

PCA Based Multivariate Statistical Diversity Analysis for Grain Yield and its Components in Wheat (*Triticum aestivum* L.)

AKSHAY HULLE, NIDHI DUBEY, HARSHAL AVINASHE*, J. PRANAY REDDY¹, SAGAR CHORMULE² AND SHARAD SACHAN³

Department of Genetics and Plant Breeding, Lovely Professional University, Phagwara-144 411 (Punjab), India
*(e-mail: havinashe@gmail.com; Mobile: 90567 87045)

(Received: February 21, 2023; Accepted: March 25, 2023)

ABSTRACT

The research work investigated the genetic diversity of 20 lines seeded in RCBD with three replications and were evaluated for 15 phenotypic characteristics at the Agriculture Research Farm, Lovely Professional University, Phagwara (Punjab). According to principal component analysis (PCA), six of the 15 (PC1 to PC6) PCs had eigen values above 1.0 and a cumulative variance of around 87.1%. These were number of productive tillers per plant, peduncle length, number of ears per plant, biological yield per plant, plant height and grain yield per plant. On this foundation of high PC1 scores, the most notable lines, MACS-6145, HI-1500 and CHIRYA-3 were selected for yield components. The outcomes of this investigation might be used as a foundation for defining and implementing subsequent wheat breeding initiatives.

Key words: Bread wheat, PCA, eigen value

INTRODUCTION

Wheat is a segmentation allohexaploid ($2n = 6x = 42$) species native to the family Poaceae. Tetraploid as well as hexaploid wheat, namely, *T. durum* and *T. aestivum*, contribute to 90% of overall worldwide wheat output (Reddy *et al.*, 2019, 2022; Kumar *et al.* 2022). “King of cereals” seems to be the title bestowed on the wheat crop based primarily on the region covered, followed by production as well as, subsequently, the dietary content that it does seem to have (Verma *et al.*, 2019; Yadav *et al.*, 2021). Wheat is the most important food crop for people all over the world, having a dominant position in Indian agriculture (Naik *et al.*, 2015), occupying 28% of the cereal land and contributing 33% of total food grain output in the country. It subsumes niacin and thiamine in fair concentration both of which primarily focus on producing that unique protein known as “Gluten” which serves as the fundamental foundation supporting that sponge cellular consistency for bread and bakery goods (Kandil *et al.* 2016; Hakimi *et al.*, 2017).

The primary goal of each and every wheat breeding effort is always to generate varieties that are high-yielding, adaptable, and resistant to insects-pests and diseases as well as other climatic influences (Reddy *et al.*, 2019). To attain this goal, there needs to be enough biological variation throughout genotypes. In a heterogeneous breeding population, principal component analysis (PCA) helps in developing the selection process for selecting better lines (Mishra *et al.*, 2015). This provides an effective strategy for identifying plant features that provide the greatest amount of output potential as well as elucidating the interactions among characteristics within a complicated number of co-combinations, which include development, performance and profitability characteristics (Lv *et al.*, 2021). It also discusses the link between eigen vectors and eigen values but rather economic yield aids in the determination of the major yield and its factors in a diversified breeding population (Reddy *et al.*, 2019). Effective parent screening is critical in bridging seedlings to improve genetic manipulation and possible yield

¹Department of Genetics and Plant Breeding, Faculty of Agriculture, Annamalai University, Annamalai-608 002 (Tamil Nadu), India.

²United Phosphorus Ltd, Hyderabad-500 016 (Telangana), India.

³Department of Agricultural Economics and Extension, Lovely Professional University, Phagwara-144 411 (Punjab), India.

growth. There are several acceptable approaches (cluster analysis, PCA and factor analysis) that exist for detecting biological variation, parental selection, mapping the road to crop development, core of existence and distinctiveness, and examining the connection between environmental and plant diversification (Reddy *et al.*, 2019). Using Principal Component Analysis (PCA), the current research was done to find potential races and key features in several wheat varieties that could be used in future crossbreeding projects.

MATERIALS AND METHODS

The experiment was conducted at the Agriculture Research Farm, Faculty of Genetics and Plant Breeding, School of Agriculture, Lovely Professional University, Phagwara (Punjab). The research included 20 wheat varieties. During *rabi* 2020-21, the experiment was designed in a randomized complete block design (RCBD) replicated three times and five rows per inclusion every replicate, with a dimension of 22.5 x 5 cm with a row distance of 3 m.

