

ASSESSMENT AND CHARACTERIZATION OF GENETIC DIVERSITY IN LENTIL (*LENS CULINARIS* MEDIK.) THROUGH K-MEANS CLUSTERING

SYED ATIQ HUSSAIN AND MUHAMMAD SAJJAD IQBAL*

*Biodiversity Informatics, Genomics and Post Harvest Laboratory,
Department of Botany, University of Gujrat, Gujrat, Pakistan*

Keyword: Lentil, Agronomic traits, K-means clustering, Class centroid

Abstract

Lentil is a major legume crop that has much potential to combat malnutrition and food security but its yield is not sufficient due to less number of varieties available for diverse ecologies. To explore new generic lines, 220 genotypes of Pakistan, USA, Syria and unknown origin were characterized for agronomic traits *viz.*, seed yield, harvest index, 100-seed weight, biological yield, plant height, lower pod height, pod size, seeds per pod, days to maturity, cooking time, number of hard seed and number of pods. K-means clustering revealed various component groups both in 2018 and 2019 with 5 clusters each. PC1 contributed 92.4% of variability while PC2 exhibited 4.1% for 2018, during 2019 PC1 90.53% and PC2 as 4%. Distance between class centroids exhibited maximum divergence group between clusters. Genotypes 6084 and 6122 (Pakistan) and 6062 (Syria) have been recommended for large scale cultivation on farmers field and breeding programs.

Introduction

Lentil (*Lens culinaris* Medik), is self-pollinating diploid ($2n = 14$), Rabi season pulse crop belonging to Fabaceae (Gupta *et al.* 2013). It is a rich source of protein (21-31%) along with complex carbohydrates (62-69%), dietary fibers (5-20%), vitamins, and minerals (Joshi *et al.* 2017, Romano *et al.* 2021) and has very little fat and fiber which play a significant role to overcome the malnutrition issues and deficiency of the micronutrients in developing regions (Thavarajah and Thavarajah 2012). It is also the best source of iron, as it helps in preventing iron deficiency (Sarker *et al.* 2018). It has a low glycemic level so physicians strongly suggested for the patients of diabetes, obesity and heart disorders (Ganesan and Xu 2017). Pakistan stands at 18th position in the world but its yield is still stagnant and needs improvement. Annual production of lentil in Pakistan is very low as compared to the neighboring countries including India, Bangladesh and Turkey. The annual lentil production of Turkey was 1.36 tons/ha, Bangladesh was 1.14 tons/ha, and India was 0.73 tons/ha in 2018. The top ten lentil producing countries in 2018 were Canada, Turkey India, USA, Kazakhstan, Nepal, Australia, Russian Federation, Bangladesh and China. The annual global yield of the lentil in 2018 was 6333,352 tons produced on about 6100,922 ha (0.96 tons/ha). In Pakistan, the annual yield of lentil was 6352 tons on 13,632 ha which was about 0.47 tons/ha in 2018. The annual yield of lentil in Syria 71,000 tons on 111,000 ha which was about 0.63 tons/ha in 2018. The USA accounts for 42.7 % of the global production of lentil in 2018 (FAOSTAT 2019).

In Pakistan, lentil production and yield are low because there is not new crop improvement programs or a way to get seeds to farmers (Ullah *et al.* 2020). Similarly available lentil germplasm have not been completely evaluated for yield and yield related traits in Pakistan (Nourin *et al.* 2019). Thus, for more efficient utilization of germplasm, the evaluation and classification of local lentils' local genetic resources is the primary step before starting any improvement program (Tripathi *et al.* 2021). Thus, there is a gap between evaluation and classification of germplasm based on agro-morphological variability to select best parents for crop improvement programs. In past, Bilal *et al.* (2018) made struggle to assort 197 genotypes through

*Author for correspondence: <drsajjad.iqbal@uog.edu.pk>.

K-means clustering. Various methods are existed for classification including hierarchical and K means clustering. K means clustering was introduced by McQueen in 1967 (Ashabi *et al.* 2020). In K means clustering an object may be assigned to a class during one iteration then change class in the following iteration, which is not possible with agglomerative hierarchical clustering for which assignment is irreversible (Sisodia *et al.* 2012). Hence an attempt was taken to assessment and characterize diverse lentil genotypes by using K means clustering approaches.

