



Studies on Genetic Diversity and Principal Component Analysis in Wheat Germplasm (*Triticum aestivum* L.)

Mohd. Khurram Niyazi^{a++}, Gideon Synrem^{a#*}
and Sharad Pandey^{bt}

^a Department of Agriculture, Himgiri Zee University, Sherpur, Chakrata Road, Dehradun (Uttarakhand)-248197, India.

^b School of Agriculture, Forestry and Fisheries, Himgiri Zee University, Sherpur, Chakrata Road, Dehradun (Uttarakhand)-248197, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/IJPSS/2023/v35i153101

Open Peer Review History:

This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: <https://www.sdiarticle5.com/review-history/100252>

Original Research Article

Received: 27/03/2023
Accepted: 29/05/2023
Published: 13/06/2023

ABSTRACT

This study aimed to assess the potential genetic diversity among wheat genotypes by using cluster analysis and principal component analysis for selection of desired parents in hybridization programmes. The study was carried out at the Crop Research Center, Department of Agriculture, Himgiri Zee University, Dehradun, Uttarakhand, during Rabi season 2021-2022 under irrigated conditions. Twenty germplasm lines of wheat (*Triticum aestivum* L.) were grown in Randomized Block Design with three replications with an objective to study the genetic variability and diversity through D^2 statistics in wheat. The Analysis of variance revealed the significant differences among

⁺⁺ M.Sc (Ag.) Genetics and Plant Breeding Student;

[#] Assistant Professor;

[†] Dean;

*Corresponding author: E-mail: gideon.synrem@hzu.edu.in;

the genotypes for all the traits studied indicating that the presence of genetic material under the study. The clustering pattern as per Tocher's optimization method obtained on the basis of magnitude of D^2 value revealed that the twenty genotypes were grouped into five clusters. The maximum intra cluster D^2 value was observed in cluster I (130.89) followed by cluster II (108.06) and cluster III (97.93). The inter cluster distance values ranged from 249.44 to 1101.59. The maximum inter cluster distance was observed between cluster II and IV (1101.59) followed by II and III (1017.07) that indicated wide divergence existed among the genotypes of these clusters. From the results obtained by Principal component analysis six principal components PC 1 to PC6, which are extracted from the original data and having latent roots greater than one thereby accounting to 84.30% of the total variation.

Keywords: *Triticum aestivum* L.; genetic diversity; intra and inter cluster distances; principal component analysis.

1. INTRODUCTION

Wheat is grown in practically all of India's states due to its high level of environmental adaptability. Wheat production has increased dramatically, from 6.60 million tons at independence to 97.44 million tons in 2016-17. During the aforementioned time period, productivity increased by 473%, from 670 to 3172 kg ha⁻¹. Despite the delayed sowing, the country recorded 30.71 million hectares of acreage in Rabi 2016-17 [1]. Only three of the seventeen different species of wheat *Triticum aestivum*, *Triticum durum*, and *Triticum dicoccum* are grown commercially worldwide. *Triticum dicoccum* only makes up a very small portion of the total area of wheat, with *Triticum aestivum* (bread wheat) taking up more than 90% of the remaining space (9–10%). According to statistics, India produced 86.50 million metric tons of wheat, ranking second in the world [2].

D^2 statistics analysis is used for selection of genetically divergent parents in hybridization program. The concept D^2 statistics was originally developed by P.C. Mahalanobis in 1928. Rao used this technique for assessment of genetic diversity in plant breeding. It is used to measure the degree of diversification and determines relative portion of each component trait to total divergence. In plant breeding, genetic diversity plays a key role because hybrids between lines of diverse origin display a greater heterosis than closely related parents.

The PCA analysis is a multivariate statistical technique that aims to analyse the relationships between a large collection of variables in terms of a relatively small set of variables or components without omitting any of the crucial details from the original data set. By looking for groups with very strong inter-correlation in a set of variables and each component explaining

percent (%) variation to the total variability, the PCA reduces a reasonably large series of data into a smaller number of components. This experiment was conducted with the objectives to assess the potential genetic diversity among wheat genotypes by using cluster analysis and principal component analysis for selection of desired parents in hybridization programmes.

2. MATERIALS AND METHODS

The present investigation was conducted at the Crop Research Centre of Himgiri Zee University, Dehradun (Uttarakhand) during rabi season of 2021- 2022.

