ELUCIDATING THE GENETIC POLYMORPHISM IN SOYBEAN GENOTYPES BASED ON MORPHOLOGICAL AND BIOCHEMICAL TECHNIQUE USING MULTIVARIATE ANALYSIS

IHSAN, M.[1] – NAZIR, N.[2] – NISAR, M.[1*] – ULLAH, S.[1] – JAN, T.[1] – AZIZ, T.[3*] – ALHARBI, M.[4] – ALSHAMMARI, A.[4] – ALASMARI, A. F.[4]

[1]*Department of Botany, University of Malakand, Dir Lower 18800, Pakistan*

[2]*Department of Biochemistry, University of Malakand, Dir Lower 18800, Pakistan*

[3]*Department of Agriculture, University of Ioannina, Arta,47132, Greece*

[4]*Department of Pharmacology and Toxicology, College of Pharmacy, King Saud University,*
P.O. Box 2455, Riyadh 11451, Saudi Arabia

[*]*Corresponding authors*
e-mail: mnshaalpk@yahoo.com, iwockd@gmail.com

(Received 17th Aug 2023; accepted 4th Mar 2024)

Abstract. Soybean (*Glycine max* (L.) Merr.) is one of the important oilseed crops that is widely cultivated around the world. It has a high amount of protein content as well as other essential vitamins that are important in our daily lives. For continuous improvement, the development of novel plant species that are resistant to biotic and abiotic stress using effective genetic diversity strategies is important. In the current study, a total of 25 diverse genotypes of soybean using 21 important agro morphological traits were studied. For qualitative traits significant level of variation was found for most of the traits. For quantitative traits, the highest coefficient of variance 32.74% was found for days to flower initiation, followed by days to flower completion 29.62%, while the lowest was found for a number of pods/plants 2.96%. Based on cluster analysis, all the genotypes were separated into two groups at 25% distance and further subdivided at 75% distance, where the genotypes NARC-2 and SWAT-84 were found the most diverse. The cluster three genotypes were found to have early mature genotypes (89 ± 2.44). Based on a number of pods/plant and yield per plant, the cluster 4 genotypes were found to have a maximum number of pods/plant (98.26 ± 32.10), and (38.66 ± 6.85). Among the studied genotypes the highest grain yield (49 g) was produced by genotype C/B 28, whereas the lowest (14.56 g) was observed for C/B 7. Principal component analysis with an eigenvalue of 1.48 accounted for the total variation of 67.73%. The total seed storage proteins analysis resulted in 13 bands, and the molecular weight ranged from 15 to 170 kDa. Two-way cluster analysis was performed and all the populations were divided into two main lineages at 25% distance and further sub divided into six subgroups at distance of a 75%, where the genotypes NARC-2 and SWAT-84 were found the most diverse genotypes. These findings provide a basis for developing elite, locally adapted soybean genotypes as well as implications for understanding the diversity and relationships among these diverse *Glycine max* genotypes.

Keywords: *genetic diversity, Glycine max, cluster analysis, PCA, SDS-PAGE*

Introduction

Soybean (*Glycine max* (L.) Merrill) is a leguminous annual crop belonging to the family Fabaceae with a chromosome number of 2n = 40. Soybean is also called the ‘‘Golden Bean’’ or ‘‘miracle legume’’ of the twenty first century because of its multiple uses (Hale, 2020). Soybean is highly favored in the human diet as it contains 40 to 42% protein, 18 to 22% oil comprising 85% unsaturated fatty acids, and it is cholesterol-free (Mirpoor et al., 2021). Soybean is one of the most essential crops for oil and protein. It is very desired for the human diet since it has the largest protein content (42%), is

DOI: http://dx.doi.org/10.15666/aeer/2205_43794394

abundant in lysine and vitamins, has 20% oil with 85% unsaturated fatty acids, and is cholesterol-free (Rahman et al., 2023). One of the most important crops farmed worldwide is soybean. Moreover, it is a significant cash crop for the industry and is used to make biofuel (Hartman et al., 2011). It also consists of many therapeutic components and has increased its importance in the industrial, agricultural, and medicinal sectors. It is one of the major sources of edible vegetable oil and proteins for livestock feed. It was revealed that soybean became somewhat popular in Khyber Pakhtunkhwa, but never picked up appreciable areas in Punjab, Sindh, and Baluchistan. According to MINFAL (2009), soybean was grown in an area of 51 hectares with a total production of 31 tons in Pakistan. It is grown in Pakistan in a variety of agro-ecological zones but is mostly grown in rainy climates. Currently, the crop is growing in a small amount in the northern region of the country for human consumption and animal feed. The soybean production in different areas of Pakistan from hilly areas of Khyber Pakhtunkhwa, FATA and Hazara Division varies between 12 to 20 mounds/acre (Khurshid et al., 2017). The reason may be a lack of knowledge or the inability to find appropriate materials with effective defenses against the negative effects of poor fertility and acidity in the soil. Therefore, finding and cultivating high-yielding cultivars ought to be a primary focus of soybean breeding. Information about genetic variety in plants is important to increasing yield potential (Ghafoor et al., 2005). Among food legumes, soybeans are acknowledged as significant crops over the world. They have vastly diverse cultivated varieties and cultivars depending on the various growing locales (Malik et al., 2009).

With effective genetic diversity approaches, it is crucial to produce new plant species that can compete with resistant lines to biotic and abiotic stress (Jan et al., 2017). Yet, by comparing/differing among several quantitative features of Soybean genotypes, it is possible to study the genetic variability among various germplasm (Ojo et al., 2012). Hence, research on genetic variety is essential to understanding variability in various germplasm and choosing better genotypes (Ihsan et al., 2021). Breeding and selecting desirable traits depend on the genetic diversity of attributes. Breeders can better comprehend the evolutionary relationships among accessions with knowledge of diversity patterns, and they can sample germplasm in a more systematic way to design better strategies for incorporating valuable diversity in their breeding operations (Ihsan et al., 2024; Bretting and Widrlechner, 2010). It is helpful to describe genetic diversity among plant species through morphological analysis. Such evaluations are simple to conduct since the attribute under investigation may be verified visually without the need of a complex apparatus (Ihsan et al., 2021). Total seed storage protein method is widely used to describe seed protein variation of crop population (Das and Mukherjee, 1995). This technique is also a promising tool for differentiating cultivars of different crop species (Jha and Ohri, 1996). Because seed storage proteins are mostly independent of environmental variation, the SDS-PAGE is an efficient and trustworthy tool for species identification (Malik et al., 2009). The plant breeders are then assisted in choosing the accessions to be employed in the planned hybridization by the characterized material (Ojo et al., 2012). Therefore, genetic diversity study plays a key role to study variability among different germplasm and to select improved genotypes (Balkaya and Ergün, 2008). The assessment of genetic diversity is important not only for crop improvement but also for efficient management and conservation of germplasm resources (Qadir et al., 2017). In the present study, we have evaluated 25 Soybean genotypes to find out genetic diversity through important morphological markers and selection of the suitable high yielding and early maturing genotypes. In this study, SDS-PAGE were used to assess the genetic variation among the studied 25 soybean genotypes.

