



ARTICLE

Estimation of Genetic Divergence and Character Association Studies in Local and Exotic Diversity Panels of Soybean (*Glycine max* L.) Genotypes

Syed Ali Zafar^{1,*}, Muhammad Aslam², Haroon Zaman Khan³, Sehrish Sarwar¹, Rao Saad Rehman⁴, Mariam Hassan¹, Ramala Masood Ahmad², Razaqat A. Gill⁵, Basharat Ali⁶, Ibrahim Al-Ashkar⁷, Abdullah Ibrahim⁷, Md Atikur Rahman⁸ and Ayman El Sabagh^{9,*}

¹Oilseeds Research Institute, Ayub Agricultural Research Institute, Faisalabad, Pakistan

²Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan

³Department of Agronomy, University of Agriculture, Faisalabad, Pakistan

⁴College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China

⁵Key Laboratory of Biology and Genetic Improvement of Oil Crops, The Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Wuhan, 430062, China

⁶Department of Agricultural Engineering, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan

⁷Department of Plant Production, College of Food and Agriculture Sciences, King Saud University, Riyadh, Saudi Arabia

⁸Grassland and Forage Division, National Institute of Animal Science, Rural Development Administration, Cheonan, Korea

⁹Department of Agronomy, Faculty of Agriculture, University of Kafrelsheikh, Kafr Elsheikh, Egypt

*Corresponding Authors: Syed Ali Zafar. Email: syedalizafar9@gmail.com; Ayman El Sabagh. Email: aymanelsabagh@gmail.com

Received: 08 November 2022 Accepted: 29 January 2023

ABSTRACT

The availability of favorable genetic diversity is a thriving vitality for the success of a breeding program. It provides a firm basis of selecting superior breeding lines for the development of high yielding crop genotypes. In this context, present investigation aimed to generate information on genetic divergence and character association in a diversity panel of 123 local and exotic soybean genotypes. Analysis of variance revealed significant response of the evaluated genotypes based on studied attributes. It depicted the probability of selecting desirable soybean genotypes by focusing on character association studies and genetic diversity analysis. Correlation analysis revealed that seed yield per plant showed high positive correlation with 100-seed weight followed by pods per plant and plant height. Furthermore, path coefficient analysis exposed that pods per plant had maximum direct contribution in seed yield per plant followed by 100-seed weight, days to flowering and SPAD measurement. Genotype named "G-10" showed maximum yield per plant followed by 24607, G-52, 24593, Arisoy, 24566, 17426, A-3127, 24570 and 24567. Genetic diversity analysis grouped the evaluated germplasm into 17 clusters. All clusters showed zero intra-cluster variability; while inter-cluster divergence ranged from 9.00 to 91.11. Cluster V showed maximum inter-cluster distance with cluster XII followed by that of between V and VIII. Moreover, cluster IV with superior genotypes (G-10, 24607, 24593 and 24566), VI (17426 and 24567), XIII (24570) and X (Arisoy and G-52) showed above mean values for most of the studied characters. Overall, the results of hybridization between the superior genotypes of these cluster pairs might be useful for soybean breeding with improving agronomic traits and adaptability.



KEYWORDS

Genetic variability; correlation; path analysis; genetic diversity analysis; soybean selective breeding

1 Introduction

Soybean (*Glycine max* L.) belongs to fabaceae family, exhibits diploid chromosome number $2n = 40$. The domestication and cultivation of soybean dates back to 5000 years ago and is thought to be originated in China [1]. The cultivated soybean is the progenitor of wild soybean, i.e., *Glycine soja*, both classified in the same genus *Glycine*. It is the leading legume crop of the world with a share of almost 50% in total legume cultivated area of the world [2]. Countries which contribute majorly in the global soybean production are United States, Brazil, Argentine, China and India [3,4]. Its seed contains high quality protein (40%) and oil contents (18%–22%) along with a significant quantity of carbohydrates (20%–30%), minerals and fiber contents (3%–6%) and 5% ash [5,6]. During 2018–19, soybean has been estimated to be grown on 124.89 million hectares globally with a total production of 348.61 million tons [7]. In Pakistan, soybean crop was introduced along with sunflower and safflower during 1960s. Unfortunately, its cultivation has been facing various setbacks predominantly due to lack of climate resilient, disease resistant and high yielding cultivars. There is a huge gap between total soybean production and the country's demand because its cultivation is restricted to the northern areas of the country [8]. Therefore, a comprehensive research is needed to pursue for the selection of superior genotypes from a genetically diverse soybean pool. It would facilitate to develop high yielding, insect-pest resistant and climate resilient soybean varieties.

Genetic diversity analysis helps to assess the existing genetic variation present within a breeding line and among the breeding lines. Furthermore, identification of various groups based on genetic similarities and dissimilarities help in devising suitable hybridization scheme. Moreover, the selection of superior parent lines can be made to utilize the available genetic variation in a more productive way [9]. Availability of genetically diverse pool of germplasm greatly enhances the selection efficiency at the initial stage of a breeding program by selecting favorable genetic combinations. Identification of various genetic patterns leads to the systematic characterization of available genetic diversity. It also facilitates plant breeders to assess the phylogenetic associations among the tested genotypes [10]. Genetic diversity analysis can be done by following various approaches, such as pedigree information, morphological and agronomic characterization, biochemical and molecular approaches for germplasm classification [11–17]. Genetic diversity analysis based on morphological and agronomic performance is an effective tool to identify the phenotypic variation in particular agro-ecological conditions [18]. Estimation of genetic divergence can be accomplished by using D^2 statistics approach [19]. It has been considered as the most efficient approach that classifies the evaluated germplasm into different clusters based on multivariate analysis.

Character association studies based on correlation and path coefficient analysis provide the basis to identify the strength of association among various yield and its related attributes [20–22]. It provides a basic insight into the elucidation of relative contribution of a particular growth attribute in the economic yield. Since plant yield is a very complex quantitative character which totally depends upon the strength of association among different growth attributes. Association studies based on only correlation analysis may mislead to select the best possible combinations due to its restricted ability of just estimating the association among different variables. While, path coefficient analysis measures the relative significance of yield attributing characters by splitting the genotypic correlation into direct and indirect effects. Considering these above issues, this study was aimed to select elite soybean genotypes accomplished

through character association studies including correlation and path coefficient analysis, and assess the magnitude of genetic diversity followed by phenotypic clustering of soybean growth attributes.

2 Materials and Methods

2.1 Collection of Germplasm

A diverse set of 123 soybean genotypes was procured (Table 1) from the national and international research institutions, e.g., Plant Genetic Resources Program (PGRP), National Agriculture Research Centre (NARC), Islamabad, Pakistan; The University of Agriculture, Peshawar, Pakistan; and Nigde Omer Halisdemir University, Nigde, Turkey.

Table 1: List of soybean germplasm evaluated in the study

Sr. no.	Accession name								
1	17418	26	24511	51	24590	76	G-10	101	G-39
2	17420	27	24512	52	24591	77	G-11	102	G-40
3	17421	28	24513	53	24592	78	G-12	103	G-41
4	17122	29	24514	54	24593	79	G-13	104	G-42
5	17423	30	24526	55	24594	80	G-14	105	G-43
6	17425	31	24527	56	24595	81	G-15	106	G-44
7	17426	32	24528	57	24597	82	G-17	107	G-45
8	17427	33	24529	58	24603	83	G-18	108	G-46
9	17431	34	24543	59	24604	84	G-19	109	G-47
10	23981	35	24544	60	24605	85	G-20	110	G-48
11	23996	36	24563	61	24606	86	G-21	111	G-49
12	24010	37	24564	62	24607	87	G-22	112	G-50
13	24488	38	24565	63	24608	88	G-24	113	G-51
14	24490	39	24566	64	33521	89	G-25	114	G-52
15	24491	40	24567	65	33522	90	G-27	115	G-53
16	24492	41	24569	66	33523	91	G-28	116	G-54
17	24494	42	24570	67	G-1	92	G-29	117	A-3127
18	24495	43	24571	68	G-2	93	G-30	118	Bravo
19	24496	44	24572	69	G-3	94	G-31	119	Ilksoy
20	24497	45	24573	70	G-4	95	G-32	120	atakisi
21	24499	46	24574	71	G-5	96	G-34	121	SA-88
22	24505	47	24576	72	G-6	97	G-35	122	Nova
23	24506	48	24577	73	G-7	98	G-36	123	Arisoy
24	24507	49	24588	74	G-8	99	G-37		
25	24508	50	24589	75	G-9	100	G-38		

2.2 Experimental Description

The study was carried out during spring season 2020–21 at the research area of Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan. The weather conditions during spring 2020–21 have been illustrated in Table 2. A pot experiment was conducted by following randomized complete block design (RCBD) with three replications per genotype. Six pots per genotype (two pots per replication and four plants per replication) were maintained. After seedling emergence only two plants per pot were allowed to grow while others were thinned out. All recommended agronomic practices including irrigation, fertilization and plant protection measures were ensured timely to raise healthy plants. The pots were irrigated on daily basis but over irrigation of pots were avoided. Total 50 g of NPK fertilizer (20:20:20) was applied to each pot and were irrigated immediately. Seeds were treated with thiophanate methyl @ 2g/kg seed to prevent possible infestation of fungal diseases. To control whitefly infestation, diafenthiuron was applied @ 2.5 ml/L. For effective management against armyworm, lufenuron was applied @ 2.5 ml/L.

