Journal of Genetics and Applied Biotechnology (JGAB)

Original Article

J Genet Appl Biotechnol, 2026: *e2026025*
DOI: https://doi.org/10.66432/29mqph86
Received: 27 Jan 2026; Revised: 09 Mar 2026; Accepted: 16 Mar 2026; Published online: 02 Apr 2026

Dissecting Key Yield Components in Soybean (*Glycine max* L.) Genotypes Through Multivariate Analysis

Ramin Munir[1], Tayyaba Sajid[1], Mehwish Kanwal[2], Muhammad Qasim[1], Zulqar Nain[1], Muhammad Rizwan Shafiq[1], Sidra Iqbal[1*]

[1]Department of Plant Breeding and Genetics, University of Agriculture Faisalabad, Constituent College Depalpur, Okara.
[2]State Key Laboratory of Green Pesticide, Centre for R&D of Fine Chemicals of Guizhou University, Guiyang, 550025, China
[*]Corresponding author: dr.sidraiqbal@uaf.edu.pk

Abstract

Soybean (*Glycine max* L. Merrill) is a highly valued leguminous oilseed crops and can contribute to meet the Pakistan's edible oil demand and reduce import bill, However, limited availability of high yielding and environmental adapted soybean genotypes is a major hinderance for its adoption by local farmers. The current study was executed to determine variability and adaptability patterns of seventy exotic soybean accessions under agro-climatic conditions of Okara, Pakistan. An augmented block design with two commercial checks (NARC-21 and Faisal Soybean) was used to study ten yield related characteristics. Result exhibited a substantial amount of diversity in the studied germplasm. Genotypic correlation and path coefficient analysis revealed that seed per pod, pods per plant, pod length, pod width and hundred-seed weight were significant contributors to seed yield. Heritability and genetic advance estimates depicted the predominance of additive gene action in pods per plant, plant height and the hundred-weight of the seed. The genotypes Toano, Glenwood, Delsoy 4500, Lawrence and DSR-262 were identified as promising parental lines for use in future crop improvement programs.

Keywords: Soybean; Genetic variability; Additive gene action; Crop improvement; Exotic accessions

How to cite this article: Munir R, Sajid T, Kanwal M, Qasim M. Nain Z, Shafiq MR, Iqbal S. Dissecting Key Yield Components in Soybean (*Glycine max* L.) Genotypes Through Multivariate Analysis. *J. Genet. Appl. Biotechnol.* 2026: e2026025. https://doi.org/10.66432/29mqph86

Introduction

Soybean (*Glycine max* L. Merrill) is an important oil and leguminous crop with great nutritional and economic importance. It is a big source of high-grade protein and consumable oil, hence a vital crop in the food as well as the feed industry. The demand of soybean meal and edible oil in Pakistan has risen in the last few years. Although the agro-climatic conditions are complimentary to soybean farming, the crop growth has not been that impressive and this is mainly because of the non-availability of genotypes that yield well and adapt to the climatic conditions of the region (1). The pressing demand of the edible oil in Pakistan is escalated by the rapid population increase and changing eating habits. Nevertheless, national demands cannot be met by local oilseed production, and it is heavily dependent on imports, which is a huge burden to the economy (2). Pakistan imported 2.717 million tonnes of edible oil with value of Rs.794 billion (US2.809 billion) and local production was only 0.471 million tonnes during 2023-24 (3).

The production of soybean is characterized by right maturity group adaptable to the agro-ecological conditions. Thus, there is a need to conduct systematic testing of genetically diverse germplasm to determine the best genotypes that fit local places. Formulation of high yielding and adaptable cultivars must involve a sound assessment of yield contributing characters.

Genetic variability can be evaluated through agronomic and morphological attributes, which is a viable method of calculating the phenotypic diversity within a given environment. Knowledge of the relationship between yield and the components of traits is important in breeding programs to ensure efficient selection. The selection of a suitable parent with additive and non-additive effect for a specific trait is the backbone for an effective breeding project through correlation analysis (4).