The observations were recorded on five randomly chosen plants of each genotype of replications and used for data analysis, with the exception of the two plants cultivated as frontier plants across both ends of the row. Characters studied were: days to 50% heading, days to maturity, plant height (cm), productive tillers per plant, peduncle length (cm), number of spikelets per ear, ear length (cm), ear weight (g), number of ears per plant, number of grains per ear, biological yield per plant (g), harvest index (g), canopy temperature, chlorophyll content and grain yield per plant (g).

The PCA was carried out to determine qualities that were responsible for the computational majority of the overall variance. Prior to starting principal component analysis, raw statistics were calibrated to obtain a mean value of zero as well as a variation of one in order to prevent disparities in measured variables. The FactoMineR statistical software tool was applied to construct the main components according to the correlation matrix.

RESULTS AND DISCUSSION

Principal component analysis (PCA) is a well-established binarization method that seeks a

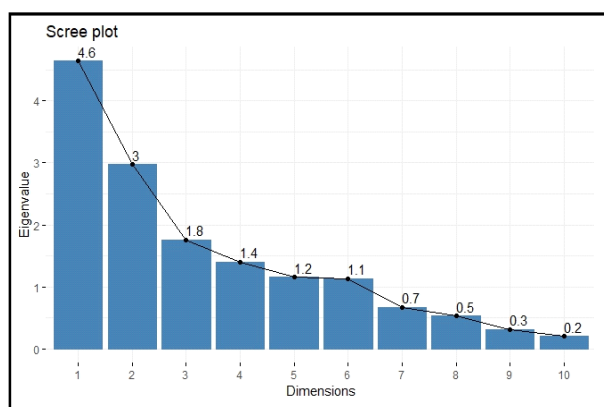
diverse set of X dimensions with the least variance or with the most statistical data. PCA is a way of attempting to reduce a complex source of statistics to a lower-dimensional form, exposing the unseen, simpler forms that exist behind it. A linear discriminant study is effective for gathering measurements for a significant number of indicator variables and then generating a smaller number of fictitious characteristics (called principal components) that reflect the bulk of the variation in the dependent factor. It is commonly used in quantitative techniques to convey data in the shortest feasible manner (Reddy, 2019). The eigen vectors and parameters from the PCA for every other individual may be utilized as variables in a quantile regression approach to account for such a core structure (Balcha *et al.*, 2022).

PCA was done in this study for 15 yield and yield-attributing characteristics in 20 wheat varieties (Table 1). The principal component analysis yielded six fundamental factors with eigen values varying from around 4.641 to 1.138 (PC 1 to PC 6). PC1 had a 4.641 eigen value, PC2 used to have a 2.972 eigen value, PC3 had a 1.763 eigen value, PC4 had a 1.397 eigen value, PC5 had a 1.155 eigen value and PC6 also had a 1.138 eigen value (Fig. 1). Every one of all six basic parameters compensated, included a significant proportion of overall variation that varied from 30.9 to 7.6%, for a total variance of 87.1%. As a result, the following six major dimensions were selected for subsequent discussion: Due to Gutten's lowest limit principle, eigen values of 1 should have been discarded. PCs with an eigen value of 1 could be excluded. The first principal component (PC1) provided for 30.9% of the variance, followed by PC2, PC3, PC4, PC5 and PC6, where they contributed 19.8, 11.8, 9.3, 7.7 and 7.6%, correspondingly, while the leading four PCs accounted for approximately 71.8% of such overall variation (Fig. 2). From the first PC, a semi-curve figure was formed, having minimal variance within every PC suggesting that characters beneath the very first PC might well become preferable (Reddy, 2019).

