Original Research Article

Diversity analysis for yield and its contributing traits in rice germplasm (Oryza sativa L.) using principal component analysis approach.

ABSTRACT

Present study was aimed to assess genetic variation among rice germplasm lines employing multivariate biometrical approach viz., Principal Component Analysis. Analysis of variance as per Augmented Block Design indicated the presence of sufficient genetic variation among rice germplasm lines, while estimates of components of variation revealed maximum contribution of genotypic variance to the phenotypic variance suggesting its exploitation through selection and hybridization. Simultaneously, correlation estimates revealed significant positive association of grain yield with panicle length, 1000 grain weight and grain length indicating suitability of these traits for indirect selection. D^2 statistics grouped germplasm lines into sixteen clusters and among these cluster I consists of 67 germplasm lines forming the largest cluster followed by cluster III (26 lines), cluster II (18 lines), cluster IV (15 lines), cluster V (13 lines), cluster VI (5 lines), cluster VIII (3 lines), cluster VII and cluster IX (2 lines each), while clusters X, XI, XII, XIII, XIV, XV, XVI (1 line each). Inter cluster distances were found to be higher than intra cluster distances and maximum inter cluster distance was observed between cluster IX and VI (73.47) followed by cluster IX and X (73.14).

Principal component analysis transformed eleven interrelated variables into four major principal components having eigen value of more than 1 thereby, indicating that these components are responsible for higher magnitude of variance in the population (76.60 per cent). Among these principal components first component accounted for 32.80 per cent of the total variation while, the second, third and fourth component explained 18.50 per cent, 13.90 per cent and 11.40 per cent of total variation, respectively. Factor loadings of the principal components revealed that principal component 1 had high positive loadings for grain length, panicle length, length/breadth ratio, 1000 grain weight and plant height whereas, principal component 2 had high positive loadings for total number of tillers per plant and number of effective tillers per plant indicating that the first two principal factors can be collectively designated as yield attributing factors. Principal component 3 had high positive loadings for grain breadth, 1000 grain weight and plant height whereas, principal component 4 had high positive loading for 1000 grain weight, days to maturity and days to 50% flowering. PCA biplots revealed that germplasm lines viz., GPL-1, GPL-4, GPL-131, GPL-128, GPL-20, GPL-127, GPL-135, GPL-100, GPL-130 were found to be superior performers for desirable traits and can be used as parents in hybridization programme.

Keywords: Genetic diversity, principal component analysis, rice germplasm

Introduction

Rice (*Oryza sativa* L., 2n=2x=24) is the predominant food crop in India and holds a cardinal place in Indian agriculture. Globally, it provides 27 per cent of dietary energy, 20 per cent of the dietary protein and 3 per cent of dietary fat (Pathak *et al*., 2019). In India during *Kharif* 2020 rice was cultivated over an area of 44.0 million hectares with production and productivity of 120.3 million tonnes and 2.73 tonnes/hectare respectively (Anonymous, 2020). In Jammu and Kashmir during 2018-19 rice was grown over an area of 262.01 thousand hectare with production and productivity of 6161 thousand quintals and 23.51 quintals per hectare respectively (Anonymous, 2019). The extent of genetic variation present in the genetic material and its efficient manipulation along with the selection of germplasm lines with all possible desirable yield and its contributing traits is the key to success of any crop improvement programme (Rai *et al.,* 2013). Thus, in order to exploit a population for trait improvement, it is necessary to understand the magnitude of variability in the population and extent to which the desirable traits are heritable. Genetic diversity analysis is used for estimating and establishing genetic relationship among germplasm collections and simultaneously, for identifying promising diverse parental combinations that will yield segregating progenies with maximum genetic variability (Islam *et al.,* 2012). Among various biometrical techniques, one of the approaches is to apply multivariate statistical tools including Principal Component Analysis (PCA) which is used to uncover similarities between variables and classify the genotypes. In the present study this technique was used to classify the relationships among the traits in a complete multi-trait system. It reduces the data with large number of correlated variables into a substantially smaller set of new variables through linear combination of the variables that accounts most of the variation present in the original variables.