Materials and Methods

Experimentation

The current experimental work was conducted at Department of Botany University of Malakand. Total of 25 soybean (*G. max*) genotypes with check variety were assessed for different morphological and biochemical (SDS-PAGE) analysis. The seeds were collected from National Agriculture Research Center Islamabad and Agriculture Research Institute Mingora Swat (*Table 1*).

Table 1. *List of Glycine max genotypes used in the present study*

S. No	Name	Passport information	S. No	Name	Passport information
1	NARC-2	NARC Islamabad	14	C/B 12	ARI Swat
2	PSC-60	NARC Islamabad	15	C/B 3	ARI Swat
3	NARC-16	NARC Islamabad	16	C/B 33	ARI Swat
4	William-82	NARC Islamabad	17	C/B 1	ARI Swat
5	Ramal-1	NARC Islamabad	18	C/B 13	ARI Swat
6	SA-72=60	NARC Islamabad	19	C/B 28	ARI Swat
7	Ajmari	NARC Islamabad	20	SWAT 84	ARI Swat
8	C/B 14	ARI Swat	21	C/B 9	ARI Swat
9	C/B 32	ARI Swat	22	C/B 7	ARI Swat
10	C/B 31	ARI Swat	23	C/B 8	ARI Swat
11	C/B 6	ARI Swat	24	C/B 11	ARI Swat
12	MKD 96	ARI Swat	25	Madyan	ARI Swat
13	C/B 17	ARI Swat			

Germination of plant materials

The experimental materials consisting 25 genotypes of soybean were sown in randomized complete block design with three replications. Each genotype was grown in a plot comprised of 3 rows each of 5-meter length and row-to-row distance of 45 cm and 15 cm between the plants was maintained. The seed were sown on 31 July 2018. Different agronomic practices (soil preparation, ploughing, leveling, irrigation, and weeding) were performed from sowing till maturity.

Morphological characterizations

During the current research study, 21 different morphological characters (15 were quantitative and six qualitative traits) were recorded following the basic descriptor of IBPGR 1985. Out of these traits, flower color, hilum color, pod color, seed color, lodging were recorded in percentage. The quantitative traits were Days to Flowering, Days to Flowering 50%, Days to Flower Completion, Days to Maturity, Primary Branches /Plant, Plant Height, Leaflet Number, Leaf Length, Leaf Width, Number of Pods/Plant, Seeds/Pod, Pod Length, Pod Width, Yield/Plant and 100 Seed weight.

Total seed protein analysis

SDS-PAGE research was carried out on all 25 genotypes of seeds used in the morphometric analysis. Five healthy and mature seeds were hand ground using a mortar and pestle. About 0.02 g of the ground seeds were weighed and put into 1.5 ml centrifuge tubes containing 400-µl protein extraction buffer. Prior to centrifugation at room temperature for ten minutes at 12,000 rpm, the powder was homogenised using a vortex for one minute (RT). Following the method of Nisar et al. (2009), 12.25% polyacrylamide gels were employed to analyse the total seed protein.

Data analysis

All the data were recorded from five guarded selected plant and were used for different statistical packages. For qualitative characters, the data frequency in percentage was calculating using Microsoft Excel. Basic statistics (Mean, Standard Deviation, Range and Coefficient of Variance %) was calculated using Microsoft Excel 2016. For quantitative data, Ward’s method was used for Cluster analysis, Pearson Correlation analysis, Principal Component Analysis were computed using SPSS version 22, PC ORD version 5 software. Similarly, for total seed protein, binary matrices data were calculated via Un-weighted pair group arithmetic average as presence and absence of bands. Two-way Cluster analysis was computed using PC ORD software.

Results

Diversity in qualitative parameters

During the present study, significant variation was found for different qualitative traits, which are present in *Table 2*. For flower color, two alleles were responsible, 13 genotypes have purple flower color, while 12 genotypes were white color flower, four different hilum color were noted in the studied genotypes, in which 8 genotypes have black color hilum, 10 genotypes were found to have brown color, 4 genotypes were observed to have buff color, and 3 were imperfect black color. For pod color, two alleles were recorded, in which 22 genotypes have brown pod color while the remaining 3 were reddish brown color pods. Three distinct types of seeds were recorded in which the 5 genotypes have small seeds, 9 genotypes have medium seed size while the 11 genotypes have large seed size. Two types of seed color were observed in the studied genotypes, which are reddish brown and yellow, 10 genotypes have reddish brown color seed, while 15 genotypes have yellow seed color. Lodging is one of the important traits, in the studied genotypes 20 genotypes were observed upright while the 5 genotypes were found lodged.

Diversity in quantitative traits

Table 3 shows the descriptive statistics; range, standard deviation, mean, and coefficient of variance percentage (CV %) of various quantitative parameters. Overall, significant variation was recorded for all the quantitative traits, the coefficient of variance ranged from 3.87 to 32.74%. The range for days to flower initiation was 36 to 40 days with mean of 38.37 ± 1.172 and CV% of 32.74, days to 50% flowering ranging from 39 to 46 days with mean of 42.52 ± 1.65 and CV% of 25.72%, maximum days to 50% flowering (46) was recorded for CB/6 2017, while minimum 39 days was recorded for genotypes William-82, days to flower completion ranged from 43.33 to 50 days with mean of 47.04 ± 1.59 and CV% of 29.62. Similarly, primary branches per plant ranged from 9.2 to 14.6 with mean of 12.19 ± 1.48 and CV% of 8.22, maximum branches/plant 14.6 was recorded for CB/33 2017, while minimum 9.2 branches was recorded for C/B 08 2017. The CV% for plant height was 5.40%, with mean value of 20.26±3.75 cm, where the tallest genotype was William-8 with the plant height of 27.27 cm, while minimum 14.07 cm was recorded for genotypes NARC-2, respectively. Furthermore, leaflet number ranged from 3 to 4.8 with CV% of 7.10, leaf length ranged from 8.68 to 130.8 cm with CV% of 11.70 and the CV% for leaf width was 6.66%. The number of pods per plant ranging from 29.8 to 117.2 pods with mean value of 68.26 ± 23.07 and CV% for the trait was 2.96%, maximum NP/P 117.2 was recorded for C/B-13, while minimum 29.8 was recorded for genotypes NARC-16. The seed per pod was ranged from 2.8 to 3.53 and CV% was 16.24%, maximum seed/pod (3.53) was recorded for NARC-16, while minimum 2.8 was recorded for genotypes Ramal-1. The pod length was ranged from 3.24 to 4.59 cm with mean value of 4.24 ± 0.28 and CV% of 15%. Similarly, the pod width ranged from 0.8 to 0.9 with CV% of 3.87, the days to maturity was ranged from 82.33 to 98 day with mean value of 91.44 ± 3.86 and CV% of 23.68, the maximum days to maturity, 98 days was recorded for MKD 96, while the lowest 39 days was recorded for genotypes NARC-16. The Y/P ranged from 14.5 to 49 g with mean value of 29.03 ± 7.49 g and CV% of 3.87, where highest yield (49 g) was recorded for C/B 28, while the lowest 14.56 g was observed for genotypes C/B 7. While the thousand seed weight ranged from 15.76 g to 21.96 with CV% of 12.88, where the genotype William-82 produced 21.96 g seed weight while C/B-6 have 15.76 g hundred seed was recorded.

