**Genetic Diversity Assessment of Wheat Genotypes in Bundelkhand Region Using Principal Component Analysis**

**Abstract**

To study the principal component analysis in eighty two wheat genotype including with three check varieties namely HD1544, DBW110 and GW322 in Augmented block design at research farm, Rani Lakshmi Bai Central Agricultural University, Jhansi, During Rabi season 2019-20. Principal component analysis is a procedure describes by Banfield (1978). PCA mostly performed on two types of data matrices viz., variance- covariance matrix and correlation matrix.The percent variance accounted for each principal component (PC) is expressed as the Eigen value divided by the sum of Eigen values.Out of fifteen, seven principal components accounted more than one eigen value viz., PC1 (2.69), PC2 (2.35), PC3 (1.76), PC4 (1.39), PC5 (1.22) and PC6 (1.14) with showed about 70.48 % variability within the traits observed for each genotype. Eigen value and percent of variance associated with every principal, slowly reduced and stopped at 7.63.The first (PC) which accounted maximum variability was more related to the traits viz., grain yield per plant, biological yield per plant and days to maturity so it must be considered. The second one (PC2) included the traits peduncle length, plant height and flag leaf length. The third principal component (PC3) accounted positive effects for 1000 grain weight, flag leaf width and grain yield per plant.The fourth principal component (PC4) was more related to the traits spike length, canopy temperature, 1000 grain weight and days to heading. The fifth principal component (PC5) accounted positive effects for chlorophyll content, canopy temperature, harvest index and spike length, whereas, the sixth principal component (PC6) were more related to flag leaf width, days to maturity, flag leaf length and biological yield per plant.

**Key words**- Variability, Principal Component Analysis, correlation and Eigen value

**Introduction**

“Wheat commonly called as the “King of Cereals” due to its prominent position in the international food grain trade and its high acreage, production, productivity. Wheat, the first domesticated food crop and since 8000 years, is the stable food for various civilizations throughout the world. It is originated from South-Western Asia and the Central Asia and Mediterranean and Ethiopian regions are centres of diversity for wheat and its related species. There are 17 different species of wheat among which three species viz,hexaploid bread wheat(*Triticum aestivum*), Triticum durum (tetraploid macaroni wheat) and tetraploid emmer wheat (*Triticum dicoccum*) are mostly cultivated and consumed throughout the world” (Mishra et al., 2019). “In world, wheat occupies an area of 217 million ha (Mha) with total production of 765 million tonnes (Mt) and productivity of 3530 kg/ha” (USDA, 2020). “China has maximum productivity followed by India, Russia and USA. In India area, production and productivity of wheat was tremendously increased since green revolution of 1967. During 2019-20, 3rd estimation India occupies an area of 30.55 million ha with total production of 107.17 million tonnes and productivity of 3508 kg/ha. Uttar Pradesh leads in area of 9.35 Mha with a production of 32.09 Mt and productivity of 3432 kg/ha in the country and Madhya Pradesh occupies wheat area of 6.02 Mha with production 18.58 Mt and productivity of 3083 kg/ha” (DAC&FW, 2020). “Wheat can be used to convert into innumerable products like chapatis, breads, cakes, biscuits, pasta and many hot and ready-to-eat breakfast foods and considered as a nature’s unique gift to the mankind” (Sharma and Gujral, 2014). “Hexaploid bread wheat genotype improvement any crop is possible through effective breeding program which generally depends on the selection of suitable genotypes, the presence of variation within population for different economic characters is required” (Islam et al., 2004). ). “PCA mostly performed on two types of data matrices viz., variance- covariance matrix and correlation matrix. A correlation matrix which standardizes the original data set is preferred with characters of different scales. A variance- covariance matrix can be used, if the characters are of same scale. In the present study, correlation matrix of traits was used to perform PCA, thereby removing the effects of scale” (Jackson, 1991).