Materials and Methods

Two hundred and twenty lentil genotypes including two control groups (Markz 2009 and Punjab 2009) were acquired in 2017 from National Gene Bank, Plant Genetic Resource Institute, National Agriculture Research Center (NARC), Islamabad, Pakistan (Table 1). Collection was representation of USA, Syria, Pakistan and unknown origin. Agro-morphological traits were assessed under field conditions of NARC, Islamabad, Pakistan located at 33.6701°N latitude North and 73.1261°E longitude East, during two consecutive years (2018 and 2019). Experiments were laid out in an augmented design according to Bilal *et al.* (2018). In each year, genotypes were planted on a line of 5 m length, with 10 cm space between plants and 30 cm line (rows) spacing. Two approved cultivars, *viz.*, Markaz 2009 and Punjab 2009, were planted as a check (control) after every twenty lines. Recommended cultural practices were followed throughout crop season.

Ten healthy plants were selected randomly to record data on per plant basis from each genotype. Before harvesting plant height (PH), number of pods per peduncle (NP), lower pod height (LPH), pod size (PS), number of seeds per pod (NSP), days to maturity (DM) were measured. After harvesting biological yield (BY), seed yield (SY), hundred seed weight (HSW), cooking time (CT) and number of hard seed per 100-seed (HS) were counted.

K mean cluster analysis by Wilks lambda was carried out through XLStat 2018 and for visualization of K mean cluster 'kmean' function in R, and ggplot2. VisualizationR package "ggplot2" is used for K-means clustering. Wilks' lambda is a test statistics used in multivariate analysis of variance (MANOVA) to test whether there are differences between the means of identified groups of subjects on a combination of dependent variables.

Results and Discussion

K-means clustering was performed for 220 lentil genotypes based on Wilks's lambda method. K-means clustering partitioned 220 genotypes in five different clusters for agronomic data of 2018 (Fig. 1A). PC1 contributed 92.4% of variability while PC2 exhibited 4.1% for characters studied during 2018. Variances between five classes were observed as cluster 1 (252.90), Cluster-2 (179.87), Cluster-3 (217.46), Cluster-4 (340.71) and cluster 5 (1289.13) (Table 2a). Cluster 1, Cluster-2, Cluster- 3, Cluster-4 and Cluster-5 contained 19, 62, 56, 73 and 10 genotypes. Cluster 4 (73) consisted of highest number of genotypes representing Pakistan (67), USA (5) and Syria (1), hence it is assumed that it is exhibiting potential as parent material for future studies, while cluster 1 have maximum number of USA (10) genotypes during 2018 (Hussain *et al.* 2022). Cluster1 consisted of those genotypes that produce lowest seed yield in both years while the Cluster-5 had genotypes with best seed yield (71-85g), biological yield (135-26g), pod size (0.48-0.61cm) and cooking time (13-15 mint) during both years. Generally, genotypes within a group showed little divergence from each other than from genotypes of different groups, as observed by Sharma *et al.* (2022). Hybridization among genotypes of same group may not be fruitful; therefore inter group crossing or selection for outperforming genotypes have been suggested.

Table 1. List of 220 diverse lentil genotypes (germplasm) with their origin evaluated during 2018 and 2019.