Correlation analysis offers data on the strength and direction of relationship between characteristics (5), and path coefficient analysis separates correlation coefficients into direct and indirect influences, which allows one to accurately detect the characteristics that greatly determine yield (6).

Yield itself is a complex character and its associated traits are highly affected by climate change (7). The available exotic germplasm can be a valuable addition to enhance quality and diverse cultivars of soybean. Estimation of variation and diversity among various accessions is possible through various uni-and multivariate analysis (8). A decent example is character association studies which can be employed under establishing successful selection approaches in any plant breeding program (6). In this respect, the current research was conducted to estimate genetic variability and the association of traits in seventy exotic soybean genotypes to screen highly yielding, disease-resistant, and well-adapted soybean genotypes under agro-climatic environment of Okara.

Materials and Methods

Experimental description

The experiment was conducted during the spring season of year 2023 at the research area of the University of Agriculture, Faisalabad, Constituent College, Depalpur, Okara, Pakistan. The experimental material consisted of seventy exotic soybean genotypes along with two check varieties, Faisal Soybean and NARC-21. The germplasm was procured from the Soybean Laboratory, CAS-AFS at University of Agriculture, Faisalabad.

Experimental design and Field management

The experiment was conducted under an augmented block design (one-directional) comprising two blocks (9). Each genotype was manually sown in three rows of 3 meters in length. Plant-to-plant spacing was maintained as 6 cm, while row spacing was set at 45 cm with blocks one meter apart.

The crop was grown under normal field conditions. A recommended fertilizer dose of 25:50:40 kg ha^{-1} of NPK was applied. All standard agronomic and cultural practices were uniformly implemented throughout the growing season to ensure optimal crop performance.

Data Collection

Data were recorded from five plants per genotype for ten quantitative traits including Germination percentage(GP), Days to flowering (DF), Days to maturity (DM), Plant height (PH), Primary branches per plant (PB), Pods per plant (PP), Pod length (PL), Pod thickness (PW), Seeds per pod(SP) and 100-seed weight (HSW).

Biometrical analysis

Data were statistically analyzed using ANOVA. Correlations among the observed traits were calculated (10). The path coefficient analysis was also performed (5).

Results

Significant Variability was recorded among the studied soybean genotypes for most of traits. The ANOVA results for germination percentage under the augmented block design are presented in Table 1. Highly significant differences among treatments ($P < 0.05$) indicated substantial variability among exotic germplasm and check varieties. The germination percentage ranged from (19.33 to 100%), with genotype CX291 exhibiting the highest GP, whereas Williams 82 showed the lowest value.

Table 2 shows the correlation coefficients between the examined characteristics. SP was significantly positively related to the PH, PL, PW, DM, PP, PP and HSW. The traits including PH, SP, PW, DM, PP and HSW also had significant and positive correlations with PL. Days to maturity were significantly positively correlated with PH, PL, PW, SP and HSW, indicating that the length of the maturity period can be favorable to the expression of yield-related characteristics. The correlation analysis highlighted the need to focus on these traits during selection process.

The path coefficient analysis suggested that the PW, PH and SP had positive direct effects on seed yield, which showed their significant roles in the enhancement of seed yield. The positive indirect effects of the seeds per pod were also exhibited using the length and width of the pod, and this implies that the enhancement of these characteristics might indirectly increase the yield (Figure 1).

