

Cluster Analysis in Emmer Wheat Germplasm using Quantitative Traits

Jatin Tanwar^{1*}, Shivani Sharma², Preeti Jakhar², Gaurav Kumar², Vikas, V.K.³, Shailendra K. Jha¹
and Jyoti Kumari²

¹ICAR-Indian Agricultural Research Institute (New Delhi), India.

²ICAR-National Bureau of Plant Genetic Resources (New Delhi), India.

³ICAR- Indian Agricultural Research Institute, Regional Station, Wellington.

(Corresponding author: Jatin Tanwar*)

(Received: 02 January 2023; Revised: 11 February 2023; Accepted: 14 February 2023; Published: 16 February 2023)

(Published by Research Trend)

ABSTRACT: Diversity analysis play key role in the germplasm evaluation which is directly correlated with the improved crop production. Different statistical methods have been used to study diversity among genotypes such as cluster analysis, principal component analysis and factor analysis. Among these techniques cluster analysis is used for the grouping of individuals based on their attributes. Emmer wheat is a rich genetic resource for the improvement of durum and bread wheat because it possesses beneficial and economically significant traits like pest and disease resistance and tolerance to abiotic stresses. Despite the substantial potential of emmer wheat, limited efforts have been made for its improvement. In the present study a total of 192 *dicoccum* wheat accessions from the National Genebank were assessed for their genetic diversity analysis using agro-morphological traits under three different locations; namely ICAR-IARI, Experimental farm, New Delhi, ICAR-NBPGR, Experimental farm, Issapur, New Delhi, and ICAR-IARI, RS, Wellington, Experimental farm, Tamil Nadu. Data were recorded for 14 different quantitative traits. Cluster analysis grouped the total 192 *dicoccum* wheat accessions into six clusters with Cluster I and VI being in contrast for several traits which may be used in selection of diverse parents for crossing program. The cluster membership of accessions revealed that Cluster I, IV and V consisted of mainly indigenous accessions whereas Cluster VI consisted of mainly exotic accessions. Cluster II and Cluster III consisted both indigenous and exotic accessions.

Keywords: Emmer wheat, Cluster analysis, Dendrogram, Quantitative traits, Germplasm categorization.

INTRODUCTION

Wheat (*Triticum aestivum* L.) is the most widely grown crop and an essential component of the global food security. India is second largest producer of wheat in the world. As per second advance estimates for 2022-23, the production of wheat (record) in the country is estimated at 1121.82 LMT which is higher by 44.40 LMT as compared to previous year's production (Ministry of Agriculture and Farmers Welfare). Mainly three species, *Triticum aestivum* L. (bread wheat), *T. durum* Desf. (*Kathia* or *macaroni* wheat) and *T. dicoccum* Schuh L. (*Khapli* or emmer wheat), are cultivated, but around 95% of the total production comes from bread wheat. With the emergence of the free-threshing and more productive durum (*T. durum* Desf.) and bread wheat (*T. aestivum* L.), emmer wheat gradually lost its importance (Shewry, 2009). Emmer wheat is a great source of genetic diversity that can be utilized to improve cultivated wheat to address the challenges of the expanding world population and climate change. Genetic diversity analysis of untapped *dicoccum* germplasm conserved in genebanks will promote their greater use by wheat breeders for crop improvement.

Analysis of genetic diversity and relationships between genotypes is a requirement for any breeding programme to be successful. A genetic diversity analysis of germplasm stored in genebanks determines its potential

