	-2.226	2.251
RIL 115-164	-1.936	-1.513	-1.610	-0.027	0.003
RIL 115-166	-3.118	2.094	0.518	-0.510	0.813
RIL 115-167	-0.510	-1.041	-1.944	0.311	1.018
RIL 115-169	3.056	1.172	-0.344	-0.596	-0.048
RIL 115-171	-0.560	-0.023	-0.905	2.133	-0.051
RIL 115-173	-2.802	0.438	-1.067	-1.033	-1.230
RIL 115-175	-0.320	-1.021	-0.615	-2.758	0.399
RIL 115-176	-3.293	-1.502	-0.899	-2.142	0.569
RIL 115-177	-1.143	0.577	0.649	-0.149	0.446
RIL 115-178	0.720	1.011	-0.468	0.062	-0.446
RIL 115-179	-1.904	0.433	-1.355	0.562	-1.435
RIL 115-181	-0.297	1.828	0.392	-0.248	1.353
RIL 115-182	2.625	0.019	-0.662	0.490	-0.429
RIL 115-183	-1.266	-1.686	-0.089	0.154	-0.867
RIL 115-184	2.681	-0.004	0.065	-0.068	0.503
RIL 115-188	0.130	0.724	-0.242	0.164	-0.063
RIL 115-193	-0.302	-0.483	-0.850	1.149	0.031
RIL 115-194	-0.652	1.072	-1.449	0.793	-0.741
RIL 115-199	-1.427	-0.616	-0.712	0.456	0.829
RIL 115-206	-4.227	1.719	-1.884	-0.416	-1.228
RIL 115-211	-1.552	-0.470	-0.239	1.483	-0.931
RIL 115-221	2.463	2.231	0.471	-0.244	-0.820
RIL 115-222	0.890	2.045	0.848	-0.548	-0.693
RIL 115-223	0.189	1.394	0.326	1.094	-0.940
RIL 115-225	0.974	-0.790	0.884	0.294	1.201
RIL 115-227	2.779	0.765	0.533	-0.947	-0.281
RIL 115-230	-2.110	2.791	0.581	-0.627	-1.579
RIL 115-233	0.999	-2.370	0.930	-1.416	1.509
RIL 115-237	2.458	-2.545	0.057	-0.627	-0.531
RIL 115-238	0.275	-0.160	-0.101	0.234	-1.734
RIL 115-239	-0.260	0.717	-1.483	-1.290	-1.523
RIL 115-243	1.342	0.952	-0.240	0.062	0.630
RIL 115-252	-2.392	-1.115	-1.105	-0.562	0.670
RIL 115-253	-3.837	-1.092	-0.747	-1.169	0.225
RIL 115-256	-0.180	-1.759	-0.496	1.090	-0.416
RIL 115-258	-0.295	1.391	0.054	-0.433	0.098
RIL 115-260	2.235	-1.784	-0.273	0.075	0.039

