



International Journal of Research in Agronomy

E-ISSN: 2618-0618
P-ISSN: 2618-060X
© Agronomy
NAAS Rating: 5.20
www.agronomyjournals.com
2025; 8(5): 415-420
Received: 02-02-2025
Accepted: 08-03-2025

Dinesh Akula
Scientist, Department of Plant
Breeding, Regional Sugarcane and
Rice Research Station, PJTAU,
Rudrur, Telangana, India

Vanisri S
Professor, Department of Plant
Breeding, Institute of
Biotechnology, PJTAU,
Hyderabad, Telangana, India

Rajender Reddy M
Scientist, Department of Plant
Breeding, Agricultural Research
Station, PJTAU, Adilabad,
Telangana, India

Balaji Naik B
Principal Scientist, Department of
Agronomy, PJTAU, Hyderabad,
Telangana, India

Santhosha Rathod
Senior Scientist, Department of
Agriculture Statistics, Indian
Institute of Rice Research,
Hyderabad, Telangana, India

Corresponding Author:
Dinesh Akula
Scientist, Department of Plant
Breeding, Regional Sugarcane and
Rice Research Station, PJTAU,
Rudrur, Telangana, India

Trait dissection for yield enhancement in *rabi* Soybean (*Glycine max* L.) using principal component and regression analyses

Dinesh Akula, Vanisri S, Rajender Reddy M, Balaji Naik B and Santhosha Rathod

DOI: <https://www.doi.org/10.33545/2618060X.2025.v8.i5f.2925>

Abstract

Soybean (*Glycine max* L.) is predominantly grown as a *kharif* crop in Telangana, where it demonstrates substantial yield potential. With its growing significance as a major oilseed crop and its proven role in enhancing the livelihoods of farmers, there is increasing interest in extending its cultivation to the *rabi* season. Despite its agronomic and economic promise, *rabi* soybean remains underexploited, primarily due to the absence of high-yielding, season-specific varieties adapted to off-season conditions. To address this gap, the present study was undertaken to dissect yield-contributing traits and identify promising genotypes for *rabi* soybean improvement. Fifty diverse soybean genotypes were evaluated during the first fortnight of November 2022 at the Regional Sugarcane and Rice Research Station, PJTSAU, Rudrur, using an alpha lattice design with two replications. Analysis of variance revealed significant genotypic variability for all the traits except seeds per pod, indicating substantial scope for genetic gain through selection. Correlation analysis, principal component analysis (PCA), and stepwise multiple regression collectively highlighted number of pods per plant as the most critical determinant of seed yield under *rabi* conditions. The traits number of pods on the main stem, plant height, days to 50% flowering, days to maturity, and 100-seed weight also exhibited strong associations with yield and should be given due consideration in selection strategies. PCA revealed that the first two principal components accounted for 57.7% of the total phenotypic variation, effectively capturing the major patterns of trait diversity. Based on multivariate analysis, five genotypes; G21, G5, G42, G25, and G3 were identified as genetically divergent and agronomically promising, offering valuable resources for parental selection and future breeding programs aimed at enhancing *rabi* soybean productivity.

Keywords: Soybean, *rabi*, yield, principal component analysis, correlation, regression