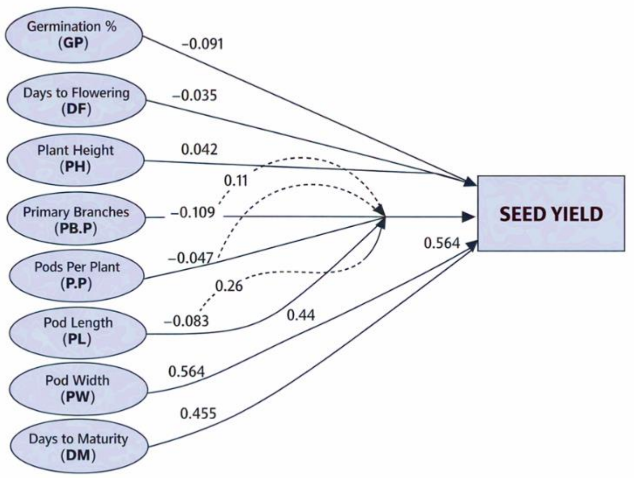

Figure 1 Path coefficient diagram illustrating the direct effects of yield-related traits on seed yield in soybean at the genotypic level.

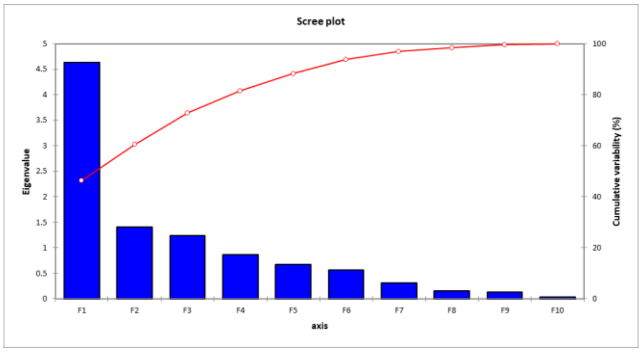

Figure 2 Scree plot of PCA between eigen values and number of component number

Table 3 represents the estimates of the phenotypic coefficient of variation (PCV), the genotypic coefficient of variation (GCV), heritability (h2), genetic advance (GA), and percentage of mean. The PCV and GCV revealed high degree of variation in the number of PP followed by PB, GP, PH, HSW, PL and PW. The variability of SP and DM was moderate, and DF recorded a low level of variability. SP and DM had the highest heritability. Most traits had a high GA suggesting that additive gene action was dominant and direct selection of these characters would be effective in the soybean improvement programs.

The Principal Component Analysis (PCA) outcomes indicated that the initial five major components accounted to 75.29 of the total variation within the genotypes. The largest percentage of proportion of variation (29.54) was explained by the first principal component (PC1) which was mostly correlated to the number SP and DM (Table 4 & 5).

The second principal component (PC2) explained 13.72 percent of the variation that was primarily determined by DF and PB. The third major principal component (PC3) had 12.45 per cent of the total variation and it was related to the GP, PH and PP. The 4th and 5th major elements accounted 10.11 and 9.47 percent of variability respectively and were principally associated with DF and the PH. According to PCA scores, genotypes Toano, Glenwood, Delsoy 4500, Lawrence and DSR-262 were found to be potential parents for use in future breeding programs (Figure 2).

Discussion

Genetic variability is prerequisite for any effective crop improvement program. In this study, we found significant differences in DM, SP, PP, PL, PW, and HSW, which are key determinants of yield potential. It has been reported that the existence of broad genetic diversity among soybean accessions allows for exploitation of genetic variability and enhancing productivity in the soybean breeding programs (10, 11).

Correlation analysis indicated that pod width, seeds per pod and days to maturity were crucial in determining the productivity of studied genotypes. A positive correlation between SP and yield observed in the current investigation is not new since other authors have reported SP as one of the strongest yield related factors in soybean (11, 12,13). In the same manner, the positive and significant correlations of PL and PW with seed yield, are the confirmation of the results of Guleria (12), who emphasized the significance of pod-related characteristics aiming at seed productivity.

Table 1 ANOVA mean squares of yield related traits.