Table 2: Mean precipitation, maximum minimum and average temperatures and relative humidity % during spring 2020–21

Months	Precipitation (mm)	Temperature (°C)			RH (%)
		Maximum	Minimum	Average	
January	0.58	15.00	10.00	12.50	80.70
February	2.29	20.30	9.10	14.70	79
March	1.80	26.00	13.80	19.90	68.50
April	1.04	35.00	20.60	27.80	42.50
May	1.26	39.00	23.90	31.45	46.5
June	1.18	42.40	27.40	34.90	47.80
July	3.32	38.00	28.00	33.00	62.70
August	2.61	38.00	28.50	33.25	72.50

Source: Agricultural Meteorology Cell (www.uaf.edu.pk/faculties/agri/depts/crop_physiology/agri_met_cell/met_bulletin.html).

2.3 Data Collection

Data were collected based on different agronomic attributes, i.e., number of leaves plant⁻¹, number of branches plant⁻¹, days to flowering initiation, days to pod initiation, number of pods plant⁻¹, days taken to maturity, number of seeds pod⁻¹. Plant height (cm) was measured at R₈ [23] growth stage through meter rod; while pod and petiole length (cm) were recorded through measuring scale and stem diameter (mm) through digital vernier caliper. SPAD measurement (cci) was taken at V₃ [24] growth stage through SPAD-502 meter. 100 seed weight (g) and seed yield plant⁻¹ (g) were recorded through electric weigh balance. Each pot with one plant was considered as one replication. Data from all replications were recorded and subjected to statistical analyses.

2.4 Statistical Analysis

The recorded data were subjected to Analysis of Variance (ANOVA) according to Steel et al. [25] followed by Least Significant Difference (LSD) test through Statistix 8.1 software. Genetic diversity analysis based on phenotypic clustering was carried out by applying D² statistics approach developed by Mahalanobis [19] through Minitab 17.1.0 software. Character association studies based on correlation and path coefficient analysis were carried out through R 3.1.0 software.

2.4.1 Estimation of Correlation Coefficients

Estimation of correlation coefficients was carried out by following [26]:

$$\text{Phenotypic correlation coefficient } (r_p) = p\text{Cov}_{xy} / \sqrt{\sigma^2_{px} \times \sigma^2_{py}}$$

$$\text{Genotypic correlation coefficient } (r_g) = g\text{Cov}_{xy} / \sqrt{\sigma^2_{gx} \times \sigma^2_{gy}}$$

where, r_p and r_g represent coefficients of phenotypic and genotypic correlation, respectively; $p\text{Cov}_{xy}$ and $g\text{Cov}_{xy}$ represent phenotypic and genotypic co-variances between x and y , respectively; σ^2_{px} and σ^2_{gx} represent phenotypic and genotypic variances for variable x , respectively and σ^2_{py} and σ^2_{gy} represent phenotypic and genotypic variances for variable y , respectively.

2.4.2 Path Coefficient Analysis

Direct and indirect effects were calculated by utilizing coefficients of genotypic correlation according to Dewey et al. [27]:

$$R_{ij} = P_{ij} + \sum r_{rk} P_{kj}$$

where, r_{ij} denotes the relationship that each independent variable “ i ” (yield related character) shows with the dependent variable “ j ” (seed yield); P_{ij} depicts the direct (diagonal) effects of the independent variable “ i ” on the dependent variable “ j ” as estimated through path analysis and denotes the additive indirect effects of a specific independent variable “ j ” from all of the other independent variable “ k ”. The effects of remaining characters are estimated as the residual effects as given by:

$$P_R = \sqrt{\left(1 - \sum P_{ij} R_{ij}\right)}$$

2.4.3 D^2 Analysis

Estimation of D^2 was carried out by following equation developed by [19]:

$$D^2 = (x - m)^T C^{-1} (x - m)$$

where,

D^2 = Mahalanobis distance

x = vector of the data

m = vector of mean values of independent variables

C^{-1} = Inverse covariance matrix of independent variables

T = Indicates that vector should be transposed

3 Results

3.1 Variability and Genetic Parameters

Present study investigated 14 different yield and its attributing characters to elucidate their relative association and the extent of genetic divergence among the evaluated soybean genotypes. The highly significant mean sum of square values for genotypes (Table 3) estimated through ANOVA revealed the availability of ample genetic variation among the evaluated genotypes (Table 3). It further encourages the selection of desirable soybean genotypes.

Table 3: Estimates of genetic variability and descriptive statistics of 123 different soybean genotypes evaluated in the study

Variable	Mean \pm S.E	MSS ($p < 0.01$)	Range	SD	Variance	GCV	PCV	CV %	Heritability
X1	58.73 \pm 0.65	465.40**	31.20–92.10	12.44	154.64	15.17	15.31	4.59	0.96
X2	43.88 \pm 0.43	195.37*	21.00–66.70	8.28	68.59	9.78	10.10	5.40	0.92
X3	10.47 \pm 0.18	36.04**	3.90–19.90	3.47	12.05	4.23	4.25	3.59	0.99
X4	17.55 \pm 0.38	1451.32*	6.00–50.33	7.32	53.53	26.78	27.04	3.70	0.97
X5	6.29 \pm 0.14	20.885**	2.00–16.00	2.68	7.20	3.22	3.26	8.65	0.96
X6	12.08 \pm 0.21	49.54**	3.97–25.76	4.08	16.60	4.97	4.98	4.29	0.98
X7	46.17 \pm 0.47	246.69**	27–65	9.08	82.44	11.08	11.13	5.15	0.99
X8	53.11 \pm 0.49	267.05*	34–76	9.45	89.27	11.53	11.58	9.96	0.99
X9	102.03 \pm 0.54	317.36**	80–130	10.29	105.91	12.56	12.63	5.71	0.95
X10	3.03 \pm 0.03	1.028**	1.55–5.05	0.60	0.36	0.71	0.73	4.73	0.94
X11	38.11 \pm 0.57	361.52**	11.00–63.00	10.98	120.50	13.39	13.48	7.53	0.99
X12	2.63 \pm 0.03	0.694 ^{NS}	1.00–4.00	0.65	0.42	0.46	0.79	1.67	0.33
X13	7.17 \pm 0.07	4.849*	3.84–11.30	1.30	1.69	1.54	1.59	4.87	0.93
X14	8.32 \pm 0.09	9.358**	3.86–13.33	1.78	3.15	2.16	2.17	3.37	0.98

Notes: X1 = Plant Height (cm); X2 = SPAD Measurement (cci); X3 = Petiole Length (cm); X4 = No. of Leaves per Plant; X5 = No. of Branches per Plant; X6 = Stem Diameter (mm); X7 = Days to Flowering Initiation; X8 = Days to Pod Initiation; X9 = Days Taken to Maturity; X10 = Pod Length (cm); X11 = No. of Pods per Plant; X12 = No. of Seeds per Pod; X13 = 100 Seed Weight (g); X14 = Seed Yield per Plant (g).

3.2 Descriptive Statistics for Morphological Attributes

Statistical analysis based on descriptive statistics (Table 3) elucidated that plant height ranged between 31.20 (G-12) to 92.10 (24566) cm with a general mean of 58.73 \pm 0.65 having coefficient of variance 5%. PCV (15.31) and GCV (15.17) were moderate having high heritability (0.96%). SPAD measurement ranged between 21.00 (24527) to 66.70 (G-41) CCI with a general mean of 43.88 \pm 0.43 having coefficient of variance 5.40%. PCV (10.10) and GCV (9.78) were found as low having heritability of 0.92%. Petiole length ranged between 3.90 (Ilksoy) to 19.90 (24588) cm with a general mean of 10.47 \pm 0.18 having coefficient of variance 3.59%. PCV (4.25) and GCV (4.23) were very low having high heritability (0.99%). Number of leaves plant⁻¹ ranged between 6.00 (23996) to 50.33 (24570) with a general mean of 17.55 \pm 0.38 having coefficient of variance 3.70%. PCV (27.04) and GCV (26.78) were high having heritability (0.97%). Number of branches plant⁻¹ ranged between 2.00 (G-4) to 16.00 (G-10) with a general mean of 6.29 \pm 0.14 having coefficient of variance 8.65%. PCV (3.26) and GCV (3.22) were very low having high heritability (0.96%). Stem diameter ranged between 3.97 (G-12) to 25.76 (24593) mm with a general mean of 12.08 \pm 0.21 having coefficient of variance 4.29%. PCV (4.98) and GCV (4.97) were very low having high heritability (0.98%). Days to flowering initiation ranged between 28 (24577) to 65 (24527) days with a general mean of 46.17 \pm 0.47 having coefficient of variance 5.15%. PCV (11.13) and GCV (11.08) were moderate having high heritability (0.99%).

Days to pod initiation ranged between 27 (24588) to 76 (24527) days with a general mean of 53.11 \pm 0.49 having coefficient of variance 9.96%. PCV (11.58) and GCV (11.53) were moderate having heritability (0.99%). Days taken to maturity ranged between 80 (G-32) to 130 (24567) days with a general mean of 102.03 \pm 0.54 having coefficient of variance 5.71%. PCV (12.63) and GCV (12.56) were moderate having high heritability (0.96%). Pod length ranged between 1.55 (24566) to 5.05 (24497) cm with a general mean of 3.03 \pm 0.03 having coefficient of variance 4.73%. PCV (0.73) and GCV (0.71) were

negligible having high heritability (0.94%). Number of pods plant⁻¹ ranged between 11.00 (G-12) to 63.00 (G-10) cm with a general mean of 38.11 ± 0.57 having coefficient of variance 7.53%. PCV (13.48) and GCV (13.39) were moderate having high heritability (0.99%). Number of seeds pod⁻¹ ranged between 1.00 (G-4) to 4.00 (17420) with a general mean of 2.63 ± 0.03 having coefficient of variance 1.67%. PCV (0.79) and GCV (0.46) were negligible having low heritability (0.33%). 100-seed weight ranged between 3.84 g (G-12) to 11.30 g (24593) with a general mean of 7.17 ± 0.07 having coefficient of variance 4.87%. PCV (1.59) and GCV (1.54) were negligible having heritability (0.93%). Seed yield plant⁻¹ (g) ranged from 3.86 (G-12) to 13.33 (G-10) with a general mean of 8.32 ± 0.09 having coefficient of variance 3.37%. PCV (2.17) and GCV (2.16) were low having high heritability (0.98%).