Introduction

Soybean (*Glycine max* L. Merr.; $2n = 40$), a photo-thermo-sensitive legume, plays a vital role in food, feed, and oilseed production. Often referred to as the "Golden Bean" and "Miracle Crop," soybean has gained popularity among Indian farmers, making it the country's second-most important oilseed crop after cotton. In 2024-25, soybean was cultivated over 126.9 lakh hectares in India, producing 133.6 lakh tonnes with an average productivity of 976 kg/ha. In Telangana, it is primarily grown as a *kharif* crop over 1.62 lakh hectares, yielding 2.38 lakh tonnes with a higher average productivity of 1502 kg/ha (Indiastat, 2024) [4]. With the expansion of irrigation infrastructure, the potential for *rabi* soybean cultivation in Telangana is increasing. Cultivating soybean during the *rabi* season offers multiple advantages, including crop diversification, reduced harvest-time rainfall risk, better seed quality, and improved germination. However, adoption remains limited, primarily due to the lack of high-performing, season-specific varieties. Limited studies have focused on identifying suitable genotypes and optimal sowing windows for *rabi* soybean cultivation. Previous research (e.g., Spoorthi *et al.*, 2024; Navya *et al.*, 2024; Mohod *et al.*, 2023; Mounika *et al.*, 2023; Lokesh *et al.*, 2020) [18, 15, 11, 14, 9] highlights significant genotype $\times$ environment interactions and the critical influence of sowing date on yield and seed quality. These studies consistently indicate that early *rabi* sowing, particularly in mid-November, enhances productivity, whereas delayed sowing results in reduced yields. Given

soybean's sensitivity to environmental cues and the complexity of yield as a polygenic trait, breeding for *rabi* adaptation requires a clear understanding of the trait architecture underlying yield. While direct selection for seed yield may be ineffective due to its polygenic nature, identifying and selecting correlated component traits can enhance breeding efficiency. Correlation analysis facilitates understanding of trait interrelationships, while Principal Component Analysis (PCA) helps in identifying the most influential traits contributing to phenotypic variation (Anderson, 1972; Morrison, 1982; Mishra *et al.*, 2017; García and García, 2010) [1, 13, 10, 3]. The present study was undertaken to identify key agronomic traits associated with seed yield in soybean genotypes under off-season conditions, using correlation and PCA as tools to inform trait-based selection for *rabi* soybean improvement.

Materials and Methods

An experimental trial was conducted at the Regional Sugarcane and Rice Research Station, PJTSAU, Rudrur, during the *rabi* season of 2022 to evaluate yield and yield-related traits in soybean. Fifty diverse soybean genotypes were sown during the first fortnight of November, each considered as an individual treatment. The experiment was laid out in an alpha lattice design with two replications to ensure precise estimation of genotypic effects while minimizing experimental error. The genotypes were randomly allocated into 10 incomplete blocks per replication, with each block comprising five genotypes. Each experimental plot consisted of two rows spaced at 45 cm between rows and 5 cm within rows. Each row measured 2 meters in length, resulting in a plot size of 1.8 m². Standard agronomic practices, as recommended by the PJTAU *Diksuchi* manual, were followed to ensure optimal crop growth and development.

Data were recorded on key agronomic and yield-related traits, including days to 50% flowering (FW), plant height (cm) (PH), number of branches per plant (BPP), number of pods on the main stem (PMS), total number of pods per plant (PPP), number of seeds per pod (SPP), 100-seed weight (g) (TW), seed yield per hectare (kg/ha) (Yha) and days to maturity (DM). Seed yield per plot was extrapolated to yield per hectare using the formula: $Yha = (\text{Plot yield in kg} \times 10,000) / 1.8 \text{ m}^2$. Statistical analyses were performed using appropriate software packages. Analysis of variance (ANOVA) was conducted using the R software "Metan" package to evaluate genotypic differences. Correlation analysis and Principal Component Analysis (PCA) were carried out using GRAPES software (Kujane *et al.*, 2019) [6] to identify trait associations and underlying patterns of multivariate variation.