Reddy *et al.* (2019) obtained a similar finding in wheat, in which out of 17, the first four major factors were PC1, PC2, PC3 and PC4, with variances of 42.74, 27.31, 7.27 and 5.6%, correspondingly, to the whole of 82.93%

Table 1. Eigen values, per cent variance and cumulative variance in wheat genotypes

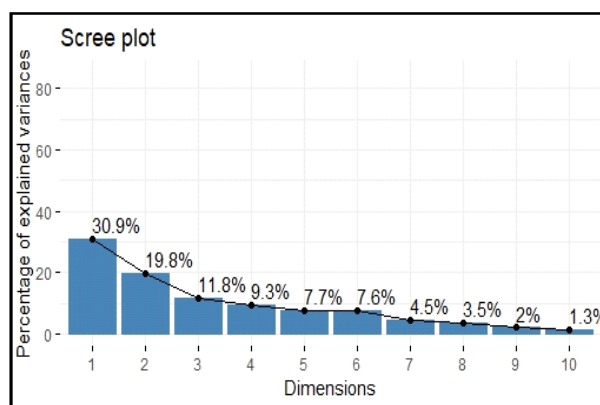
Principal components	Eigen values	Proportion	Cumulative proportion
PC1	4.641	30.9	30.9
PC2	2.972	19.8	50.8
PC3	1.763	11.8	62.5
PC4	1.397	9.3	71.8
PC5	1.155	7.7	79.5
PC6	1.138	7.6	87.1
PC7	0.675	4.5	91.6
PC8	0.528	3.5	95.1
PC9	0.307	2	97.2
PC10	0.202	1.3	98.5
PC11	0.11	0.7	99.2
PC12	0.084	0.6	99.8
PC13	0.019	0.1	99.9
PC14	0.009	0.1	100
PC15	0.001	0	100

**Fig. 1.** Eigen values scree plot of wheat genotypes.

accumulated variance. According to Ozukum *et al.* (2019), the top five PCs accounted for 22.14, 20.19, 13.67, 11.09 and 8.04% of the entire genetic variation with 75.15%, respectively.

Table 2. Principal component values of rotation component matrix

Characters	Principal components					
	PC1	PC2	PC3	PC4	PC5	PC6
Days to 50% heading	-0.028	-0.342	-0.173	-0.215	0.478	-0.276
Days to maturity	0.183	-0.426	0.141	-0.025	0.214	-0.313
Plant height	0.211	-0.418	0.161	-0.161	-0.119	0.252
Productive tillers/plant	0.379	-0.172	-0.099	0.216	-0.016	0.136
Peduncle length	0.287	-0.119	0.291	0.249	-0.018	0.474
No. of spikelets/ear	-0.343	-0.333	-0.047	0.132	-0.058	0.034
Ear length	-0.216	-0.197	-0.266	0.168	0.126	0.54
Ear weight	-0.063	-0.033	0.011	0.751	0.325	-0.195
No. of ears/plant	0.384	-0.134	-0.154	-0.175	0.06	0.143
No. of grains/ear	-0.348	-0.325	-0.058	0.137	-0.061	0.044
Biological yield/plant	0.302	-0.21	-0.312	0.021	-0.252	-0.254
Harvest index	0.151	0.385	-0.218	0.08	0.256	0.128
Canopy temperature	0.113	0.083	0.452	-0.117	0.577	0.085
Chlorophyll content	-0.126	0.007	-0.483	-0.282	0.341	0.259
Grain yield/plant	0.339	0.096	-0.384	0.245	0.028	-0.102

**Fig. 2.** Cumulative variance scree plot of wheat genotypes.

Within same PC1, qualities with the highest utter and total numbers closest to one had a greater effect on the cluster than those with the generally lowest scores closest to zero (Adilova *et al.*, 2020). As a result, rather than the same huge contribution of a small number of attributes, the same divergence of varieties in the various clusters was because of the combined influence of those many phenotypes. Tables 2 and 3 show the rotated component matrix principal component values and genotyping principal component scores.