SOV	DOF	GP	DF	DM	PH	PB/P	PP	SP	PL	PW	HSW
Blocks	2	17.71	8.31	1465.87	8.31	6.53	4242.90	0.57	10.38	0.30	11.34
Treatments	69	34.80*	4.96*	797.15**	4.96*	4.16*	2552.30**	0.26**	4.74*	0.26*	3.81*
Checks	1	3.60	0.06	0.00**	0.06	0.17	42.70	0.00**	1.96	0.02	1.37
Test entries + Test entries vs Checks	68	35.60*	5.03*	808.88**	5.03*	4.22*	2589.20**	0.26***	4.78*	0.26*	3.85*
Error	2	5.00	0.06	0.00	0.06	0.17	19.50	0.00	0.24	0.01	0.18

** Highly significant at P< 0.01; * significant at P < 0.05; NS non-significant

GP: germination%, DF: days to flowering, DM: days to maturity PH: Plant height, PB: Primary branches, PP: pods per plant, SP: Seeds per pod, PL: Pod length, PW: Pod width, HSW: hundred seed weight.

Table 2 Genotypic correlation Analysis of 10 characters in 70 soybean genotypes

Trait	GP	DF	PH	PB/P	PP	PL	PW	SP	DM	SW
GP	1.00									
DF	-0.16 ^{NS}	1.00								
PH	0.06 ^{NS}	0.11 ^{NS}	1.00							
PB/P	-0.20 ^{NS}	0.20 ^{NS}	0.21 ^{NS}	1.00						
PP	0.13 ^{NS}	-0.02 ^{NS}	0.29*	0.02 ^{NS}	1.00					
PL	-0.08 ^{NS}	0.06 ^{NS}	0.30*	0.00 ^{NS}	0.43**	1.00				
PW	-0.10 ^{NS}	-0.04 ^{NS}	0.20 ^{NS}	-0.09 ^{NS}	0.46**	0.78**	1.00			
SP	-0.25*	-0.02 ^{NS}	0.32**	-0.07 ^{NS}	0.36**	0.84**	0.82**	1.00		
DM	-0.20 ^{NS}	-0.13 ^{NS}	0.22 ^{NS}	-0.14 ^{NS}	0.40**	0.86**	0.85**	0.95**	1.00	
SW	-0.22 ^{NS}	-0.07 ^{NS}	0.22 ^{NS}	-0.17 ^{NS}	0.32**	0.69**	0.84**	0.82**	0.82**	1.00

* Significant at P < 0.05, ** Significant at P < 0.01, ^{NS} = non-significant

GP: germination%, DF: days to flowering, DM: days to maturity PH: Plant height, PB: Primary branches, PP: pods per plant, SP: Seeds per pod, PL: Pod length, PW: Pod width, SW: Seed weight.

Table 3. Estimation of PCV, GCV, ECV, Heritability, Genetic Gain and GAM for growth, yield and its attributing traits of seventy soybean genotypes

Traits	**σ2e**	**σ2g**	**σ2p**	**ECV**	**GCV**	**PCV**	**Hbs**	**GA**	**GAM**
Germination%	5.90	347.81	353.71	3.74	28.74	28.98	0.98	38.09	58.71
Days to flowering	0.29	4.98	5.27	1.17	4.86	5.00	0.94	4.46	9.74
Plant height	0.53	317.81	318.34	0.97	23.83	23.85	0.99	36.69	49.05
Primary branches plant^{-1}	0.28	4.00	4.29	8.62	32.30	33.43	0.93	3.98	64.30
Pods per plant	0.43	2605.88	2606.32	0.55	42.60	42.60	0.99	105.14	87.75
Pod Length	0.17	4.76	4.93	3.81	20.15	20.51	0.96	4.41	40.80
Pod Width	0.02	0.26	0.28	6.35	20.10	21.08	0.90	1.00	39.49
Seeds per pod	0	0.27	0.27	0.00	18.22	18.22	1.00	1.08	37.54
Days to Maturity	0.00	838.44	838.44	0.00	17.27	17.27	1.00	59.64	35.59
100 Seed Weight	0.12	3.94	4.07	3.77	21.09	21.42	0.96	4.02	42.77