Results and Discussion

Analysis of variance (ANOVA)

The evaluation of 50 soybean genotypes during the *rabi* season revealed substantial phenotypic variation across key agronomic and yield-related traits, reflecting the genetic diversity present in the panel. Analysis of variance (ANOVA) revealed highly significant differences among genotypes for all traits except seeds per pod (SPP), underscoring the existence of considerable genetic variability and the potential for effective selection (Table 1). These findings align with previous reports (Sivabharathi *et al.*, 2022) [17] and reinforce the reliability of trait-based selection in soybean improvement programs. Broad-sense heritability estimates were high for most traits, including days to 50% flowering (0.96), plant height (0.94), branches per plant (0.98), pods on the main stem (0.94), total pods per plant (0.92), seed

yield (0.92), 100-seed weight (0.70), and days to maturity (0.97). These values suggest that the observed phenotypic variation is largely attributable to genetic factors, and that selection based on phenotype would be highly effective for these traits. Similar trends were reported by Spoorthi *et al.* (2024) [18] in *rabi*-season evaluations of 15 soybean genotypes. Conversely, seeds per pod exhibited low heritability (0.21), indicating a higher influence of environmental factors on this trait and reduced potential for genetic gain through selection. Traits with high heritability and moderate to low coefficients of variation, such as pods per plant, plant height, seed yield, and days to maturity, are ideal candidates for phenotypic selection in breeding programs aiming to improve soybean performance under off-season conditions.

Trait Interrelationships and Correlation Analysis

Understanding the interrelationships among agronomic traits is fundamental to designing effective breeding strategies for yield improvement. In this study, Pearson's correlation coefficients were estimated among nine key traits and seed yield across 50 soybean genotypes grown under *rabi* season conditions (Table 2). The correlation matrix revealed several significant positive associations that offer valuable insights for trait-based selection. Seed yield per hectare (Yha) exhibited significant positive correlations with days to 50% flowering (FW; $r = 0.27$), plant height (PH; $r = 0.30$), branches per plant (BPP; $r = 0.318$), pods on the main stem (PMS; $r = 0.406$), and total pods per plant (PPP; $r = 0.424$). These associations suggest that genotypes expressing increased values for these traits tend to have superior yield performance.

Plant height was positively and significantly correlated with BPP ($r = 0.344$), PMS ($r = 0.48$), and PPP ($r = 0.675$), indicating that taller plants may possess more branching and pod-bearing capacity. Notably, PMS showed strong associations with both PPP ($r = 0.609$) and seed yield ($r = 0.406$), reinforcing its value as an indirect selection criterion for enhancing productivity. A moderate but significant positive correlation was also observed between days to 50% flowering and days to maturity, as well as with seed yield, suggesting that late-maturing genotypes may exploit the full length of the growing season to accumulate biomass and achieve higher yields under *rabi* conditions. These findings are consistent with earlier reports by Lodhi *et al.* (2023) [8] and Banerjee *et al.* (2022) [2], who also documented positive relationships between yield and component traits such as flowering time, plant stature, and pod number.

Among all the traits, total pods per plant (PPP) showed the strongest positive correlation with seed yield ($r = 0.424$), underscoring its role as a primary yield-contributing trait in *rabi* soybean. This highlights the importance of selecting genotypes with higher pod counts as a direct strategy to enhance productivity under off-season conditions. In conclusion, the positive intercorrelations among FW, PH, BPP, PMS, PPP, and Yha suggest that simultaneous selection for these traits could be an effective breeding approach for yield enhancement. Emphasizing PPP in selection schemes is particularly crucial, given its consistent and strong association with seed yield in the present study.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a robust multivariate statistical technique widely used to reduce dimensionality and extract key patterns from complex datasets. In the present study, PCA was performed on nine agro-morphological traits recorded across 50 soybean genotypes evaluated under *rabi* conditions. Of the nine principal components (PCs), only the first three had

eigenvalues greater than 1.00, cumulatively explaining 69.2% of the total phenotypic variation (Table 3). The remaining six components, each with eigenvalues below 1.00, contributed 31.8% of the variance, suggesting that the majority of the trait variation could be captured within the first three components.