PC scores were generated in six principal components within individual genotypes, which were then used to determine which varieties performed better under different variations of morphological variables. A significant PC score associated with a unique genotype in a specific PC suggested higher values for just the parameters in the same genotype that this component represents. As a result, individual values may be exploited to

Table 3. PCA scores of wheat genotypes

Characters	Principal components					
	PC1	PC2	PC3	PC4	PC5	PC6
PBW-590	0.469	0.34	-0.144	1.859	0.892	0.02
HP-1761	-1.788	0.439	0.968	1.168	1.1	-1.153
UP-262	0.46	-0.161	-1.523	0.308	-0.399	-1.041
DBW-14	0.217	0.541	-0.009	1.316	-0.009	1.4
MACS-6145	1.788	-2.075	0.398	0.715	2.036	0.437
HI-1500	1.812	-0.956	0.784	0.397	-1.777	-0.13
DBW-16	0.653	1.012	-1.049	-0.876	1.911	-2.074
WL-711	-0.911	0.194	1.082	-0.646	0.288	0.616
K-9644	-1.233	-0.184	0.024	1.108	0.077	0.175
K-0307	-0.239	-0.867	-0.587	-2.059	0.868	1.416
GW-366	0.218	-0.633	0.094	0.285	-0.667	-0.771
HUW-213	-1.593	-1.302	0.748	-0.79	-0.48	-1.179
RAJ-4037	0.497	0.268	0.661	-1.836	0.173	-0.206
HS-240	-1.182	-0.678	0.028	0.409	-0.49	1.196
CHIRYA-3	1.006	1.202	0.646	-0.495	-0.435	0.701
PBW-527	0.026	0.949	-2.11	0.369	-1.226	-0.195
K-1006	-0.361	-1.502	-1.005	-0.372	-1.089	-0.401
LOK-1	0.695	1.451	2.148	-0.154	-0.713	-0.562
KRL-19	-0.35	1.132	-0.866	-0.325	0.563	1.798
GW-322	-0.183	0.83	-0.289	-0.38	-0.626	-0.047

generate exact screening indicators, the magnitude of which could be determined through the amount of variance per PC.

In the current research, PC1 (Figs. 3 and 4) had the greatest positive genotype scores, especially regarding MACS-6145, HI-1500 and CHIRYA-3, and the lowest negative genotype scores for HP-1761, K-9644 and HUW-213. It showed that perhaps the genotypes had good performance with these important traits: plant height, productive tillers per plant, peduncle length, number of spikelets per ear, ear length, number of ears per plant, number of grains per ear, biological yield per plant, and grain yield per plant. In PC2, those genotypes

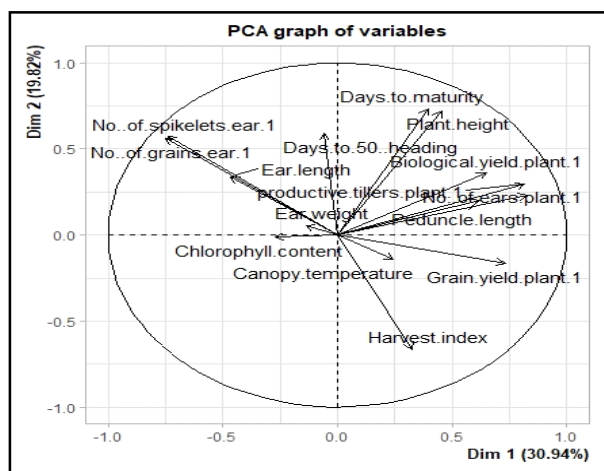


Fig. 3. PC1 biplot on various characters in wheat genotypes.

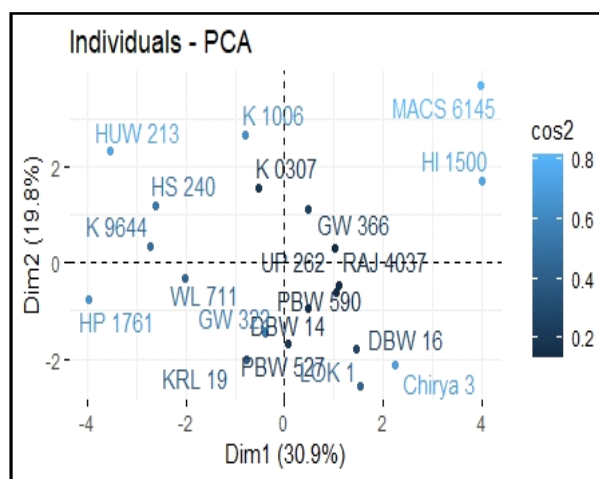


Fig. 4. PC1 graph of genotypes in wheat.

