IJPS INTERNATIONAL JOURNAL OF PLANT SCIENCES
ISSN-0973-1547 Volume 17 | Issue 2 | July, 2022 | 191-196
DOI: 10.15740/HAS/IJPS/17.2/191-196
Visit us - www.researchjournal.co.in

RESEARCH ARTICLE

Genetic diversity using principal component analysis and hierarchical cluster analysis in rice

■ S. C. Talekar, M. Vani Praveena and R. G. Satish

SUMMARY

A set of 100 germplasm lines with four checks *viz*., BPT-5204, PSB-68, Siri1253 and MGD-101 were evaluated in augmented block design during *Kharif* 2020. The observations were documented for 5 quantitative traits *viz*., days to 50% flowering, panicle length, number of panicles per square meter, 1000 grain weight and grain yield by principal component analysis and cluster analysis to determine the relationship and genetic divergence among the individuals. The cumulative variance of 55.60% was explained by 1st two principal components (PC1 and PC2) with eigen values greater than 1. Component 1 with variance of 32.10% had contribution from days to 50% flowering, panicle length, panicles per square meter and grain yield while principal component 2 accounting 23.50% total variability has contribution from days to 50% flowering and panicles per square meter. The remaining variability of 17.68%, 16.10% and 10.60% was consolidated in PC3, PC4 and PC5. Results from cluster analysis grouped 100 germplasm lines into four clusters with minimum individuals constituted in cluster 1 and maximum individuals were found in cluster 4. The lines in cluster 1 (2.62) showed maximum divergence followed by cluster 3 (2.23). The maximum inter cluster Euclidean distance was observed between clusters 2 and cluster 3 followed by cluster 1 and cluster 2 giving a scope for selection of parents for hybridization programme from these clusters to realize high genetic variation and novel combinations for yield increment.

Key Words : Principal component analysis, Hierarchical cluster analysis, Rice

How to cite this article : Talekar, S.C., Vani Praveena, M. and Satish, R.G. (2022). Genetic diversity using principal component analysis and hierarchical cluster analysis in rice. *Internat. J. Plant Sci.*, **17** (2): 191-196, **DOI: 10.15740/HAS/IJPS/17.2/191-196,**

Article chronicle : Received : 28.04.2022; **Revised :** 06.05.2022; **Accepted :** 07.06.2022

MEMBERS OF THE RESEARCH FORUM

Author to be contacted :
S. C. Talekar, All India Co-ordinated Maize Improvement Project, Main Agricultural Research Station, University of Agricultural Sciences, **Dharwad (Karnataka) India**
Email : talekarsc@uasd.in

Address of the Co-authors:
M. Vani Praveena and R.G. Satish, Department of Genetics and Plant Breeding, Acharya N.G. Ranga University of Agriculture, **Bapatla (A.P.) India**
Email : vanipraveenamadhunapantula@gmail.com

R.G. Satish, All India Coordinated Rice Improvement Project, Agricultural Research Station, **Mugad (Karnataka) India**

Rice is the major cereal crop with 95% of rice production observed in Asian continent and consumed by more than 4 billion people in the world. In India, rice is grown in an area of 43 million hectares, with 112 mt production and 2.6 tha^{-1} productivity (Pathak *et al.,* 2020). The Indus valley civilization is credited with the domestication of rice in India. In Kerala, rice is planted below sea level; most rice-growing locations of the country are located at or near sea level, as well as at heights above 2000 metres in Kashmir. However, rice is cultivated in all corners of the world

except Antarctica (Rathna Priya *et al.,* 2019). After China, India is the world's second largest producer of rice, accounting for 22.5 per cent of global rice production. Rice grain contains highest lysine content (4%) and endosperm consists of water soluble albumin, salt soluble globulin, 8% alcohol soluble prolamine. Rice protein is preferable over other cereal proteins since it has balanced amino acid profile due to more lysine content (Julino, 1993).