The experiment was laid in Randomized Block Design (RBD) with three replications with 20 entries. The seeds of wheat were sown as per the treatment in each row of 3m length. The row-to-row distance of 23 cm and plant to plant distance of 10 cm were kept for the experiment. Five randomly selected plants were tagged to record observations for all the character from each treatment except for days to booting, flowering, heading and maturity for 50 % and 100 % data were taken in plot basis.

2.1 Experimental Materials

The experimental material for the present investigation comprised of twenty genotypes obtained from different State Agriculture University of the country.

2.2 Observations Recorded

Five competitive plants were randomly selected, prior to flowering from each plot in each replication and were tagged with proper information for identification. Later, the data on twenty characters were recorded on these randomly selected plants. The observations were

taken on the following characters viz. days to emergence, days to 50% booting, days to 50% heading, days to 50% flowering, days to 100% flowering, grain filling period, spike length, awn length, plant height, leaf area index, flag leaf area per plant, leaf relative water content, peduncle length, days to maturity, number of productive tillers per plant, biological yield per plant, spike weight per plant, grains count per spike, 1000 seed weight, grain yield per plant.

Table 1. List of wheat genotype

S. No.	Genotypes
1	GW-322
2	HI-8737
3	RAJ-4037
4	GW-273
5	RAJ-4238
6	HD-3226
7	HI-8759
8	DBW-303
9	DBW-187
10	HI-1544
11	DBW-222
12	LOK-1
13	VARDAN GOLD
14	SARDAR-303
15	WH-711
16	PBW-550
17	DBW-17
18	PBW-502
19	PBW-343
20	HD-2967

2.3 Data Analysis

Analysis of variance (ANOVA)(Ronald Fisher,1918), genetic diversity(P.C. Mahalanobis,1928), principal component analysis(Karl Pearson,1901). Statistical analysis of characteristic data was processed by Windostat Version 9.2 from indostat services, Hyderabad Licensed to Dept. of Plant Breeding and Genetics Punjab Agriculture University Ludhiana.

3. RESULTS AND DISCUSSION

3.1 Analysis of Variance

The results of analysis of variance are presented in Table 2. The variance due to genotypes were significant for all the characters observed, indicating that the genotypes selected for the present study were genetically divergent. The experimental findings were well supported by the results obtained by Bhanupriya [3], Devesh [4].

3.2 Cluster Analysis of Genotype

The clustering pattern obtained on the basis of magnitude of D^2 value was made and present in Table 3 and clustering as per Tocher method. The 20 genotypes were grouped into five clusters. It is revealed from Table that cluster I (DBW-187, HI-1544, DBW-303, GW-273, DBW-222, DBW-17, VARDAN GOLD) had highest number of genotypes (7) followed by cluster III (WH-711, PBW-550, RAJ-4238, SARDAR 303, PBW-343, LOK-1, HD-2967) along with 7 genotypes. The cluster II (GW-322, HD-3226, PBW-502, RAJ-4037), IV (HI-8737) and cluster V (HI-8759).

The intra and inter cluster D^2 values are presented in Table 4. The intra-cluster distance values ranged from 0 to 130.89. The maximum intra cluster D^2 value was observed in cluster I (130.89) followed by cluster II (108.06) and cluster III (97.93). The inter cluster distance values ranged from 249.44 to 1017.07. The maximum inter cluster distance was observed between cluster II and III (1017.07) followed by II and IV (1101.59) that indicated wide divergence existed among the genotypes of these clusters. From the studies, inter cluster distance was more than the intra cluster distances. The least value of intercluster D^2 value was observed in between cluster IV and V (249.44) suggested that the genotype in one cluster is in close proximity with the genotype in the other cluster of pair. Hence, genotypes from both clusters may not be useful in breeding program.

The cluster mean for the twenty characters studied in wheat is given in Table 5. The lowest mean value for days to 50 % flowering (111.92 days) was recorded in cluster II, while for days to maturity was (145.90 days) in cluster III. The highest mean values were recorded by cluster V for awn length with two genotypes (HI-8737, HI8759), the highest grain count per spike was recorded (58.51) in cluster I, the maximum plant height was seen in cluster I, cluster II for the maximum biological yield (37.93); cluster V showed the highest leaf relative water content in percentage (85.02). The highest 1000 seeds weight was seen in cluster IV (43.67) while high grain yield was also seen in cluster IV (56.00). The experimental results were supported by the similar findings of Poudel et al. [5], Singh et al. [6], Aashu et al. [7], Geeta et al. [8].