RIL 115-265	1.242	-0.414	-1.230	-0.756	0.593
RIL 115-267	-0.413	-1.627	0.794	-1.451	0.452
RIL 115-268	-1.997	-1.485	-0.123	-0.504	-0.800
RIL 115-269	-0.596	-0.831	-0.789	-0.525	-1.397
RIL 115-272	0.188	-2.666	1.200	0.645	-0.080
RIL 115-273	2.466	0.087	0.532	-0.201	-0.712
RIL 115-274	2.515	-2.486	0.961	0.588	-0.742
RIL 115-277	1.333	-2.395	-0.425	0.677	-0.374
RIL 107-38	-0.142	0.615	-0.685	1.587	0.523
RIL 107-63	2.412	1.812	0.755	0.313	1.032
RIL 107-77	1.411	1.838	1.245	-1.006	0.260
RIL 107-99	2.328	-0.764	1.036	1.492	-0.349
RIL 107-166	-0.073	-2.105	-0.037	-1.156	-0.779
RIL 107-24	-2.121	0.340	2.781	-4.039	-1.213
RIL 107-31	-0.384	-1.000	2.103	0.060	-1.652
RIL 107-35	-1.301	-0.419	0.096	0.514	-0.079
RIL 107-43	3.321	0.768	0.731	-0.636	-0.686
RIL 107-44	2.580	-0.846	0.528	-0.264	-0.364
RIL 107-46	1.660	-1.711	1.400	1.100	-0.826
RIL 107-57	1.844	0.128	0.361	-0.346	-1.116
RIL 107-68	2.055	0.971	0.197	-2.311	-0.196
RIL 107-71	0.092	0.039	1.123	0.346	-1.627
RIL 107-72	2.360	0.206	0.205	-1.447	-0.199
RIL 107-73	0.477	1.041	1.333	0.872	-0.950
RIL 107-74	1.897	-0.831	0.586	-1.731	-1.407
RIL 107-75	1.601	-1.822	-0.200	1.195	-0.232
RIL 107-79	0.860	1.307	-0.074	-0.231	-0.384
RIL 107-80	2.119	0.348	-0.922	-1.725	-0.895
RIL 107-87	-0.292	1.280	-0.018	0.814	-1.102
RIL 107-88	1.808	-2.101	1.077	1.058	-0.249
RIL 107-96	3.863	1.620	0.168	-0.086	-0.527
RIL 107-103	2.546	0.579	0.262	0.305	0.729
RIL 107-128	-0.624	4.133	0.118	0.101	0.310
RIL 107-131	-0.064	1.571	0.284	0.506	-0.629
RIL 107-137	1.298	2.388	0.670	-0.097	-1.183
RIL 107-139	-0.605	4.991	-0.244	0.553	-1.159
RIL 107-151	0.666	0.668	-0.244	0.960	0.560
RIL 107-161	-0.208	2.044	-0.267	0.395	0.726
JS 97-52	-2.396	1.022	1.074	-0.280	1.208
NRC-37	-0.415	-1.117	2.515	0.275	0.434
JS-335	-2.728	0.295	2.191	1.394	0.094

Table 5. Rotated component matrix for different traits of soybean RILs

Traits	PC1	PC2	PC3	PC4	PC5
Days to 50 % Flowering	0.026	-0.217	-0.686	-0.340	0.351
Days to Maturity	0.061	0.643	-0.132	-0.189	-0.083
No. of Primary branches per plant	0.282	-0.028	0.299	-0.300	0.768
No. of pods per plant	0.333	-0.213	-0.223	-0.372	-0.418
No. of seeds per plant	0.427	-0.112	-0.213	-0.109	-0.189
Plant height	0.144	-0.089	0.558	-0.576	-0.243
100 Seed weight	0.444	0.029	0.002	0.262	-0.051
Biological yield	0.169	0.635	-0.098	-0.118	0.033
Harvest Index	0.412	-0.208	0.083	0.361	0.053
Seed yield	0.454	0.157	0.017	0.250	0.065

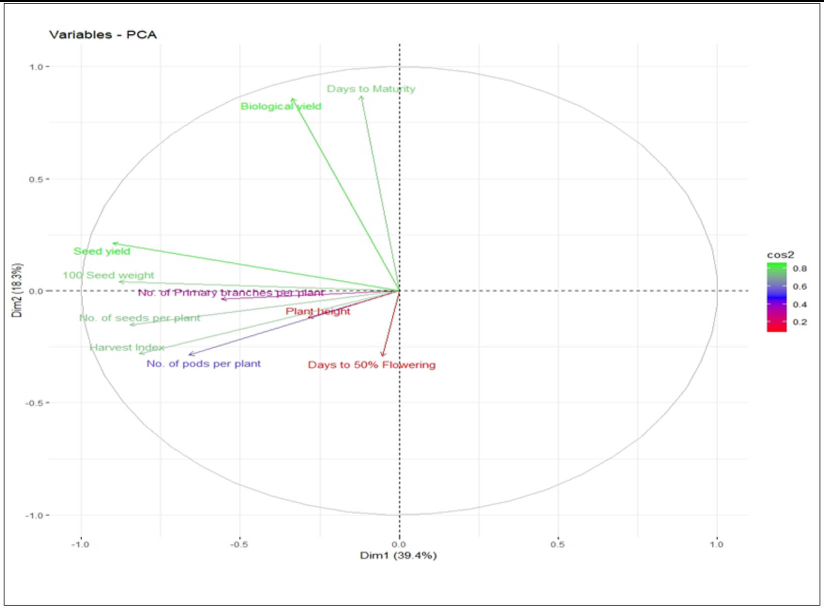

Fig. 2. Correlated response of PCs variables.