3.3 Correlation Analysis

Correlation coefficients for all yield and its contributing characters are shown in Table 4 and Fig. 1. Plant height showed significant and positive correlation for most of the characters under study, i.e., petiole length (0.336), number of leaves per plant (0.642), number of branches per plant (0.414), stem diameter (0.576), days to flowering initiation (0.902), days to pod initiation (0.837), days taken to maturity (0.938), number of pods per plant (0.951), number of seeds per pod (0.278), 100 seed weight (0.262) and seed yield per plant (0.485). Plant height showed significant but negative correlation with pod length (0.485). On the other hand, SPAD measurement showed non-significant correlation with plant height (0.017). Data regarding SPAD measurement showed non-significant correlation with rest of the studied characters. Pod length showed significant but negative correlation with most of the characters under study, i.e., number of leaves per plant (-0.161), number of branches per plant (-0.160), stem diameter (-0.265), days to flowering initiation (-0.277), days to pod initiation (-0.234), days taken to maturity (-0.332), number of pods per plant (-0.309), 100-seed weight (-0.127). Number of seeds per pod showed significant positive correlation with pod length (0.444). On the other hand, pod length showed non-significant correlation with SPAD measurement (0.062). Petiole length showed positive correlation with most of the characters under study, i.e., plant height (0.336), pod length (0.166), number of leaves per plant (0.298), days to flowering initiation (0.343), days to pod initiation (0.324), days taken to maturity (0.344), number of pods per plant 0.344, number of seeds per pod (0.384) and seed yield per plant (0.130). On the other hand, petiole length showed negative correlation with SPAD measurement (-0.061) and 100-seed weight (-0.056).

Table 4: Genotypic and phenotypic correlation coefficients among the recorded agronomic attributes in the study

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14
X1 G	1	0.017	-0.331*	0.333*	0.639*	0.395*	0.571*	0.896*	0.831*	0.932*	0.946*	0.160	0.254*	0.476*
P		0.018	-0.345*	0.336*	0.641*	0.414*	0.576*	0.902*	0.837*	0.938*	0.951*	0.278*	0.262*	0.485*
X2 G		1	0.048	-0.057	0.048	0.053	0.020	-0.006	0.055	-0.005	0.039	-0.056	0.118*	0.090
P			0.062	-0.061	0.051	0.043	0.021	-0.005	0.057	-0.004	0.041	-0.058	0.132*	0.094
X3 G			1	0.157*	-0.155*	-0.145*	-0.252*	-0.265*	-0.224*	-0.317*	-0.296*	0.241*	-0.112*	-0.041
P				0.166*	-0.161*	-0.160*	-0.265*	-0.277*	-0.234*	-0.332*	-0.309*	0.444*	-0.127*	-0.044
X4 G				1	0.297*	-0.110*	0.087	0.337*	0.319*	0.341*	0.341*	0.214*	-0.054	0.126*
P					0.298*	-0.118*	0.089	0.343*	0.324*	0.344*	0.344*	0.381*	-0.056	0.130*
X5 G					1	0.316*	0.379*	0.582*	0.496*	0.581*	0.613*	0.116*	0.276*	0.372*
P						0.333*	0.382*	0.586*	0.500*	0.586*	0.616*	0.203*	0.284*	0.378*
X6 G						1	0.364*	0.342*	0.277*	0.342*	0.393*	-0.001	0.347*	0.392*
P							0.384*	0.362*	0.290*	0.362*	0.416*	-0.022	0.372*	0.421*

(Continued)

Table 4 (continued)

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14
X7 G							1	0.459*	0.399*	0.503*	0.515*	-0.028	0.278*	0.351*
X7 P								0.466*	0.404*	0.510*	0.521*	-0.049	0.292*	0.359*
X8 G								1	0.932*	0.988*	0.864*	0.158*	0.194*	0.418*
X8 P									0.946*	0.960*	0.873*	0.266*	0.199*	0.427*
X9 G									1	0.908*	0.811*	0.142*	0.165*	0.382*
X9 P										0.918*	0.819*	0.245*	0.172*	0.390*
X10 G										1	0.889*	0.141*	0.199*	0.409*
X10 P											0.896*	0.250*	0.205*	0.416*
X11 G											1	0.164*	0.281*	0.539*
X11 P												0.272*	0.289*	0.548*
X12 G												1	-0.009	0.178*
X12 P													-0.023	0.280*
X13 G													1	0.859*
X13 P														0.876*
X14 G														1
X14 P														

Notes: G = Genotypic Correlation; P = Phenotypic Correlation; X1 = Plant Height (cm); X2 = SPAD Measurement (cci); X3 = Pod Length (cm); X4 = Petiole Length (cm); X5 = No. of Leaves per Plant; X6 = No. of Branches per plant; X7 = Stem Diameter (mm); X8 = Days to Flowering Initiation; X9 = Days to Pod Initiation; X10 = Days taken to Maturity; X11 = No. of Pods per Plant; X12 = No. of Seeds per Pod; X13 = 100 Seed Weight (g); X14 = Seed Yield per Plant (g).

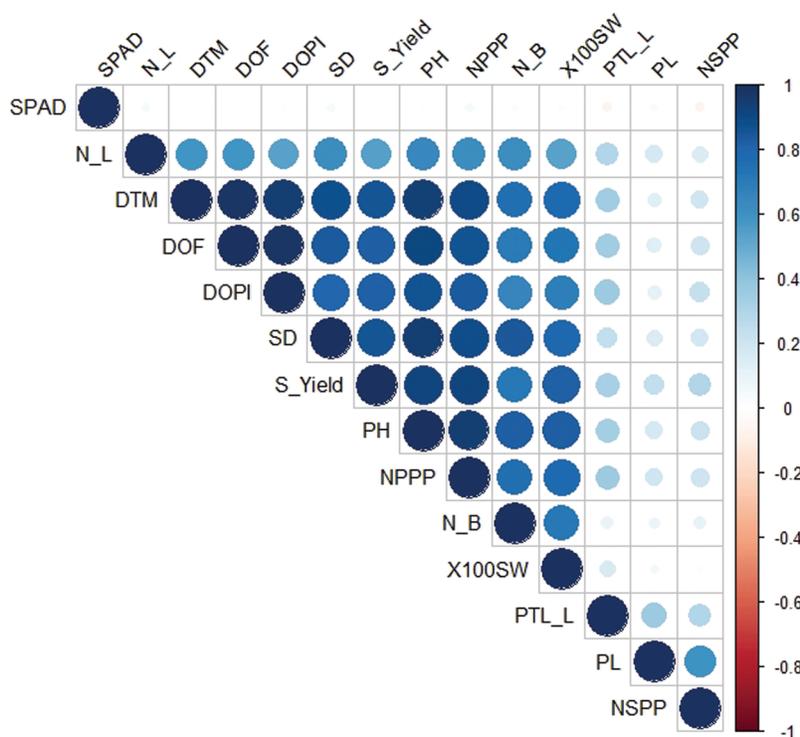


Figure 1: Correlogram for the 14 agronomic attributes recorded in the study. SPAD = SPAD measurements; N_L = No. of leaves per plant; DTM = Days taken to maturity; DOF = Days taken to flowering initiation; DOPI = Days to pod initiation; SD = Stem diameter (mm); S_Yield = Seed yield per plant (g); PH = Plant height (cm); NPPP = No. of pods per plant; N_B = No. of branches per plant; X100 SW = 100 seed weight (g); PTL_L = Petiole length (cm); PL = Pod length (cm); NSPP = No. of seeds per pod

Number of leaves per plant showed significant positive correlation with most of the characters under study, i.e., plant height (0.641), petiole length (0.298), number of branches per plant (0.333), stem diameter (0.382), days to flowering initiation (0.586), days to pod initiation (0.50), days taken to maturity (0.586), number of pods per plant (0.616), number of seeds per pod (0.203), 100-seed weight (0.284) and seed yield per plant (0.378). On the other hand, number of leaves per plant showed non-significant correlation with SPAD measurement (0.051); while, pod length was negatively correlated with number of leaves per plant (-0.161). Number of branches per plant showed significant positive correlation with most of the characters under study, i.e., stem diameter (0.384), days to flowering initiation (0.362), days to pod initiation (0.290), days taken to maturity (0.362), number of pods per plant (0.416), 100-seed weight (0.372) and seed yield per plant (0.421). Number of branches per plant were observed to be negatively correlated with number of seeds per pod (-0.049). Stem diameter was observed to be positively correlated with seed yield per plant (0.359), days to pod initiation (0.404), days taken to maturity (0.510), number of pods per plant (0.521), 100-seed weight (0.291) and days to flowering initiation (0.466). While, number of seeds per pod is negatively correlated (-0.049) with stem diameter. Days to flowering initiation is positively correlated with majority of the traits under study, i.e., seed yield (0.427), days taken to maturity (0.980), 100-seed weight (0.199), number of pods (0.873), number of seeds per pod (0.266) and days to pod initiation (0.946). Days to pod initiation is positively associated with seed yield (0.390), number of pods (0.819), 100-seed weight (0.172), number of seeds per pod (0.245) and days taken to maturity (0.918). Number of pods per plant showed significant positive association with seed yield (0.548), 100-seed weight (0.289) and number of seeds (0.272). Number of seeds per pod is negatively correlated with 100-seed weight (-0.023); while positively associated with seed yield (0.280). 100-seed weight was observed to be positively correlated (0.876) with seed production.