G. No.	Source	G. No.	Source	G. No.	Source	G. No.	Source	G. No.	Source
5472	Pakistan (Sialkot)	5511	Pakistan (Muzaffargarh)	5571	Pakistan (Jacobabad)	5636	Pakistan (Layyah)	5669	Pakistan (Layyah)
5474	Pakistan (Gujranwala)	5512	Pakistan (Muzaffargarh)	5575	Pakistan (Ghotki)	5637	Pakistan (Hyderabad)	5670	Pakistan (Layyah)
5475	Pakistan (Gujranwala)	5517	Pakistan (Muzaffargarh)	5576	Pakistan (Ghotki)	5638	Pakistan (Hyderabad)	5671	Pakistan (Layyah)
5476	Pakistan (Gujranwala)	5518	Pakistan (Muzaffargarh)	5580	Pakistan (Muzaffargarh)	5639	Pakistan (Hyderabad)	5672	Pakistan (Layyah)
5477	Pakistan (Gujranwala)	5527	Pakistan (Sialkot)	5581	Pakistan (Multan)	5640	Pakistan (Hyderabad)	5673	Pakistan (Narawal)
5478	Pakistan (Sheikhupura)	5529	Pakistan (Kasur)	5583	Pakistan (Muzaffargarh)	5643	Pakistan (Thatta)	5677	Pakistan (Narawal)
5479	Pakistan (Lahore)	5530	Pakistan (Muzaffargarh)	5584	Pakistan (Muzaffargarh)	5647	Pakistan (Thatta)	5677	Pakistan (Narawal)
5480	Pakistan (Kasur)	5531	Pakistan (Muzaffargarh)	5590	Pakistan (Muzaffargarh)	5650	Pakistan (Thatta)	5679	Pakistan (Narawal)
5481	Pakistan (Kasur)	5532	Pakistan (Jhang)	5593	Pakistan (Nawabshah)	5652	Pakistan (Hyderabad)	5680	Pakistan (Narawal)
5482	Pakistan (Lahore)	5535	Pakistan (Jhang)	5595	Pakistan (Kharan)	5653	Pakistan (Layyah)	5683	Pakistan (Narawal)
5483	Pakistan (Sheikhupura)	5537	Pakistan (Gujrat)	5598	Pakistan (Kharan)	5654	Pakistan (Layyah)	5684	Pakistan (Narawal)
5484	Pakistan (Sheikhupura)	5538	Pakistan (Thatta)	5600	Pakistan (Punjab)	5655	Pakistan (Bhakkar)	5685	Pakistan (Narawal)
5485	Pakistan (Kasur)	5549	Pakistan (Badin)	5610	Pakistan (Khuzdar)	5656	Pakistan (Bhakkar)	5686	Pakistan (Narawal)
5486	Pakistan (Kasur)	5550	Pakistan (Hyderabad)	5621	Pakistan (Faisalabad)	5657	Pakistan (Bhakkar)	5687	Pakistan (Narawal)
5488	Pakistan (Okara)	5553	Pakistan (Hyderabad)	5622	Pakistan (Faisalabad)	5658	Pakistan (Bhakkar)	5688	Pakistan (Sialkot)
5489	Pakistan (Okara)	5555	Pakistan (Hyderabad)	5623	Pakistan (Faisalabad)	5659	Pakistan (Bhakkar)	5689	Pakistan (Sialkot)
5491	Pakistan (Sahiwal)	5556	Pakistan (Hyderabad)	5624	Pakistan (Rawalpindi)	5660	Pakistan (Bhakkar)	5689	Pakistan (Sialkot)
5493	Pakistan (Sahiwal)	5561	Pakistan (Sanghar)	5625	Pakistan (Muzaffargarh)	5661	Pakistan (Bhakkar)	5690	Pakistan (Sialkot)
5494	Pakistan (Okara)	5562	Pakistan (Sanghar)	5626	Pakistan (Muzaffargarh)	5664	Pakistan (Layyah)	5691	Pakistan (Sialkot)
5500	Pakistan (Khanewal)	5563	Pakistan (Sanghar)	5628	Pakistan (Rajapur)	5665	Pakistan (Layyah)	5692	Pakistan (Sialkot)
5501	Pakistan (Khanewal)	5564	Pakistan (Sanghar)	5630	Pakistan (Rajapur)	5666	Pakistan (Layyah)	5693	Pakistan (Sialkot)
5506	Pakistan (Muzaffargarh)	5565	Pakistan (Hyderabad)	5634	Pakistan (Layyah)	5667	Pakistan (Layyah)	5694	Pakistan (Sialkot)
5510	Pakistan (Muzaffargarh)	5570	Pakistan (Larkana)	5635	Pakistan (Larkana)	5668	Pakistan (Layyah)	5695	Pakistan (Sialkot)
5696	Pakistan (Narawal)	5751	Pakistan (Jhelum)	6015	USA	6075	Pakistan (Bahawalnagar)	6123	Pakistan (Sheikhupura)
5698	Pakistan (Sialkot)	5753	Pakistan (Rawalpindi)	6017	USA	6076	Pakistan (Bahawalnagar)	6124	Pakistan (Sheikhupura)

(Contd.)