for use in breeding programmes, which could ultimately result in increased crop production. Geographic diversity, release sites, and ploidy levels are only a few examples of factors that influence genetic diversity. As a result, statistical techniques should constitute the foundation for genotype characterisation. D² statistics and hierarchical Euclidean cluster analysis are two examples of the statistical techniques that have been created to evaluate genetic diversity. These techniques compare or contrast traits to find genetic differences depending on the interaction of several economically significant factors. Cluster analysis, PCA and factor analysis, are some important methods used for genetic diversity analysis, parental selection, and to study genotype environment interaction (Bhatt, 1970; Carves *et al.*, 1987; Mohammdi and Prasanna 2003; Kumar *et al.*, 2020). Wild emmer wheat has a high allelic diversity for several important traits, including agronomic characteristics, grain quality, and resistance to biotic and abiotic stresses (Nevo *et al.*, 2012; Fadida-Myers *et al.*, 2022). A large number of genes and QTLs that are valuable for wheat improvement have been identified in the wild emmer gene pool and mapped (Balla *et al.*, 2022a; Balla *et al.*, 2022b). Accurate information on nature and degree of genetic divergence helps the plant breeder to select genetically diverse parents for the hybridization programme (Arunachalam 1981). Cluster analysis algorithms can be used to determine relationships among germplasm

material and it is helpful to identify variables that can be classified into main groups and subgroups based on similarity and dissimilarity, it is also useful for parental selection in breeding programs (El- Deeb and Mohamed 1999) and crop modelling (Jaynes *et al.*, 2003). It allows mixing of both qualitative and quantitative information for selection and data reduction via similarity coefficient and then grouping of similar genotypes in one cluster (Peters and Martinelli 1989). The analysis of genetic association among breeding material and the categorization of germplasm require the application of various statistical approaches. A study was conducted to investigate relationship between wheat grain yield and its components (Leilah and Al-Khateeb 2005). In addition, 113 barley accessions were evaluated using cluster and principal component analysis (Ahmad *et al.*, 2008). Variability in seventy wheat genotypes was compared, utilizing cluster analysis for eight variables (Ali *et al.*, 2008). 94 bread wheat inbred lines were divided into three groups based on cluster analysis using the Wards algorithm and squared Euclidean distances, more diversity was found in peduncle length than in the number of spikes, flag leaf area and grain yield (Aharizad *et al.*, 2012). Soleymanifard *et al.* (2012) found that spikes m², 1000 grain weight, and plant height were responsible for 75% of the variation in grain yield of durum wheat genotypes. Present study was conducted for analysis of the extent of genetic variation and relationships among 192 *dicoccum* wheat accessions based on different quantitative traits using cluster analysis.

MATERIALS AND METHODS

In the present study, experimental material for genetic diversity analysis included 192 *dicoccum* or emmer wheat accessions (135 indigenous and 57 exotic collection) conserved in the national gene bank. These accessions comprised varieties, landraces, germplasm collection and elite lines assembled from different geographical regions. The present study for diversity analysis was carried out at three locations; namely ICAR-IARI, Experimental farm, New Delhi, located at 28.40° N latitude and 77.12° E longitude. ICAR-NBPGR, Experimental farm, Issapur, New Delhi, located at 28.24° N altitude and 76.50° E longitude and ICAR-IARI, RS, Wellington, Experimental farm, Tamil Nadu, located at 11.36° N latitude and 76.78° E longitude. The experiment was conducted in Augmented Block Design (ABD) with four checks (DDK 1025, DDK 1029, HW 1093 & MACS-2971) randomised in each block of size 32 during *Rabi* 2019-20. Data was recorded on the 14 quantitative traits, days to 75% spike emergence (DSE), days to 90% maturity (DM), plant height (PH), peduncle length (PDL), effective tillers per plant (EFT), spike length (SL), spikelets per spike (SLS), grains per spike (GRS), grain weight per spike (GWS), 1000-grain weight (TGW), grain length (GL), and grain width (GW), grain length width ratio (GLWR), grain yield per meter row length (GY). After compilation of data descriptive statistics analysis was done for the estimation of mean,

range, variance, standard error and coefficient of variation (CV). Pooled mean data were used for cluster analysis. Cluster analysis was done using Ward's method, a hierarchical agglomerative method using Euclidean distance. Here, grouping is archived in such a way that within-group variance is minimized, hence this method is also known as Ward's minimum variance method (Ward, 1963). The number of clusters was determined at the point where the total within-cluster variance of the clusters fell significantly to lower variance.