3.4 Path Coefficient Analysis

Path coefficient analysis was carried out to assess the relative contribution of each studied parameter towards seed yield per plant based on computed indirect and direct impacts (Table 5 and Fig. 2). Results based on path coefficient analysis exhibited positive and direct contribution of plant height in seed yield per plant (0.024). Similarly, plant height showed the highest and the lowest indirect positive contribution in seed yield per plant via SPAD measurement (0.719) and petiole length (0.009) respectively. SPAD measurement showed neither significant direct nor indirect contribution in seed production as correlation analysis exhibited non-significant relation with rest of the other variables under study. Pod length showed negative direct contribution in the seed yield per plant (-0.156). It exhibited maximum and minimum positive indirect contribution in seed production using SPAD measurement (0.301) and number of branches per plant (0.002), respectively. Petiole length showed direct positive effect on seed production per plant (0.026). It exhibited highest and lowest positive indirect contribution towards seed production using number of branches per plant (0.003) and SPAD measurement (0.316), respectively. Number of leaves showed negative direct contribution in seed production per plant (-0.092). Contrary to this, number of leaves exhibited maximum and minimum indirect and positive contribution in seed production by number of pods (0.383) and petiole length (0.008), respectively.

Table 5: Path coefficient analysis showing direct (diagonal) and indirect effects of different quantitative traits on seed yield/plant

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	Genotypic corr. with yield
X1	0.024	0.719	-0.028	0.009	-0.059	0.025	0.043	0.326	0.011	-0.487	0.589	0.093	0.370	0.916
X2	0.000	0.268	-0.004	-0.002	-0.005	0.001	0.002	-0.002	0.001	0.002	0.026	-0.019	0.009	-0.007
X3	0.004	0.301	-0.156	0.010	-0.016	0.002	0.007	0.049	0.002	-0.064	0.120	0.263	0.024	0.243

(Continued)

Table 5 (continued)

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	Genotypic corr. with yield
X4	0.008	0.316	-0.057	0.026	-0.027	0.003	0.011	0.124	0.004	-0.178	0.213	0.129	0.072	0.329
X5	0.016	0.250	-0.028	0.008	-0.092	0.020	0.028	0.212	0.006	-0.304	0.383	0.066	0.239	0.553
X6	0.020	0.156	-0.012	0.002	-0.058	0.031	0.039	0.258	0.008	-0.392	0.467	0.040	0.316	0.718
X7	0.023	0.513	-0.024	0.007	-0.057	0.026	0.046	0.301	0.010	-0.454	0.550	0.080	0.353	0.860
X8	0.022	0.343	-0.021	0.009	-0.054	0.022	0.038	0.361	0.012	-0.508	0.541	0.089	0.323	0.834
X9	0.020	0.597	-0.022	0.009	-0.046	0.019	0.035	0.342	0.013	-0.476	0.508	0.082	0.303	0.785
X10	0.023	0.680	-0.019	0.009	-0.054	0.024	0.040	0.354	0.012	-0.519	0.556	0.084	0.348	0.857
X11	0.023	0.380	-0.030	0.009	-0.057	0.023	0.041	0.315	0.010	-0.465	0.620	0.091	0.345	0.926
X12	0.007	-0.071	-0.122	0.010	-0.018	0.004	0.011	0.096	0.003	-0.130	0.169	0.335	0.009	0.374
X13	0.020	0.246	-0.008	0.004	-0.050	0.022	0.037	0.265	0.009	-0.410	0.486	0.007	0.440	0.822

Notes: X1 = Plant Height (cm); X2 = SPAD Measurement (cci); X3 = Pod Length (cm); X4 = Petiole Length (cm); X5 = No. of Leaves per Plant; X6 = No. of Branches per Plant; X7 = Stem Diameter (mm); X8 = Days to Flowering Initiation; X9 = Days to Pod Initiation; X10 = Days Taken to Maturity; X11 = No. of Pods per Plant; X12 = No. of Seeds per Pod; X13 = 100 Seed Weight (g).

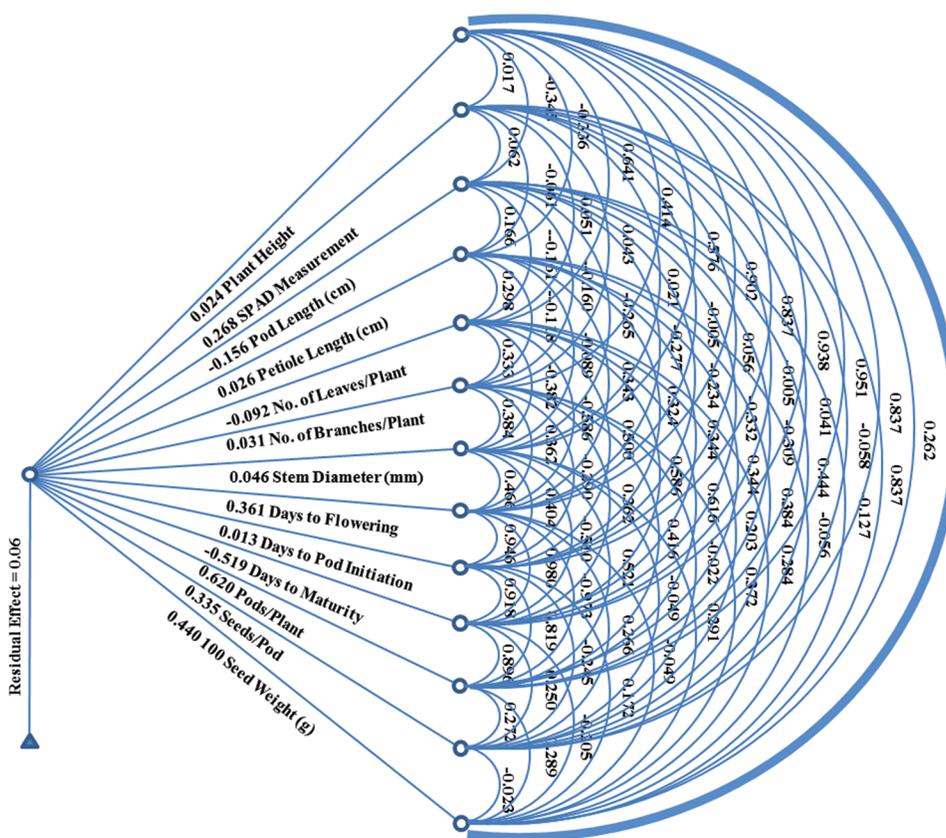


Figure 2: Genotypic path diagram for seed yield/plant (g)

Number of branches exhibited positive direct contribution in seed yield per plant (0.031). Maximum and minimum positive indirect contribution of number of branches for seed production was observed by 100-seed weight (0.316) and petiole length (0.002), respectively. Stem diameter exhibited positive direct effects on seed production per plant (0.046). The highest and the lowest positive indirect effect of stem diameter on seed production was recorded by days taken to pod initiation (0.010) and number of pods per plant (0.550), respectively. Days taken to flowering initiation exhibited direct and positive impacts on seed yield per plant (0.361). Maximum and minimum positive indirect effect of days to flowering initiation on seed production were recorded by number of pods and petiole length (0.009), respectively. Days to pod initiation showed positive direct contribution in seed production per plant (0.013). The highest and the lowest positive indirect effect of days to pod initiation on seed production was recorded by using SPAD measurement (0.597) and petiole length (0.009), respectively. In this study, data regarding days taken to maturity showed negative direct effect on seed production per plant (-0.519). Maximum and minimum positive indirect effect of days taken to maturity on seed production were computed by the help of SPAD measurement (0.680) and petiole length (0.009), respectively. Number of pods showed positive and direct contribution in seed yield (0.620). The highest and the lowest positive indirect contribution of number of pods in seed production was seen via SPAD measurement (0.380) and petiole length (0.009), respectively. Number of seeds per pod showed positive direct effects on seed production (0.335). Maximum and minimum positive indirect effect of number of seeds per pod on seed production was seen by days taken to pod initiation (0.003) and number of pods (0.169). 100-seed weight showed positive direct effect on seed yield per plant (0.440). The highest and the lowest positive indirect effect of 100-seed weight on seed production was seen by plant height (0.020) and number of pods per plant (0.486) respectively. The residual effects in the present study were estimated (0.06) depicting 6% of the total generated variation due to possible environment factors.