**Material and Methods**

Eighty two genotypes including three checks (HI1544, DBW110 and GW322) observed its fifteen quantitative traits for principal component analysis and character association. This investigation was conducted at Research Farm, Rani Lakshmi Bai Central Agricultural University, Jhansi (Uttar Pradesh) during Rabi, 2019-20 in Augmented Design in five blocks. Seeds of each genotype were sowed in unit plot size of 3 meter long with number of accession in each block is sixteen. The standard agronomic practices were followed for enhanced crop of wheat and competitive crop stand. The observation were recorded on five randomly selected competitive plants from each genotype on fifteen agro-morphological and physiological characters viz; days to 50 per cent heading, days to maturity, canopy temperature (oC), chlorophyll content, flag leaf width (cm), flag leaf length (cm), tillers per meter, peduncle length (cm), plant height (cm), spike length (cm), awn length (cm), 1000-grain weight (g), grain yield per plant (g), biological yield per plant (g) and harvest index (%). Chlorophyll content is measured by SPAD-502 chlorophyll meter as well as canopy temperature was measured by using a hand held infrared thermometer. The analysis of variance for quantitative character was carried out as per standard statistical procedure for Augmented Randomized Complete Block Design (ARCBD) as given by Federer (1956). Principal component analysis is a procedure describes by Banfield (1978). PCA mostly performed on two types of data matrices viz., variance- covariance matrix and correlation matrix. A correlation matrix which standardizes the original data set is preferred with characters of different scales. A variance- covariance matrix can be used, if the characters are of same scale. In the present study, correlation matrix of traits was used to perform PCA; thereby removing the effects of scale (Jackson, 1991).Data matrix was used to compute the Eigen value and Eigen vector. Eigen value is defined as the amount of total variation that is displaced on principal components. The percent variance accounted for each principal component (PC) is expressed as the Eigen value divided by the sum of Eigen values.

**Result and Discussion**

The principal component analysis for fifteen quantitative characters of wheat genotypes as presented in the Table 1(b). Out of fifteen, seven principal components accounted more than one Eigen value viz., PC1 (2.69), PC2 (2.35), PC3 (1.76), PC4 (1.39), PC5 (1.22) and PC6 (1.14) with showed about 70.48 % variability within the traits observed for each genotype. Eigen value and percent of variance associated with every principal, slowly reduced and stopped at 7.63. The results have been presented in Table 1. (a) and 1. (b).The first principal component had maximum contribution of 17.96% with Eigen value 2.69 towards total variation present within the genotypes in a population. PC1 was more associated to grain yield per plant (0.851), followed by biological yield per plant (0.818), tillers per meter (0.668), plant height (0.290) days to heading (0.258), days to maturity (0.187).The second principal component exhibited for 15.66% of the total variation with eigen value 2.35 within the genotypes in the population. In PC2 Peduncle length (0.922) followed by plant height (0.817), flag leaf length (0.463), awn length (0.094), tillers per meter (0.092) and one thousand grain weight (0.015) were the maximum contributed characters.The third principal component showed 11.77% of the total variation with eigen value1.76. PC3 was more related to 1000 grain weight (0.593), flag leaf width (0.438), grain yield per plant (0.397), flag leaf length (0.384), harvest index (0.367), biological yield per plant (0.276), awn length (0.152), spike length (0.143) and peduncle length (0.094). Fourth principal component contributed 9.28% of the total variation with Eigen value 1.39. PC4 was more associated to spike length (0.621) followed by canopy temperature (0.493), one thousand grain weight (0.472), days to heading (0.389), biological yield per plant (0.212), plant height (0.205) and grain yield per plant (0.080).The fifth principal component exhibited 8.15% of the total variation with eigen value1.22. PC5 was more related to chlorophyll content (0.500) followed by canopy temperature (0.303), harvest index (0.294), spike length (0.231), flag leaf length (0.224), grain yield per plant (0.132), peduncle length(0.071), plant height (0.011) and biological yield per plant (0.007).The six principal components showed 7.63% variation with eigen value 1.14. PC6 exhibited positive value for flag leaf width (0.624) followed by days to maturity (0.550), flag leaf length (0.424), biological weight per plant (0.237), spike length (0.210),days to heading (0.157), chlorophyll content (0.152), peduncle length (0.061) and tillers per meter (0.038).

Results obtained from PCA on the correlation matrix of the traits suggests the reduction of the dimensionality of the data by creating several sets of significant principal components having Eigen value more than one. In the present investigation, PCA was performed for all fifteen quantitative traits of wheat genotypes as indicated in table 1.(b)

. It has been observed that out of eight, first five principal components with Eigen value more than one viz. PC1 (2.69), PC2 (2.35), PC3 (1.76), PC4 (1.39) PC5 (1.22) and PC6 (1.14) contributed 70.48 per cent towards the total variation among the traits studied for each genotype. The principal components with Eigen value less than one is considered as non-significant and therefore, not used for determining variance present in the population.

“Principal component analysis (PCA) reflects the importance of the largest contributor to the total variation at each axis of differentiation. The Eigen values are often used to determine how many factors to retain. The PCA was performed for all the morphological traits of wheat genotypes as indicated in the table 1(a). Out of fifteen traits under study, six principal components exhibited more than one Eigen value and showed about 70.48% variability among the traits studied for each genotype. So, these six principal components were given due to importance for the further explanation. The PC1 and PC2 showed exhibited maximum variability whereas, PC5 showed minimum variability among the genotypes for the traits under study” (Poonia et al.2023). According to Chahal and Gosal (2002), characters with largest absolute value closer to unity within the first principal component influence the clustering more than those with lower absolute value closer to zero. Therefore, in the present study, differentiation of the genotypes into different clusters was because of relatively high contribution of few characters rather than small contribution from each character.