Materials and methods

The present study was carried out during *Kharif* 2019 at Sher-e-Kashmir University of Agricultural Sciences and Technology, Chatha, Jammu. The material for the study consisted of 155 rice germplasm lines and 3 locally adapted varieties as checks. The experiment was laid out in Augmented Block Design in five blocks with replicated checks, having plot size of 1.6 m^2. Row to row and plant to plant spacing was kept 20×15 cm while, standard agronomic and plant protection practices as per package and practice were adopted to raise a good crop. Five

plants per germplasm line were randomly selected and tagged to record data on grain yield and its attributing traits viz., days to 50 per cent flowering, days to maturity, plant height(cm), total number of tillers per plant, number of effective tillers per plant, panicle length, grain yield, 1000 grain weight, grain length, grain breadth and length/breadth ratio. Mahalanobis (1936) D^2 statistic was used to estimate the genetic diversity between populations and D^2 values were clustered using Tocher's method as described by Rao (1952). Principle component analysis (PCA) was carried out following standard procedure of PCA given by Pearson (1901) and Hotelling (1933

Results and Discussion

Analysis of variance revealed significant differences among germplasm lines for all the traits revealing wide range of variation among them thereby, suggesting that there is copious scope of selection and hybridization. These results are in agreement with the studies conducted by Lingaiah *et al.* (2014) and Devi *et al.* (2015) for the traits studied by them in their respective studies. For identification of genetically diverse parents and in order to determine the relative proportion of component traits to total divergence Mahalanobis D^2 statistic is one of the potent tools. In addition, Tocher's method (Rao, 1952) also serves as a tool for clustering of genotypes in different clusters based on their D^2 values. In the present study, 155 rice germplasm lines were grouped into sixteen clusters (Table 1 and Fig. 1). Cluster I was the largest among all comprising of 67 germplasm lines followed by cluster III with 26 germplasm lines, cluster II with 18 germplasm lines, cluster IV with 15 germplasm lines, cluster V with 13 germplasm lines, cluster VI with 5 germplasm lines, cluster VIII with 3 germplasm lines cluster VII and cluster IX with 2 germplasm lines each, while, clusters X, XI, XII, XIII, XIV, XV, XVI consisted of one germplasm line each. The pattern of distribution of germplasm lines into various clusters was found to be random revealing non parallelism between geographical and genetic diversity. Germplasm lines from same geographical origin had fallen into different clusters indicating that the genetic diversity found among the genotypes belonging to same geographic origin might be due to natural/artificial selection, exchange of breeding material and environmental variation. Other researchers' viz., Devi *et al.* (2015) and Shivani *et al*. (2018) reported similar results, thereby, suggesting that grouping of materials of similar origin into different clusters is an indication of broad genetic base of the genotypes belonging to that origin. The principal component analysis (Table 2 and 3) was performed for eleven yield and

yield component traits and principal components with higher eigen values and variables which had high factor loading were considered as best representative of system attributes. In our study, first four principal components had eigen value greater than one and they cumulatively explained 76.63 per cent of the total variation present in the original data set. So, these four principal components were considered important for further explanation. The first principal component explained 32.8 per cent while, the second, third and fourth principal component exhibited 18.5 per cent, 13.9 per cent and 11.4 per cent variability, respectively among the germplasm lines for the traits under study. Similar results were observed by Ashfaq *et al.* (2012), Khare *et al.* (2014), Ravikumar *et al.* (2015) and Pachauri *et al.* (2017) in their respective studies. The first principal component accounts for as much of the variability in the data as possible and each succeeding component accounts for as much of the remaining variability as possible. Scree plot explains the percentage of variation associated with each principal component and is obtained by drawing a graph between principal component numbers (X- axis) and percentage of variation explained (Y-axis). The Principal Component 1 showed 32.8 per cent variability with eigen value 3.61 which then declined gradually. From the graph, it is clear that the maximum variation was observed in Principal Component 1(Fig.2). The result of the PCA explained the genetic diversity of rice germplasm lines. Eigen values assess the importance and role of each component to total variation, while the factor loading indicates the scale of contribution of every origin variable with which each principal component is associated. Within each principal component, only highly loaded factors or traits were retained for further explanation. Component matrix revealed that Principal Component 1 showed high positive loading for grain length (0.821), length/breadth ratio (0.815), panicle length (0.813) and plant height (0.453), 1000 grain weight (0.488), grain yield (0.288) whereas, it showed high negative loadings for days to 50 per cent flowering (-0.621), days to maturity (-0.516), grain breadth (-0.517). Principal Component 2 enabled high positive loading for total number of tillers per plant (0.894) and number of effective tillers per plant (0.892), length/breadth ratio (0.366) and grain length (0.323). As a result, the first two principal components that explained about 52.13 per cent of the total variation can be concluded to differentiate rest of the germplasm lines on the basis of yield attributing traits (Fig.