3.5 D^2 Statistics

To assess the extent of genetic variability soybean genotypes were categorized into various clusters. It allowed optimizing the hybridization pattern among best performing entries of genetically distinct groups. A core set of 123 soybean genotypes was grouped into 17 different clusters. Among 17 clusters, clusters III and IV were the biggest with 18 genotypes in each followed by clusters VII and II with 14 genotypes in each. Clusters VIII and XI contained 11 genotypes in each; while clusters X and XVI contained six genotypes in each cluster. Moreover, clusters V and IX contained five genotypes in each followed by cluster VII with 4 genotypes whereas clusters including 1, XIV and XVII contained two genotypes in each cluster and cluster I was solitary as it contained only one genotype. The distribution of genotypes and clustering pattern is illustrated in Fig. 3 and Table 6. Average D^2 values of inter and intra-cluster distances are demonstrated in Table 7. All 17 clusters in the study showed zero intra-cluster differences. Differences among the clusters ranged from 9.00 to 91.11 inter-cluster distances. Cluster I exhibited minimum inter-cluster distance with cluster IV (13.19) and maximum inter-cluster distance with cluster V (67.72). Cluster II exhibited minimum inter-cluster distance with cluster III (12.19) and maximum inter-cluster distance with cluster VII (49.67). Cluster III exhibited minimum inter-cluster distance with cluster XI (14.12) and maximum inter-cluster distance with cluster IX (46.97). Cluster IV exhibited minimum inter-cluster distance with cluster VIII (9.00) and maximum inter-cluster distance with cluster V (60.05). Cluster V exhibited minimum inter-cluster distance with cluster IV (18.16) and maximum inter-cluster distance with cluster XII (91.11). Cluster VI exhibited minimum inter-cluster distance with cluster IX (17.33) and maximum inter-cluster distance with cluster IV (76.88). Cluster VII exhibited minimum inter-cluster distance with cluster XVI (10.72) and maximum inter-cluster distance with cluster XII (70.99).

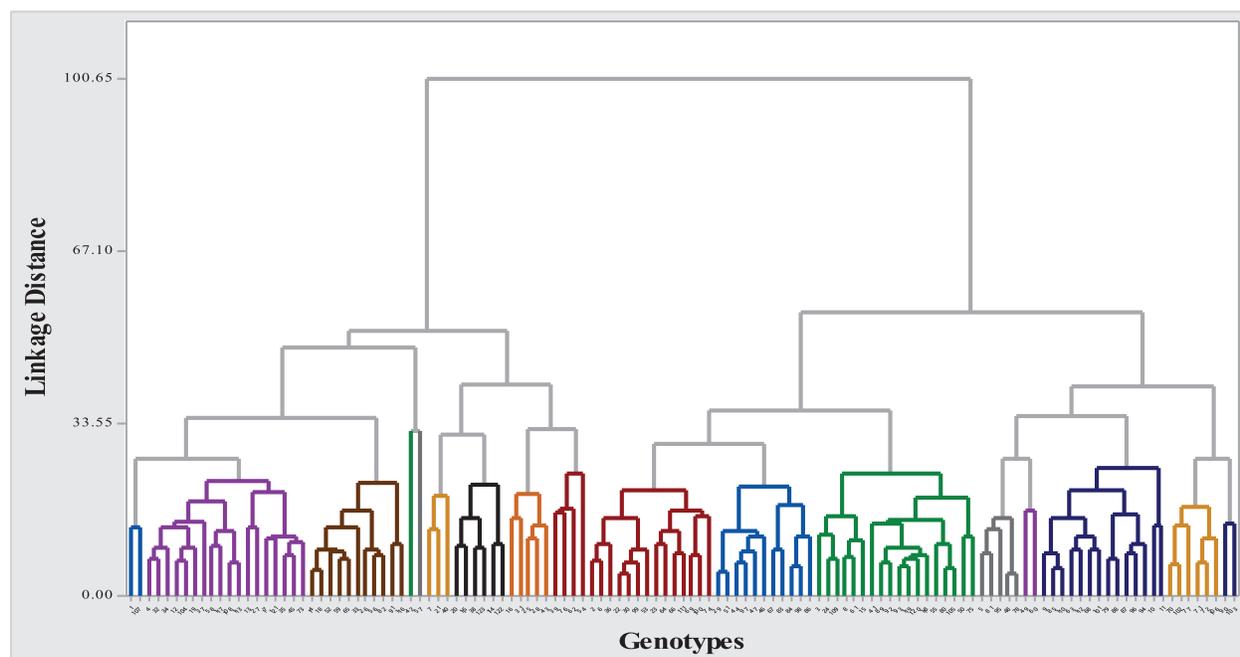


Figure 3: Dendrogram of 123 soybean genotypes based on morphological, phenological and yield attributing characters

Table 6: Distribution of 123 soybean genotypes in different clusters groups

Sr. no.	Cluster	No. of genotypes	Names of genotypes
1	I	2	17418 and G-45
2	II	14	24606, 24491, 24569, G-25, G-29, G-30, Ilksoy, atakisi, Bravo, 24594, G-14, G-43, 24589 and G-9
3	III	18	17122, 24528, 24543, 24010, G-42, 24496, 24564, 24603, A-3127, G-46, G-51, 24488, 24512, 24494, SA-88, 24544, 24573 and G-7
4	IV	18	24566, G-10, 24607, 24593, 17420, 17425, 24563, 24505, 24526, G-37, 24592, 24506, 33521, 33523, G-49, G-3, G-38 and G-8
5	V	5	24492, 24527, 24508, 24513 and 24571
6	VI	3	17426, 24499 and 24567
7	VII	14	17431, G-20, G-48, 24608, G-50, G-2, G-39, G-13, G-24, G-22, G-34, G-31, 23981, 23996
8	VIII	11	24490, 24495, 24591, 24604, 33522, 24529, 24595, G-17, G-28, G-54 and 24511
9	IX	5	17423, G-15, G-32, 24577 and G-12
10	X	6	24497, G-53, 24565, Arisoy, Nova and G-52
11	XI	11	24514, 24590, 24572, G-35, 24574, 24576, G-1, G-18, G-19, G-36 and G-21
12	XII	4	17421, 24507, G-47 and 17427

(Continued)

Sr. no.	Cluster	No. of genotypes	Names of genotypes
13	XIII	1	24570
14	XIV	2	24588 and 24605
15	XV	1	24597
16	XVI	6	G-4, G-40, G-11, G-5, G-6 and G-44
17	XVII	2	G-27 and G-41

Table 7: Average intra and inter-cluster D^2 values of soybean genotypes

Clusters	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII
I	0.00	30.50	30.50	13.19	67.72	23.71	49.86	21.30	19.09	23.98	41.95	31.21	37.92	61.84	26.20	51.85	59.92
II		0.00	12.19	18.74	46.16	40.46	27.33	15.93	45.39	35.03	13.74	49.67	48.19	39.68	34.97	25.55	30.02
III			0.00	20.84	39.63	42.60	20.22	20.36	46.97	41.25	14.12	52.43	54.93	35.97	35.63	22.33	32.59
IV				0.00	60.05	24.53	40.86	9.00	27.17	22.10	30.29	34.16	38.57	54.14	27.18	41.89	48.29
V					0.00	81.28	21.04	59.35	85.50	80.12	36.39	91.11	90.64	18.16	66.99	22.45	34.29
VI						0.00	62.28	25.54	17.33	20.53	51.06	25.12	41.63	76.88	43.26	62.88	67.89
VII							0.00	39.49	66.75	60.41	17.10	70.99	71.47	21.24	49.23	10.72	24.67
VIII								0.00	31.91	22.19	26.51	35.51	40.24	54.19	32.15	39.46	43.40
IX									0.00	20.69	56.94	18.36	33.79	79.43	35.43	68.51	74.50
X										0.00	46.79	20.17	24.18	72.77	33.51	60.17	61.84
XI											0.00	59.52	58.57	31.57	42.46	15.77	19.76
XII												0.00	27.87	84.55	36.44	73.07	75.81
XIII													0.00	79.88	32.04	72.10	71.96
XIV														0.00	57.33	21.81	30.94
XV															0.00	53.31	58.54
XVI																0.00	17.51
XVII																	0.00

Cluster XIII exhibited minimum inter-cluster distance with cluster XI (26.51) and maximum inter-cluster distance with cluster XVII (43.40). Whereas, cluster IX minimum inter-cluster with cluster XII (18.36) and exhibit maximum inter-cluster distance with cluster XIV (79.43). Cluster X exhibited minimum inter-cluster distance with cluster XII (20.17) and maximum inter-cluster distances with cluster XIV (72.77) while cluster XI exhibited minimum inter-cluster distance with cluster XVI (15.77) and maximum inter-cluster distance with cluster XII (59.52). On the other hand, XII showed minimum inter-cluster distances with cluster XVIII and maximum inter-cluster distance with cluster IV (84.55) while cluster XIII showed minimum inter-cluster distances with cluster XV (32.04) and highest inter-cluster distance with cluster XIV (79.88). Cluster XIV exhibited minimum inter-cluster distances with cluster XVI (21.81) and maximum inter-cluster distances with cluster XV (57.33). Means of each cluster for 14 different agronomic attributes are illustrated in the [Table 8](#). Cluster I comprised of only two genotypes which showed above mean values for all agronomic attributes except SPAD measurement. Cluster II contained 14 genotypes with above mean values of agronomic parameters, i.e., days taken to flowering initiation, petiole length, number of leaves per plant, SPAD measurement, days to pod initiation, and number seed per pod. Cluster III contained 18 soybean genotypes which showed more than mean values

for number of seeds per pod, days to flowering initiation and number of leaves per plant. Cluster IV exhibit 18 evaluated genotypes which showed above average performance for rest of the studied agronomic characters except SPAD measurement. Cluster V contained five soybean genotypes and exhibit no parameter with above average mean cluster value for the studied agronomic parameters.