Table 2. *Qualitative characters of 25 soybean genotypes used in the present study*

Genotypes	Flower color	Hilum color	Pod color	Seed size	Seed color	Lodging
NARC-2	White	Buff	Brown	Small	yellow	Upright
PSC-60	White	Buff	Brown	Small	Yellow green	Upright
NARC-16	Purple	Black	Redish brown	Large	Yellow green	Upright
William-82	White	Intermediate black	Brown	Small	yellow	Upright
Ramal-1	White	Buff	Brown	Large	Yellow green	Upright
SA-72=60	Purple	Black	Brown	Medium	Yellow green	Upright
Ajmari	White	Buff	Brown	Medium	Yellow	Upright
C/B 2017 14	Purple	Brown	Brown	Medium	Yellow	Upright
C/B 2017 32	Purple	Brown	Brown	Small	Yellow green	Upright
C/B 2017 31	White	Black	Brown	Large	Yellow green	Lodged
C/B 2107 6	Purple	Brown	Brown	Large	Yellow green	Upright
MKD 96	White	Brown	Brown	Large	yellow	Upright
C/B 2017-17	Purple	Buff	Brown	Medium	Yellow green	Upright
C/B 2017-12	Purple	Intermediate black	Brown	Large	Yellow green	Upright
C/B 2017-3	Purple	Brown	Redish brown	Small	Yellow	Lodged
C/B 2017-33	White	Black	Brown	Medium	Yellow	Upright
C/B 2017-1	White	Black	Brown	Large	Yellow	Upright
C/B 2017-13	White	Black	Brown	Large	Yellow green	Upright
C/B 2017-28	White	Black	Brown	Medium	Yellow green	Upright
Swat 84	White	Black	Brown	Medium	Yellow	Upright
C/B 2017-9	Purple	Brown	Brown	Large	Yellow	Upright
C/B 2017-7	Purple	Buff	Redish brown	Small	Yellow	Upright
C/B 2017-8	Purple	Brown	Brown	Medium	Yellow	Upright
C/B 2017-11	Purple	Intermediate black	Brown	Large	Yellow	Upright
Madyan	Purple	Black	Brown	Medium	Yellow green	Upright

Table 3. *Descriptive statistics of 25 soybean genotypes using 15 quantitative traits*

Traits	Mean + SD	Range	CV%
Days to flower initiation	38.37 ± 1.172	36-40	32.74
Days to Flowering (50%)	42.52 ± 1.65	39-46	25.72
Days to flower completion	47.04 ± 1.59	43.33-50	29.62
Primary branches/plant	12.19 ± 1.48	9.2-14.6	8.22
Plant height (cm)	20.26 ± 3.75	14.07-27.27	5.40
Leaflet number	3.57 ± 0.50	3-4.8	7.10
Leaf length (cm)	11.75 ± 1.00	8.68-13.08	11.70
Leaf width (cm)	6.49 ± 0.97	4.74-8.96	6.66
Number of pods/plant	68.26 ± 23.07	29.8-117.2	2.96
Seed/plant	3.05 ± 0.19	2.8-3.53	16.24
Pod length (cm)	4.24 ± 0.28	3.24-4.59	15.00
Pod width (cm)	0.88 ± 0.04	0.8-0.99	19.97
Days to maturity	91.44 ± 3.86	82.33-98	23.68
Yield/plant (g)	29.03 ± 7.49	14.56-49	3.87
Hundred seed weight (g)	18.91 ± 1.47	15.76-21.96	12.88

Correlation analysis

Correlation analysis was performed using different quantitative traits (*Table 4*). Days to 50% flowering was strongly correlated with DFI (r = 0.77**), DFC was strongly significant DFI (r = 0.78**) and DF 50% (r = 0.91**). DM was found significantly correlated with DFI (r = 0.62**) and DF 50% (r = 0.60**). Similarly primary branches per plant was found significantly correlated with days to 50% flowering (r = 0.36*), PH with PB/P (r = 0.36*). Furthermore, LW were found negatively correlated with the leaflet number (r = -0.35*). S/P were found negatively correlated with PH (r = -0.38*), leaflet number (r= - 0.22*). Pod length was found negatively correlated with DFI (r = -0.20*), DM (r = -0.32*) leaflet number (r = -0.38*) and number of branches per plant (r = -0.35*). PW were found negatively correlated with leaflet number (r = -0.26*) and also with seed per pods (-0.20). Yield per plant was significantly correlated with LW (r = 0.58**) and NP/P (r = 0.82**).

Principal component analysis

Principal component analysis showed significant divergence among the 25 genotypes of *soybean.* PCA was analysed on 15 quantitative characters. The eigenvalue of 1.48 accounted for 67.73% of the total variation (*Table 5*). The 1st PC accounted for 27.92% out of the total variations and they were associated with DFI (0.41), DF 50% (0.39), DFC (0.41) and DM (0.32), while PL (-0.16) and PW (-0.11) were associated negative. In PC2 the total variation was 44.93%, and it was found that the association of leaf width (0.47), pod width (0.33) and thousand seed weight (0.30) were found positive weight while days to maturity (-0.33), leaflet number (-0.30) were found negatively contribution. In PC3 the total variation was 56.68% and it was observed that contribution of plant height (0.63), pod width (0.32) and primary branches/plant (0.22) were recorded positively, while the seed/pod (-0.54) and leaf length (-0.21) were negatively weight.