with both the greatest positive ratings were DBW-16, CHIRYA-3, LOK-1 and KRL-19, whereas the genotypes with the least negative values were MACS-6145, HUW-213 and K-1006. The genotypes were shown to exhibit a high value considering parameters such as days to 50% heading, days to maturity, plant height, number of spikelets per ear and harvest index. PC3 had the greatest positive genotype scores, WL-711 and LOK-1, and the lowest negative genotype scores, UP-262, DBW-16, PBW-527 and K-1006. It showed that such genotypes had high values for peduncle length, ear length, biological yield per plant, canopy temperature, chlorophyll content and grain yield per plant. PC4 contained the greatest

positive genotype scores, including PBW-590, HP-1761, DBW-14 and K-9644, as well as negative genotype scores, including K-0307 and RAJ-4037. It suggested that the single genotype had a high character ear weight value. In PC5, the top positive genotype values were HP-1761, MACS-6145 and DBW-16, although the largest negative genotype values were HI-1500, PBW-527 and K-1006. It showed that indeed, genotypes had high values for traits such as days to 50% heading, ear weight, canopy temperature and chlorophyll content. In PC6, the genotypes with the greatest positive ratings were DBW-14, K-0307, HS-240 and KRL-19, whereas the genotypes with the lowest negative values were HP-1761, UP-262, DBW-16 and HUW-213. It showed that certain genotypes had a high value for characteristics such as days to 50% heading, days to maturity, peduncle length and ear length.

Reddy *et al.* (2019) observed analogous findings through an experiment in wheat crop, revealing variance in days to 50% heading, days to maturity, overall plant height. Kumar *et al.* (2016) and Devesh *et al.* (2019) also found diversity for the same i.e. number of grains per spike, number of spikelets per spike, spike length, plant height, days to heading, days to flowering, protein content and yield per plant. A biplot was created using the initial two PCs (Fig. 5). A biplot is used to illustrate overall correlations between genotypes and quantifiable attributes. Both the first and second PC biplots described 42.4% of these variables within the overall strongest discriminatory parameters, which were plant height, number of spikelets per ear and number of grains per ear. The genotypes identified close to the progenitor possessed comparable genetic traits, however, this same

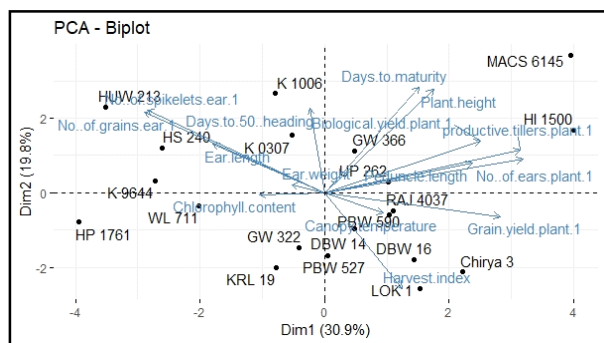


Fig. 5. Biplot between PCs showing contribution of genotypes and traits responsible for variability in wheat.

genotype obtained farther away from the source were deemed unconnected. As a result, in wheat breeding programme, several different genotypes might be employed as potential parents for blending to form heterotic populations.

CONCLUSION

The analysis found that mainly six PCs (PC1 to PC6) out of 15 demonstrated upwards of 1.0 eigen value but also around 87.1% overall variability. The findings revealed a high level of genetic variability in the analyzed physical features. These mentioned factors may well be examined when choosing high-yielding wheat genotypes: number of productive tillers per plant, peduncle length, number of ears per plant, biological yield per plant, plant height and grain yield per plant. Three notable lines for yield and quality attributes were discovered based on strong PC1 scores: MACS-6145, HI-1500 and CHIRYA-3. Utilizing these wheat genotypes, this variation might be used to generate varieties with various desired high-yielding qualities via selection and/or hybridization.