3). Principal Component 3 exhibited high positive loading for grain breadth (0.674), 1000 grain weight (0.406), plant height (0.379), grain yield (0.361) total number of tillers per plant (0.284), number of effective tillers per plant (0.286) and high negative loading for length/breadth ratio (-0.428), days to maturity (-0.354) and days to 50 per cent flowering (-0.365). Principal Component 4 had high positive loading for 1000 grain weight (0.619), days to maturity (0.598), days to 50 per cent flowering (0.457) and grain breadth (0.328), grain length (0.314) and grain yield (0.174). Similar results were also reported by Mahendran *et al.* (2015) The prominent traits contributing maximum variability and desegregating in different principal components have the tendency to remain together which may be kept into consideration during utilization of these characters in crop improvement programme as a donor for the associated traits.

The contribution of first two Principal Components to the total variability was maximum (51.3 per cent), thus these two were plotted to reveal the relationship between them (Fig.4). Germplasm line GPL-35, GPL-10, GPL-4, GPL-128, GPL-32, GPL-130, GPL-11, GPL-131, GPL-36, GPL-100, GPL-127, GPL-139, GPL-1, GPL-126, GPL-135, and GPL- 20 clustered towards better side of Principal Component 1. Germplasm line GPL-124, GPL-4, GPL-20, GPL-87, GPL-129, GPL-135, GPL-108, GPL-99, GPL-125, GPL-128, GPL-61, GPL-100, GPL-127, GPL-1, GPL-130, GPL-131 clustered towards better side of Principal Component 2 (Fig.4). Since these two principal components contribute more than 50 per cent of the total variability along with high positive loading of yield attributing traits, it can be concluded that germplasm lines viz., GPL-1, GPL-4, GPL-131, GPL-128, GPL-20, GPL-127, GPL-135, GPL-100, GPL-130 which clustered towards the better side of both Principal Component 1 and Principal Component 2, were superior collectively for yield and its attributing traits among all the lines studied.

References

Anonymous. 2020. Annual Report 2020.ICAR- National Rice Research Institute Cuttack, India.

Anonymous. 2019. Digest of statistics, Directorate of Economics and Statistics, J&K Govt.

Ashfaq, M., Khan, A. S., Khan, S. H. U. and Ahmad, R. 2012. Association of various morphological traits with yield and genetic divergence in rice (*Oryza Sativa* L.). *International Journal of Agriculture and Biology*, **14**: 55-62.

Devi, G. N., Babu, V. R., Padmavathi, G. and Sunitha, T. 2015. Genetic Diversity in Grain Quality Traits of Rice Genotypes. *Journal of Rice Research*, **8**(1): 6-14.

Hotelling H. 1933. Analysis of a complex of statistical variables into principal components. J Educ Psychol. 25: 417-441

Islam, A.S.M., Ali, R.M., Gregorio., B.G and Islam, M.R. 2012. Genetic diversity analysis of stress tolerant rice (*Oryza sativa* L.). African Journal of Biotechnology, **11**(85):15123-129.

Khare, R., Singh, A. K., Eram, S. and Singh, P. K. 2014. Genetic variability, association and diversity analysis in upland rice (*Oryza sativa* L). *SAARC Journal of Agriculture*, **12**(2): 40-51.

Lingaiah, N., Venkanna, V. and C. Cheralu. 2014. Genetic variability analysis in rice (*Oryza sativa* L.). *International Journal of Pure and Applied Bioscience*. **2**(5): 203-204.

Mahendran, R., Veerabadhiran, P., Robin, S., & Raveendran, M. 2015. Principal component analysis of rice germplasm accessions under high temperature stress. *International Journal of Agricultural Science and Research*, **5**(3): 355- 360.

Mvuyekure, S. M., Sibiya, J., Derera, J., Nzungize, J. and Nkima, G. 2018. Application of principal components analysis for selection of parental materials in rice breeding. *Journal of Genetics & Genomic Sciences*, **3**: 010.