Table 4. *Correlation analysis of 25 soybean genotypes using 15 quantitative traits*

Traits	DFI	DF50%	DFC	DM	PB/P	PH	LLN	LL	LW	NP/P	S/P	PL	PW	Y/P
DF50%	0.77**													
DFC	0.78**	0.91**												
DM	0.62**	0.60**	0.60**											
PB/P	0.25	0.36*	0.31	0.05										
PH	0.13	0.06	-0.14	0.05	0.36*									
LLN	-0.18	-0.12	-0.10	0.05	-0.16	0.15								
LL	0.03	-0.12	-0.09	-0.14	0.05	-0.03	-0.16							
LW	0.09	0.09	0.13	-0.11	0.43*	-0.02	-0.35*	0.15						
NP/P	0.49**	0.29	0.40*	0.36*	0.27	-0.04	-0.03	0.02	0.47**					
S/P	-0.12	0.05	0.14	-0.26	0.21	-0.35*	-0.21	0.25	0.04	-0.04				
PL	-0.19	-0.12	-0.15	-0.31	-0.07	0.13	-0.37*	0.11	0.01	-0.35*	0.12			
PW	-0.18	-0.17	-0.24	-0.45*	0.19	0.23	-0.19	0.07	0.40*	-0.03	-0.20	0.28		
Y/P	0.42*	0.34*	0.42*	0.22	0.32	-0.29	-0.11	0.17	0.58**	0.82**	0.20	-0.19	-0.08	
HSW	0.09	-0.15	-0.14	-0.11	0.23	0.17	-0.17	0.05	0.33*	0.37*	-0.01	-0.02	-0.12	0.36*

Note: DFI- Days to flower initiation, DF50%-Days to 50% flowering, DFC- Days to flower completion, DM-Days to maturity, PB/P- Primary branches/plant, PH-Plant height, LLN-Leaflet number, LL- Leaf length, LW- Leaf width, NP/P- Number of pod/plant, S/P- Seed/pod, PL- Pod length, PW- Pod width, Y/P-Yield/plant, HSW- Hundred seed weight

DOI: http://dx.doi.org/10.15666/aeer/2205_43794394

Table 5. *Principal component analysis of soybean through 15 quantitative traits*

AXIS	**PC1**	**PC2**	**PC3**	**PC4**
% of Variance	27.92	17.01	11.75	11.04
Cum. % of Variance	27.92	44.93	56.68	67.73
Eigenvalue	3.31	2.31	1.81	1.48
Days to flower initiation	0.41	-0.11	0.11	0.15
Days to 50% flowering	0.39	-0.16	0.01	0.32
Days to flower completion	0.41	-0.15	-0.09	0.25
Days to maturity	0.32	-0.33	0.087	-0.05
Primary branches/plant	0.22	0.26	0.229	0.15
Plant height	-0.02	0.02	0.63	0.12
Leaflet number	-0.07	-0.30	0.15	-0.38
Leaf length	-0.03	0.24	-0.21	0.05
Leaf width	0.18	0.47	0.08	-0.05
Number of pod/plant	0.35	0.17	0.05	-0.34
Seed/pod	0.02	0.17	-0.54	0.15
Pod length	-0.16	0.19	-0.01	0.47
Pod width	-0.11	0.33	0.32	0.23
Yield/plant	0.35	0.26	-0.17	-0.25
Thousand seed weight	0.08	0.30	0.12	-0.34

Similarly, in PC4 the total variation was 67.73%, and it was found that the contribution of days to 50% flowering (0.32), days to flower completion (0.25) pod length (0.47), pod width (0.23) were contributing positively, while leaflet number (-0.38), number of pod per plant (0.34) and thousand seed weight (-0.34) were contributing negatively in PC4. As the results of scatter plot defined by the two PCS, PC2 shows the variation of 44.93% and PC3 56.68% of the total variations to find out the genetic variation and geographical association among the genotypes. The separation reveled that all the genotypes were scattered in all the quarters, which shows the high level of genetic variation. Based on the scatterplot, the genotypes C/B-11, C/B-9C/B-17 were recorded early days to flowering, PSC-60, C/B-14 and C/B-8 were early days to flower completion. The genotype C/B 28 and C/B-1 were recorded maximum seed yield/plant (*Fig. 1*).

Cluster analysis

Ward’s method was used for cluster analysis to find out there relationship among the genotypes, and all the genotypes were divided into two groups. The cluster 1 consist of eleven genotypes, cluster two had one genotype, Cluster 3 having six genotypes. The 4th cluster consist of four genotypes. Cluster 5 have three genotypes (*Fig. 2*).

Mean performance based on cluster analysis

Mean performance for all the cluster were calculated and presented in table 3, where the cluster 3 genotypes were found early days to flowering (37.25 ± 1.25) and days to flower completion (41.75 ± 2.87), while cluster 4 genotypes were recorded late days to flower initiation (39.66 ± 0.57) and (44.33 ± 1.52). The cluster three genotypes were found early mature genotypes (89 ± 2.44) while cluster 2 genotypes were found late mature genotypes (93.16 ± 3.86). Similarly, the genotypes of cluster 4 were found to have maximum number of branches per plant (13.13 ± 1.17), whereas the cluster two genotypes were found the tallest genotypes (21.15 ± 3.86). Based on number of pods/plant and yield per plant, the cluster 4 genotypes were observed to have maximum number of pods (98.26 ± 32.10), and yield per plant (38.66 ± 6.85). Furthermore, for thousand seed weight the cluster 4 genotypes were also found to have maximum thousand seed weight (19.166 ± 3.13).

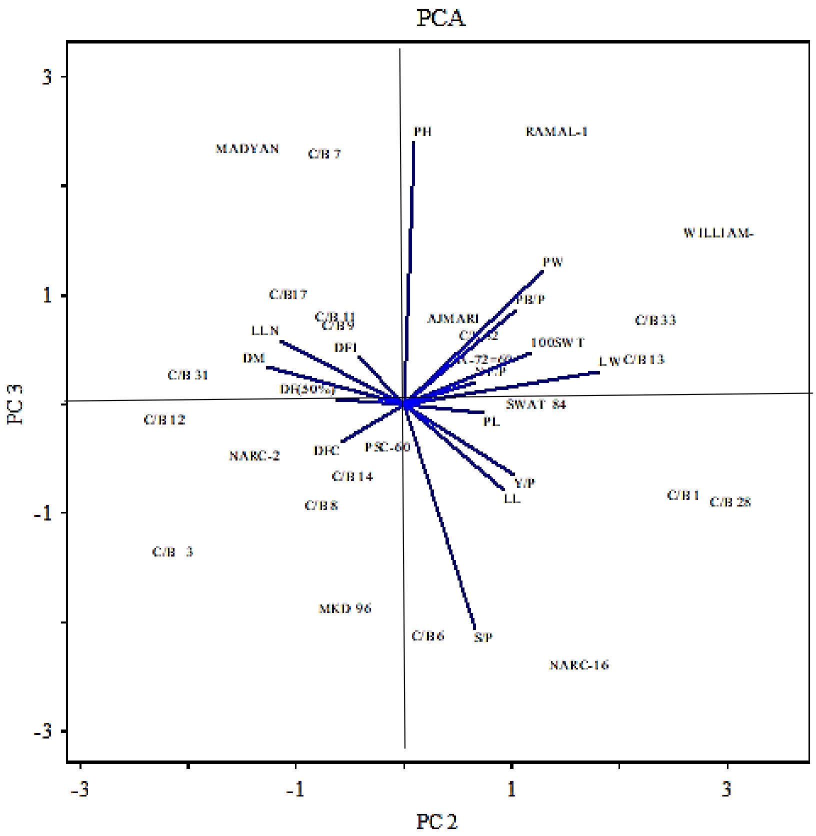

Figure 1. *Scatter plot diagram of PC2 and PC3 of different quantitative parameters of soybean genotypes*

Total seed storage proteins

During the study of total seed storage proteins, a total of 13 polypeptides bands were recorded with the molecular weight ranged from 15 to 170 kDa (*Fig. 3*). The highest genetic diversity (64%) was recorded for band 1, 7, 8 and 9, followed by band 2, band 8 and band 13 with the genetic diversity of (48%), whereas the lowest (32%) was observed for band 3, band 4 and band 12.