(Contd.)

G. No.	Source	G. No.	Source	G. No.	Source	G. No.	Source	G. No.	Source
5700	Pakistan (Sialkot)	5772	Pakistan (Panjgur)	6037	Syria	6077	Pakistan (Bahawalnagar)	6125	Pakistan (Narawal)
5712	Pakistan (Narawal)	5773	Pakistan (Kharan)	6038	Syria	6078	Pakistan (Khaneval)	23776	Pakistan
5716	Pakistan (Narawal)	5856	USA	6041	Syria	6080	Pakistan (Narawal)	23777	Pakistan
5717	Pakistan (Gujranwala)	5861	USA	6042	Syria	6081	Pakistan (Narawal)	23779	Pakistan
5723	Pakistan (Layyah)	5979	USA	6043	Syria	6082	Pakistan (Narawal)	23781	Pakistan
5724	Pakistan (Layyah)	5981	USA	6045	Syria	6083	Pakistan (Narawal)	23787	Unknown
5726	Pakistan (Layyah)	5982	USA	6046	Syria	6084	Pakistan (Narawal)	24783	Unknown
5727	Pakistan (Layyah)	5988	USA	6047	Syria	6085	Pakistan (Narawal)	24784	Unknown
5729	Pakistan (Layyah)	5993	USA	5856	USA	6086	Pakistan (Narawal)	24785	Unknown
5730	Pakistan (Layyah)	5995	USA	6052	Syria	6087	Pakistan (Narawal)	24786	Unknown
5737	Pakistan (Chakwal)	5996	USA	6054	Syria	6089	Pakistan (Narawal)	24787	Unknown
5739	Pakistan (Muzaffargarh)	5999	USA	6058	Syria	6090	Pakistan (Narawal)		
5741	Pakistan (Bahawalpur)	6002	USA	6060	Syria	6092	Pakistan (Rawal pindi)		
5742	Pakistan (Bahawalpur)	6003	USA	6062	Syria	6093	Pakistan (Rawal pindi)		
5744	Pakistan (Faisalabad)	6005	USA	6064	Syria	6097	Pakistan (Chakwal)		
5745	Pakistan (Faisalabad)	6008	USA	6066	Pakistan (Bahawalpur)	6099	Pakistan (Rawal pindi)		
5746	Pakistan (Lahore)	6010	USA	6067	Pakistan (Bahawalpur)	6101	Pakistan (Rawal pindi)		
5747	Pakistan (Kasur)	6011	USA	6068	Pakistan (Bahawalpur)	6104	Pakistan (Rawal pindi)		
5748	Pakistan (Gujranwala)	6012	USA	6069	Pakistan (Bahawalpur)	6114	Pakistan (Bahawalnagar)		
5749	Pakistan (Gujranwala)	6013	USA	6073	Pakistan (Bahawalnagar)	6116	Pakistan (Bahawalnagar)		
5750	Pakistan (Jhelum)	6014	USA	6074	Pakistan (Bahawalnagar)	6122	Pakistan (Sheikhupura)		