Pachauri, A. K., Sarawgi, A. K., Bhandarkar, S. and Ojha, G. C. 2017. Agro- morphological characterization and morphological based genetic diversity analysis of rice (*Oryza sativa* L.) germplasm. *Journal of Pharmacognosy and Phytochemistry*, **6**(6): 75-80

Pathak, H., Voleti, S. R., Meera Shaik, N., Tripathi, R., Sailaja, B., Nayak, A. K., Subba Rao, L. V., Mondal, B., Reddy, J. N. and Mohapatra, T. 2019. Reorientation of All India Coordinated Crop Improvement Projects: The Case of Rice. NRRI Research Bulletin No. 18.

Pearson K. 1901. On lines and planes of closest fit to systems of points in space. Philos Mag A. 6: 559-572.

Rai, K. N., Govindaraj, M. and Rao, A. S. 2012. Genetic enhancement of grain iron and zinc content in pearl millet. Quality Assurance and Safety of Crops and Food, 4(3):119-125.

Ravikumar, B. N. V. S. R., Kumari, P. N., Rao, P. V. R., Rani, M. G., R., Satyanarayana, P. V., Chamundeswari, N., Vishnuvardhan, K. M., Suryanarayana, Y., Bharathalakshmi, M., and Reddy, A.V. 2015. Principal component analysis and character association for yield components in rice (*Oryza sativa* L.) cultivars suitable for irrigated ecosystem. *Current Biotica*, **9** (1): 25-35.

Shivani, Dwivedi, D. K., Husain, R., Gyanendra, K. and Khan, N. A. 2018. Genetic divergence for yield and other quantitative traits in rice (*Oryza sativa* L.). *International Journal of Current Microbiology and Applied Sciences*, **7**(1): 1201-1207.

Table 1: Distribution of germplasm lines in different clusters

Cluster no.	Number of germplasm lines	Nomenclature
I	67	GPL-99,GPL-108,GPL-107,GPL-120,GPL-116,GPL-109,GPL-29,GPL-23,GPL-51,GPL-97,GPL-144,GPL-148,GPL- 151,GPL-2,GPL-98,GPL-45,GPL-5,GPL-85,GPL-143,GPL-92,GPL-103,GPL-130,GPL-110,GPL-80,GPL-28,GPL-111,GPL-93,GPL-7,GPL-123,GPL-39,GPL-83,GPL-24,GPL-133,GPL-6,GPL-4,GPL-43,GPL-17,GPL-41,GPL-37,GPL-15,GPL-100,GPL-31,GPL-54,GPL-150,GPL-33,GPL-22,GPL-58,GPL-115,GPL-136,GPL-30,GPL-21,GPL-84,GPL-42,GPL-64,GPL-19,GPL-131,GPL-53,GPL-66,GPL-50,GPL-119,GPL-46,GPL-20,GPL-32,GPL-90,GPL-72,GPL-125,GPL-152
II	18	B-564, JB- 129, GPL-11, GPL-12, GPL-35, GPL-126, GPL-102, GPL-18, GPL-10, GPL-8, GPL-1, GPL-9, GPL-137, GPL-3, GPL-112, GPL-149, GPL-34, GPL-67
III	26	GPL-61, GPL-62, GPL-48, GPL-49, GPL-95, GPL-76, GPL-113, GPL-114, GPL-56, GPL-65, GPL-52,GPL-69,GPL-38,GPL- 40,GPL-27,GPL-75,GPL-63,GPL-55,GPL-36,GPL-77,GPL-70,GPL-121,GPL-73,GPL-89,GPL-78,GPL-81
IV	15	GPL-26, GPL-141, GPL-104, GPL-57, GPL-71, GPL-60, GPL-91, GPL-68, GPL-118, GPL-106, GPL-74,GPL-82,GPL-140,GPL-132,GPL-25
V	13	GPL-139, GPL-145, GPL-101, GPL-13, B-370, GPL-153, GPL-14, GPL-135, GPL-142, GPL-147, GPL-146, GPL-16, GPL-154
VI	5	GPL-47, GPL-44, GPL-86, GPL-87, GPL-96
VII	2	GPL-128, GPL-134
VIII	3	GPL-88, GPL-94, GPL-59
IX	2	GPL-138, GPL-155
X	1	GPL-79
XI	1	GPL-105
XII	1	GPL-117
XIII	1	GPL-122
XIV	1	GPL-124
XV	1	GPL-127