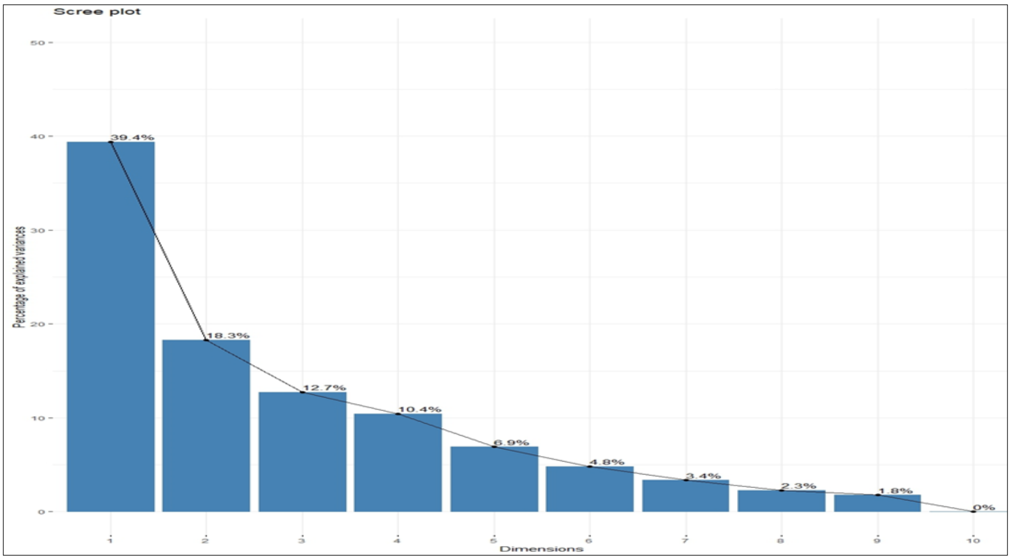

Fig. 3. Screen plot showing Eigen value and percentage of cumulative variability.

95.93 % of the variance. These findings provide valuable guidance for future soybean breeding programs and facilitate the selection of genotypes with desirable traits, ultimately contributing to soybean improvement.

Biochemical analysis

A thorough biochemical investigation of the genotypes of soybeans in this study showed notable differences in important compositional characteristics, such as the amount of moisture, oil, protein and ash in the seeds (Table 6). RIL 115 -71 had the highest oil content at 21.224 %, followed by RIL 115-62 at 20.424 % and RIL 115-159 at 19.721 %. This level of oil content significantly surpasses that of other genotypes. On the other end of the spectrum, the genotypes with the most inferior seed oil content were RIL 107-72 and RIL 115-160, denoting 16.432 % and 16.569 %, respectively. Regarding seed protein content, genotypes RIL 115-92 (42.467 %), RIL 115-163 (42.218 %) and RIL 115-159 (41.597 %) displayed the highest quantities compared to all other studied genotypes. Conversely, the genotypes RIL 107-72 (34.489 %), RIL 115-160 (34.587 %) and RIL 115-64 (35.711 %) recorded the lowest seed protein content. Assessing ash content revealed that genotype JS-335 recorded a maximum value of 6.83 %, with RIL 115-71 (6.486 %) and RIL 115-178 (5.81 %) trailing behind, significantly surpassing other genotypes. The lowest ash content was found in genotypes RIL 115-161 (3.72 %), RIL 115-64 (4.753 %) and RIL 115-238 (4.763 %). Since biochemical analysis was conducted on RILs exhibiting superior performance across various traits, those with enhanced biochemical composition can be further utilized for breeding