G. No., represents genotype number

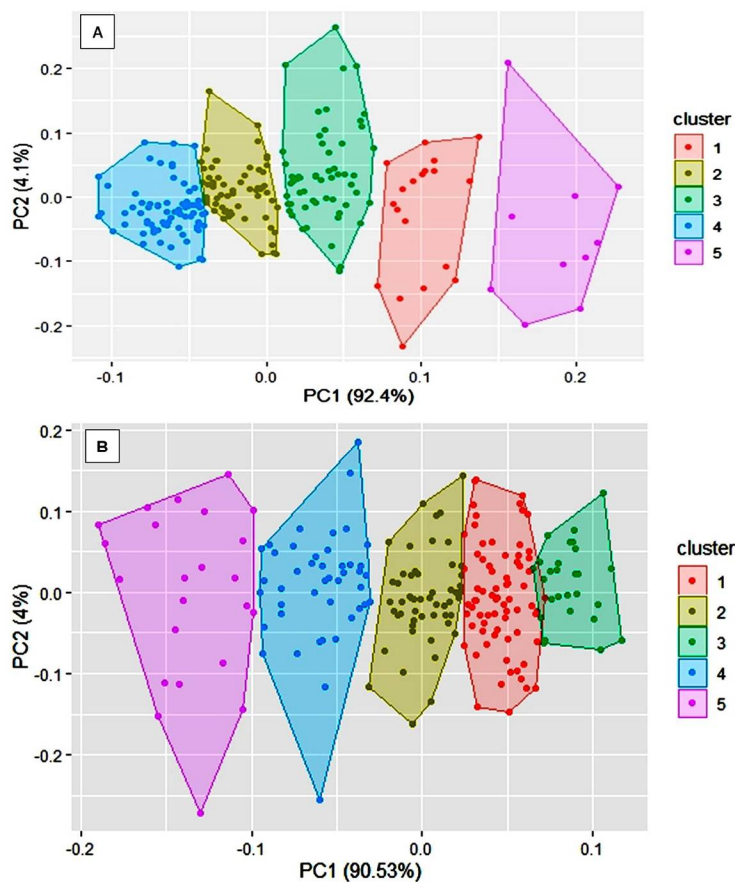


Fig. 1. K-mean clustering distribution of 220 lentil genotypes during 2018 (A) and 2019 (B) into various clusters based on class centroid and characters performance.

Results showed that distance between class centroid exhibited highest divergence groups between classes. Maximum distance between class centroid was observed in class 1-5 (138.70), Class 2-5 (112.63), Class 3-5 (85.02), and Class 4-5 (62.09) (Table 2b). Therefore, genotypes present in class 5 could be used for crossing with genotypes of different classes to make new possible combinations with highest potential for agronomic performance. Variation within group might be due to genetic variation of genotypes collected from USA, Syria or Pakistan, other reasons could be environmental factors especially temperature (Bharadwaj *et al.* 2001). Large inter-cluster distance between genotypes of any two clusters showed that genotypes falling in such clusters would be more distant genetically. K means clustering for cultivation year 2019 also represented 220 lentil genotypes in five different clusters (Fig. 1B). Here PC1 contributed 90.53% whereas 4% by PC2 during 2019. Variance between five clusters were observed as cluster 1, 2, 3, 4 and cluster 5 had variance values 454.98, 171.94, 130.63, 138.64 and 201.81. Cluster 1, 2, 3, 4 and cluster 5 have 73, 53, 27, 62 and 48 lentil genotypes, respectively. Highest number of genotypes was observed in cluster 1 (73) and 4 (62) (Table 2b). Distance between class centroids exhibited maximum divergence group between clusters. Maximum distance between class centroid was observed in cluster 1-5 (103.40), cluster 2-5 (66.36) cluster 3-5 (40.24) and cluster 4-4 (20.15). Class centroids of each trait highlighted variations.

Table 2a. K-mean clustering based on Wilks' lambda classifying each genotype, along with variance within class and lowest, highest and average distance from centroid for agronomic traits in 2018 and 2019 for lentil.