The per capita net availability of rice increased from 70.1 kg/year in 2019 to 73.4 kg/year in 2020. From 1960 to 2016 the volume of international rice marketing has almost increased by six-fold from nearly 8 million tonnes to 44 million tonnes annually. Plant disease and pest outbreaks are becoming more common, resulting in lower crop yields, endangering global food security. The important diseases include chickpea dry root rot (Talekar *et al.,* 2017), sorghum downey mildew (Sharma *et al.,* 2010), chickpea phyllody (Balol *et al.,* 2021), pigeonpea wilt (Saxena *et al.,* 2012), groundnut leaf spots (Kolekar *et al.,* 2016) and bud necrosis (Balol and Patil, 2014) and sunflower necrosis (Sundaresha *et al.*, 2012) affecting the yield levels significantly. To maintain current food self-sufficiency and fulfil future food and export demands, production must increase roughly at a rate of 1.5 mt per year by 2050, and productivity must increase by about 30% *i.e.*, from current level of 2.56 t/ha to 3.25 t/ha. (Pathak *et al.,* 2021). Hence, there is an urgent need to increase the rice production with limited land resource using high yielding varieties/hybrids using heterosis breeding and other novel breeding methods (Padmavathi, 2012).

The amount of genetic variability present in the population is the primary requirement for developing high yielding rice varieties outperforming present day cultivated varieties. Thus, studies on genetic diversity to know the nature and degree of divergence for yield traits will help in designing efficient breeding programme by facilitating the selection of elite parental lines which can be further used in hybridization programme to develop superior and high yielding varieties. Principle component analysis (PCA) is an effective approach for quantifying genetic divergence among germplasm lines with respect to characters Beena *et al.,* 2020.Also cluster analysis offers a best method for selecting parents to realize high heterosis to strengthen hybrid breeding programme in rice. With this background the present study was carried out to understand and classify the magnitude and nature of genetic diversity among the studied germplasm lines.

MATERIAL AND METHODS

Experiment material for the present study comprised of 100 germplasm lines and four checks *viz.*, MGD-101, Siri1253, BPT5204 and PSB-68. The genotypes were sown in Agriculture Research Station, Mugad, University of Agricultural Sciences, Dharwad. Test entries along with checks were sown at a spacing of 20×10 cm in augmented Block Design with four blocks, wherein each block comprised of 25 genotypes and four checks were repeated in each block. The data was recorded on a plot and plant basis at various agronomic stages of the crop that were appropriate for each measured trait. Data related to days to 50% flowering, panicle length, panicles per square metre, 1000-grain weight and grain yield was collected and analysed to draw valid conclusions.

PCA analysis was done as per the methodology suggested by Massay (1965) and Jolliffie (1986). Cluster analysis was done using mean data following ward method of diversity analysis. Principal component analysis and cluster analysis was performed in R software version 4.1.3.

RESULTS AND DISCUSSION

The results obtained from the present investigation as well as relevant discussion have been summarized under following heads :

Principle component analysis:

Polygenic variation is evaluated by looking at the relative contribution of each character to total divergence. Principal component analysis is a non-parametric, multivariate technique that analyses a data Table 1 in which several inter-correlated quantitative dependent variables describe observations to extract the important information (Nachimuthu *et al.,* 2014). The first 2

Table 1: Eigen values, Proportion of variance and cumulative effect of five different principal components

Components	PC1	PC2	PC3	PC4	PC5
Eigen value	1.60	1.17	0.88	0.80	0.53
Proportion	32.10	23.50	17.68	16.10	10.60
Cumulative	32.10	55.60	73.30	89.40	100

components in principal component analysis contributed 55.60% of variability with eigen values more than 1 (Table 1).The eigen values greater than 1 in PCA suggests that the identified traits had a significant impact on the phenotype of the population and exhibit more variation among the germplasm lines (Christiana *et al.,* 2021). PC 1 with eigen value of 1.60 contributed 32.10% to the total variability, PC2 with eigen value of 1.17 contributed 23.50% of total variation (Fig. 1), while PC3, PC4 and PC5 with eigen values of less than 1 contributed 17.68%, 16.10% and 10.60% respectively to the total genotypic variation. Similar results with six principal components were reported by Suneetha (2018) and Pathak *et al.* (2018). The contribution of traits studied to different principal components was presented in Table 2. Principal component 1 accounting 32.10% to the total variability (Fig. 1) has contribution from days to 50% flowering (0.35), panicle length (0.55), panicles per square meter (0.50) and grain yield (0.50), while principal component 2 accounting 23.6% total variability has contribution only from days to 50% flowering (0.48) and panicles per square meter (0.26). From the Table 1 it is clear that PC 1 reported maximum variation than other principal components studied. Hence individuals selected from this component (PC1) will be desirable in breeding programmes for improving the traits contributing high variability *viz*., panicle length, panicles per square metre