Table 6. Biochemical analysis of 31 RILs and 2 Checks

S. No.	Name of RILs	Seed moisture %	Seed Oil %	Seed protein N %	Ash %
1	RIL 115-43	6.341	17.458	41.592	5.01
2	RIL 115-45	6.326	19.685	41.543	5.625
3	RIL 115-62	5.004	20.424	38.089	5.469
4	RIL 115-64	5.634	16.769	35.711	4.753
5	RIL 115-69	6.026	17.223	39.839	4.963
6	RIL 115-71	5.49	21.224	40.592	6.486
7	RIL 115-91	4.98	18.275	35.90	5.434
8	RIL 115-92	3.84	17.448	42.467	5.72
9	RIL 115-93	6.29	17.03	38.089	5.53
10	RIL 115-140	4.206	16.642	37.084	5.55
11	RIL 115-159	4.474	19.721	41.597	5.42
12	RIL 115-160	4.526	16.569	34.587	5.33
13	RIL 115-161	3.998	18.063	37.144	3.72
14	RIL 115-162	4.72	16.581	40.278	5.295
15	RIL 115-163	6.282	16.745	42.218	5.14
16	RIL 115-164	7.16	16.915	36.338	5.335
17	RIL 115-171	4.65	17.415	38.089	5.25
18	RIL 115-176	5.4	19.532	41.467	5.505
19	RIL 115-177	6.85	18.396	39.084	5.22
20	RIL 115-178	4.50	19.113	40.278	5.81
21	RIL 115-179	5.546	18.496	37.214	5.43
22	RIL 115-238	6.238	17.753	36.391	4.763
23	RIL 115-253	6.56	19.385	38.089	5.165
24	RIL 107-99	4.882	19.00	39.278	5.24
25	RIL 107-166	5.79	16.89	40.164	5.305
26	RIL 107-71	4.376	18.541	39.364	5.77
27	RIL 107-72	5.497	16.432	34.489	5.36
28	RIL 107-73	6.28	17.723	39.687	4.97
29	RIL 107-87	6.249	16.424	40.456	5.19
30	RIL 107-128	3.98	19.739	39.486	5.25
31	RIL 107-161	3.456	20.436	40.014	5.315
32	NRC-37	5.13	17.185	37.648	5.085
33	JS-335	5.40	17.01	41.219	6.83

programs and industrial applications, particularly in nutrient enhancement and biofortification initiatives. Similar studies suggest that this variation in these genotypes' oil, seed protein and ash content suggests diversity in their genetic structure and potential adaptations according to their growing environment (46-48).

Conclusion

This research advances soybean genetics and yield optimization by identifying key traits for improved production. Correlation analysis showed strong positive links between seed yield and factors like harvest index, 100 seed weight, seed count and biological yield, highlighting the role of resource allocation and seed size in yield enhancement. Path analysis further pinpointed harvest index and biological yield directly impacting seed production. Principal component analysis simplified trait complexity, with PC1 being the most influential (including yield-related traits). These findings guide soybean breeding by prioritizing traits in PC1, adding efforts to boost yield and support global food security amidst climate challenges.

Furthermore, investigating biochemical traits, including seed moisture, oil, protein and ash content, highlighted significant variations among genotypes, emphasizing the potential applications of different soybean lines in various industries. The impact pathways identified in this study highlight the direct influence of harvest index and biological yield on seed production, emphasizing their role in resource allocation efficiency. Biochemical trait variations suggest applications in industry and targeted breeding for improved soybean adaptability. These findings provide a framework for selecting high-yielding genotypes to enhance global food security. Future research should focus on stress tolerance and adaptability across diverse agro-climatic conditions.

Abbreviations

AICRP: All India Coordinated Research Project; RIL: Recombinant Inbred Lines; PCA: Principal Component Analysis; PC: Principal Component; Dff: Days to 50 % Flowering; dtm: Days to Maturity; nbp: No. of Primary branches per plant; npopp: No. of pods per plant; nspp: No. of seeds per plant; ph: Plant height; hsw: 100-Seed weight; by: Biological yield; hi: Harvest Index

Authors' contributions

RM was responsible for drafting and writing the manuscript. MS, MT, PA and YS contributed to the study's conceptualisation and provided essential project supervision. RS conducted the data analysis, while SS[1] and SS[2] were instrumental in preparing the research framework. All authors critically reviewed and approved the final version of the manuscript. [SS[1] stands for Sanjeev Sharma and SS[2] stands for Sachin Shukla].