The first (PC) which accounted maximum variability was more related to the traits viz., grain yield per plant, biological yield per plant, tillers per meter, plant height, days to heading and days to maturity so it must be considered. The second one (PC2) included the traits peduncle length, plant height, flag leaf length, awn length, tillers per meter and 1000 grain weight. The third principal component (PC3) accounted positive effects for one thousand grain weight, flag leaf width, grain yield per plant, flag leaf length, harvest index, biological yield per plant, awn length, spike length and peduncle length.The fourth principal component (PC4) were more related to the traits spike length, canopy temperature, 1000 grain weight, days to heading, biological yield per plant, plant height and grain yield per plant. The fifth principal component (PC5) accounted positive effects for chlorophyll content, canopy temperature, harvest index, spike length, flag leaf length, grain yield per plant, peduncle length, plant height and biological yield per plant, whereas, the sixth principal component (PC6) were more related to flag leaf width, days to maturity, flag leaf length, biological yield per plant, spike length, days to heading, chlorophyll content, peduncle length and tillers per meter.The PCA helps in identifying the traits contributing maximum towards the existing variability such as grain yield per plant (0.851), peduncle length (0.922), 1000 grain weight (0.593), spike length (0.621), chlorophyll content (0.500), flag leaf width (0.624) in PC1, PC2, PC3, PC4, PC5 and PC6.

**Table 1 (a) Principal components of wheat genotypes with Eigen values and variance**

|  |  |  |  |
| --- | --- | --- | --- |
| Principalcomponent | Eigenvalue | Variance **(%)** | Cumulativevariance **(%)** |
| PC1 | 2**.**695 | 17**.**96 | 17**.**96 |
| PC2 | 2**.**350 | 15**.**66 | 33**.**63 |
| PC3 | 1**.**766 | 11**.**77 | 45**.**40 |
| PC4 | 1**.**393 | 9**.**28 | 54**.**69 |
| PC5 | 1**.**224 | 8**.**15 | 62**.**84 |
| PC6 | 1**.**145 | 7**.**63 | 70**.**48 |
| PC7 | 0**.**933 | 6**.**22 | 76**.**70 |
| PC8 | 0**.**727 | 4**.**84 | 81**.**55 |
| PC9 | 0**.**676 | 4**.**50 | 86**.**05 |
| PC10 | 0**.**605 | 4**.**03 | 90**.**09 |
| PC11 | 0**.**537 | 3**.**58 | 93**.**67 |
| PC12 | 0**.**499 | 3**.**32 | 96**.**99 |
| PC13 | 0**.**316 | 2**.**10 | 99**.**10 |
| PC14 | 0**.**143 | 0**.**89 | 99**.**99 |
| PC15 | 0**.**001 | 0**.**06 | 100 |

**Table 1 (b) Studies on principal component for fifteen characters in wheat**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Characters | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
| Days to 50**%**heading | 0**.**258 | **-**0**.**387 | **-**0**.**227 | 0**.**389 | **-**0**.**465 | 0**.**157 |
| Daystomaturity | 0**.**187 | **-**0**.**025 | **-**0**.**626 | **-**0**.**050 | **-.**095 | 0**.**550 |
| Chlorophyllcontent | **-**0**.**211 | **-**0**.**439 | **-**0**.**185 | **-**0**.**162 | 0**.**500 | 0**.**152 |
| Canopytemperature | **-**0**.**344 | 0**.**090 | **-**0**.**254 | 0**.**493 | 0**.**303 | **-**0**.**028 |
| Flagleafwidth | **-**0**.**330 | **-**0**.**242 | 0**.**438 | **-**0**.**113 | **-**0**.**179 | 0**.**624 |
| Flagleaf length | **-**0**.**147 | 0**.**463 | 0**.**384 | **-**0**.**348 | 0**.**224 | 0**.**424 |
| Tillerspermeter | 0**.**668 | 0**.**092 | **-**0**.**320 | **-**0**.**068 | **-**0**.**053 | 0**.**038 |
| Pedunclelength | 0**.**073 | 0**.**922 | 0**.**094 | **-**0**.**021 | 0**.**071 | 0**.**061 |
| Plantheight | 0**.**290 | 0**.**817 | **-**0**.**101 | 0**.**205 | 0**.**011 | **-**0**.**046 |
| Spike length | **-**0**.**385 | **-**0**.**030 | 0**.**143 | 0**.**621 | 0**.**231 | 0**.**210 |
| Awnlength | **-**0**.**300 | 0**.**094 | 0**.**152 | **-**0**.**287 | **-**0**.**601 | **-**0**.**119 |
| 1000grainweight | **-**0**.**165 | 0**.**015 | 0**.**593 | 0**.**472 | **-**0**.**219 | **-**0**.**079 |
| Biologicalyield**/**plant | 0**.**851 | **-**0**.**191 | 0**.**397 | 0**.**080 | 0**.**132 | 0**.**075 |
| Grainyieldperplant | 0**.**818 | **-**0**.**026 | 0**.**276 | 0**.**212 | 0**.**007 | 0**.**237 |
| Harvestindex | **-**0**.**337 | **-**0**.**392 | 0**.**367 | **-**0**.**214 | 0**.**294 | **-**0**.**274 |