XVI	1	GPL-129

Table 2: Eigen values, proportion and cumulative proportion of variation

S. no.	Eigen value	Proportion	Cumulative Proportion
1	**3.611**	**0.328**	**0.328**
2	**2.04**	**0.185**	**0.513**
3	**1.527**	**0.139**	**0.652**
4	**1.255**	**0.114**	**0.766**
5	0.894	0.081	0.848
6	0.838	0.076	0.924
7	0.405	0.036	0.961
8	0.262	0.024	0.985
9	0.158	0.014	0.999
10	0.005	0.0005	0.999
11	0.001	0.00009	1.00

Table 3: Factor loadings of principal components

Trait	Principal Components			
	PC 1	**PC 2**	**PC 3**	**PC4**
Days to 50% flowering	-0.621	0.227	-0.365	0.457
Days to maturity	-0.516	0.197	-0.354	0.598
Plant height	0.453	-0.086	0.379	-0.026
Total number of tillers per plant	-0.286	0.894	0.284	-0.183
No. of effective tillers per plant	-0.289	0.892	0.286	-0.178
Panicle length	0.813	0.003	0.105	0.054
Grain yield	0.288	0.064	0.361	0.174
1000 grain weight	0.488	0.121	0.406	0.619
Grain length	0.821	0.323	-0.130	0.314
Grain breadth	-0.517	-0.295	0.674	0.328
Length/Breadth ratio	0.815	0.366	-0.428	0.02