3.3 Percent Contribution of Different Characters to Genetic Diversity

The utility of D² analysis was enhanced by its application to estimate the relative contribution of the various plant characters to genetic divergence. The percent contribution of twenty character studied was given in Table 6. It was observed that days to 100% maturity

(73.16%), biological yield (7.89%), contributed highest for divergence followed by awn length (8.42), peduncle length (3.68%), leaf relative water content (2.63%), 1000 seed weight (1.05%) and many character shows 0% contribution for genetic divergence. The results were supported by the findings by Mostafa et al. [9], Shashi et al. [10] Pooja et al. [11].

Table 2. Analysis of variance (ANOVA) of twenty wheat genotypes for twenty characters

S. N	Characters	Mean sum of squares			CV (%)
		Treatment (df=19)	Replication (df=2)	Error (df=38)	
1.	Days to emergence	0.43**	0.20	0.16	6.29
2.	Days to 50% Booting	65.43***	4.11	10.02	3.33
3.	Days to 50% Heading	27.29***	6.11	6.87	2.53
4.	Days to 50% Flowering	28.43***	2.51	8.14	2.52
5.	Days to 100% Flowering	19.92***	0.65	4.03	1.72
6.	Grain Filling Period	39.82***	3.61	4.80	7.36
7.	Spike Length	4.42***	0.89	0.45	5.83
8.	Awn Length	17.56***	0.33	0.17	6.07
9.	Plant Height	102.83**	316.88	38.73	6.61
10.	Leaf Area Index	0.14***	0.52	0.03	11.04
11.	Flag Leaf Area	29.23***	3.19	3.79	6.48
12.	Leaf Relative Water Content	69.79***	2.02	3.84	2.37
13.	Peduncle Length	9.69***	0.29	0.17	2.94
14.	Days to 100% Maturity	13.70***	4.06	1.27	0.77
15.	Productive Tillers per Plant	0.38**	0.016	0.15	9.14
16.	Biological Yield per Plant	58.34***	4.132	0.80	2.47
17.	Spikes weight per Plant	23.21***	0.98	1.42	8.69
18.	Grains Count per Spike	125.33***	34.70	12.52	6.63
19.	1000 Seed Weight	30.11***	7.71	2.85	4.71
20.	Grain Yield per Plant	41.21***	1.55	1.90	2.94

df =Degrees of freedom, **=significant at 0.01% probability level, ***=significant at 0.001% probability level.
CV= Coefficient of Variation

Table 3. Composition of D² clusters for wheat genotypes

Cluster No	Genotypes	Number of genotypes
Cluster – I	DBW-187, HI-1544, DBW-303, GW-273, DBW-222, DBW-17, VARDAN GOLD	7
Cluster – II	GW-322, HD-3226, PBW-502, RAJ-4037	4
Cluster – III	WH-711, PBW-550, RAJ-4238, SARDAR 303, PBW-343, LOK-1, HD-2967	7
Cluster – IV	HI-8737	1
Cluster – V	HI-8759	1

Table 4. Average intra (diagonal bold) and inter cluster distance (D²) (between the clusters) of wheat genotypes

	Cluster – I	Cluster – II	Cluster – III	Cluster – IV	Cluster – V
Cluster – I	130.89	310.95	395.40	597.85	372.10
Cluster – II		108.06	1017.07	1101.59	525.79
Cluster – III			97.93	335.70	523.97
Cluster – IV				0.00	249.44
Cluster – V					0.00

Table 5. Cluster wise mean performance for different quantitative traits of wheat genotypes

S. N.	Group	Characters									
		Days to emergence (days)	Days to 50 % Booting (days)	Days to 50% heading (days)	Days to 50% Flowering (days)	Days to 100% Flowering (days)	Grain Filling Period (days)	Spike Length (cm)	Awn Length (cm)	Plant Height (cm)	Leaf Area Index (m ²)
1.	Cluster – I	6.48	94.29	103.00	112.62	116.24	29.76	12.54	6.00	97.88	1.75
2.	Cluster– II	6.42	93.33	102.75	111.92	115.42	32.42	11.01	6.33	94.90	1.64
3.	Cluster– III	6.57	95.29	103.29	113.48	117.48	28.43	11.70	6.07	90.95	1.50
4.	Cluster– IV	6.00	101.00	105.33	113.67	117.33	30.33	8.82	13.30	92.33	1.73
5.	Cluster– V	6.00	97.00	107.00	116.33	120.00	28.00	9.57	13.70	87.87	1.49

Table 5. Continued...