RESULTS AND DISCUSSION

Hierarchical Cluster Analysis based on pooled data.

Cluster analysis groups individuals based on their attributes, so that individuals with identical attributes are grouped into the same cluster. Clustering methods can be categorized in two groups, namely Hierarchical and Non-hierarchical clustering methods. Hierarchical clustering methods are more employed in research, and in analysing genetic diversity. This approach proceeds either through a series of successive fusions or through a series of successive groupings of individuals. The most related individuals are grouped first, and these initial classes are combined according to their similarities. When the similarity declines, both classes are merged into a single cluster. This method is termed as agglomerative method. Cluster analysis results can be represented as a two-dimensional diagram known as dendrogram. There are several hierarchical agglomerations widely used in clustering techniques. In this case, Ward's method, a hierarchical agglomerative method was used using Euclidean distance. Here, grouping is archived in such a way that within-group variance is minimized, hence this method is also known as Ward's minimum variance method (Ward, 1963). The number of clusters was determined at the point where the total within-cluster variance of the clusters fell significantly to lower variance.

Emmer wheat accessions were grouped into six clusters (Fig. 1). Cluster I comprised 25 accessions, cluster II had 55 accessions, cluster III grouped 59 accessions, cluster IV grouped 31 accessions, cluster V had 12 and cluster VI comprised 14 accessions (Table 1). The cluster membership of accessions revealed that Cluster I, IV and V consisted of mainly indigenous accessions whereas Cluster VI consisted of mainly exotic accessions. Cluster II and Cluster III consist of both indigenous and exotic accessions. Further when mean performance of accessions were considered, cluster I comprised of accessions having low plant height and high number of grains per spike, high thousand grain weight and grain length. Cluster II consisted of accessions with high grain weight per spike. Accessions having highest number of effective tillers per plant were grouped in Cluster III and highest grain length width ratio was observed among accessions grouped in Cluster IV. Accessions with early spike emergence and maturity were found in Cluster V. High mean performance for peduncle length, spike length, spikelets per spike, grain width and grain yield of one metre row length was observed in cluster VI (Table 2). Two-way

clustering i.e. dendrogram with heat map also depicted the grouping of genotypes with respect to traits in different clusters. Red colour depicted higher values of

trait mean and blue colour depicted lower value of trait mean (Fig. 1).

Table 1: Cluster Mean and Standard Deviation values of different agro-morphological traits.

Cluster	No of Accessions	Parameters	DSE	DM	EFT	PH	PDL	SL	SLS	GRS	GWS	TGW	GL	GW	GLWR	GY
1	25	Mean	87.33	118.46	13.81	93.16	26.23	10.11	21.85	41.88	1.83	39.89	8.01	2.35	3.51	25.29
		SD	1.61	2.91	1.80	9.54	2.40	0.75	1.61	4.17	0.33	3.84	0.37	0.22	0.40	5.74
2	55	Mean	89.34	121.05	13.73	103.24	28.18	8.93	19.58	36.60	1.92	37.32	7.86	2.17	3.72	26.03
		SD	1.09	3.14	1.81	6.06	3.27	0.71	1.44	3.89	0.53	5.15	0.35	0.17	0.27	5.02
3	59	Mean	86.47	117.51	14.80	105.86	29.56	8.79	19.34	33.59	1.99	36.53	7.98	2.33	3.48	24.12
		SD	2.08	2.35	2.04	4.57	3.04	0.70	1.73	3.67	0.49	4.16	0.28	0.13	0.23	7.26
4	31	Mean	87.41	117.07	13.72	107.71	25.99	9.11	20.10	31.63	1.53	30.24	7.70	2.07	3.86	18.06
		SD	1.30	3.25	2.23	5.41	3.18	0.85	2.26	5.66	0.47	3.99	0.23	0.16	0.30	7.00
5	12	Mean	84.78	114.87	10.38	99.41	27.24	8.48	16.34	29.17	1.15	30.25	6.84	2.39	2.91	21.18
		SD	4.31	5.98	1.99	5.45	5.47	1.07	1.15	7.36	0.37	5.74	0.35	0.20	0.28	7.96
6	14	Mean	92.02	122.39	10.64	120.94	35.07	10.42	22.73	39.48	1.42	33.20	7.34	2.43	3.11	27.51
		SD	3.74	3.84	1.47	9.98	5.52	1.36	2.40	8.17	0.36	5.33	0.45	0.24	0.32	9.47

Table 2: Germplasm accession grouped in different clusters.