**Graph 1(a) Scree plot graph represent of different principal component with their Eigen value.**



**Conclusion:**

In summary, the Principal Component Analysis (PCA) successfully identified the traits that play a major role in the observed variability of the plant characteristics. The first principal component (PC1) emphasized critical traits, including grain yield per plant, biological yield per plant, and plant height, which are important for future research. The other components, from PC2 to PC6, highlighted additional significant traits like peduncle length, spike length, chlorophyll content, and days to maturity, all of which affect plant growth and yield. This analysis offers valuable insights into the factors that drive plant performance, particularly grain yield, and serves as a foundation for enhancing crop breeding and selection methods. In PCA analysis, the magnitude of variation for a particular trait is meaningful, rather than the sign (+/-) which only shows the direction of variability. Similar findings for PCA were also reported by Bhanupriya et al. (2014), Pooja et al. (2018), Panotra et al. (2018), Hazra et al. (2019), Devesh et al. (2019) and Fouad (2020) where weight of grains per spike, grain yield per plant, grain yield quintal per hectare, harvest index, biological yield respectively, showed the maximum variations in wheat.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.

2.

3.

**References:**

Banfield, C. F. (1978). Principal Component Analysis for genstat. *Journals of Statistics and Computer Simulation*, *6*, 211-222.

Bhanupriya, B., Satyanarayana, N., Mukherjee, S., & Sarkar, K. (2014). Genetic diversity of wheat genotypes based on principal component analysis in Gangetic alluvial soil of West Bengal. *Journal of Crop Weed* *10*(2), 104-107.

Chahal, G.S. and Gosal, S.S. (2002). Principles and procedures of plant breeding: Biotechnological and conventional approaches. Alpha Science Int'l Ltd.

DAC&FW, (2020). Agricultural Statistics at a glance 2019. Department of Agriculture, Cooperation & Farmer Welfare. Government. 45-92.

Devesh, P., Moitra, P. K., Shukla, R. S., & Pandey, S. (2019). Genetic diversity and principal component analyses for yield, yield components and quality traits of advanced lines of wheat. *Journal ofPharmacognosy & Phytochemistry*, *8*(3), 4834-4839.

Hazra, S., Chatterjee, S., Kumar, V. U., Mukherjee, S., & Maji, A. (2019). Multivariate analysis in bread wheat genotypes grown under late sown condition. *International Journal of Crop Science,* *7*(4), 2549-2552.

Federer, M. T. (1961). Augmented design with one way elimination of heterogeneity. *Cornell University, Ithava, New York, USA.*

Foud, H. M. (2020). Principal component and cluster analysis to estimate genetic diversity in Bread Wheat (*Triticum astivum* L.) genotypes. *Journal of Plant Production*, *11*(4), 325-331.

Islam, M.R., Faruquei, M.A.B., Bhuiyan, M.A.R., Biswas, P.S., & Salam, M.A. (2004). Genetic diversity in irrigated rice. *Pakistan Journal of Biological Sciences*, 2, 226-229.

Jackson, J. K. (1991). A user’s guide to Principal Component. *John Wiley and Sons Inc., New York.*

Jackson, J. K. (1991). A user’s guide to Principal Component. *John Wiley and Sons Inc., New York.*

Panotra, N., Sharma, A., Bhat, M. I. J., & Sharma, R. (2018). Principal component analysis of yield characteristics of wheat (*Triticum aestivum* L. emend Fiori and Paol) in intermediate zone of Jammu & Kashmir. *International Journal of Crop Science*, *6*(4), 892-895.

Pooja., Dhanda, S. S., &Beniwal,(2018).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 Pure Applied Bioscience, (6)2, 242-247.*

Sharma, P., Gujral, H.S., 2014. Anti-staling effects of b-glucan and barley flour in wheat flour chapatti. *Food Chemistry*, 145, 102–108.

Poonia, M. K., Kumar, A., Kumar, V., Bhanu, A. N., & Kumar, S. (2023). Genetic variability and character association analysis for seed yield and its attributes in wheat (Triticum aestivum L.). Int. J. Plant Soil Sci, 35(9), 123-131.