S. N.	Group	Characters									
		Flag Leaf Area (cm ²)	Leaf Relative Water Content (%)	Peduncle Length (cm)	Days to 100% Maturity (days)	Productive Tillers Per Plant	Biological Yield per Plant (g)	Spikes Weight Per Plant (g)	Grains Count Per Spike	1000 Seed Weight (g)	Grain Yield per Plant (g)
1.	Cluster – I	30.17	83.82	14.79	146.00	4.33	37.23	14.70	58.51	34.48	44.76
2.	Cluster– II	29.48	81.72	14.19	147.83	4.17	37.93	13.43	51.38	34.92	47.25
3.	Cluster– III	29.80	82.88	13.20	145.90	4.38	34.57	14.00	51.87	36.00	48.10
4.	Cluster– IV	33.61	75.33	16.91	147.67	4.33	33.73	10.67	45.27	43.67	56.00
5.	Cluster– V	30.01	85.02	13.48	148.00	4.67	37.40	9.40	44.13	41.00	53.00

Table 6. Percentage contribution of twenty characters towards genetic divergence of twenty wheat genotypes

S. No	Characters	Times ranked 1st	Contribution (%)
1.	Days to emergence	0	0.11%
2.	Days to 50% Booting	0	0%
3.	Days to 50% Heading	0	0%
4.	Days to 50% Flowering	0	0.23%
5.	Days to 100% Flowering	0	0%
6.	Grain Filling Period	0	0%
7.	Spike Length	0	0%
8.	Awn Length	16	8.42%
9.	Plant Height	0	0.13%
10.	Leaf Area Index	0	0%
11.	Flag Leaf Area	0	0%
12.	Leaf Relative Water Content	5	2.63%
13.	Peduncle Length	7	3.68%
14.	Days to 100% Maturity	139	73.16%
15.	Productive Tillers per Plant	3	1.58%
16.	Biological Yield per Plant	15	7.89%
17.	Spikes weight per Plant	1	0.53%
18.	Grains Count per Spike	0	0%
19.	1000 Seed Weight	2	1.05%
20.	Grain Yield per Plant	2	1.05%
			100%