PC1 accounted for 43.4% of the total variation and was primarily influenced by number of pods per plant (PPP, 17.34%), pods on the main stem (PMS, 15.6%), and days to 50% flowering (FW, 15.13%), indicating that these traits are major contributors to genotypic differentiation under *rabi* conditions. PC2 explained 14.2% of the variability, with major contributions from days to maturity (DM, 21.05%) and seed yield (Yha, 18.35%), suggesting its association with maturity duration and overall productivity. PC3, contributing 11.5% of the variation, was predominantly driven by 100-seed weight (TW, 64.6%), underscoring the importance of seed weight in yield determination and seed quality.

The Principal Component Analysis (PCA) revealed the contribution of various traits to eight principal components (PCs), each representing distinct underlying patterns in the data (Table 4). Days to 50% flowering (FW) loads heavily on PC1 (15.13%), PC2 (14.40%), and PC8 (17.08%), indicating its broad influence across several components. Seed yield per hectare (Yha) showed strong associations with PC2 (18.35%), PC3 (13.28%), and PC8 (22.88%), suggesting that yield is influenced by a combination of factors related to maturity, seed weight, pods per plant and days to 50% flowering traits.

Plant height (PH) and branches per plant (BPP) contribute significantly to PC5 (32.13% and 28.54%, respectively) and PC6 (45.10% and 34.38%), highlighting that these components capture plant phenology. Pods on the main stem (PMS) is predominantly represented by PC7 (69.85%), indicating it is largely independent from other traits. Pods per plant (PPP) has its highest contribution in PC8 (35.95%), with additional influence in PC1 and PC2, showing its relevance to both general productivity and phenology.

Seeds per pod (SPP) is strongly associated with PC4 (56.99%), which appears to represent reproductive efficiency. The 100 seed weight (TW) is almost exclusively captured by PC3 (64.69%), suggesting that seed size is a distinct and independent component in the trait structure. Finally, days to maturity (DM) is most influential in PC2 (21.05%), with notable contributions to PC4 (12.65%) and PC8 (13.28%), reflecting its multifactorial role in soybean development and yield. Overall, the PCA results indicates that seed yield is driven by multiple interacting components primarily maturity, pods per plant, 100 seed weight,

and flowering traits.

The scree plot (Figure 1) displayed a steep decline in eigenvalues after the fourth component, confirming that most of the variation is captured within the first four PCs. The biplot derived from PC1 and PC2 (Figure 2) effectively illustrated the distribution of genotypes in the trait space. Traits such as PPP, PMS, FW, DM, and Yha were positioned far from the origin, indicating their strong contribution to genotypic variance. In contrast, traits like seeds per pod (SPP) and TW were located closer to the origin, suggesting lower influence on total variation (Figure 3). These results are in agreement with previous studies by Sivabharathi *et al.* (2022)^[17] and Vijayakumar *et al.* (2022)^[19], which also emphasized the importance of yield-related traits in explaining genotypic variation in soybean. Cos^2 values, representing the squared coordinates of genotypes on the principal components, were used to assess genetic divergence. Genotypes G21, G5, G42, G25, and G3 exhibited the highest divergence as indicated by their position farthest from the origin and high cos^2 values and may perform exceptionally well in specific environments. Conversely, G28, G36, G31, and G10 were located near the origin, indicating lower genetic variability and are likely more stable across conditions. These findings highlight the potential of highly divergent genotypes as promising parental lines for future breeding programs targeting improved performance under *rabi* conditions.

Regression analysis

Regression analysis is a fundamental statistical method used to investigate the relationship between a dependent variable and one or more independent variables. It provides insights into how variations in an independent variable impact the expected value of the dependent variable, assuming all other factors remain constant. In addition to prediction, regression analysis can reveal potential causal relationships between variables (Montgomery *et al.*, 2012)^[12]. This technique is widely applied across various disciplines, including agriculture (Jaysawal *et al.*, 2023)^[5]. In the present analysis (Table 5), the regression results highlight the influence of different traits on the dependent variable. Specifically, the trait "pods per plant" (PPP) exhibited a coefficient of 45.873, indicating a strong positive relationship. This suggests that a one-unit increase in PPP corresponds to a 45 kg increase in the dependent variable, assuming all other factors are held constant. Similarly, the "100-seed weight" (TW) showed a coefficient of 108.013, indicating that each unit increase in TW is associated with a 108 kg increase in the dependent variable.