Sr. No.	Cluster	Germplasm Accession
1.	Cluster 1	DDK 1029, DDK 1025, HW 1093, IC566241, IC535302, IC138898, IC535118, IC47048, EC577409, EC299240, IC252504, IC416358, EC609395, IC535106, IC402045, IC551396, IC551397, IC551398, IC551400, IC113725, IC591073, IC593663, IC593664, IC584049, IC443708
2.	Cluster 2	IC118774, IC531559, IC535301, IC277713, IC47800, EC11071, IC531969, IC47022B, EC12941, EC11074, EC6909, EC577398, EC11386, EC6838, EC6839, EC577412, IC448026, IC35170, EC299171, EC11389, EC577410, IC535070, IC535123, IC535131, IC535133, IC535134, IC535136, IC535138, IC535139, IC539302, IC443709, IC138474, IC530555, IC547564, IC566241, EC577401, EC577404, EC06838, EC06839, EC06840, EC06845, EC06900, IC535071, IC535078, IC535083, IC535088, IC535090, IC535093, IC535105, IC535108, IC212168, IC551399, EC590345, IC35167
3.	Cluster 3	EC11232, EC11073, EC577411, EC12954, IC138418, IC35171, IC35174, IC47021, IC47026, IC47034, IC47035, IC47037, IC47049, IC47545, IC47548, IC535304, IC47040, IC112083, EC06902, EC006903, EC06908, EC06909, EC06910, EC06912, EC08479, EC08572, EC577399, IC535072, IC535074, IC535081, IC535082, IC535092, IC535113, IC535115, IC535116, IC535117, IC535120, IC535125, IC535137, IC128392, EC577904, IC138331, IC138371, IC534012, IC534016, IC534586, IC534587, IC534621, IC138896-A, IC252503, IC138472, IC78706, EC577407, EC577406, IC138471, IC138475, IC138897, IC212164, IC212165
4.	Cluster 4	IC535079, IC535140, IC535141, IC535142, IC128425, IC138455, EC12565, IC533783, IC534018, IC118729, IC138900, IC252486, IC35091, EC11072, IC535076, IC535085, IC535086, IC535097, IC535112, IC535124, IC535126, IC535129, IC535130, IC535143, IC535144, IC535150, IC535153, IC402012, IC402018, IC402020, IC535073
5.	Cluster 5	IC28596, IC32513, IC35093, IC35097, IC35119, IC32502, EC578111, IC78699, EC299114, EC299208, IC118763, IC118765
6.	Cluster 6	EC578064, EC299111, EC299157, EC577932, EC577960, IC138450, EC519491, EC540809, EC540812, EC540813, EC299074, EC299211, EC577402, IC534960