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None

References

1. Mishra R, Tripathi MK, Tripathi N, Singh J, Yadav PK, Sikarwar RS, et al. Breeding for major genes against drought stress in soybean. In: Tripathi MK, Tripathi N, editors. Advances in Plant Biotechnology. Puducherry, India: Cornous Publications LLP; 2024. p. 22–68. Available from: https://doi.org/10.37446/volbook032024/22-68
2. Mishra R, Tripathi MK, Sikarwar RS, Singh Y, Tripathi N. Soybean (*Glycine max* L. Merrill): A multipurpose legume shaping our world. Plant Cell Biotechnol Mol Biol. 2024;25(3–4):17–37. Available from: https://doi.org/10.56557/pcbmb/2024/v25i3-48643
3. Mishra R, Tripathi M, Tripathi N, Singh J, Tiwari S. Nutritional and anti-nutritional factors in soybean. Acta Sci Agricult. 2024;8(11):46–63. Available from: https://doi.org/10.31080/ASAG.2024.08.1432
4. The Soybean Processors Association of India (SOPA); 2025. Available from: https://www.sopa.org/
5. Deshmukh R, Sonah H, Patil G, Chen W, Prince S, Mutava R, et al. Integrating omic approaches for abiotic stress tolerance in soybean. Front Plant Sci. 2014;5:244. Available from: https://doi.org/10.3389/fpls.2014.00244
6. Shelke DB, Chambhare MR, Nikalje GC, Nikam TD. Improvement of soybean crop for yield, stress tolerance and value-added products using a transgenic approach. Adv Agricult. 2023;2023(1):8166928. Available from: https://doi.org/10.1155/2023/8166928
7. Rahman SU, McCoy E, Raza G, Ali Z, Mansoor S, Amin I. Improvement of soybean: A way forward transition from genetic engineering to new plant breeding technologies. Mol Biotechnol. 2023;65(2):162–80. Available from: https://doi.org/10.1007/s12033-022-00456-6
8. Shaibu AS, Ibrahim H, Miko ZL, Mohammed IB, Mohammed SG, Yusuf HL, et al. Assessment of the genetic structure and diversity of soybean (*Glycine max* L.) germplasm using diversity array technology and single nucleotide polymorphism markers. Plants. 2021;11(1):68. Available from: https://doi.org/10.3390/plants11010068
9. Mishra R, Tripathi MK, Shrivastava MK, Amrate PK. 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; 2024. p. 1–21. Available from: https://doi.org/10.37446/volbook032024/1-21
10. Karikari B, Chen S, Xiao Y, Chang F, Zhou Y, Kong J, et al. Utilization of interspecific high-density genetic map of RIL population for the QTL detection and candidate gene mining for 100-seed weight in soybean. Front Plant Sci. 2019;10:1001. Available from: https://doi.org/10.3389/fpls.2019.01001
11. Priyadarshan PM. Recombinant inbred lines. In: Plant Breeding: Classical to Modern. Singapore: Springer Singapore; 2019. p. 257–68. Available from: https://doi.org/10.1007/978-981-13-7095-3_13
12. Cao Y, Li S, Chen G, Wang Y, Bhat JA, Karikari B, et al. Deciphering the genetic architecture of plant height in soybean using two RIL populations sharing a common M8206 parent. Plants. 2019;8(10):373. Available from: https://doi.org/10.3390/plants8100373
13. Mahalakshmi PJ, Satish Y, Ramana JV, Raju MRB. Correlation and path coefficient analysis for yield and its component traits in advanced breeding lines of rice (*Oryza sativa* L.). J Adv Biol Biotechnol. 2024;27(9):769–77. Available from: https://doi.org/10.9734/jabb/2024/v27i91350
14. David CC, Jacobs DJ. Principal component analysis: A method for determining the essential dynamics of proteins. Methods Mol Biol.