Table 1: Analysis of variance for yield and yield contributing traits

Trait	Mean sum of squares				Mean	CV (%)	h ²
	Genotype (df=49)	Replication (df=1)	Block (Rep) (df=18)	Error (df=31)			
FW	22.66***	1.44	0.37	0.834	38.2	2.39	0.96
PH	137.76***	17.61	12.86	8.761	29.2	10.15	0.94
BPP	1.16***	0.44	0.11	0.12	3.3	10.51	0.9
PMS	58.5***	0.8	4.1	3.47	16.2	11.46	0.94
PPP	168.8***	0.17	15.1	13.129	32.1	11.3	0.92
SPP	0.12 ^{ns}	0.16	0.03	0.094	2.3	13.18	0.21
TW	3.51***	0.04	1.37	1.07	11.9	8.64	0.7
Yha	366097.9***	693.44	66154.91*	30253.4	1649.5	10.54	0.92
DM	56.43***	0.49	1.0	1.578	88.5	1.42	0.97

***: significance at p=0.001 %. ns: non-significant

FW: days to 50% flowering, PH: plant height (cm), BPP: number of branches per plant, PMS: number of pods on main stem, PPP: number of pods per plant, SPP: number of seeds per

pod, TW: 100 seed weight (gm), Yha: Seed yield (kg/ha) and DM: Day to maturity. CV: Coefficient of variation, h²: Broad sense heritability.

Table 2: Pearson correlation between yield and yield attributing traits.

	FW	PH	BPP	PMS	PPP	SPP	TW	Yha	DM
FW	1								
PH	0.46***	1							
BPP	0.29**	0.34***	1						
PMS	0.40***	0.48***	0.42***	1					
PPP	0.42***	0.47***	0.45***	0.60***	1				
SPP	0.21*	0.26**	0.01	0.10	0.18	1			
TW	-0.20*	-0.23*	-0.09	-0.14	-0.22*	-0.10	1		
Yha	0.27**	0.30**	0.31**	0.40***	0.67***	0.17	0.11	1	
DM	0.81***	0.41***	0.19*	0.40***	0.26**	0.22*	-0.13	0.21*	1

*** Correlation is significant at 0.001 level (two tailed)

** Correlation is significant at 0.01 level (two tailed)

* Correlation is significant at 0.05 level (two tailed)

FW: days to 50% flowering, PH: plant height (cm), BPP: number of branches per plant, PMS: number of pods on main

stem, PPP: number of pods per plant, SPP: number of seeds per pod, TW: 100 seed weight (gm), Yha: Seed yield (kg/ha) and DM: Day to maturity.

Table 3: Eigen value, percentage of variance and cumulative proportion of the principal component.

Princomp	Eigenvalue	Percentage of variance	Cumulative percentage of variance
PC1	3.9	43.4	43.4
PC2	1.3	14.2	57.6
PC3	1.1	11.5	69.2
PC4	1.0	10.4	79.7
PC5	0.6	6.7	86.4
PC6	0.5	5.2	91.7
PC7	0.4	4.7	96.4
PC8	0.2	2.2	98.6
PC9	0.1	1.3	100.0