σ2g=Genotypic variance, σ2p=Phenotypic variance, σ2e =Environmental variance, Hbs= Heritability broad sense GCV = Genotypic coefficient of variation, PCV =Phenotypic coefficient of variation, GA =Genetic advance, GAM =Genetic advance as percent of mean

Table 4 Eigen values, % variance and cumulative variance obtained from PCA among seventy soybean genotypes for seed yield

Traits	**Principal Component**	**Eigen values %**	**Variation**	**Cumulative %**
Germination percentage	PC1	3.25	29.54	29.54
Days to flowering	PC2	1.51	13.72	43.26
Plant height	PC3	1.37	12.45	55.71
Number of pods per plant	PC4	1.11	10.11	65.82
Primary branches of plant	PC5	1.04	9.47	75.29
Pod width	PC6	0.82	7.43	82.71
Pod length	PC7	0.80	7.24	89.95
Leaf area	PC8	0.54	4.92	94.87
Seeds per pod	PC9	0.33	3.00	97.87
Days to maturity	PC10	0.19	1.71	99.58
100 seed weight	PC11	0.05	0.42	100.00

Tables 5a: PC values of rotation component matrix for ten variables of seventy genotypes of soybean

Traits	**PC1**	**PC2**	**PC3**	**PC4**
Germination percentage	-0.08	-0.44	0.62	0.15
Days to flowering	-0.01	0.55	0.00	0.80
Plant height	0.16	0.32	0.54	-0.12
Primary branches/ plant	-0.03	0.62	0.16	-0.54
No. of pods per plant	0.24	-0.04	0.50	0.09
Pod length	0.41	0.06	0.05	0.06
Pod width	0.42	-0.07	-0.03	0.05
Seeds per pod	0.51	0.03	-0.10	-0.04
Days to maturity	0.51	-0.08	-0.12	-0.06

Table 5b: Interpretation of rotated matrix for the traits having values >0.5 in each PCs

PC	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**
Traits	Seeds per pod	Days to flowering	Germination percentage	Days to flowering	Plant height
	Days to maturity	Primary branches per plant	Plant height	-	-
	-	-	No. of pods per plant	-	-

The positive correlation between SP and DM was found to be exceptionally high. The correlation can be interpreted to imply that the genotypes with a comparatively longer growth period can store more photosynthates and commit more assimilates to reproductive structures. This correlation of maturity period and the constituents of the yield have been outlined by Lin and Nelson (14), who established a close relationship between developmental characteristics like flowering and maturity period with yield formation in soybean. Recent findings have also highlighted that the phenological characteristics are significant in the management of resource distribution and reproduction success in legume crops in different agro-ecological conditions.

Even though there was a positive relationship between PP and yield, analysis of path coefficients indicated that there was a negative direct relationship (−0.0465) between both traits. These inconsistencies between correlation and direct effects depict the inability to use correlation coefficients alone to explain the interrelationships between traits. The significance of path coefficient analysis in discerning between a direct and indirect effect was initially explained by Dewey and Lu (5), who presented that correlation could mask the actual contribution of a given set of traits to yield. The analysis using path coefficient also indicated that PW had the greatest positive direct influence on seed yield suggesting it to be a key yield-definitive character in soybean. Similar findings have been brought up by studies reporting that seeds per pod and pod-related characteristics comprise significant yield-influencing factors of soybean via direct and indirect impact. Recent research work has also emphasized the need to use component trait analysis to increase selection efficiency in soybean improvement programs (15, 16).

Plant height had a moderate positive direct impact on yield as well as an indirect impact of a positive nature through PW and SP. This indicates that PH is a helpful factor in the development of yield, which probably helps to improve canopy structure and photosynthetic capacity. Tall plants can capture lighter and generate more assimilates that are then transferred to growing pods and seeds. Lin and Nelson (14) also reported similar associations between plant structure and reproductive performance explaining structural characteristics of plants as important determinants of the reproductive development of soybean.