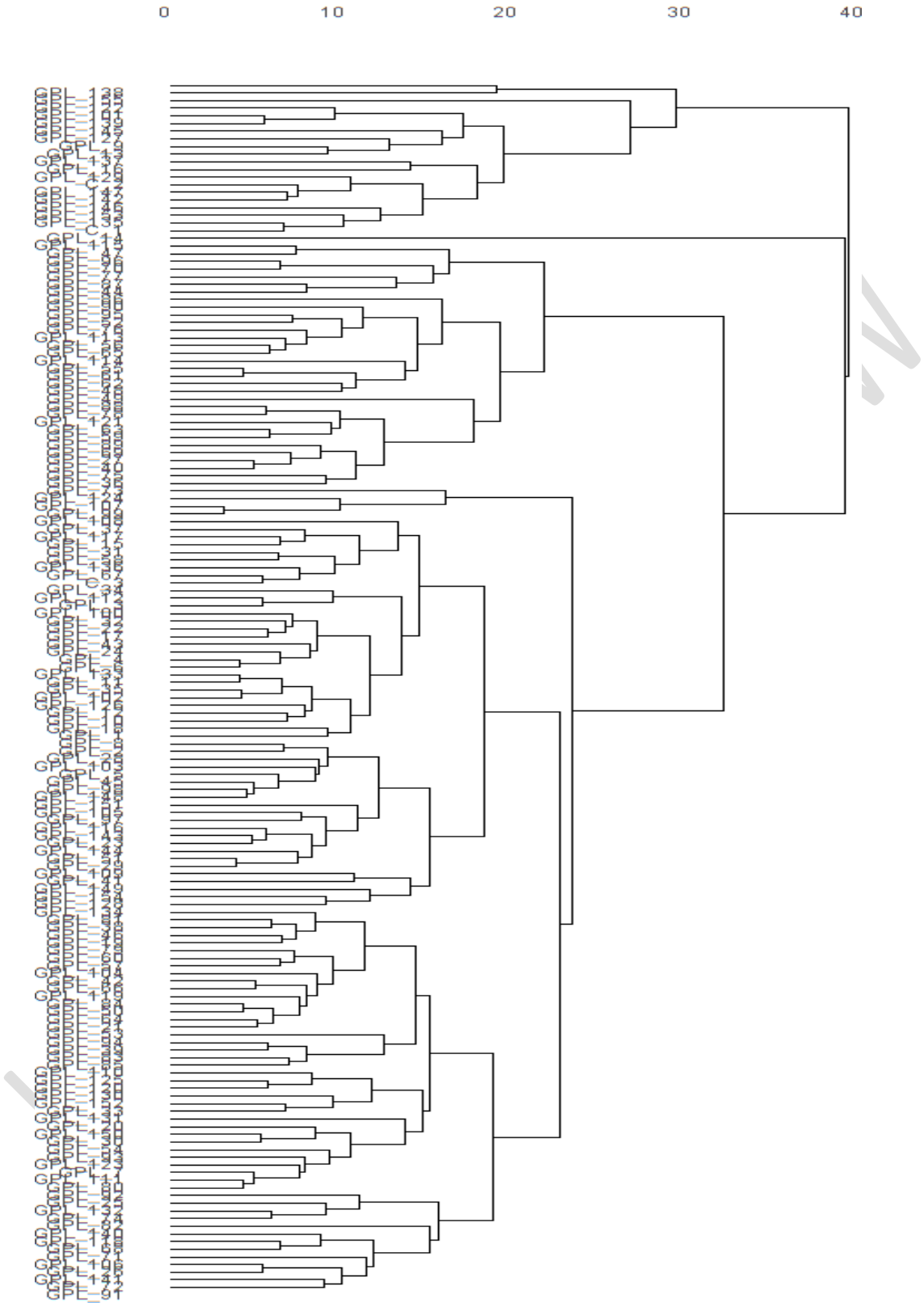

Fig. 1: Dendrogram exhibiting distribution of germplasm lines in clusters

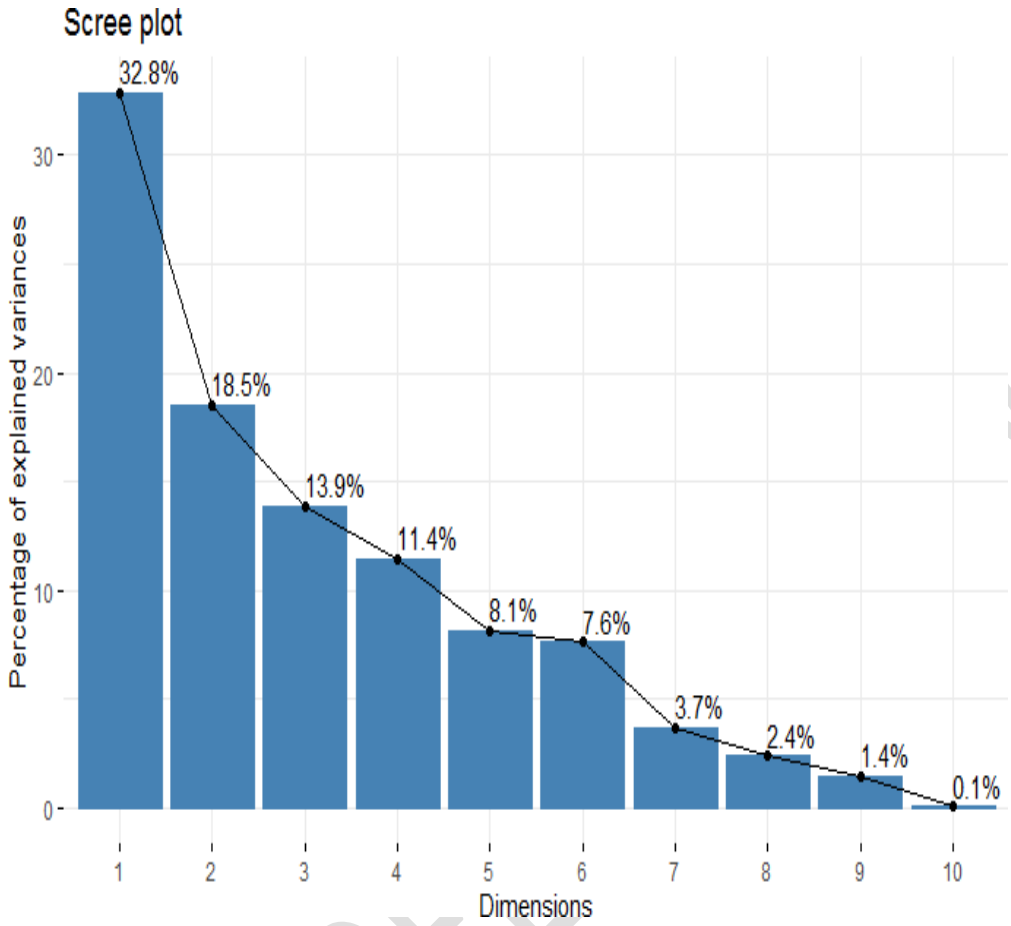

Fig. 2: Scree plot showing Principal Components and percentage of variation explained