REFERENCES

- Adilova, S., Qulmamatova, D., Baboev, S., Bozorov, T. and Morgunov, A. (2020). Multivariate cluster and principal component analyses of selected yield traits in uzbek bread wheat cultivars. *Am. J. Plant Sci.* **11**: 903-912. doi: 10.4236/ajps.2020.116066.
- Balcha, U., Mekbib, F. and Lule, D. (2022). Cluster and principal component analysis among bread wheat (*Triticum aestivum*L) genotypes in mid rift valley of Oromia, Ethiopia. *Adv. Crop Sci. Tech.* **10**: 525. doi: 10.4172/2329-8863.100052.
- 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. *J. Pharmacogn. Phytochem.* **8**: 4834-4839.
- Hakimi, N., Lavanya, G. R., Babu, G. S. and Ahmad, A. (2017). Genetic variability, correlation and path coefficient analysis for metric traits in wheat (*Triticum aestivum* L.). *Int. J. Multidisciplinary Res. Develop.* **4**: 354-358.
- Kandil, A. A., Sharief, A. E. and Gomaa, H. (2016). Estimation of general and specific combining ability in bread wheat (*Triticum aestivum* L.). *Int. J. Agr. Res.* **8**: 37-44.

- Kumar, J., Kumar, A., Singh, S. K., Singh, L., Chaudhary, M., Kumar, S. and Singh, S. K. (2016). Principal component analysis for yield and its contributing traits in bread wheat (*Triticum aestivum*) genotypes under late sown condition. *Curr. Adv. Agric. Sci.* **8**: 55-57. doi: 10.5958/2394-4471.2016.00011.3.
- Kumar, M., Avinash, H., Dubey, N., Goutam, U., Janeja, H. S., Reddy, J. P. and Chowdhury, S. (2022). Studies on combining ability and authentication of F_1 hybrids by using STS markers in bread wheat (*Triticum aestivum* L.). *Int. J. Agric. Stat. Sci.* **18**: 265-272.
- Lv, C., Lu, W., Quan, M., Xiao, L., Li, L., Zhou, J., Li, P., Zhang, D. and Du, Q. (2021). Pyramiding superior haplotypes and epistatic alleles to accelerate wood quality and yield improvement in poplar breeding. *Indus. Crops Prod.* **171**: 113891. <https://doi.org/10.1016/j.indcrop.2021.113891>.
- Mishra, C. N., Tiwari, V., Satish-Kumar, V. G., Kumar, A. and Sharma, I. (2015). Genetic diversity and genotype by trait analysis for agromorphological and physiological traits of wheat (*Triticum aestivum* L.). *Sabrao. J. Breed. Genet.* **47**: 40-48.
- Naik, V. R., Biradar, S. S., Yadawad, A., Desai, S. A. and Veerasha, B. A. (2015). Study of genetic variability parameters in bread wheat (*Triticum aestivum* L.) genotypes. *Res. J. Agric. Sci.* **6**: 123-125.
- Ozukum, W., Avinash, H., Dubey, N., Kalubarme, S. and Kumar, M. (2019). Principal component analysis for yield and yield attributing traits in bread wheat (*Triticum aestivum* L.) genotypes under normal sown condition. *Think India J.* **22**: 705-710.
- Reddy, J. P. (2019). Genetic divergence and principal component analysis in bread wheat (*Triticum aestivum* L.) genotypes. M. Sc. thesis, Lovely Professional University, Phagwara, Punjab.
- Reddy, J. P., Dubey, N., Avinash, H., Kumar, M. and Kalubarme, S. (2019). Genetic diversity based on PCA for grain yield and its components in wheat (*Triticum aestivum* L.) genotypes in Punjab region. *Think India J.* **22**: 934-938.
- Reddy, J. P., Kumar, M., Dubey, N., Avinash, H., Rohith, K. and Kalubarme, S. (2022). Genetic divergence for grain yield and its components in bread wheat (*Triticum aestivum* L.). *e J. Plant Breed.* **13**: 258-261.
- Verma S. P., Pathak V. N. and Verma O. P. (2019). Interrelationship between yield and its contributing traits in wheat (*Triticum aestivum* L.). *Int. J. Curr. Microbiol. App. Sci.* **8**: 3209-3215.
- Yadav, P. K., Tiwari, S., Kushwah, A., Tripathi, M. K., Gupta, N., Tomar, R. S. and Kandalkar, V. S. (2021). Morpho-physiological characterization of bread wheat genotypes and their molecular validation for rust resistance genes Sr2, Sr31 and Lr24. *Proc. Ind. Nat. Sci. Acad.* **87**: 534-545.