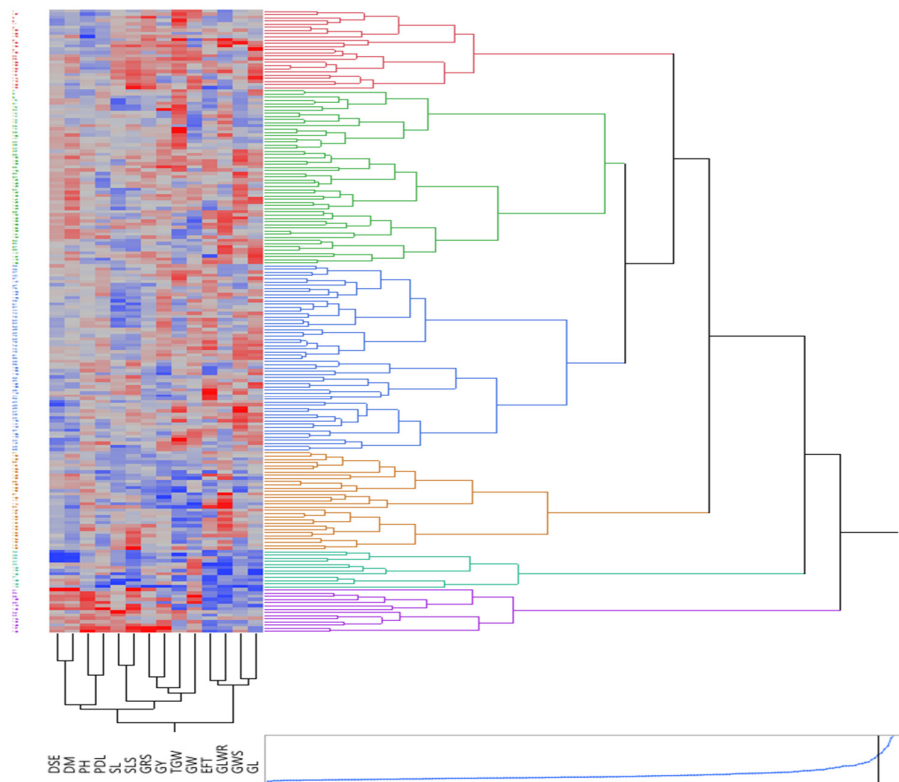


Fig. 1. Dendrogram and heat map based on 14 quantitative traits for 192 *dicoccum* wheat accession.

Various techniques have been used successfully for analysing genetic diversity to create diverse genotypes. Among these techniques the most frequent and successful is the agro-morphological characterization, which is commonly used for estimation of genetic variation in most of the breeding programs (Phougat *et al.*, 2017). Cluster analysis divided genotypes into various groups and sub groups based on morphological differences not on geographical affinity. A combination of genotypes with various taxonomic features were placed in clusters IV and V, which displayed a significantly greater degree of clear separation than clusters I, II, and III (Devesh *et al.*, 2019).

The main objective of evaluation of *dicoccum* wheat accessions for various agro-morphological traits is to classify these accessions for different breeding programs. The presence of genetic variability for various traits among accessions is a pre-requisite for breeding high yielding varieties, which leads to expression of various yield contributing characters. To start any breeding programme, information on the nature and magnitude of genetic variability is of utmost importance because presence of considerable amount of variability in the basic genetic material ensures better chances of evolving desired plant types. The estimates of Phenotypic Coefficient of Variation (%) are helpful to determine the variation for a specific trait in germplasm to improve the hexaploidy or tetraploid wheat. In the present study, high percent coefficient of variation (%CV) was observed for GWS, GY and moderate coefficient of variation for EFT, GRS, TGW, PDL, SLS, SL and GLWR whereas low % coefficient of variation was observed for GW, GL, DM and DSE. Similar results have been reported by Phougat and Verma (2022); Phougat *et al.* (2017); Jain *et al.* (2017); Preeti *et al.* (2016); Yadav *et al.* (2014); Bhushan *et al.* (2013) for days to heading, days to maturity, plant height (cm), grain yield per plot (g). Though cluster analysis grouped genotypes together with greater morphological similarity, the clusters did not necessarily include all genotypes from same origin. Ali *et al.* (2021); Ahmad *et al.* (2008), reported lack of association between agro-morphological traits and origin. In our study cluster analysis grouped the accessions into six clusters. Cluster III was the largest (59 acc.) followed by Cluster II (55 acc.), Cluster IV (31 acc.), Cluster I (25 acc.), Cluster VI (14 acc.) and Cluster V was the smallest one (12 acc.). Cluster I, Cluster IV and Cluster V consisted of mainly indigenous accessions whereas Cluster VI consisted of mainly exotic accessions. Cluster I had accessions with low plant height and high grains per spike, high thousand grain weight and grain length. On the other hand, accessions with high peduncle length, spike length, spikelets per spike, grain width and grain yield was grouped in cluster VI. We found that Cluster I and VI being contrast for several traits may be used in selection of diverse parents for crossing program.