Conversely, GP and PB plant had negative direct effects on yield although positive indirect effects were observed. This depicts that an increase in these traits may not necessarily translate to increase in seed yield. These adverse direct impacts of some vegetative characteristics have been reported in past studies in which over-grown vegetation has diverted the assimilates to other parts of the body at the expense of the reproductive parts ultimately causing low production of seeds. Harer and Deshmukh (16) also raised similar observations and recommended that soybean productivity may be maximized only when vegetative growth is balanced with reproductive growth.

The correlation and path coefficient analyses were also strengthened by the multivariate analysis in the present study. Principal component analysis (PCA) summarized the variability observed among the assessed genotypes and helped in determining the key characteristics that may be related to overall genetic variation. Attributes like pod width, seeds per pod, pod length and days to maturity were strongly related to yield performance. PCA and biplot analysis are multivariate statistical methods that have been extensively applied in breeding legumes to define high-performing genotype and to illustrate complex traits associations.

Multivariate methods like PCA analysis could effectively determine stable and high yielding genotypes in legume crops. Additionally, assessment of the heritability and environmental impacts on the yield traits have reported that both genetic factors and environmental interaction are key determinants of crop performance (2, 17). These quantitative methods of analysis are thus combined to offer a strong platform for yield determinant traits and the use of the best genotypes.

Altogether, the joint interpretation of the correlation, path coefficient, and multivariate analysis indicated that the most appropriate selection criteria to enhance the yield of soybean under agro-climatic conditions of Okara were pod width, seeds per pod and days to maturity. Not only did these traits show high positive associations with yield, but they had significant direct effects in determining seed productivity. The results comply with the already reported studies that have highlighted the significance of the yield component analysis in coming up with effective breeding strategies of soybean (17, 18, 19, 20). Therefore, these characteristics can be effective for the selection in soybean breeding programs, focusing on the development of high-yielding well adapted cultivars.

Conclusion

Significant genetic differences were found among the soybean accessions offering a great range of selection in breeding initiatives. The analysis based on correlation and the path coefficient revealed the importance of pod width, seeds per pod, and days to maturity as relevant variables in predicting seed yield, which means that they are useful as a predictive variable in selecting seedlings. Multivariate analysis (PCA) also supported the significance of these traits in the characterization of genetic variation of genotypes. Altogether, the results suggest the importance of pod-related characteristics in improving yield and provide a powerful rationale behind using trait-based selection methods to produce high-yielding soybean cultivars that can be cultivated in Pakistan, which would lead to the decrease in imports bills and the increase in agricultural sustainability.

Acknowledgement: The authors sincerely acknowledge the Soybean Laboratory at CAS-AFS, University of Agriculture Faisalabad, for providing the experimental material and technical facilities necessary to conduct this research. Special thanks are extended to Dr. Zaheer Ahmed for his valuable guidance, technical support, and cooperation throughout the experimental work.

Disclaimer: None.

Conflicts of Interest: The authors declare no conflict of interest.

Source of Funding: None

Ethics approval statement: Not Applicable

Authors' Contribution Statements: S.I and Z.A. conceptualized and designed the study. R.M. and T.S conducted the experiment, collected data, and performed the statistical analysis. R.M and M.Q drafted the manuscript. M. K. and M.R.S reviewed and edited the manuscript. T.S., Z.N. assisted in field management, data recording, and data compilation. S.I. supervised the research work and provided technical guidance throughout the study.