On the basis of total seed storage proteins, two-way cluster analysis was performed, and the genotypes were divided into two main groups at a 25% distance and further sub divided into six subgroups at 75% Euclidian distance, where the genotype NARC-2 and SWAT-84 were found the most diverse genotypes. The 1st group consist of 16 genotypes. The 2nd groups comprised of 3 genotypes. The 3rd group consist of 2 genotypes, the 4th group had 1 genotype, while the 5th groups consist of three genotypes. The last 6th group comprised of one genotype namely SWAT 84 (*Fig. 4*).

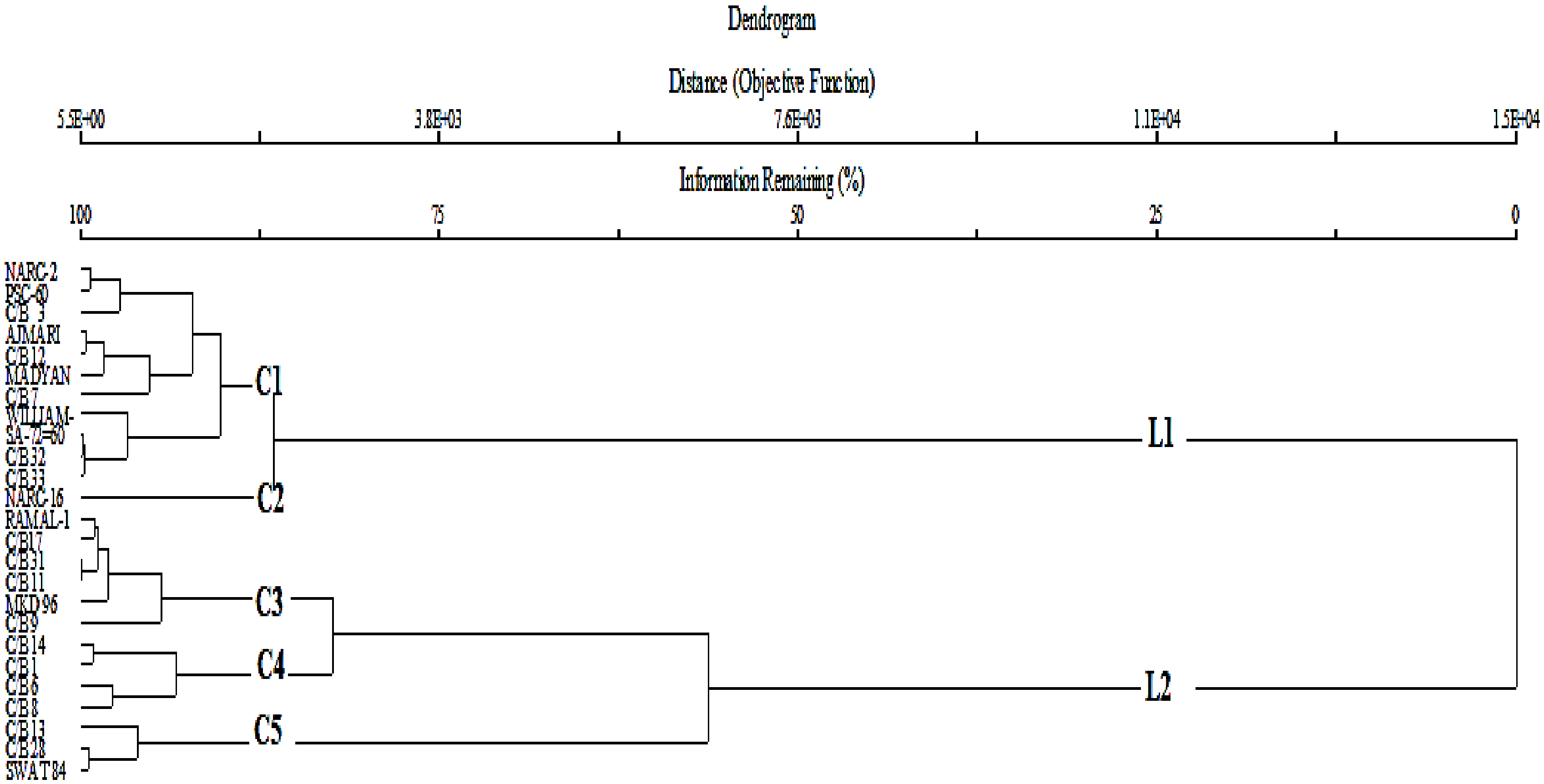

Figure 2. *Cluster dendrogram of 25 soybean genotypes based on 15 agro morphological parameters*

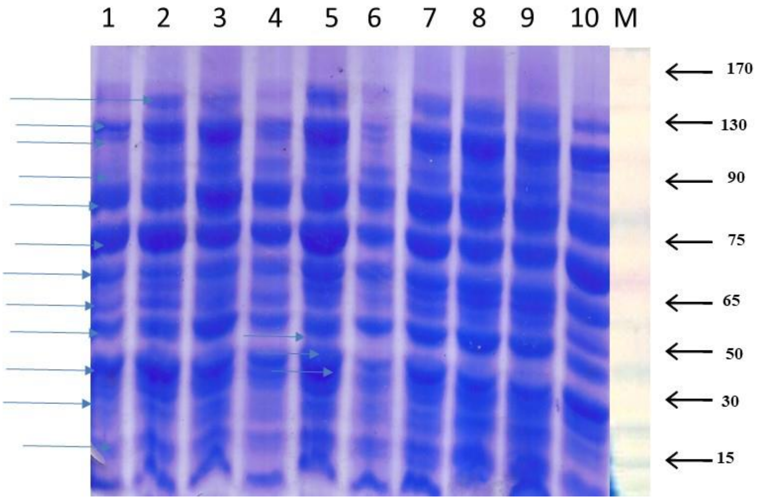

Figure 3. *Gel electrophoretic banding pattern of 25 soybean genotypes, numbering shows 1-NARC-2, 2-PSC-60, 3-NARC-16, 4-WILLIAM-82, 5-RAMAL-1, 6-SA-72=60, 7 AJMARI, 8-C/B 14, 9-C/B 32, 10-C/B 31, M shows Molecular marker*