2014;1084:193–226. Available from: https://doi.org/10.1007/978-1-62703-658-0_11

15. Ghodrati GR, Sekhavat R, Mahmood SH, Gholami A. Evaluation of correlations and path analysis of components seed yield in soybean. Int J Agricu Res Rev. 2013;3(4):795–800.
16. Koraddi S, Basavaraja GT, Immadi S, Kumar AG. Correlation and path coefficient analysis for yield and its component in soybean [*Glycine max* (L.) Merill]. The Bioscan. 2015;10(4):2065–67.
17. Chavan BH, Dahat DV, Rajput HJ, Deshmukh MP, Diwane SL. Correlation and path analysis in soybean. Int Res J Multidiscip Stud. 2016;2(9):122–29.
18. Balla MY, Ibrahim SE. Genotypic correlation and path coefficient analysis of soybean [*Glycine max* (L.) Merill.] for yield and its component. Agric Res Technol. 2017;7(3):01–05.
19. Shree Y, Ram S, Bhushan S, Verma N, Ahmad E, Kumar S. Correlation between yield and yield attributing traits in soybean (*Glycine max* (L.) Merill). J Pharmacog Phytochem. 2018;6(4):298–301.
20. Desissa DH, Gemechu HL. Correlation and path analysis studies among yield and yield-related traits in soybean (*Glycine max* (L.) Merrill) varieties. Int J Plant Breed Crop Sci. 2018;5(1):352–60.
21. Mehra S, Shrivastava MK, Amrate PK, Yadav RB. Studies on variability, correlation coefficient and path analysis for yield-associated traits in soybean [*Glycine max* (L.) Merrill]. J Oilseeds Res. 2020;37(1):56–59. Available from: https://doi.org/10.56739/jor.v37i1.136390
22. Shekhar GC, Prashant M, Pushpendra SM, Lokesh K, Shrotia PK, Singh K, et al. Correlation and path analysis in soybean [*Glycine max* (L.) Merill]. Int J Curr Microbiol App Sci. 2018;7(09):1232–29. Available from: https://doi.org/10.20546/ijcmas.2018.709.147
23. Pawar MG, Chaudhary SB, Pawar VS, Chavan SB. Correlation coefficient and path analysis study in different soybean genotypes based on yield and yield-contributing traits. Int J Curr Microbiol App Sci. 2020;9(9):434–44. Available from: https://doi.org/10.20546/ijcmas.2020.909.055
24. Lyimo LD, Tamba MR, Madege RR. Effects of genotype on yield and yield components of soybean (*Glycine max* (L.) Merrill). Afr J Agric *Res*. 2017;12(22):1930–39. Available from: https://doi.org/10.5897/AJAR2017.12177
25. Painkra P, Shrivatava R, Nag SK, Kute I. Correlation analysis for seed yield and its attributing traits in soybean (*Glycine max* L. Merrill). Int J Curr Microbiol *Appl Sci*. 2018;7(04):2034–40. Available from: https://doi.org/10.20546/ijcmas.2018.704.234
26. Alpana P, Pushpendra K, Singh K, Gupta MK, Bhareti P. Morphological characterization and genetic divergence study in soybean. Soybean Res. 2015;13(2):01–08.
27. Kuswantoro H, Adie MM, Putri PH. Genetic variability, heritability and genotypic correlation of soybean agronomic characters. Buletin Palawija. 2021;19(2):117. Available from: https://doi.org/10.21082/bulpa.v19n2.2021.p117-125
28. Mahbub MM, Shirazy BJ. Evaluation of genetic diversity in different genotypes of soybean [*Glycine max* (L.) Merrill]. Am J Plant Biol. 2016;1(1):24–29. Available from: https://doi.org/10.11648/j.ajpb.20160101.14
29. Jain S, Srivastava SC, Singh SK, Indapurkar YM, Singh BK. Studies on genetic variability, character association and path analysis for yield and its contributing traits in soybean [*Glycine max* (L.) Merrill]. Legume Res- An Indian Journal. 2015;38(2):182–84. Available from: https://doi.org/10.5958/0976-0571.2015.00031.4
30. Tigga A. Determination of significant characters for improving seed yield in soybean *(Glycine max* L. Merrill) via correlation and path coefficient analysis. Int J Agric Environ Biotechnol. 2021;14(3):299–305. Available from: https://doi.org/10.30954/0974-1712.03.2021.2
31. Nagarajan D, Kalaimagal T, Murugan E. Correlation and path coefficient analysis for yield and yield attributes in soybean, *Glycine max* (L.) Merrill. Int J Farm Sci. 2015;5(3):28–34.