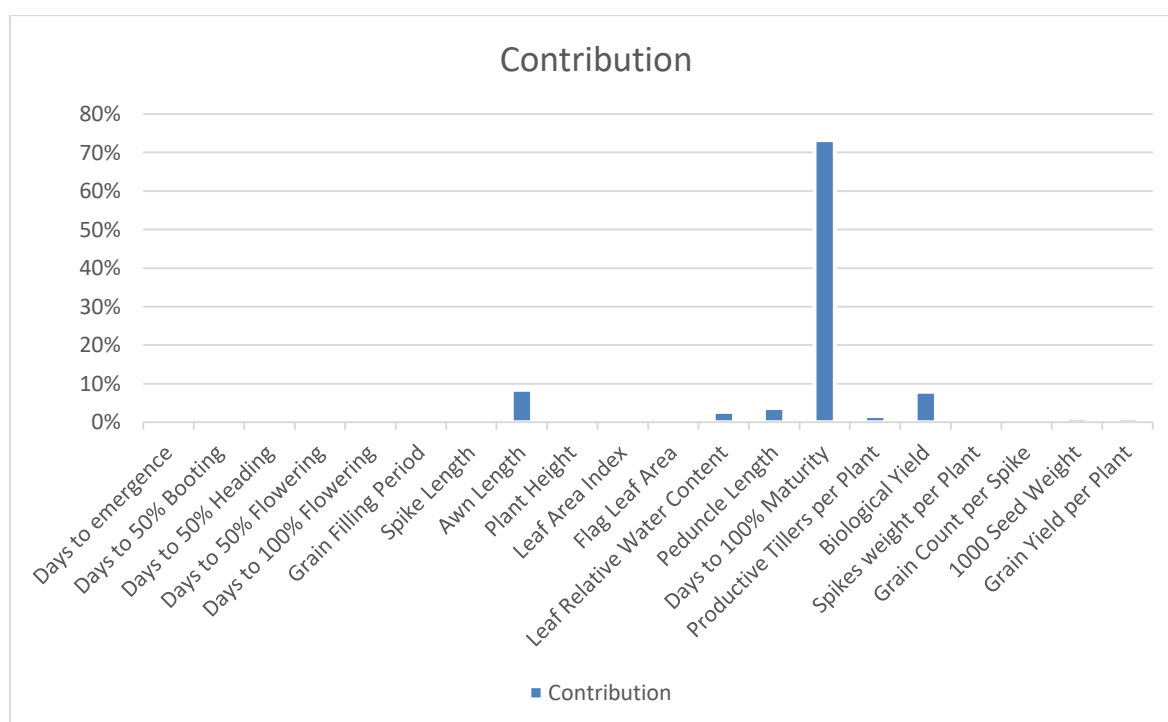


Fig. 1. Contribution of different characters towards genetic divergence of twenty wheat genotypes for twenty characters