Table 8: Mean values of 14 characters in 17 clusters in evaluated soybean genotypes

Cluster	No. of genotypes	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14
I	2	67.48	32.07	3.27	11.17	24.33	7.00	13.45	58.50	64.50	112.50	44.50	2.83	7.50	9.49
II	14	55.69	50.23	3.01	10.71	19.26	5.43	11.17	46.43	54.36	101.86	36.50	2.81	6.87	8.22
III	18	55.54	39.99	2.96	9.77	14.30	5.89	11.10	44.33	52.11	99.17	35.57	2.80	6.62	7.75
IV	18	64.28	41.18	3.17	12.03	20.35	7.00	13.89	53.17	60.22	108.39	44.18	2.67	7.43	9.22
V	5	34.77	35.11	2.94	7.51	7.80	3.00	5.30	30.60	39.00	84.20	14.80	2.33	5.07	5.09
VI	3	76.34	42.57	2.88	15.28	11.33	8.67	16.83	59.67	67.67	120.67	54.00	2.67	9.13	10.98
VII	14	46.72	39.30	2.90	8.40	11.60	4.21	8.29	35.14	42.71	89.93	28.36	2.48	6.47	6.98
VIII	11	65.59	48.14	3.13	10.17	16.64	7.09	13.96	50.82	57.64	107.55	44.83	2.49	7.93	9.00
IX	5	80.05	33.74	2.98	13.08	25.20	9.40	17.46	61.60	70.00	120.20	53.20	2.93	8.60	10.64
X	6	75.01	51.89	3.43	11.37	27.83	9.83	18.87	58.17	64.83	115.50	52.50	2.78	8.63	10.56
XI	11	52.88	50.85	3.12	10.55	14.73	5.00	10.25	39.55	46.55	95.00	34.82	2.55	6.50	7.46
XII	4	88.25	41.17	3.33	10.33	29.25	13.00	22.25	56.50	61.75	118.75	60.50	2.83	10.25	12.21
XIII	1	79.77	52.70	3.58	15.03	50.00	11.00	14.90	57.00	63.00	112.00	57.00	3.00	9.88	11.23
XIV	2	37.73	39.53	3.03	16.78	22.00	3.00	5.65	32.00	40.50	87.00	16.00	2.50	4.82	4.82
XV	1	66.77	31.53	2.28	11.67	43.00	8.00	14.60	47.00	52.00	107.00	48.00	2.00	8.38	8.38
XVI	6	44.70	47.75	2.84	7.81	10.61	4.17	7.83	36.67	46.17	91.67	23.83	2.56	6.55	6.56
XVII	2	46.88	63.48	2.30	6.30	13.83	4.00	8.15	34.50	40.50	90.50	25.50	2.00	6.61	6.63
Mean	7.24	58.69	43.84	3.04	10.47	17.57	6.27	12.08	46.18	53.61	102.03	38.20	2.64	7.17	8.32

Notes: X1 = Plant Height (cm); X2 = SPAD Measurement (cci); X3 = Pod Length (cm); X4 = Petiole Length (cm); X5 = No. of Leaves per Plant; X6 = No. of Branches per Plant; X7 = Stem Diameter (mm); X8 = Days to Flowering Initiation; X9 = Days to Pod Initiation; X10 = Days Taken to Maturity; X11 = No. of Pods per Plant; X12 = No. of Seeds per Pod; X13 = 100 Seed Weight (g); X14 = Seed Yield per Plant (g).

Cluster VI contained three genotypes and exhibit above average cluster means for all of the recorded agronomic parameters except SPAD measurement, pod length and number of leaves per plant. Cluster VII contained 14 genotypes with no parameter having above average cluster means. Cluster VIII contained 11 genotypes and exhibited above average cluster means for most of the characters except petiole length, number of leaves per plant and number of seeds per pod. Cluster IX contained five soybean genotypes and showed above average cluster means for most of the characters under study except SPAD measurement and pod length. Cluster X contained six soybean genotypes which showed above average cluster means for all agronomic parameters in the study except number of seeds per pod. Cluster XI covered 11 soybean genotypes with above average cluster means contained by SPAD measurement, pod length and petiole length. Cluster XII contained four soybean genotypes which showed above average cluster mean with rest of the recorded parameters except SPAD measurement and petiole length. Moreover, cluster XIII contained only one soybean genotype with above average cluster means with all of the studied agronomic characters. Cluster XIV contained four soybean genotypes and showed above average cluster means for only petiole length and number of leaves per plant. Cluster XV contained only one soybean genotype which showed above average cluster means for all characters except SPAD measurement, pod length, days taken to pod initiation, number of seeds per pod. Cluster XVI contained six soybean genotypes which showed above average cluster means only for SPAD

measurement taken in the study. Cluster XVII contained only two soybean genotypes with above average cluster means only for SPAD measurement.

4 Discussion

Utilization and conservation of gene diversity is a thriving vitality to develop crop genotypes that give high yield [28]. Analysis of variance (ANOVA) is a statistical method that uncovers the variability present among the treatment variables [29]. Analysis of variance (ANOVA) depicted significant genetic diversity in the evaluated soybean varieties for all recorded parameters excluding number of seeds per pod. Hence, variation exposed by ANOVA encourages the effective selection on the base of evaluated growth and yield parameters in order to initiate an effective soybean breeding program. Results of the study conducted by Sulistyono et al. [30] coincide with our findings as there is existence of ample variation. Descriptive statistics gives an obvious middle ground for effective data visualization based on quantitative description of different variables [31]. Various estimates of variability in the present study indicated non-significant environmental influence on evaluated soybean genotypes for most of the characters. A slight difference between genotypic coefficient of variation (GCV) and phenotypic (PCV) for almost all studied traits except days taken to maturity depicted the minimal environmental influence. Because of high susceptibility of soybean to photoperiod, there is a difference between GCV and PCV for days taken to maturity [32,33]. Moreover, high heritability estimates for most of the studied characters suggested the high transmissibility of genes associated with superior agronomic attributes. It also depicted the minimal environmental influence on the genetic expression of growth attributes [34]. Similar heritability estimates were seen in soybean [35] and maize [36], which proved to be of immense help in effective selection of superior agronomic attributes.

An adequate value of coefficient of variation (CV) percentage indicates the degree of precision with which the treatments are compared. In plant sciences, CV % is an important index for the assessment of yield stability based on agronomic performance [37]. Outcomes of this research show high CV % for most of the studied characters viz., 100 seed weight, No. of leaves, petiole length, No. of pods, days to pod initiation, SPAD measurement, stem diameter, days taken to flowering initiation, plant height, No. of branches, seed yield, pod length, and days taken to maturity. High heritability estimates for all of the studied agronomic characters could facilitate effective selection of the desirable soybean genotypes. Results of the descriptive statistics are in concordance with research conducted by Karunakar et al. [38] and Kumari et al. [39]. Selection relies on the level of the correlation between two traits in order to pinpoint pleiotropic effects and genetic linkage [20]. Correlation studies help pick good parental genotypes from the breeding population [40,41]. Generally, positive correlation depicts the direct relation among two variables while negative correlation coefficient leads to an inverse relation among two variables [42]. Several studies reported the significance of correlation analysis in indicating superior agronomic attributes to select high yielding genotypes in different oilseeds, i.e., in soybean [21,43,44], sesame [45–47] and canola [48,49]. In this research, correlation analysis revealed slight higher phenotypic correlation coefficients as compared to genotypic ones. Significant positive correlations among most of the studied characters with yield were recorded. Same result was observed by Ghodrati et al. [50] in soybean where most of the studied characters were positively correlated. Plant height is an integral phenotypic and growth attribute in plant breeding with which most of the other plant characteristics are associated. It provides an insight for estimating biological as well as economic yield. Plant height showed high positive relationship with almost all of the studied characters which directly depicted its usefulness for the selection of desirable soybean genotypes with improved agronomic performance. Similar association pattern of plant height with various yield and its attributing characters was observed by Mahbub et al. [51] except the association of 100-seed weight; which had a negative correlation with plant height in their findings while positively correlated in our study.

Utilization of SPAD readings is a simple and relatively reliable way for the instant assessment of chlorophyll contents without damaging the plant leaf [52]. Previously, Cruz et al. [53] argued that non-significant correlation might also lead to an effective interpretation of the degree of association based on magnitude and direction among different variables. In this research, SPAD measurements exhibited non-significant positive relation with many characters under observation whereas non-significant negative relation with number of seeds per pod, days to maturity and petiole length. Seed yield showed negative relation with pod length in this research. This significant decrease in pod length with increased yield might be due to increased number of pods per plant thus indicating the phenomenon of genetic determinacy. Similar results were observed in a study conducted by Akram et al. [54]. In contrast, Mahbub et al. [22] observed highly positive relationship between seed yield and pod length in soybean.

Selection of desirable genotypes from a genetically diverse pool of germplasm is the final goal of a soybean breeding program. Plant production is a complicated quantitative trait that relies on the relationship between growth parameters. Since correlation analysis can only estimate the relationship between variables, it may confound breeders to choose the optimal combination. In contrast, by decomposing the genotypic association into direct and indirect effects path coefficient analysis assesses the importance of yield-attributing traits. Therefore, a path coefficient analysis was performed to calculate the exact role that each trait had in the plant production. Path coefficient analysis has shown that most of the characters exhibit positive contribution in the seed production. High positive effect on seed production was estimated for plant height, 100-seed weight, number of branches per plant, days taken to flowering initiation, number of seeds per pod, SPAD measurement, stem diameter, number of pods per plant, petiole length and days taken to pod initiation. Similar results for 100-seed weight, number of pods per plant and number of branches per plant were individually revealed previously [44,55]. Positive effects of 100-seed weight, plant height and days taken to flowering initiation on seed yield coincides with the results concluded by Ghodrati et al. [50]. Whereas, Chavan et al. [56] revealed that plant height exhibited negative direct contribution in seed yield. Similarly, positive direct effects for number of branches per plant are in contradiction with the study conducted by Malek et al. [57].