CONCLUSIONS

Cluster analysis grouped the total 192 accessions into six clusters. Cluster III is the largest (59 acc.) and V is

the smallest one (12 acc.). Cluster I, Cluster IV and Cluster V consisted of mainly indigenous germplasm whereas Cluster VI consisted of mainly exotic germplasm. Cluster I had accessions with low plant height and high grains per spike, high thousand grain weight and grain length. Moreover, high peduncle length, spike length, spikelets per spike, grain width and grain yield of one metre row length was observed in cluster VI. Thus, Cluster I and VI being contrast for several traits may be used in selection of diverse parents for crossing program.

Acknowledgement. The author wishes to thank the ICAR-Indian Agricultural Research Institute (IARI) for all the support during research as well as the facility to carry out the experimental work and also thank Indian Council of Agriculture Research (ICAR) for the financial support under Junior Research Fellowship to the student during research work.

Conflict of Interests. None.

REFERENCES

- Aharizad, S., Sabzi, M., Mohammadi, S.A. and Khodadadi, E. (2012). Multivariate analysis of genetic diversity in wheat (*Triticum aestivum* L.) recombinant inbred lines using agronomic traits. *Annals of Biological Research*, 3(5), 2118–2126.
- Ahmad, Z., Ajmal, S. U., Munir, M., Zubair, M. and Masood, M. S. (2008). Genetic diversity for morpho-genetic traits in barley germplasm. *Pakistan Journal of Botany*, 40(3), 1217–1224.
- Ali, N., Hussain I., Ali, S., Khan N. U. and Hussain, I. (2021). Multivariate analysis for various quantitative traits in wheat advanced lines. *Saudi Journal of Biological Sciences*, 28(1), 347-352.
- Ali, Y., Atta, B.M., Akhter, J., Monneveux, P. and Lateef, Z. (2008). Genetic variability, association and diversity studies in wheat (*Triticum aestivum* L.) germplasm. *Pakistan Journal of Botany*, 40(5), 2087–2097.
- Arunachalam, V. A. (1981). Genetic distances in plant breeding. *Indian Journal of Genetics and Plant Breeding*, 4, 226- 236.
- Balla, M. Y., Gorafi, Y. S. A., Kamal, N. M., Abdalla, M. G. A., Tahir, I. S. A. and Tsujimoto, H. (2022a). Exploiting wild emmer wheat diversity to improve wheat A and B genomes in breeding for heat stress adaptation. *Frontiers in Plant Science*, 13.
- Balla, M. Y., Gorafi, Y. S. A., Kamal, N. M., Abdalla, M. G. A., Tahir, I. S. A. and Tsujimoto, H. (2022b). Harnessing the diversity of wild emmer wheat for genetic improvement of durum wheat. *Theoretical and Applied Genetics*, 135(5), 1671-1684.
- Bhatt, G. M. (1970). Multivariate analysis approach to selection of parents for hybridization aiming at yield component in self-pollinated crops. *Australian Journal of Agricultural Research*, 21, 1-7.
- Bhushan, B., Bharti, S., Ojha, A., Pandey, M., Gourav, S. S., Tyagi, B. S. and Singh, G. (2013). Genetic variability, correlation coefficient and path analysis of some quantitative traits in bread wheat. *Journal of wheat research*, 5(1), 21-26.
- Carves, B. F., Smith, E. L. and England, H. O. (1987). Regression and cluster analysis of environmental responses of hybrid and pure line winter wheat cultivars. *Crop Science*, 27, 659-664.
- Devesh, P., Moitra, P. K., Shukla, R. S. and Pandey, S. (2019). Genetic diversity and principal component analyses for yield, yield components and quality traits