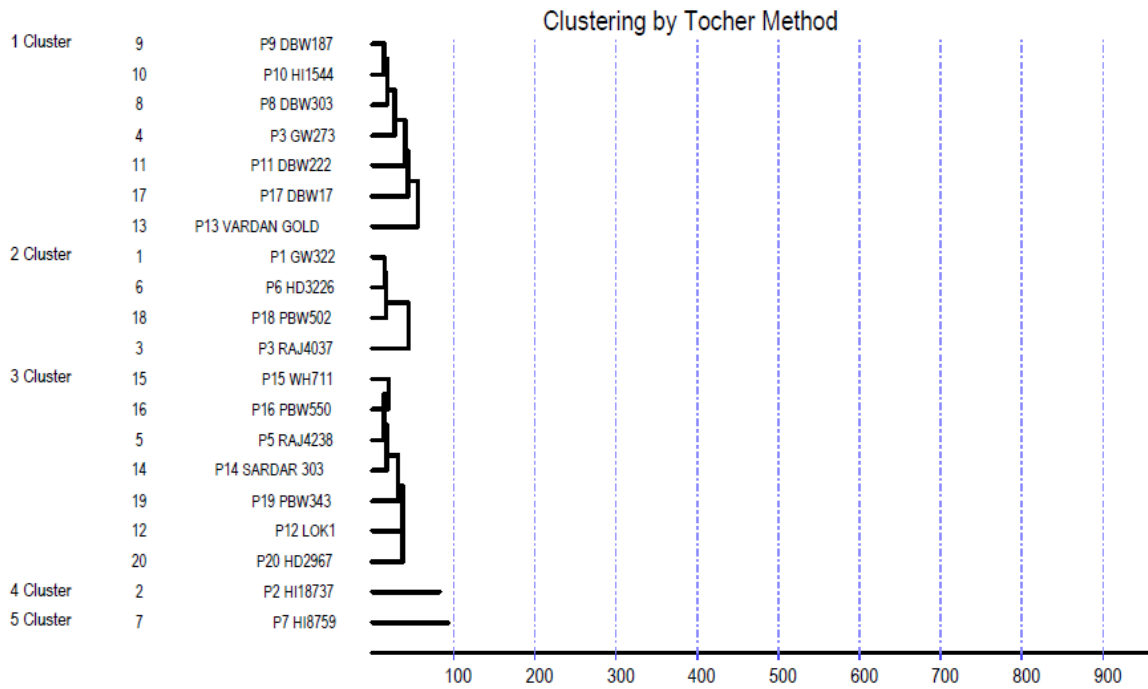


Fig. 2. Clustering by tocher method

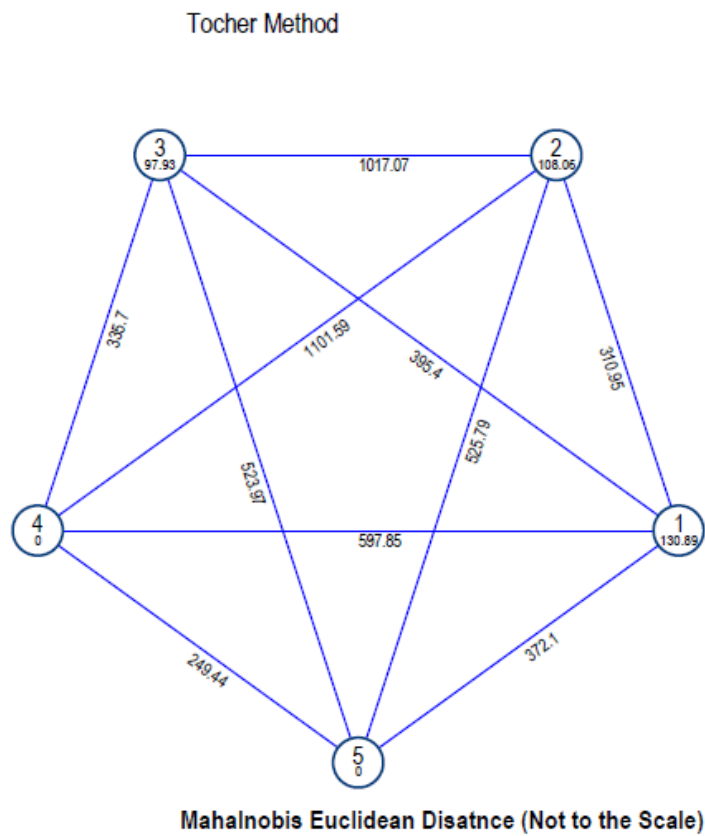


Fig. 3. Mahalanobis Euclidian²distance graph

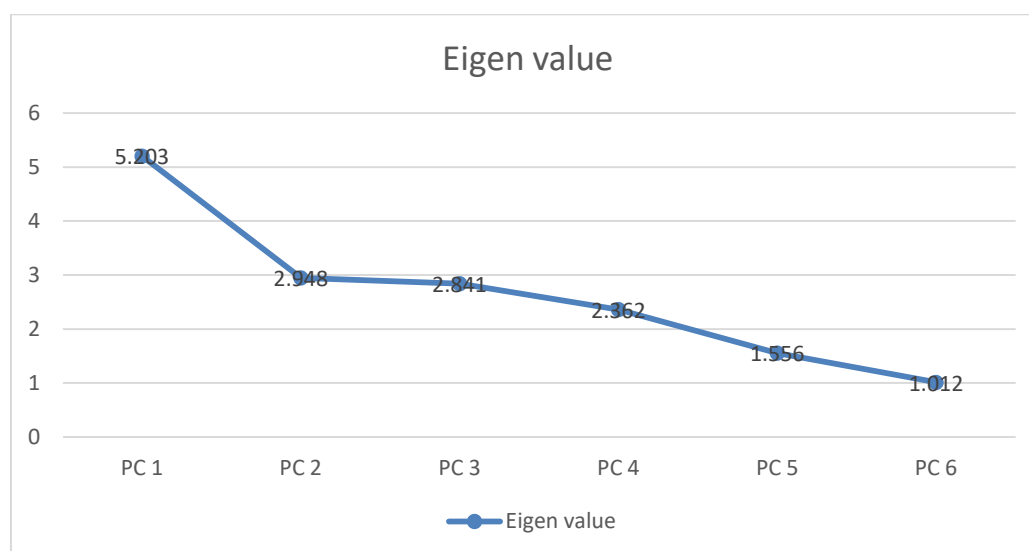


Fig. 4. Eigen value graph

Table 7. Principal component matrix of six PCs

PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Spike length	1000 seed weight	Days to 50% flowering	Emergence days	Days to 50% booting	Leaf area index
Biological yield	Grain yield per plant	Days to 100% flowering	Days to 50% heading	Leaf relative water content	Flag leaf area
Spikes weight per plant		Plant height	Grain filling period	Days to 100% maturity	Peduncle length
			Productive tillers per plant		
			Grain count per spike		

Table 8. Eigen values and variability explained by each principal component (PCs)

Principal Component	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Eigen Value (Root)	5.203	2.948	2.841	2.362	1.556	1.012
% Var. Exp.	26.015	14.743	14.205	11.812	7.783	5.064
Cum. Var. Exp.	26.015	40.759	54.964	66.777	74.560	79.625

3.4 Principal Component Analysis (PCA)

From six principal components PC 1 to PC6, which are extracted from the original data and having latent roots greater than one, accounting 84.30% of the total variation which suggested that these principal component scores might be used to summarize the original 20 variables in any further analysis of the data.

Accordingly, the first principal component had high positive coefficients for spike length, biological yield and spikes weight per plant. The major contributing characters for the diversity in the second principal component (PC2) were

1000 seed weight and grain yield per plant, while third principal component (PC3) was days to 50% flowering, days to 100% flowering and plant height. Fourth principal component contributed diversity for the character days to emergence, days to heading, grain filling period, productive tillers per plant and grain count per spike. For fifth principal component was days to 50% booting, leaf relative water content and days to 100% maturity and for sixth principal component leaf area index, flag leaf area, peduncle length.