32. Kumar R, Singh V, Singh SK, Singh BP. Character association, path coefficient analysis and genetic diversity in soybean [*Glycine max* (L.) Merrill*]*. Legume Res. 2018;41(1):81–88.
33. Yahaya SU, Ankrumah E. Character association and path coefficient analysis for yield components and grain yield in soybean [*Glycine max* (L.) Merrill]. Legume Res. 2017;40(5):231–39. Available from: https://doi.org/10.18805/lr.v0i0.8408
34. Berhanu H, Tesso B, Lule D. Correlation and path coefficient analysis for seed yield and yield-related traits in soybean [*Glycine max* (L.)] genotypes. Plant Sci. 2021;9(4):106–10. Available from: https://doi.org/10.11648/j.plant.20210904.15
35. Dangi V, Sharma KK. Genetic variability, correlation and path analysis for grain yield and its components in soybean. Int J Curr Microbiol Appl Sci. 2021;10(3):1776–82. Available from: https://doi.org/10.20546/ijcmas.2021.1003.221
36. Massay HA. Principal components analysis. *J Am Stat Assoc*. 1965;60(309):234–56. https://doi.org/10.1080/01621459.1965.10480787
37. Birla D, Ramgiry SR. AMMI analysis to comprehend genotype-by-environment (G × E) interactions in rainfed grown soybean [*Glycine max* (L.) Merrill]. Indian J Agric Res. 2015;49(1):39–45. Available from: https://doi.org/10.5958/0976-058X.2015.00005.0
38. Liying HE, Chengjie Y, Shouchang H, Yinjie S. Correlation, clustering and principal component analysis of primary agronomic traits of vegetable soybean. Hortic Sci. 2018;30(1):50–57. Available from: https://doi.org/10.3969/j.issn.1004-1524.2018.01.07
39. Singh PK, Shrestha J, Kushwaha UKS. Multivariate analysis of soybean genotypes. J Agric Natural Res. 2020;3(1):69–76. Available from: https://doi.org/10.3126/janr.v3i1.27092
40. El-Hashash EF. Genetic diversity of soybean yield based on cluster and principal component analyses. J Adv Biol Biotechnol. 2016;10(3):1–9. Available from: https://doi.org/10.9734/JABB/2016/29127
41. Verma V, Shrivastava MK, Mehra S, Amrate PK, Yadav RB. Estimation of genetic parameters for yield-associated traits and principal components in advanced breeding lines of soybean (*Glycine max* L. Merrill). Int J Curr Microbiol Appl Sci. 2021;10(01):2704–10. Available from: https://doi.org/10.20546/ijcmas.2021.1001.314
42. Dubey N, Avinashe HA, Shrivastava AN. Principal component analysis in advanced genotypes of soybean [*Glycine max* (L.) Merrill] over seasons. Plant Arch. 2018;18(1):501–06. Available from: https://doi.org/10.20546/ijcmas.2021.1003.019
43. Fuzy A, Ramona, Kovacs, Cseresnyes I, Paradi I, Bettina TS, Kelemen. Selection of plant physiological parameters to detect stress effects in pot experiments using principal component analysis. Acta Physiol Plant. 2019;41:56. Available from: https://doi.org/10.1007/s11738-019-2842-9
44. Goonde DB, Ayana NG. Genetic diversity and character association for yield and yield-related traits in soybean (*Glycine max* L.) genotypes. J Agric Sci Food Res. 2021;12:280.
45. Aulia R, Amanah HZ, Lee H, Kim MS, Baek I, Qin J, Cho BK. Protein and lipid content estimation in soybeans using Raman hyperspectral imaging. Front Plant Sci. 2023;14. Available from: https://doi.org/10.3389/fpls.2023.1167139
46. Etiosa O, Chika N, Benedicta A. Mineral and proximate composition of soybean. Asian J Phys Chem Sci. 2018;4(3):01–06. Available from: https://doi.org/10.9734/AJOPACS/2017/38530
47. Obua T, Sserumaga JP, Awio B, Nganga F, Odong TL, Tukamuhabwa P, Tusiime G, Mukasa SB, Nabasirye M. Multi-environmental evaluation of protein content and yield stability among tropical soybean genotypes using GGE biplot analysis. Agronomy. 2021;11(7):1265. Available from: https://doi.org/10.3390/agronomy11071265
48. Žilić SM, Barać MB, Pešić MB, Mladenović Drinić SD, Ignjatović-Micić DD, Srebrić MB. Characterization of proteins from the kernel of different soybean varieties. J Sci Food Agric. 2011;91(1):60–67. Available from: https://doi.org/10.1002/jsfa.4148

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.