Clusters	CS	Var. C	Min DC	Averag e DC	Max. DC	Class centroids for 2018												
						SY	HSW	HI	BY	PH	LPH	PS	SP	DM	CT	HS	NP	
1	19	252.90	5.16	14.56	31.76	53.10	1.61	22.29	39.80	47.08	15.36	0.38	2.00	185.70	14.30	0.50	2.00	
2	62	179.87	3.52	12.36	31.42	41.25	1.78	30.11	137.29	43.54	14.80	0.41	2.02	184.98	14.34	0.44	2.05	
3	56	217.46	3.15	13.04	30.92	16.54	1.66	22.81	73.21	46.08	15.24	0.35	1.98	185.44	14.53	0.27	2.00	
4	73	340.71	6.15	16.88	35.84	28.21	1.71	28.08	101.63	44.04	14.89	0.36	2.00	183.52	14.71	0.43	2.03	
5	10	1289.13	15.34	33.26	60.72	42.55	1.84	23.61	180.25	44.10	14.52	0.48	2.00	185.05	14.85	0.30	2.20	
Class centroids for 2019																		
1	73	454.98	4.88	19.55	34.87	26.66	26.76	1.68	100.06	50.22	15.07	0.39	2.04	184.96	14.57	0.36	2.02	
2	53	171.94	3.68	11.83	30.59	35.10	27.16	1.76	129.64	50.54	14.55	0.37	2.02	184.51	14.58	0.33	2.02	
3	27	130.63	3.02	10.37	28.71	19.62	25.77	1.66	76.72	49.17	15.35	0.39	2.08	185.07	14.66	0.32	1.97	
4	62	138.64	4.28	10.86	22.42	12.84	23.00	1.55	57.03	44.49	14.97	0.37	2.08	184.75	14.60	0.48	2.00	
5	48	201.81	3.62	13.16	25.61	53.61	33.15	1.90	163.16	49.78	14.26	0.35	2.00	183.50	13.88	0.38	2.00	

Cluster size (CS), Within-class variance Var. (c), lowest distance to centroid (Min DC), average distance to centroid (average DC), maximum distance to centroid (Max. DC), SY; seed yield, HI; harvest index, HSW; hundred seed weight, BY; biological yield, PH; plant height, LPH; lower pod height, PS; pod size, SP; seeds per pod, DM; days to maturity, CT; cooking time, HS; number of hard seed, NP; number of pods.

Table 2b. Distance between class centroid for agronomic traits in 2018 and 2019.

Years/clusters	Distance between class centroid for agronomic traits in 2018 and 2019			
	1	2	3	4
2018	27.18			
2019	37.44			
2018	53.76	27.87		
2019	63.42	26.31		
2018	79.02	52.24	26.63	
2019	83.45	46.33	20.13	62.09
2018	138.70	112.63	85.02	20.15
2019	103.40	66.36	40.24	

It was noted that no geographical relationship among the clusters was detected as genotypes were independently showing their hidden potential. According to Bilal *et al.* (2018) clusters having genotypes with analogous source representing that there is no such relation between the characters adapted by the genotype and the geographical scattering of the genotype (Singh *et al.* 2010). The tendency of genetic diversity amid genotypes in current study specifies that geographical segregation might not be the only reason (Kumar *et al.* 2012). So, the genetic diversity can be the appearance of certain other interventions as like, germplasm used, genetic drift, artificial assortment, natural dissimilarity, etc., as supported by Egea *et al.* (2017). It can be accomplished that genetic diversity could be given preference over geographical distribution during the selection of parents used for breeding (Iqbal *et al.* 2021a, 2021b).

Table 3. Genotypes with high seed yield (g) per plant along with observed quantitative traits 2018 and 2019.

G/ years	SY (g)		HSW (g)		HI		BY (g)		PH (cm)		LPH (g)	
	2018	2019	2018	2019	2018	2019	2018	2019	2018	2019	2018	2019
6084	85	76	1.75	1.76	40	47	210	161	42	50	15	14
6062	73	69	2.1	2.4	54	46	135	149	35	69	24	17
6122	71	68	1.68	1.76	29	50	264	137	51	47	14	16
6058	67	67	1.63	1.73	49	41	146	162	35	52	16	14
6087	67	62	1.67	1.56	48	34	140	183	47	37	15	13
5689	66	60	1.67	1.61	33	34	202	179	44	45	15	17
6042	63	58	1.87	1.82	26	50	246	115	44	65	20	13
6074	62	57	1.78	1.89	54	30	115	187	50	40	15	12
5664	59	53	1.36	1.57	39	52	152	102	46	45	14	17
5687	58	53	1.67	1.71	29	36	200	148	61	56	12	13
G/ years	PS (cm)		SP (g)		DM (g)		CT (g)		HS		NP (g)	
	2018	2019	2018	2019	2018	2019	2018	2019	2018	2019	2018	2019
6084	0.48	0.32	2	2	184	191	15	13	2	0	2	2
6062	0.42	0.36	2	2	186	182	13	13	0	1	2	2
6122	0.61	0.34	2	2	183	183	14	16	0	1	2	2
6058	0.38	0.36	2	2	186	189	17	14	1	0	2	2
6087	0.36	0.40	2	2	188	179	15	11	0	2	2	2
5689	0.44	0.34	2	2	183	188	15	16	1	0	3	2
6042	0.34	0.32	2	2	183	183	14	13	0	0	2	3
6074	0.54	0.36	3	2	183	184	14	13	0	0	2	2
5664	0.42	0.42	2	2	183	184	11	14	0	0	2	2
5687	0.54	0.34	2	2	184	183	16	15	1	0	2	2