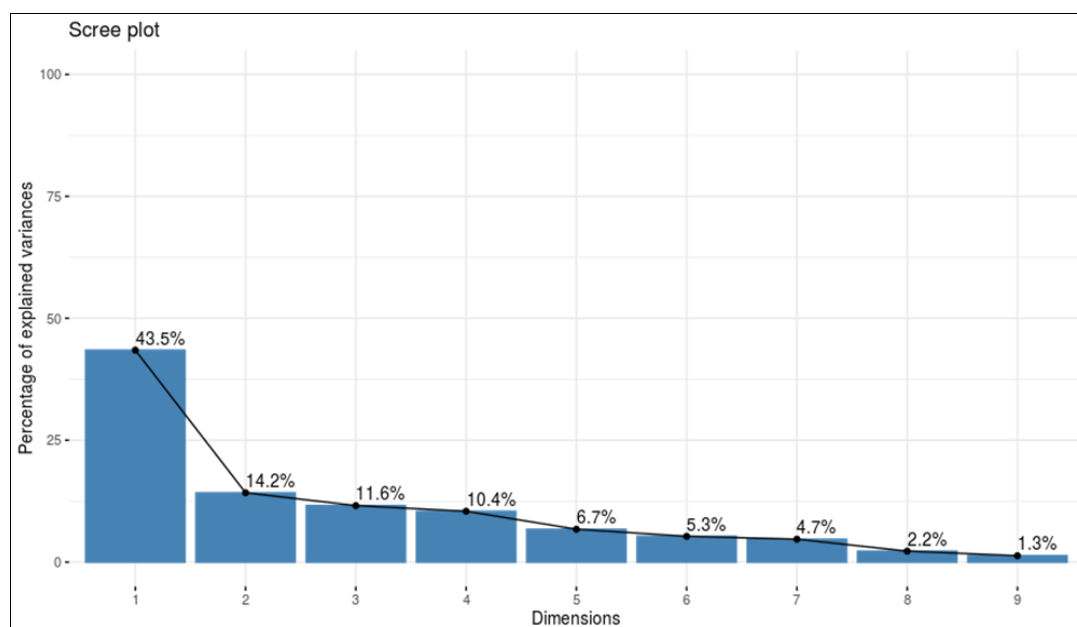
Table 4: percent contribution of variables on Principal components

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
FW	15.129	14.395	2.508	9.324	2.333	1.493	3.208	17.084
PH	13.493	1.396	1.108	0.886	32.134	45.104	5.814	0.044
BPP	9.414	14.462	2.57	7.871	28.542	34.375	1.581	1.183
PMS	15.689	3.452	1.443	0.69	0.562	2.981	69.849	1.345
PPP	17.338	9.714	1.169	2.984	8.769	0.011	0.912	35.95
SPP	3.937	10.373	6.396	56.996	5.593	13.221	3.426	0.042
TW	2.939	6.811	64.688	3.765	8.209	1.507	3.213	8.191
Yha	10.259	18.347	13.284	4.831	12.03	0.727	11.91	22.881
DM	11.803	21.049	6.835	12.653	1.828	0.58	0.086	13.279

Table 5: Regression Analysis for dependent (Seed yield) and independent variable among 50 soybean genotypes evaluated in *rabi* season.

	Estimate	Std. Error	t value	P-value
(Intercept)	-1816.916	949.605	-1.913	0.063
FW	-16.230	25.469	-0.637	0.527
PH	0.202	6.582	0.031	0.976
BPP	8.430	83.325	0.101	0.920
PMS	-4.154	14.365	-0.289	0.774
PPP	45.873	8.653	5.301	0.000
SPP	142.863	188.675	0.757	0.453
TW	108.013	34.108	3.167	0.003
DM	11.809	15.330	0.770	0.445

Regression co-efficients with P-value <0.05 are significant (alpha = 0.05)

**Fig 1:** Scree plot of principal component analysis of soybean germplasm accessions between percentage of variance and principal components