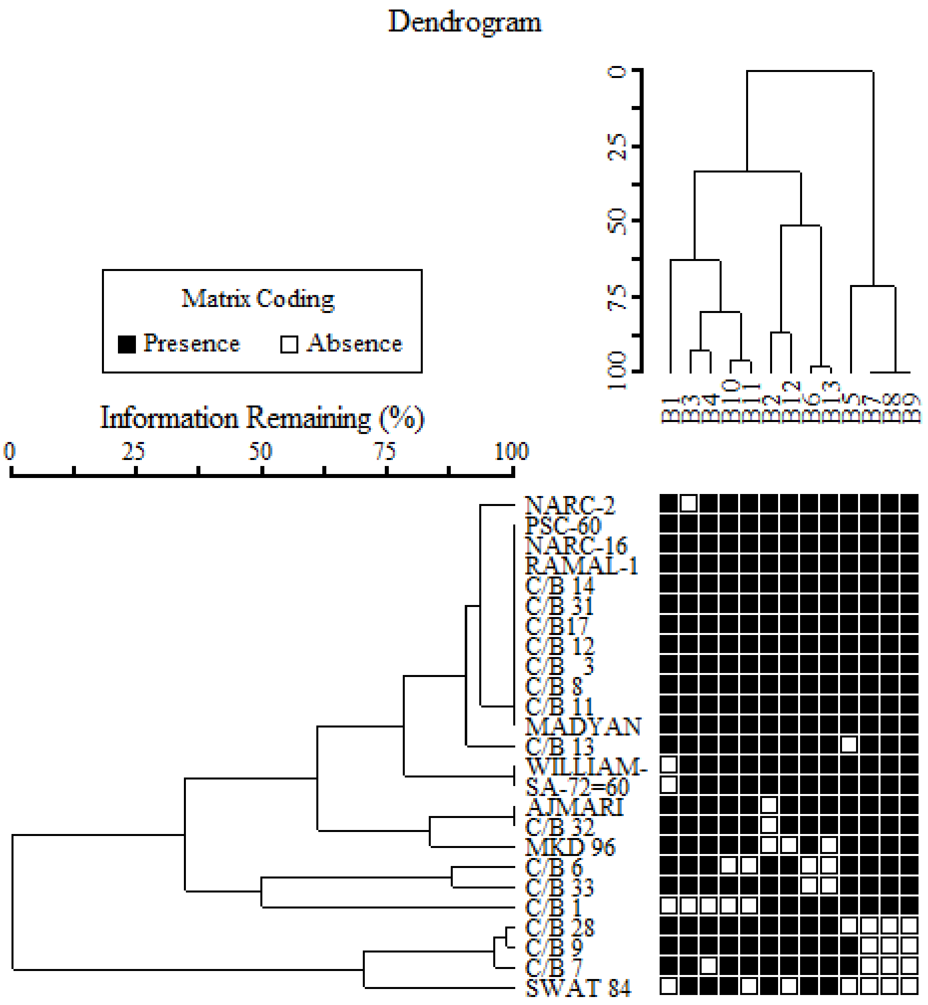

Figure 4. *Two-way cluster dendrogram of 25 soybean genotypes based on SDS-PAGE analysis*

Discussion

To develop any breeding programme, it is necessary to examine the available material to determine which genotypes are suitable for breeding. Agronomic characteristics have been exploited as a primary goal in the development of novel breeding materials. The determination of correlation helps in the decision-making process when it comes to selecting strategies. Diversity studies are important in breeding programmes since they help to improve a population's genetics. Plant breeders are very interested in genetic variation both within and between species of population crops (Ihsan et al., 2021). When analysing the genetic makeup of germplasm, an estimation of the degree of variation within and across populations of a species is helpful. The evaluation and classification of crop plant variability may be accomplished through the estimation of agro-morphological diversity (Tahir and Karim, 2011). Making plans for the evaluation, collection, and storage of various genotypes that are used in various parts of the world is therefore essential (Martins et al., 2006).

Soybean is one of the most important cultivated crops in the globe for its agro-economic importance and wide application in agriculture and industry. In the present study, a number of useful variations were found in different qualitative traits. Two types of flower color were found in the studied genotypes. The differences in flower color among genotypes are due to differences in gene expression. The present study work is in close agreement with Malik et al. (2006), where the contributors found purple and white colored flowers. Four distinct types of hilum color were noted in the studied genotypes, whereas two types of pod color were recorded. The present study agrees with Wang et al. (2002) and Malik et al. (2007), where the contributors found significant differences in seed color.

For quantitative traits, significant variation was found for most of the traits. The highest coefficient of variance was found for days to flower initiation, followed by days to flower completion, while the lowest was found in number of pods/plants and SY/P. Among the studied genotypes, three genotypes were found early flowering. The present investigation is comparable with the previous literature of Malik et al. (2007). Maturity played a significant role in the development of any crop. However, early maturity and high yielding genotypes are preferred in crops plants. In the present study, days to Maturity ranged from 82 to 98 days. The late mature genotype was MKD 96, which takes 98 days to reach the stage of maturity while NARC-16 genotype was found early mature genotype. The current research work is comparable with Arshad et al. (2006) and Ashraf et al. (2010), where the researchers noted a considerable variance in the number of days till soybean maturity. The development and expansion of the plant, which has a direct impact on the yield, is greatly influenced by the number of branches per plant. The yield of the fodder crop is directly impacted by the amount of branches/plant. Maximum branches per plant was recorded for CB/33 2017. The plant height ranged from 14.07 to 27.27 cm, where the tallest genotype was William-8, while the dwarf genotype was NARC-2. Similarly, Akande et al. (2007), and Iqbal et al. (2008), also observed significant divergence in plant height among *G. max* genotypes.

The number of pods/plants plays a vital role in soybean production. In the present study, a significant variation was recorded for pods/plant. The pod length ranged from 3.24 to 4.59 cm. The present results are comparable with Doulat Baig et al. (2018) where the contributors found significant diversity in pods per plant. The study is comparable with that of Ashraf et al. (2010) and Doulat Baig et al. (2018) where the contributors found significant diversity in soybean genotypes. Seed yield is one of the most important traits as seed is one of the most edible and nutritious parts of the plant. All field crops' goals and products are seed yield. The seed yield ranged from 14.56 g to 49 g. Maximum yield/plant was recorded for C/B 28, and such genotypes may be used in future breeding practices. The present study is comparable with that of Storck et al. (2008) and Iqbal et al. (2010), where the authors found significant variations among soybean genotypes.

Correlation analysis is one of the most important statistical package for understanding the relationship between different traits and for the selection of most important characters. Determination of correlation coefficients between various characters helps to obtain best combinations of attributes of a crop for obtaining higher return per unit area (Kumar et al., 2015). Among the studied genotypes, significant level of correlation was found for days to flowering 50% with days to flower initiation, days to flower completion with days to flower initiation and days to flowering 50%. Days to maturity was correlated with days to flower initiation. Furthermore, leaf width was found negatively correlated with the leaflet number. Seed per pod were found negatively correlated with plant height, leaflet number. Y/P was significantly correlated with leaf width and number of pod/plants. The present study is comparable with that of Chamundeswari and Aher (2003), Malik et al. (2011), and Mukhekar et al. (2004), where the authors found significant correlation for different parameters.