References

1. Khurshid, H., D. Baig, S. Jan, M. Arshad and M. Khan. 2017. Miracle crop the present and future of soybean production in Pakistan. Journal of Biology and Medicine. 2:89-191.
2. Ullah, H., M. Rahman, M. Hussain, M. Zafar and I.A. Khalil. 2012. GGE biplot analysis for yield stability in mungbean (Vigna radiata L.). Pakistan Journal of Botany. 44: 1537–1544.
3. Government of Pakistan. 2024. Pakistan Economic Survey 2023–24. Ministry of Finance, Islamabad, Pakistan, pp. 21–43.
4. Rehman, A., A. Bashir, G. Sarwar, J. Yinhua, X. Du, and M. T. Azhar. 2020 Genetic Assessment of Oil Contributing Traits in Upland Cotton Gossypium hirsutum L.). J Cotton Res. 2020;3:1–9.
5. Dewey, D.R. and K.H. Lu. 1959. A correlation and path coefficient analysis of components of crested wheatgrass seed production. Agronomy Journal. 51:515-518.
6. Iqbal, Z., M. Arshad, M. Ashraf, T. Mahmood and A. Waheed. 2008. Evaluation of soybean (*Glycine max* (L.) Merrill) germplasm for some important morphological traits using multivariate analysis. Pakistan Journal of Botany. 40:2323-2328.
7. Perveen S, Fatima A, Zia A, Rehman AU. 2026. Correlation analysis of several economic traits of upland cotton (Gossypium hirsutum L.). Journal of Genetics and Applied Biotechnology. 2026: e2026020.
8. Sarwar, M. K. S., M. Usman, S. Marium, M. Zeeshan and M. Y. Ashraf. 2026. Biochemical Strategies of Cotton Defense: Osmolyte Accumulation and Stability Enhance Resistance to Cotton Leaf Curl Virus. Journal of Genetics and Applied Biotechnology, 1(1), e2026019.
9. Federer, W.T., 1956. Augmented (or hoonuiaku) designs. Hawaiian Planters' Record, 55, pp.191–208.
10. Steel, R.G.D., Torrie, J.H. and Dickey, D.A., 1997. Principles and procedures of statistics: A biometrical approach. 3rd ed. New York: McGraw-Hill, pp.172–177.
11. Jain, R. K. A. Joshi, H. R. Chaudhary, A. Dashora and C. L. Khatik. 2018. Study on genetic variability, heritability and genetic advance in soybean (*Glycine max* (L.) Merrill). Legume Research – An International Journal. 41(4):532-536.
12. Guleria, S., N. Sharma and R. Kumar. 2019. Genetic variability, correlation and path coefficient analysis for yield and yield contributing traits in soybean. Legume Research – An International Journal. 42: 506–511.
13. Bisinotto, F.F., O.T. Hamawaki, A.P.O. Nogueira, R.L. Hamawaki, J.S. Glansenapp and C.L. Hamawaki. 2017. Path analysis and traits correlation in soybean. Bioscience Journal, 33(1), 27–33.
14. Lin, M.S. and R.L. Nelson. 1988. Relationship between plant height and flowering date in determinate soybean. Crop Sciences. 28:27-30.
15. Das, P.K., P.K. Jena and B. Mohanty. 1989. Correlation and path coefficient analysis in soybean. Indian Journal of Agricultural Sciences. 59: 121–123
16. Harer, P.N. and M.P. Deshmukh. 1992. Correlation and path coefficient analysis in soybean. Journal of Maharashtra Agricultural Universities. 17: 74–76.
17. Harer PN, Deshmukh RB. Genetic variability, correlation and path coefficient analysis in soybean (*Glycine max* (L.) Merrill). Soybean Res. 1992; 9:65–71
18. Ullah, H., M. Rahman, M. Hussain and I.A. Khalil. 2011. Estimation of heritability and genetic advance for yield related traits in mungbean (Vigna radiata L.). Pak. J. Bot. 43: 1221–1225.
19. Jain, P., S. Kumar and R. Singh. 2018. Genetic variability, correlation and path coefficient analysis for yield and yield attributes in soybean (*Glycine max* L.). Journal of Pharmacognosy and Phytochemistry. 7: 2023–2026.
20. Mili, K.N., Shirazy, B.J. and Mahbub, M.M., 2017. Evaluation of genetic diversity in soybean (*Glycine max* (L.) Merrill) genotypes based on agronomic traits. International Review of Scientific Synthesis, 20, pp.92–98.