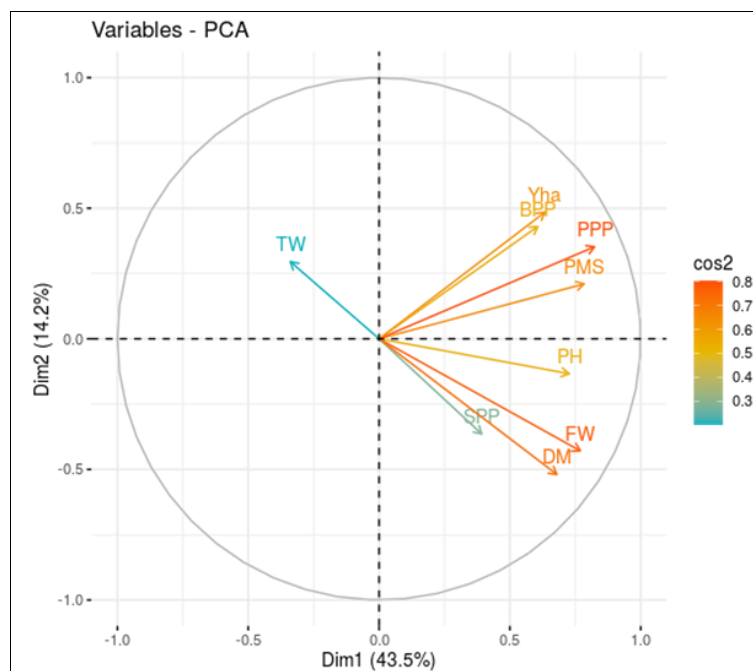


Fig 3: Variables plot with contribution of quantitative characters to the total divergence.

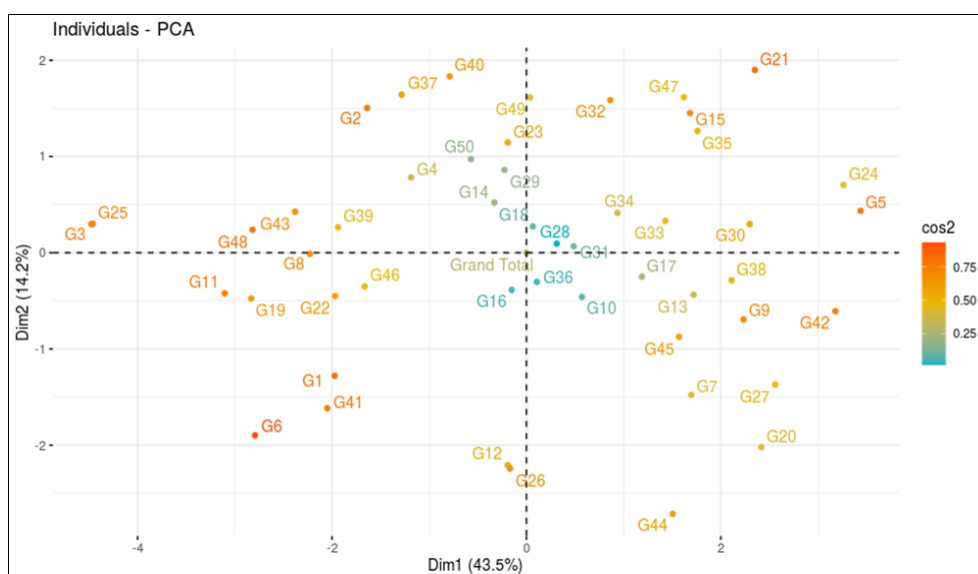


Fig 2: Genetic divergence of 50 soybean germplasm accessions in biplot with cos2 loadings. Genotypes in red color are diverse parents.

Conclusion

In conclusion, key traits such as pods per plant, pods on the main stem, plant height, and days to 50% flowering were identified as the primary factors influencing soybean yield. These findings suggest that targeting these traits in breeding programs could lead to substantial improvements in soybean yield under *rabi* conditions. Genotypes G21, G5, G42, G25, and G3 exhibited considerable genetic divergence, making them promising candidates for future crop improvement initiatives.