PCA is a powerful technique for data reduction which removes interrelationships among components. Results reported by various researchers showed multivariate analysis to be a valid system to deal with germplasm collection. Kumar et al. (2015) conducted average linkage cluster and principal component analyses and reported the utility of these results in preservation and utilization of germplasm. In the present study, 4 Principal component analysis showed significant divergence among the 25 genotypes of *soybean* with eigenvalue of 1.48 accounted for 67.73% of the total variation. Based on cluster analysis, Ward’s method was used to find out their relationship among the genotypes, where all the genotypes were divided into two main lineages and further sub divided into 5 sub-groups. The grouping pattern of the genotypes showed no parallelism between genetic divergence and geographical distribution of the genotypes (Das et al., 2000). Mean performance for all the cluster were calculated, where the cluster 3 genotypes were found early days to flowering and days to flower completion. The cluster 4 genotypes was found in maximum number of pods and yield per plant. These genotypes from cluster analysis could be utilize directly or included in hybridization program for development of lines suitable for such region. Soybean breeders can use the genetic patterns discovered in this study of soybean germplasm to make better decisions from the enormous number of accessions available. The clustering pattern showed no obvious association between genetic diversity and spatial diversity (Sihag et al., 2004). One cluster had genotypes from the same eco-geographic region, and another contained genotypes from separate eco-geographic regions.

Biochemical analysis (SDS-PAGE) contributes significantly to the assessment of genetic diversity. Biochemical analysis has become more common in recent decades for finding genetic diversities among plant germplasms, and it is more helpful than traditional morphological analysis (Ihsan et al., 2021). SDS-PAGE is one of the most useful procedures for protein profiling, and it is commonly used to explore genetic differences among various crops (Ghafoor and Ahmad, 2005). In the present study SDS-PAGE analysis was carried out and a total of 13 polymorphic bands were recorded, ranged from 15 to 170 kDa. Based on total seed storage proteins, two-way cluster analysis was performed where the genotypes were split into two groups and six subgroups, where the genotype NARC-2 and SWAT-84 were found the most diverse genotypes. Because total seed proteins are essentially independent of abiotic factors, SDS-PAGE provides an easy approach to find distinct species. Moreover, many researchers have consistently evaluated soybean germplasm using total seed protein markers (Patil et al., 2017). SDS-PAGE has been proposed as a reliable method to assess soybean cultivars because biochemical markers are highly polymorphic and do not affect electrophoretic patterns by environmental factors (Shahnejat-Bushehri et al., 2000; Shah et al., 2011). Using the use of SDS-PAGE, Iqbal et al. (2008) evaluated 139 accessions from various sources and reported 17 bands. According to the study of Malik et al. (2009), the contributors found a total of 26 polypeptide bands in soybean genotypes. Shahnejat-Bushehri et al. (2000), used polyacrylamide gel electrophoresis to study twenty-one soybean cultivars and observed that seed storage protein banding patterns grouped soybean cultivars into four separate groupings.

Conclusions

For any breeding programme, a genetic diversity estimation is important. To enhance the genetics of the plant, plant breeders employ the diversity included in a vast number of soybean germplasm samples originating from different agro-ecologies. To develop an improved variety of particular germplasm, variations in morphological and biochemical parameters, which are the fundamental criteria for successful breeders, are examined. Soybean production in Pakistan is low, and it is planted only in marginal fields for domestic consumption. The present study was conducted to analyze soybean genotypes to select high yielding and early mature genotypes for use of soybean crop development programs. The genotypes *C/B 2017 28, SWAT 84, C/B 2007, MKD 96,* C/B 2017 1, *C/B 2017 32* and *C/B 2017 6* were found the high yielding genotypes and *NARC-16, WILLIAM-82, SA-72=60*, and *C/B 2017 14* were found early mature genotypes and recommended for future breeding. These high yielding and adaptable Soybean genotypes will offer new opportunities to small farmers whose land remain non-cultivated after wheat crop in the Kharif season.

Funding. The authors greatly acknowledge and express their gratitude to the Researchers Supporting Project number (RSP2024R462), King Saud University, Riyadh, Saudi Arabia.

Acknowledgments. The authors would also like to thank the Director, of Bio-Resources Conservation Institute (BCI), National Agriculture Research Centre, Islamabad for providing the seeds and facilitate to conduct of the present research work.

Conflicts of Interest. The authors declare that they have no conflict of interest.

REFERENCES

[1] Akande, S. R., Owolade, O. F., Ayanwole, J. A. (2007): Field evaluation of soybean varieties at Ilorin in the southern guinea savanna ecology of Nigeria. – African Journal of Agricultural Research 2(8): 356-359.

[2] Arshad, M., Ali, N., Ghafoor, A. (2006): Character correlation and path coefficient in soybean *Glycine max* (L.) Merrill. – Pakistan Journal of Botany 38(1): 121-130.

[3] Ashraf, M., Iqbal, Z., Arshad, M., Waheed, A., Glufran, M. A., Chaudhry, Z., Baig, D. (2010): Multi-environment response in seed yield of soybean [*Glycine max* (L.) Merrill], genotypes through GGE biplot technique. – Pak. J. Bot 42(6): 3899-3905.

[4] Balkaya, A., Ergün, A. (2008): Diversity and use of pinto bean (Phaseolus vulgaris) populations from Samsun, Turkey. – New Zealand Journal of Crop and Horticultural Science 36(3): 189-197.

[5] Bretting, P., Widrlechner, M. P. (2010): Genetic markers and plant genetic resource management. – Plant Breed. Rev. 13: 11.

[6] Chamundeswari, N., Aher, R. P. (2003). Character association and component analysis in soybean [*Glycine max* (L.) Merrill]. – Ann. Biol. 19(2): 199-203.

[7] Das, S., Mukherjee, K. K. (1995): Comparative study on seed proteins of Ipomoea. – Seed Science and Technology 23(2): 501509.

[8] Das, S. P., Harer, P. N., Biradar, A. B. (2000): Genetic divergence and selection of genotypes in soybean. – Journal of Maharashtra Agricultural Universities 25(3): 250-252.

[9] Baig, D., Khurshid, H., Arshad, M., Jan, S. A., Khan, M. A., Nawaz, N. (2018): Evaluation of soybean genotypes for yield and other economically important traits under rainfed condition. – PAB Pure and Applied Biology 7(1).

[10] Ghafoor, A., Zahoor, A., Qureshi, A. S., Bashir, M. (2002): Genetic relationship in *Vigna mungo* (L.) Hepper and *V. radiate* (L.) R. Wilczek based on morphological traits and SDS-PAGE. – Euphytica 123: 367-2002.

[11] Ghafoor, A., Ahmad, Z. (2005): Diversity of agronomic traits and total seed protein in black gram *Vigna mungo* (L.) Hepper. – Acta Biologica Cracoviensia Series Botanica 47(2): 69-75.