Days taken to maturity number of leaves per plant and pod length negatively and directly contributed in seed yield per plant. This proposed that selection on the basis of such characters might reduce plant yield in soybean thus should not be included in the selection criteria. Similarly, Malik et al. [21] and Mahbub et al. [22] individually studied negative impacts of days taken to maturity on seed production; thus are in concordance with the results of the present study. On the other hand, Balla [43] showed high positive impact of days taken to maturity on seed production. Therefore, it might be possible that maturity index in soybean showed higher environmental influence as compared to the other reproductive growth attributes. Negative impacts of pod length on seed production in this research contradict with the study conducted by Arshad et al. [55] where it showed positive direct contribution in the seed yield. Results of the path coefficient analysis in this research concluded that selection based on various reproductive growth attributes i.e., number of pods per plant followed by days to flowering and 100-seed weight might lead to development of soybean genotypes with superior agronomic performance.

In plant breeding, identification of various groups based on genetic similarities and dissimilarities play a pivotal role in devising suitable hybridization scheme and the selection of superior parental lines [9]. Identification of various genetic patterns leads to the systematic characterization of available genetic diversity and facilitate plant breeders to assess the phylogenetic associations among the tested genotypes [10]. Estimation of genetic divergence can be effectively accomplished by using D^2 statistics. Several researchers applied D^2 statistics to classify evaluated genotypes into various groups or clusters in different legumes including soybean [58,39], chickpea [59], garden pea [60], lentils [61] and alfalfa [62]. In the present study, D^2 statistics approach developed previously [19], and it was followed for phenotypic clustering of the evaluated soybean diversity panel based on various agronomic attributes. The soybean

diversity panel was grouped into 17 clusters. All 17 clusters showed zero intra-cluster variability on the basis of which it can be assumed that negligible amount of variability has been existed within cluster. However, effective results for inter-cluster variation were observed where differences among the clusters ranged from 9.00 to 91.11. Cluster V gave maximal inter-cluster distance with cluster XII which showed that hybridization between the genotypes of these two genetically divergent clusters might produce extraordinary results, i.e., increased seed yield and adaptability. Considering the cluster mean analysis, it has been observed that cluster V contained early maturing genotypes as compared to rest of the clusters. Similarly, cluster XVII contained genotypes with highest SPAD readings depicting the higher photosynthetic efficiency in soybean and other plant species [63–66]. Cluster XII showed higher cluster means for No. of pods per plant along with 100-seed weight and seed yield per plant while the clustered genotypes took higher number of days to mature. Same outcomes were observed previously in soybean [58,39] where increased cluster mean for days taken to maturity resulted into higher seed yield per plant.

5 Conclusions

This study implies the high genetic divergence, growth and yield related attributes in soybean genotypes. Higher correlation coefficients coupled with considerable direct and indirect effects could be helpful for selecting superior yield and its attributing descriptors. Therefore, selection of desirable genetic combinations could be possible to develop soybean genotypes with superior agronomic performance. The exotic soybean genotypes also showed significant genetic potential regarding their utilization in the breeding program. Phenotypic clustering accomplished through cluster analysis effectively grouped the evaluated soybean diversity panel into various clusters. Genotypes which classified into those cluster groups having above mean value for maximum recorded parameters could be used for developing an effective soybean hybridization program. The findings of this study and the traits associated genotypic variations in soybean can be utilized by the breeders and farmers for improving agricultural traits in soybean.

Acknowledgement: The authors extend their appreciation to Researchers Supporting Project No. (RSP2023R298), King Saud University, Riyadh, Saudi Arabia.

Funding Statement: This research was funded by the Researchers Supporting Project No. (RSP2023R298), King Saud University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Li, J. H., Wang, E. T., Chen, W. F., Chen, W. X. (2008). Genetic diversity and Potential for promotion of plant growth detected in nodule endophytic bacteria of soybean grown in Heilongjiang Province of China. *Soil Biology and Biochemistry*, 40, 238–246.
2. Herridge, D. F., Peoples, M. B., Boddey, R. M. (2008). Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil*, 311, 1–8.
3. Biate, D. L., Kumar, L. V., Ramadoss, D., Kumari, A., Naik, S. et al. (2014). Genetic diversity of soybean root nodulating bacteria. *Bacterial Diversity in Sustainable Agriculture*, 1, 131–145.
4. Chang, W. S., Lee, H. I., Hungria, M. (2015). Soybean production in the Americas. In: *Principles plant-microbe interactions*, pp. 393–400. Cham, Switzerland: Springer.
5. Liu, J., Zhu, J. K. (1997). Proline accumulation and salt-stress-induced gene expression in a salt-hypersensitive mutant of *Arabidopsis*. *Plant Physiology*, 114, 591–596.
6. Carrera, C., Martínez, M. J., Dardanelli, J., Balzarini, M. (2011). Environmental variation and correlation of seed components in nontransgenic soybeans: Protein, oil, unsaturated fatty acids, tocopherols, and isoflavones. *Crop Science*, 51, 800–809.

7. FAOSTAT (2019). World soybean production statistics. <http://www.fao.org/faostat/en/#data/QC>
8. Khurshid, H., Baig, D., Jan, S. A., Arshad, M., Khan, M. A. (2017). Miracle crop: The present and future of soybean production in Pakistan. *MOJ Biology and Medicine*, 2, 189–191.
9. Bhakuni, V., Shukla, P. S., Singh, K., Singh, V. K. (2017). Morphological characterization and assessment of genetic variability in soybean varieties. *International Journal of Current Microbiology and Applied Sciences*, 6, 361–369.
10. Naik, S. M., Madhusudan, K., Motagi, B. N., Nadaf, H. L. (2016). Diversity in soybean (*Glycine max* L.) accessions based on morphological characterization and seed longevity characteristics. *Progressive Research—An International Journal*, 11(3), 377–381.
11. Masoudi, B., Bihanta, M. R., Babaei, H. R., Peyghambari, S. A. (2008). Evaluation of genetic diversity for agronomic, morphological and phenological traits in soybean. *Seed and Plant Journal*, 24(3), 413–427.
12. Salimi, S., Lahiji, H. S., Abadi, G. M., Salimi, S., Moradi, S. (2012). Genetic diversity in soybean genotypes under drought stress condition using factor analysis and cluster analysis. *World Applied Sciences Journal*, 16, 474–478.
13. Malik, R., Tiwari, R., Arora, A., Kumar, P., Sheoran, S. et al. (2013). Genotypic characterization of elite Indian wheat genotypes using molecular markers and their pedigree analysis. *Australian Journal of Crop Sciences*, 7, 61–67.
14. Follmann, D. N., Souza, V. Q., Cargnelutti, A., Demari, G. H., Nardino, M. et al. (2019). Agronomic performance and genetic dissimilarity of second-harvest soybean cultivars using REML/BLUP and Gower's algorithm. *Bragantia*, 18, 197–207.
15. Mukuze, C., Tukamuhabwa, P., Maphosa, M., Dari, S., Dramadri, I. O. et al. (2020). Genetic diversity analysis among soybean genotypes using SSR markers in Uganda. *African Journal of Biotechnology*, 19, 439–448.
16. Hwang, T. Y., Gwak, B. S., Sung, J., Kim, H. S. (2020). Genetic diversity patterns and discrimination of 172 korean soybean (*Glycine max* (L.) merrill) varieties based on SSR analysis. *Agriculture*, 10, 77.
17. Arslan, E., Mutlu, E. G., Dursun, O., Bagci, S. A. (2020). Comparative analysis of agronomic traits and ISSR method among some soybeans [*Glycine Max* (L.) Merr.] genotypes. *Kahramanmaraş Sütçü İmam Üniversitesi Tarım ve Doğa Dergisi*, 23, 687–696.
18. Kaur, G., Joshi, A., Jain, D., Choudhary, R., Vyas, D. (2016). Diversity analysis of green gram (*Vigna radiata* (L.) Wilczek) through morphological and molecular markers. *Turkish Journal Agriculture and Forestry*, 40, 229–240.
19. Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India*, vol. 2, pp. 49–55.
20. Sakai, K. I., Suzuki, A. (1964). Induced mutation and pleiotropy of genes responsible for quantitative characters in rice. *Radiation Botany*, 4, 141–151.
21. Malik, M. F., Ashraf, M., Qureshi, A. S., Ghafoor, A. (2007). Assessment of genetic variability, correlation and path analyses for yield and its components in soybean. *Pakistan Journal of Botany*, 39, 405.
22. Mahbub, M. M., Rahman, M. M., Hossain, M. S., Mahmud, F., Kabir, M. M. (2015). Genetic variability, correlation and path analysis for yield and yield components in soybean. *American-Eurasian Journal Agriculture and Environment Sciences*, 15(2), 231–236.
23. Diondra, W., Ivey, S., Washington, E., Woods, S., Walker, J. et al. (2008). Is there a correlation between plant height and yield in soybean. *Reviews Biology and Biotechnology*, 7, 70–76.
24. Hassanijalilian, O., Igathinathane, C., Doetkott, C., Bajwa, S., Nowatzki, J. et al. (2020). Chlorophyll estimation in soybean leaves infield with smartphone digital imaging and machine learning. *Computers and Electronics in Agriculture*, 174, 105433. <https://doi.org/10.1016/j.compag.2020.105433>
25. Steel, R. G., Torrie, J. H. (1980). *Principles and procedures of statistics: A biometrical approach*. New York: McGraw-Hill.
26. Cohen, J., Cohen, P., West, S. G., Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences*, 3rd edition. Routledge. <https://doi.org/10.4324/9780203774441>
27. Dewey, D. R., Lu, K. (1959). A correlation and path-coefficient analysis of components of crested wheatgrass seed production. *Agronomy Journal*, 51(9), 515–518. <https://doi.org/10.2134/agronj1959.00021962005100090002x>