G. represents genotype. SY; seed yield, HSW; hundred seed weight, HI; harvest index, BY; biological yield, PH; plant height, LPH; lower pod height, PS; pod size, SP; seeds per pod, DM; days to maturity, CT; cooking time, HS; number of hard seed, NP; number of pods.

As far as characterization is concerned, 100-seed weight is used as an indicator of seed yield and quality. HSW showed wide variability and genotypes were grouped in cluster 1 and 4 during 2018 while in 2019 into cluster 5. Such variability can be accommodated in future lentil hybridization. Likewise, biological yield increases the photosynthetic area of the plants which resultantly enhances food quality, seed yield and HSW that exhibit significance in breeding.

Genotypes with the highest values of class centroids i.e BY (180.25), PS (0.48), CT (14.85), and NP (2.20) were grouped into cluster 1, while cluster 2 consisted of those genotypes which have the highest PH (47.08), LPH (15.36), MD (185.70) and HS (0.50) values of class centroids during 2018 (Table 3). Similarly, in 2019 genotypes with the highest-class centroid values for SY (50.56), HSW (1.88), and BY (159.43) in cluster 5 and genotypes with the most elevated PH (50.84), NP (2.02) were placed in cluster 2. Further, plant height is a significant yield-contributing trait which promotes vegetative growth, and branching provides an opportunity to enhance a photosynthetic area, pods, seed yield, and plant biomass (Iqbal and Akbar 2021). Time of maturity is considered an important trait with its effective mechanism against early maturity, drought and seed set before onset of terminal drought is an important trait in grain legumes. The cluster 3 consisted as of those genotypes which have highest DM (187.18) and HS (0.44) while genotypes having less DM (183.47) and HS (0.30) grouped in cluster 5. Similarly, genotypes which have highest LPH (15.30) and CT (14.72) as well as PS (0.39) and SP (2.09) were gathered in cluster 2. Low cooking time is also very important trait for consumer point of view.

Based on two years performance it may be concluded that lentil genotypes viz., 6084, 6062, 6122 are best outcome of seed yield (71-85 g), biological yield (135-264 g), pod size (0.48-0.61 cm), and cooking time duration (13-15 min). Therefore, it is recommended for multi-purpose uses in breeding programs to develop high yielding lentil varieties. Further, genotypes showed highest within class variances with respect to lowest, average, and maximum distance to centroid which is recommended for future crop improvement under diverse climates.

Acknowledgements

Authors are grateful to National Gene Bank, Plant Genetic Resource Institute, NARC, Islamabad, Pakistan, for providing lentil germplasm and field facilities.