[12] Hale, B. M. (2020): Induction and Characterization of Microspore Embryogenesis in Soybean (*Glycine Max* [L.] Merrill). – Arkansas State University.

[13] Hartman, G. L., West, E. D., Herman, T. K. (2011): Crops that feed the World 2. Soybean worldwide production, use, and constraints caused by pathogens and pests. – Food Security 3: 5-17.

[14] Ihsan, M., Khan, A., Nazir, N., Nisar, M., Jan, T., Ullah, S., Aziz, T., Al-Asmari, F., Alshareef, S. A., Aljabri, M., Sameeh, M. Y. (2024). Evaluation of the durum wheat landrace genetic diversity using agro analysis and its benefit for human health. Italian Journal of Food Science 36(1): 15-27.

[15] Ihsan, M., Nazir, N., Ghafoor, A., Khalil, A. A. K., Zahoor, M., Nisar, M., Khames, A., Ullah, R., Shah, A. B. (2021): Genetic diversity in local and exotic *Avena sativa* L. (Oat) germplasm using multivariate analysis. – Agronomy 11(9): 1713.

[16] Iqbal, Z., Arshad, M., Ashraf, M., Mahmood, T., Waheed, A. (2008): Evaluation of soybean [*Glycine max* (L.) Merrill] germplasm for some important morphological traits using multivariate analysis. – Pakistan Journal of Botany 40(6): 2323-2328.

[17] Jan, S. A., Bibi, N., Shinwari, Z. K., Rabbani, M. A., Ullah, S., Qadir, A., Khan, N. (2017): Impact of salt, drought, heat and frost stresses on morphobiochemical and physiological properties of Brassica species: an updated review. – JPAA 2(1).

[18] Jha, S. S., Ohri, D. (1996): Phylogenetic relationships of *Cajanus cajan* (L.) Millsp. (pigeonpea) and its wild relatives based on seed protein profiles. – Genetic Resources and Crop Evolution 43: 275-281.

[19] 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 Biol Med 2(1): 189-191.

[20] Kumar, A., Pandey, A., Aochen, C., Pattanayak, A. (2015): Evaluation of genetic diversity and interrelationships of agro-morphological characters in soybean (*Glycine max*) genotypes. – Proceedings of the National Academy of Sciences, India Section B: Biological Sciences 85: 397-405.

[21] Malik, M. F. A., 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(2): 405-413.

[22] Malik, M. F. A., Qureshi, A. S., Ashraf, M., Khan, M. R., Javed, A. (2009): Evaluation of genetic diversity in soybean (*Glycine max*) lines using seed protein electrophoresis. – Australian Journal of Crop Science 3(2): 107-112.

[23] Malik, M. F. A., Ashraf, M., Qureshi, A. S., Khan, M. R. (2011): Investigation and comparison of some morphological traits of the soybean populations using cluster analysis. – Pakistan Journal of Botany 43(2): 1249-1255.

[24] Martins, S. R., Vences, F. J., de Miera, L. S., Barroso, M. R., Carnide, V. (2006): RAPD analysis of genetic diversity among and within Portuguese landraces of common white bean (*Phaseolus vulgaris* L.). – Scientia Horticulturae 108(2): 133-142.

[25] MINFAL. (2009): Ministry of Food, Agriculture and Livestock, Government of Pakistan.

[26] Mirpoor, S. F., Giosafatto, C. V. L., Porta, R. (2021): Biorefining of seed oil cakes as industrial co-streams for production of innovative bioplastics. A review. – Trends in Food Science & Technology 109: 259-270.

[27] Mukhekar, G. D., Bangar, N. D., Lad, D. B. (2004). Character association and path coefficient analysis in soybean [Glycine max (L.) Merrill]. – J. Mah. Agril. Univ. 29(3): 256-258.

[28] Nisar, M., Ghafoor, A., Khan, M. R., Subhan, M. (2009): Genetic similarity of Pakistani pea (*Pisum sativum* L.) germplasm with world collection using cluster analysis and Jaccard's similarity index. – Journal of the Chemical Society of Pakistan 31(1): 138-144.

[29] Ojo, D. K., Ajayi, A. O., Oduwaye, O. A. (2012): Genetic relationships among soybean accessions based on morphological and RAPDs techniques. – Pertanika Journal of Tropical Agricultural Science 35(2): 237-248.

[30] Patil, G., Mian, R., Vuong, T., Pantalone, V., Song, Q., Chen, P., Shannon, G. J., Carter, T. C., Nguyen, H. T. (2017): Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. – Theoretical and Applied Genetics 130: 1975-1991.

[31] Qadir, A., Ali, N., Jan, S. A., Rabbani, M. A., Shaheen, A., Khurshid, H. (2017): Estimating phenotypicdiversity in diverse Fenugreek (*Trigonella foenum-graceum* L.) population through morphometric analysis. – International. Journal of Biosciences 11(1): 274-283.

[32] Rahman, S. U., McCoy, E., Raza, G., Ali, Z., Mansoor, S., Amin, I. (2023): Improvement of soybean; A way forward transition from genetic engineering to new plant breeding technologies. – Molecular Biotechnology 65(2): 162-180.

[33] Shah, S. M. A., Rahman, H., Abbasi, F. M., Rabbani, M. A., Khan, I. A., Shinwari, Z. K., Shah, Z. (2011): Inter specific variation of total seed protein in wild rice germplasm using SDS-PAGE. – Pak. J. Bot 43(4): 2147-2152.

[34] Shahnejat-Bushehri, A. A., Abd-Mishani, C., Yazdi-Samadi, B., Sayed-Tabatabaei, B. E. (2000): Variety-specific electrophoretic profiles of soybean cultivars. – Iranian Journal of Agricultural Sciences 31(1): 55-61.

[35] Sihag, R., Hooda, J. S., Vashishtha, R. D., Malik, R. P. S. (2004): Genetic divergence in soybean (*Glycine max* (L.) Merrill). – Annals of Biology 20(1):17-21.

[36] Smith, S. E., Guarino, L., Al Doss, A., Conta, D. M. (1995): Morphological and agronomic affinities among Middle Eastern alfalfas accessions from Oman and Yemen. – Crop Sci 35: 1188-1194.

[37] Storck, L., Cargnelutti Filho, A., Lúcio, A. D. C., Missio, E. L., Rubin, S. D. A. L. (2010): Experimental precision evaluation of soybean cultivar yield trials. – Ciência e Agrotecnologia 34: 572-578.

[38] Tahir, N. A. R., Karim, H. F. H. (2011): Determination of genetic relationship among some varieties of chickpea (*Cicer arietinum* L.) in Sulaimani by RAPD and ISSR markers. – Jordan Journal of Biological Sciences 4(2): 77-86.

[39] Wang, K., Li, F., Norishiko, K., Yoshihito, T. (2002): Investigation and comparison of some characters of the natural wild soybean populations collected from China's Hebei Province and northeastern Japan. – Chinese Journal of Oil Crop Sciences 22(4): 17-22.