On the basis of PCA, most of the important yield and yield attributing traits were present in PC4, PC3, PC2, and PC1. High PC score for a

Table 9. Studies on principal components for 20 varieties on 20 characters in wheat

Variables	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Days to emergence	0.244	0.233	0.063	0.256	0.084	0.152
Days to 50% Booting	-0.178	0.041	0.047	-0.264	0.418	0.109
Days to 50% Heading	0.062	0.0608	0.196	0.323	-0.462	-0.199
Days to 50% Flowering	-0.020	-0.041	0.491	0.075	0.036	-0.000
Days to 100% Flowering	-0.159	0.254	0.406	0.129	0.127	-0.029
Grain Filling Period	-0.002	0.006	-0.306	0.446	0.116	0.237
Spike Length	0.332	-0.072	0.269	-0.061	0.018	-0.266
Awn Length	-0.387	0.019	-0.061	0.045	-0.185	-0.125
Plant Height	-0.268	-0.156	0.259	0.060	-0.301	-0.053
Leaf Area Index	0.347	-0.056	0.120	-0.120	-0.008	0.364
Flag Leaf Area	0.027	0.277	0.263	0.200	-0.101	0.615
Leaf Relative Water Content	-0.056	0.044	-0.057	0.393	0.534	-0.233
Peduncle Length	-0.193	-0.457	0.071	0.103	0.020	0.294
Days to 100% Maturity	-0.080	0.260	0.222	-0.261	0.277	-0.004
Productive Tillers per Plant	-0.218	-0.428	0.096	0.141	0.138	0.135
Biological Yield	0.230	-0.099	-0.235	0.131	-0.059	-0.028
Spikes weight per Plant	0.306	-0.019	0.023	-0.284	-0.065	-0.015
Grain Count per Spike	0.210	-0.041	0.206	0.323	0.152	-0.317
1000 Seed Weight	-0.068	0.458	-0.246	0.096	-0.131	0.026
Grain Yield per Plant	-0.365	0.279	0.005	-0.085	-0.083	-0.067

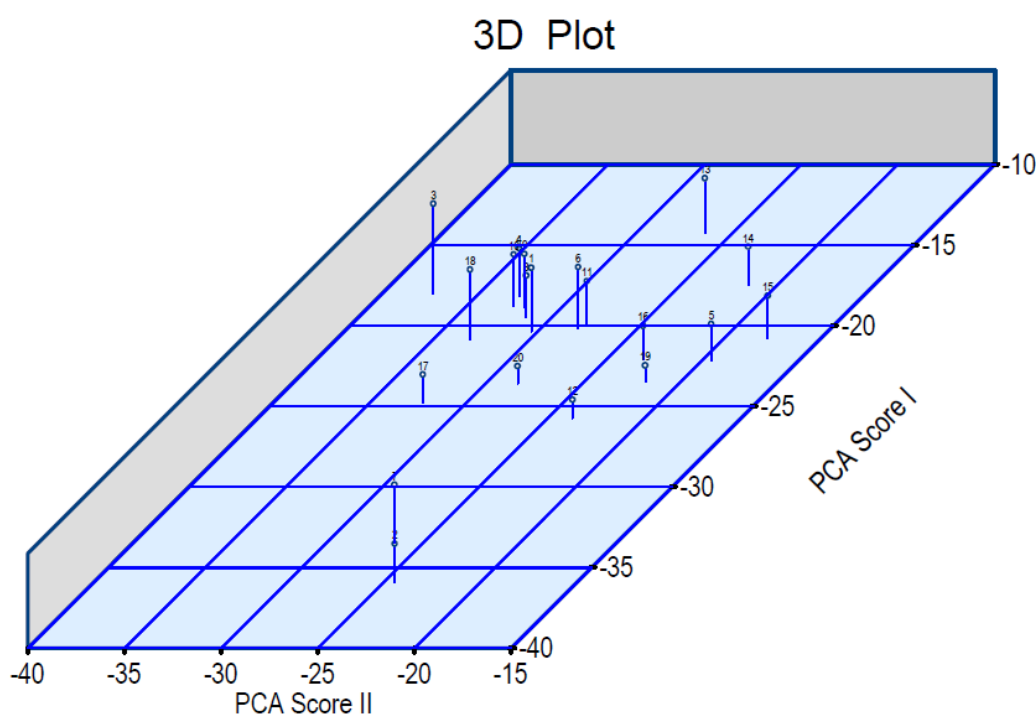


Fig. 5. Principal component analysis graph

particular genotype in a particular component denotes high values for the variables in that particular genotype. The experimental results were well supported by the similar findings by Bhanupriya [3], Narinder et al. [12] and Pandey et al. [13,14,15].