References

1. Anderson TW. An introduction to multivariate analysis. New Delhi: Wiley Eastern Pvt Ltd; 1972.
2. Banerjee J, Shrivastava MK, Amrate PK, Singh Y, Upadhyay A, Soni M. Genetic variability and association of yield contributing traits in advanced breeding lines of soybean. *Electron J Plant Breed*. 2022;13(2):597-607.
3. García MFP, García PIP. Principal component analysis applied to filtered signals for maintenance management. *Qual Reliab Eng Int*. 2010;26(6):523-7.
4. Indiatat. Agricultural Statistics at a Glance 2024. Directorate of Economics and Statistics, Department of Agriculture Cooperation and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India.
5. Jaysawal P, Chandra U, Shukla G. Crop modelling with regression analysis: Promoting agriculture. *Sci Innov Mag*. 2023;1:63-7.
6. Kujane K, Sedibe MM, Mofokeng A. Genetic diversity analysis of soybean [*Glycine max* (L.) Merr.] genotypes making use of SSR markers. *Aust J Crop Sci*. 2019;13(7):1113-9.
7. Kumar S. Study of genetic diversity and stability in groundnut (*Arachis hypogaea* L.) genotypes [dissertation]. Sabour: Department of Plant Breeding and Genetics, BAU; 2019.
8. Lodhi BS, Kumar P, Chouhan M, Rajpoot A, Jha A. Comprehensive genetic analysis of yield and yield-related traits in soybean germplasms for enhanced crop

- improvement. *Int J Plant Soil Sci.* 2023;35(22):9-17.
9. Lokesh K, Basavegowda GC, Shekar M, Muniswamy BS, Ganiger N. Integrated approach for maximization of seed yield and quality in soybean (*Glycine max* L.). *Pharma Innov.* 2020;9(7):518-21.
 10. Mishra SP, Sarkar U, Taraphder S, Datta S, Swain D, Saikhom R, *et al.* Multivariate statistical data analysis-principal component analysis (PCA). *Int J Livest Res.* 2017;7(5):60-8.
 11. Mohod NB, Verma AK, Nemade S, Paslawar AN, Magar AS, Langangmeilu G. Effect of different sowing dates and spacing on soybean varieties for seed quality during rabi season. *Int J Environ Clim Change.* 2023;13(11):2380-4.
 12. Montgomery DC, Myers RH, Vining GG, Robinson TJ. Generalized linear models: with applications in engineering and the sciences. New Jersey: John Wiley & Sons; 2012.
 13. Morrison DE. Multivariate statistical methods. 2nd ed. 4th print. New York: McGraw Hill Kogakusta Ltd; 1987.
 14. Mounika U, Balaji Naik B, Ramanjaneyulu AV, Pavan Chandra Reddy K, Yakadri M. Optimum sowing time and suitable cultivars for seed quality parameters of soybean (*Glycine max* L.) during off-season in Northern Telangana Zone. *Int J Environ Clim Change.* 2025;15(1):448-62.
 15. Navya K, Parimala K, Reddy MR, Padmasri A. Effect of planting dates and varieties on seed quality parameters of off-season soybean (*Glycine max* L.). *Biol Forum Int J.* 2022;14(3):143-6.
 16. Reddy BR, Pandey M, Singh J, Singh PM, Rai N. Principal component analysis and stability of genotypes in French bean (*Phaseolus vulgaris* L.). *Legume Res.* 2024;47(5):745-50.
 17. Sivabharathi R, Muthuswamy A, Anandhi K, Karthiba L. Genetic diversity studies of soybean [*Glycine max* (L.) Merrill] germplasm accessions using cluster and principal component analysis. *Legume Res Int J.* 2023;1(6).
 18. Spoorthi B, Naidu G, Deshpande S, Mummigatti U. Impact of seasons on genetic variability parameters in soybean (*Glycine max* L. Merrill). *J Farm Sci.* 2024;37(1):1-6.
 19. Vijayakumar E, Sudhagar R, Vanniarajan C, Ramalingam J, Allan V, Senthil N. Estimating the breeding potency of a soybean core set. *Int J Agric Biol.* 2022;27(3):184-92.