References

- Ashabi A, Sahibuddin SB and Haghghi MS. 2020. The systematic review of K-means clustering algorithm. In: 2020 the 9th International Conference on Networks, Communication and Computing. pp. 13-18.
- Bharadwaj C, Satyavathi CT and Subramanyam D 2001. Evaluation of different classificatory analysis methods in some rice (*Oryza sativa*) collections. Indian J. Agri. Sci. **71**: 123-125.
- Bilal M, Jamil M, Ali A, Ilyas MK and Ghafoor A 2018. Multivariate analysis of genetic diversity in lentil (*Lens culinaris*) germplasm. Fresenius Environ. Bull. **27**: 2151-2159.
- Egea LA, Mérida-García R, Kilian A, Hernandez P and Dorado G 2017. Assessment of genetic diversity and structure of large garlic (*Allium sativum*) germplasm bank, by diversity arrays technology “genotyping-by-sequencing” platform (DArTseq). Front. Genet. **8**: 98.
- FAOSTAT 2019. <http://www.fao.org/faostat/en/#data>. Accessed 26 May 2021.
- Ganesan K, and Xu B 2017. Polyphenol-rich lentils and their health promoting effects. Int. J. Molecul. Sci. **18**: 2390.
- Gupta DS, Thavarajah D, Knutson P, Thavarajah P, McGee RJ, Coyne CJ and Kumar S 2013. Lentils (*Lens culinaris* L.), a rich source of folates. J. Agri. Food Chem. **61**: 7794-7799.
- Hussain SA, Iqbal MS, Akbar M, Arshad N, Munir S, Ali MA, et al. (2022) Estimating genetic variability among diverse lentil collections through novel multivariate techniques. PLoS ONE **17**(6): e0269177.
- Iqbal MS and Akbar M 2021. Phenotypic diversity analysis and performance of elite lines of *Brassica napus* L. Bangladesh J. Bot. **50**:1-6.
- Iqbal MS, Akbar M, Akhtar S, Fatima S, Ghafoor A 2021a. Multivariate analysis and its elaboration for characterizing cowpea landraces. Legume Res. **44**: 1-7.

- Iqbal MS, Hussain SA, Arshad N, Munir S, Ali MA, Masood H and Ghafoor A 2021b. Exploring new sources of viral resistance in cowpea germplasm and its elaboration through multivariate analysis. *Legume Res.* **44**: 673-678.
- Joshi M, Timilsena Y and Adhikari B 2017. Global production, processing and utilization of lentil: A review. *J. Integra. Agri*, **16**(12): 2898-2913.
- Kumar J, Basu PS, Srivastava E, Chaturvedi SK, Nadarajan N and Kumar S 2012. Phenotyping of traits imparting drought tolerance in lentil. *Crop Past. Sci.* **63**: 547-554.
- Nourin A, Kiran A, Kaukab S, Rehman A, Saeed MS, Tahir A and Khan E 2019. Evaluation of lentil gene pool for yield and some yield related attributes. *Univer. J. Agri. Res.* **7**: 32-62.
- Romano A, Gallo V, Ferranti P and Masi P 2021. Lentil flour: Nutritional and technological properties, in vitro digestibility and perspectives for use in the food industry. *Current Opin. Food Sci.* **40**: 157-167.
- Sarker A, Rizvi AH and Singh M 2018. Genetic variability for nutritional quality in lentil (*Lens culinaris* Medikus subsp. *culinaris*). *Legume Res.* **41**: 363-368.
- Sharma R, Chaudhary L, Kumar M, Yadav R, Devi U and Kumar V. 2022. Phenotypic diversity analysis of *Lens culinaris* Medik., accessions for selection of superior genotypes. *Sustainability*, **14**: 5982.
- Singh G, Aggarwal N and Khanna V 2010. Integrated nutrient management in lentil with organic manures, chemical fertilizers and biofertilizers. *J. Food Legumes* **23**: 149-151.
- Sisodia D, Singh L, Sisodia S and Saxena . 2012. Clustering techniques: a brief survey of different clustering algorithms. *Int. J. Latest Trends Eng. Tech.* **1**: 82-87.
- Thavarajah D and Thavarajah P 2012. Lentils (*Lens culinaris* L.): a superior whole food solution to micronutrient malnutrition. *Micronutrients sources, properties, and health effects*. Nova Science Publishers Inc., New York, 179-191.
- Tripathi K, Kumari J, Gore, PG, Mishra DC, Singh AK, Mishra GP and Sarker A 2021. Agro-morphological characterization of Lentil germplasm of Indian National Gene bank and development of a core set for efficient utilization in Lentil improvement programs. *Front. Plant Sci.* **12**.
- Ullah A, Shah TM and Farooq M 2020. Pulses production in Pakistan: status, constraints and opportunities. *Int. J. Plant Produc.* **14**: 549-569.

(Manuscript received on 30 June, 2021; revised on 27 August, 2022)