4. CONCLUSION

From the experimental results it can be revealed that the twenty wheat genotypes showed sufficient genetic variability and crossing between the genotypes belong to Cluster II and

IV can be suggested for development of potential hybrids as the inter cluster distance was the largest indicating wide genetic diversity. Findings from PCA revealed that first four principal's components were related to various morphological and physiological traits in wheat mostly associated with grain yield, biological yield and 1000 seed weight genotypes and also these traits can identify the diverse genotypes which could be employed in hybridization programme for improvement of wheat.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Annual Report, ICAR- Indian Institute of Wheat and Barley Research, Karnal - 132001, Haryana, India; 2016-17.
2. FAO Statistical data. Available:<http://faostat.fao.org>
3. Bhanupriya Das B, Satyanarayan NH, Mukharjee S, Sarkar KK. Genetic diversity of wheat genotypes based on principal component analysis in Gangetic alluvial soil of West Bengal. *Journal of Crop and Weed*. 2014;10(2):104-107.
4. Bhanupriya Das B, Satyanarayan NH, Mukharjee S, Sarkar KK. Genetic diversity of wheat genotypes based on principal component analysis in Gangetic alluvial soil of West Bengal. *Journal of Crop and Weed*. 2014;10(2):104-107.
5. Poudel A, Thapa DB, Sapkota M. Assessment of genetic diversity of bread wheat (*Triticum aestivum* L.) genotypes through cluster and principal component analysis. *International Journal of Experimental Research and Review*. 2017;11(1):1-9.
6. Singh G, Kumar P, Kumar R, Gangwar KL. Genetic diversity analysis for various morphological and quality traits in bread wheat (*Triticum aestivum* L.). *Journal of Applied and Natural sciences*. 2018;10(1): 24-29.
7. Aashu Solanki YPS, Divya P, Tanya B. Estimating genetic diversity of bread wheat (*Triticum aestivum* L.). In different environments. *Electronic Journal of Plant Breeding*. 2022;13(2):399-409.
8. Geeta K, Shukla RS, Devesh P. Genetic diversity analysis in bread wheat (*Triticum aestivum* L E.M. Thell.) for quantitative and physiological traits under normal sown condition. *The Pharma Innovation*. 2022; 11(8):574-577.
9. Mostafa Khodadadi, Mohammad Hossein Fotokian Mohammad Miransari. Genetic diversity of wheat (*Triticum aestivum* L.) genotypes based on cluster and principal component analyses for breeding strategies. *Australian Journal of Crop Sciences*. 2011;5(1):17-24.
10. Shashi kala S. Kolakar RR. Hanchinal and Sadashiv nadukeri. Analysis of genetic diversity in wheat genotypes. *International Journal of Agricultural Sciences*. 2014; 10(1):142-145.
11. Pooja, Dhanda SS, Beniwal RS. Genetic diversity based on cluster and principal component analyses for yield and its contributing characters in RILs of bread wheat (*Triticum aestivum* L.). *International Journal of Pure and Applied Bioscience*. 2018;6(2):242-247.
12. Narinder P. Amitesh SM. IqbalJB, Rohit S. Principal component analysis of yield characteristics of wheat (*Triticum aestivum* L.) in intermediate zones of Jammu and Kashmir. *International Journal of Chemical Studies*. 2018;6(4):892-895.
13. Pandey J, Tripathi RM, Singh A, Singh V. and Mishra KU. Genetic diversity and principal component analyses for yield and yield components of advanced lines of wheat (*Triticum aestivum* L.). *The Pharma Innovation Journal*. 2021;10(7):998-1003.
14. Das S, Das SS, Chakraborty I, Roy N, Nath MK, Sarma D. Principal component analysis in plant breeding. *Biomolecule Reports*. 2017;2456-8759.
15. Maurya DM, Singh dp. Genetic divergence in rice. *Indian Journal of Genetics and Plant Breeding*. 1977;37(3):395-402.

© 2023 Niyazi et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here:

<https://www.sdiarticle5.com/review-history/100252>