- of advanced lines of wheat. *Journal of Pharmacognosy and Phytochemistry*, 8(3), 4834-4839.
- El- Deeb, A. A. and Mohamed, N. A. (1999). Factor and cluster analysis for some quantitative characters in sesame (*Sesamum indicum* L.). The Annual Conference ISSR, Cairo University, 4-6 December, Vol. 34, Part (II).
- Fadida-Myers, A., Fuerst, D., Tzuberi, A., Yadav, S., Nashef, K., Roychowdhury, R., & Ben-David, R. (2022). Emmer wheat eco-geographic and genomic congruence shapes phenotypic performance under Mediterranean climate. *Plants*, 11(11), 1460.
- Jain, N., Marker, S. and Kanwar, S. (2017). Assessment of genetic variability in late sown wheat (*Triticum aestivum* L.) germplasm. *Chemical Science Review and Letters*, 6(22), 1295-1300.
- Jaynes, D. B., Kaspar, T. C., Colvin, T. S. and James, D. E. (2003). Cluster analysis of spatio temporal corn yield pattern in an Iowa field. *Agronomy Journal*, 95(3), 574-586.
- Kumar, V., Pandey S., Shukla, R. S. and Rani, A. (2020). Multivariate Analysis for Morphological traits in newly developed Cytoplasmic lines of wheat. *Int.Arch. Appl. Sci. Technol.*, 11(1), 116-121.
- Leilah, A. A. and Al-Khateeb, S. A. (2005). Statistical analysis of wheat yield under drought conditions. *Journal of Arid Environments*, 61, 483-496.
- Mohammdi, S. A. and Prasanna, B. H. (2003). Analysis of genetic diversity in crop plant salient statistical tools and considerations, *Crop Science*, 43(4), 1235-1248.
- Nevo, E., Fu, Y. B., Pavlicek, T., Khalifa, S., Tavasi, M. and Beiles, A. (2012). Evolution of wild cereals during 28 years of global warming in Israel. *Proceedings of the National Academy of Sciences*, 109(9), 3412-3415.
- Peters, J.P. and Martinelli, J.A. (1989). Hierarchical cluster analysis as a tool to manage variation in germplasm collections. *Theoretical and Applied Genetics*, 78, 42-48.
- Phougat, D., Panwar, I. S., Saharan R. P., Singh V. and Godara, A. (2017). Genetic diversity and association studies for yield attributing traits in bread wheat [*Triticum aestivum* (L.) em. Thell]. *Research on crops*, 18(1), 139-144.
- Phougat, D. and Verma, A. (2022). Diversity in Bread Wheat genotypes (*Triticum aestivum*L.) deciphered by Multivariate Biplot analysis. *Biological Forum – An International Journal*, 14(1), 1610-1616.
- Preeti, Singh, V., Yashveer, S., Panwar, I. S. and Phougat, D. (2016). Evaluation of phenotypic and molecular diversity in wheat cultivars and advanced breeding lines of wheat in India. *Research on crops*, 17(3), 605-612.
- Shewry, P. R. (2009). Wheat. *Journal of Experimental Botany*, 60, 1537-1553.
- Soleymanifard, A., Naseri, R. and Meysam, M. (2012). The study genetic variation and factor analysis for agronomic traits of Durum wheat genotypes using cluster analysis and path analysis under drought stress condition in western of Iran. *International Research Journal of Applied & Basic Sciences*, 3(3), 479-485.
- Ward, H. J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236-244.
- Yadav, K. S., Singh, K. A., Baghel, S. S., Jarman, M. and Singh, A. K. (2014). Assessment of genetic variability and diversity for yield and its contributing traits among CIMMYT based wheat germplasm. *Journal of wheat research*, 6(2), 154-159.

How to cite this article: Jatin Tanwar, Shivani Sharma, Preeti Jakhar, Gaurav Kumar, Vikas, V.K., Shailendra K. Jha and Jyoti Kumari (2023). Cluster Analysis in Emmer Wheat Germplasm using Quantitative Traits. *Biological Forum – An International Journal*, 15(2): 556-560.