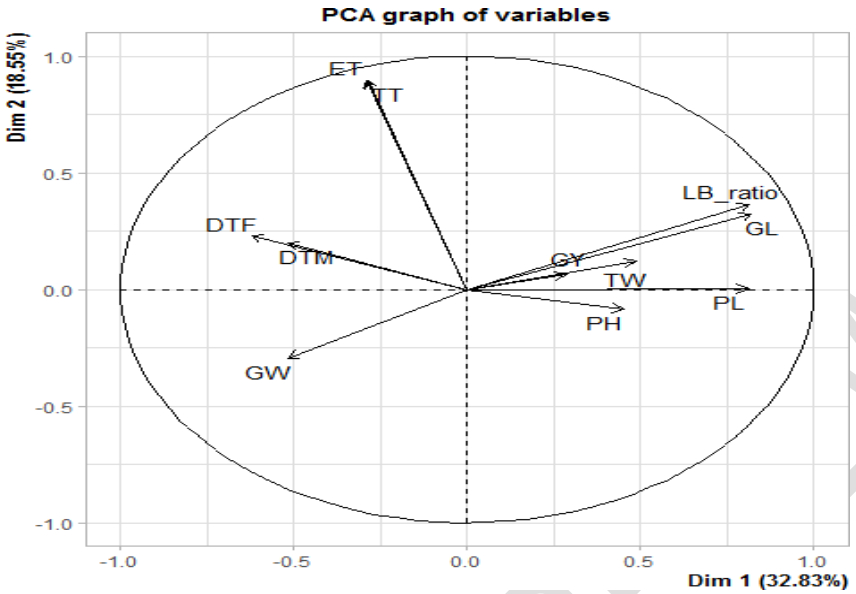

Fig. 3: PCA graph of different traits for first two principal components

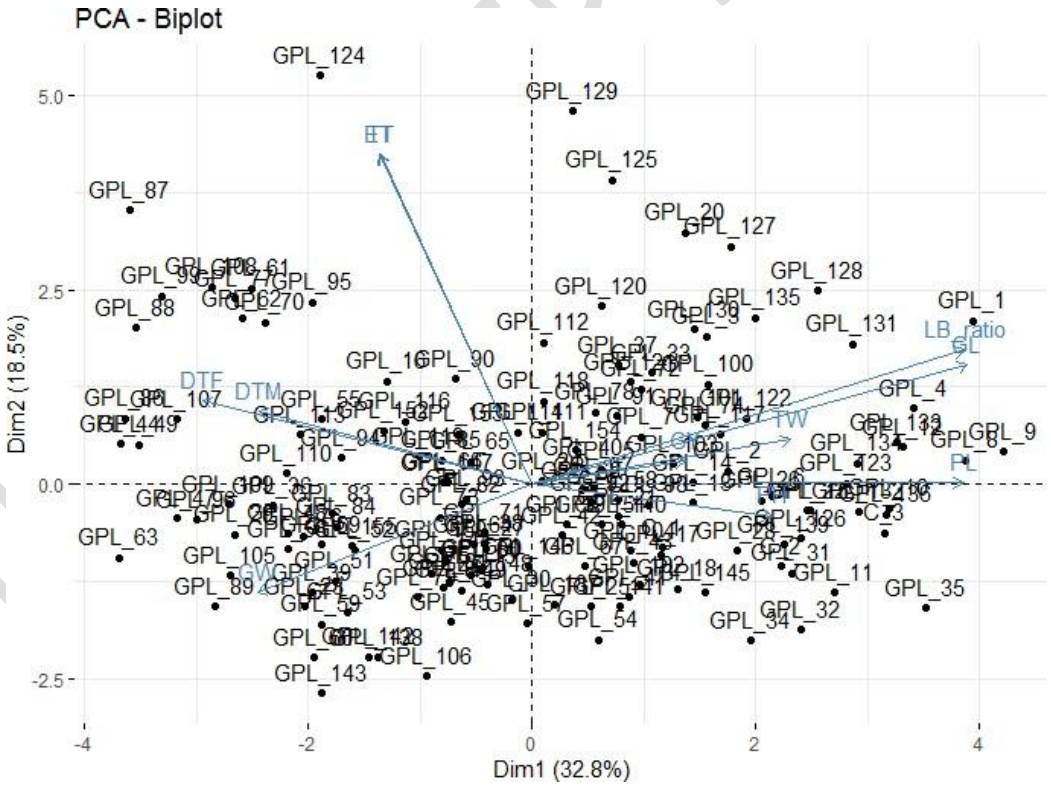

Fig. 4: Distribution and grouping of germplasm lines across first two principal components