28. Hufford, M. B., Berny, Y., Teran, J. C., Gepts, P. (2019). Crop biodiversity: An unfinished magnum opus of nature. *Annual Review of Plant Biology*, 29(1), 727–751. <https://doi.org/10.1146/annurev-arplant-042817-040240>
29. Sawyer, S. F. (2009). Analysis of variance: The fundamental concepts. *Journal of Manual and Manipulative Therapy*, 17(2), 27–38. <https://doi.org/10.1179/jmt.2009.17.2.27E>
30. Sulisty, A., Sari, K. P. (2018). Correlation, path analysis and heritability estimation for agronomic traits contribute to yield on soybean. *IOP Conference Series: Earth and Environmental Science*, 102, 12034. <https://doi.org/10.1088/1755-1315/102/1/012034>
31. Proches, S. (2016). Descriptive statistics in research and teaching: Are we losing the middle ground? *Quality and Quantity*, 50, 2165–2174.
32. Nico, M., Miralles, D. J., Kantolic, A. G. (2019). Natural post-flowering photoperiod and photoperiod sensitivity: Roles in yield-determining processes in soybean. *Field Crops Research*, 231, 141–152.
33. Yang, W., Wu, T., Zhang, X., Song, W., Xu, C. (2019). Critical photoperiod measurement of soybean genotypes in different maturity groups. *Crop Sciences*, 59, 2055–2061.
34. Saravanan, K. R., Vishnupriya, V., Prakash, M., Anandan, R. (2019). Variability, heritability and genetic advance in tomato genotypes. *Indian Journal of Agricultural Research*, 53, 92–95.
35. Mesfin, H. H. (2018). Path analysis, genetic variability and correlation studies for soybean (*Glycine max* (L.) Merrill) for grain yield and secondary traits at Asosa, Western Ethiopia. *Greener Journal of Plant Breeding and Crop Science*, 6, 35–46.
36. Chavan, S., Bhadru, D., Swarnalatha, V., Mallaiiah, B. (2020). Studies on genetic parameters, correlation and path analysis for yield and yield attributing traits in sweet corn (*Zea mays* L.). *International Journal of Current Microbiology and Applied Sciences*, 9, 1725–1734.
37. Ray, D. K., Gerber, J. S., MacDonald, G. K., West, P. C. (2015). Climate variation explains a third of global crop yield variability. *Nature Communications*, 6, 1–9.
38. Karunakar, J., Preethi, T. L., Boopathi, N. M., Pugalandhi, L., Hepziba, S. J. (2018). Genetic variability, correlation and path analysis in Moringa (*Moringa oleifera* L.). *Journal of Pharmacognosy and Phytochemistry*, 7, 3379–3382.
39. Kumari, S., Sreenivasa, V., Lal, S. K., Singh, S. K., Singh, K. P. (2019). Analysis of genetic diversity of 120 genotypes of *Glycine max* (L.) merrill by using D² analysis. *Journal of Pharmacognosy and Phytochemistry*, 8, 1324–1329.
40. Abdelghany, A. M., Zhang, S., Azam, M., Shaibu, A. S., Feng, Y. et al. (2020). Profiling of seed fatty acid composition in 1025 Chinese soybean accessions from diverse ecoregions. *Crop Journal*, 8, 635–644.
41. Yu, K., Wang, H., Liu, X., Xu, C., Li, Z. et al. (2020). Large-scale analysis of combining ability and heterosis for development of hybrid maize breeding strategies using diverse germplasm resources. *Frontiers in Plant Science*, 11, 660.
42. Nogueira, A. P., Sedyama, T., de Sousa, L. B., Hamawaki, O. T., Cruz, C. D. et al. (2012). Path analysis and correlations among traits in soybean grown in two dates sowing. *Bioscience Journal*, 28, 877–888.
43. Balla, M. Y. (2017). Ibrahim, S.E. Genotypic correlation and path coefficient analysis of soybean [*Glycine max* (L.) Merr.] for yield and its components. *Agriculture Research and Technology*, 7, 1–5.
44. Machado, B. Q., Nogueira, A. P., Hamawaki, O. T., Rezende, G. F., Jorge, G. L. (2017). Phenotypic and genotypic correlations between soybean agronomic traits and path analysis. *Genetics and Molecular Research*, 16(2), gmr16029696.
45. Teklu, D. H., Kebede, S. A., Gebremichael, D. E. (2014). Assessment of genetic variability, genetic advance, correlation and path analysis for morphological traits in sesame genotypes. *Asian Journal of Agricultural Research*, 8, 181–194.
46. Fazal, A., Mustafa, H. S., Hasan, E., Anwar, M., Tahir, M. H. (2015). Interrelationship and path coefficient analysis among yield and yield related traits in sesame (*Sesamum indicum* L.). *Nature Science*, 13, 27–32.
47. Navaneetha, J. S., Murugan, E., Parameswari, C. (2019). Correlation and path analysis for seed yield and its components in sesame (*Sesamum indicum* L.). *Electronic Journal Plant Breeding*, 10, 1262–1268.

48. Rameeh, V. (2016). Multivariate analysis of some important quantitative traits in rapeseed (*Brassica napus*) advanced lines. *Journal of Oilseed Brassica*, 1, 162–169.
49. Tariq, H., Tanveer, S. K., Qamar, M., Javaid, R. A., Vaseer, S. G. (2020). Correlation and path analysis of *Brassica napus* genotypes for yield related traits. *Life Science Journal*, 17, 22–34.
50. Ghodrati, G. R., Sekhavat, R., Mahmoodi, S. H., Gholami, A. (2013). Evaluation of correlations and path analysis of components seed yield in soybean. *International Journal of Agriculture*, 3, 795.
51. Mahbub, M. M., Shirazy, B. J. (2016). Evaluation of genetic diversity in different genotypes of soybean (*Glycine max* (L.) Merrill). *American Journal of Plant Biology*, 25, 24–29.
52. d'Oliveira, P. S., Suzuki, L. S., da-Costa Carneiro, J., Machado, J. C., Prates, J. F. et al. (2020). Use of SPAD index in elephant grass pre-breeding. *Ciências Agrárias*, 41, 1047–1052.
53. Cruz, C. D., Regazzi, A. J., Carneiro, P. C. S. (2012). *Modelos biométricos aplicados ao melhoramento genético*. 4th edition. Viçosa: UFV.
54. Akram, R. M., Fares, W. M., Fateh, H. S., Rizk, A. M. (2011). Genetic variability, correlation and path analysis in soybean. *Egyptian Journal of Plant Breeding*, 15, 89–102.
55. Arshad, M., Ali, N., Ghafoor, A. (2006). Character correlation and path coefficient in soybean *Glycine max* (L.) Merrill. *Pakistan Journal of Botany*, 38, 121.
56. Chavan, B. H., Dahat, D. V., Rajput, H. J., Deshmukh, M. P., Diwane, S. L. et al. (2016). Correlation and path analysis in soybean. *Studies*, 1, 2.
57. Malek, M. A., Rafii, M. Y., Afroz, S. S., Nath, U. K., Mondal M. (2014). Morphological characterization and assessment of genetic variability, character association, and divergence in soybean mutants. *Scientist World Journal*, 1, 14–20.
58. Dubey, N., Avinash, H. A., Shrivastava, A. N. (2018). Evaluation of genetic diversity among soybean [*Glycine max* (L.)] genotypes using multivariate analysis. *Plant Archeology*, 18, 908–912.
59. Singh, R. P., Singh, I., Singh, S., Sandhu, J. S. (2012). Assessment of genetic diversity among interspecific derivatives in chickpea. *Journal of Food Legumes*, 25, 150–152.
60. Sekhon, B. S., Sharma, A., Katoch, V., Kapila, R. K., Sood, V. K. et al. (2019). Assessment of genetic diversity in advanced breeding lines derived from intraspecific crosses of garden pea (*Pisum sativum* L.). *Legume Research International Journal*, 42, 145–152.
61. Satpathy, S., Debnath, S. (2020). Assessment of genetic divergence for quantitative traits in *Lens culinaris*. *International Journal of Current Microbiology and Applied Sciences*, 9, 252–261.
62. Rahman, M. A., Alam, I., Kim, Y. G., Ahn, N. Y., Heo, S. H. et al. (2015). Screening for salt-responsive proteins in two contrasting alfalfa cultivars using a comparative proteome approach. *Plant Physiology and Biochemistry*, 89, 112–122.
63. Islam, M. R., Kamal, M. M., Hossain, M. F., Hossain, J., Azam, M. G. et al. (2023). Drought tolerance in Mung bean is associated with the genotypic divergence, regulation of proline, photosynthetic pigment and water relation. *Phyton*, 92, 955–981.
64. Rahman, M. A., Kabir, A. H., Song, Y., Lee, S. H., Hasanuzzaman, M. et al. (2021). Nitric oxide prevents Fe deficiency-induced photosynthetic disturbance and oxidative stress in alfalfa by regulating Fe acquisition and antioxidant defense. *Antioxidants*, 10, 1556.
65. Kabir, A. H., Ela, E. J., Bagchi, R., Rahman, M. A., Peiter, E. et al. (2023). Nitric oxide acts as an inducer of Strategy-I responses to increase Fe availability and mobilization in Fe-starved broccoli (*Brassica oleracea* var. oleracea). *Plant Physiology and Biochemistry*, 194, 182–192.
66. Rahman, M. A., Woo, J. H., Lee, S. H., Park, H. S., Kabir, A. H. (2022). Regulation of Na^+/H^+ exchangers, Na^+/K^+ transporters, and lignin biosynthesis genes, along with lignin accumulation, sodium extrusion, and antioxidant defense, confers salt tolerance in alfalfa. *Frontiers in Plant Science*, 13, 1041764.