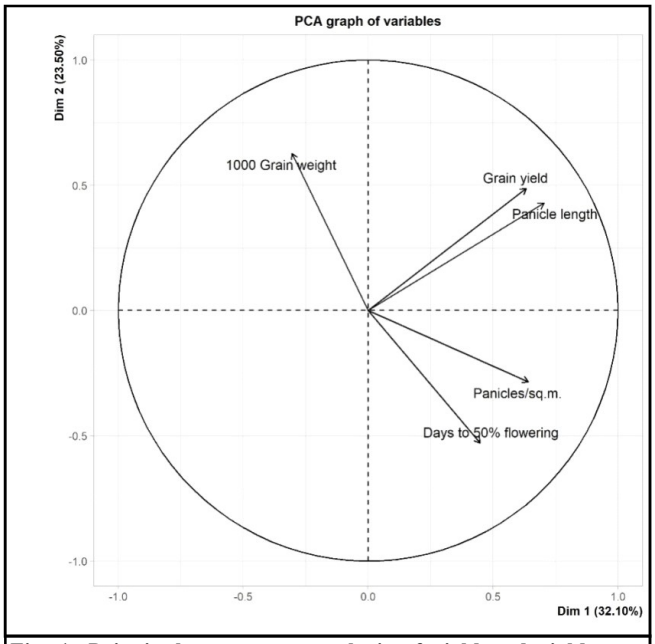

Fig. 1: Principal component analysis of yield and yield related traits in rice

Table 2: Contribution of five quantitative traits to different principal components

Characters	PC1	PC2	PC3	PC4	PC5
Days to 50% flowering	0.35	0.48	-0.70	0.16	-0.34
Panicle length	0.55	-0.39	0.31	-0.04	-0.65
Panicles per square meter	0.50	0.26	0.08	-0.73	0.36
1000 Grain weight	-0.24	-0.57	-0.59	-0.49	-0.10
Grain yield	0.50	-0.45	-0.22	0.43	0.55

Table 3 : Average intra cluster and inter cluster euclidean distance values

Clusters	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	2.62			
Cluster 2	4.49	2.10		
Cluster 3	3.01	4.64	2.23	
Cluster 4	3.55	3.90	3.08	2.20

Table 4: Cluster means for yield and yield related traits in rice

Characters	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Days to 50% flowering	101.20	105.52	86.61	102.46
Panicle length	24.00	22.53	21.47	19.34
Panicles per square meter	280.20	307.89	269.30	270.63
1000 Grain weight	25.00	18.26	24.96	21.10
Grain yield	5665.00	2412.96	1562.68	1625.18

and grain yield. Similar results were reported by Jeevanapriya *et al.* (2019), Rahangdale *et al.* (2021) and Kumari *et al.* (2021).

Cluster analysis:

There are four clusters which could be framed from cluster analysis (Fig. 2). Out of 100 germplasm lines, the cluster 1 is made of individuals five individuals GPM 4, GPM 8, GPM 11, GPM 5 and GPM 6. Cluster 2 have twenty seven individuals, cluster 3 is formed with twenty three individuals while cluster 4 have rest of forty five germplasm lines along with four checks used in the study. In this study, the genotypes in cluster 1 are characterised by early maturity, high grain yield, long panicle length and medium 1000-grain weight. Cluster 2 is characterized by individuals which are early maturing types with medium panicle length and low 1000-grain weight. The individuals in cluster 3 are characterized by very early flowering and medium 1000-grain weight. The individuals in cluster 4 are grouped based on early flowering, short panicle length and more panicles per square meter.

The average intra-cluster and inter-cluster Euclidean distances were estimated using ward's minimum variance and the results were tabulated in Table 3. Among the four clusters, cluster 1 (2.62) has maximum intra-cluster distance followed by cluster 3 (2.23), cluster 4 (2.20) and cluster 2 (2.10). The occurrence of maximum intra-cluster distance indicated the possibility of high genetic diversity among the genotypes within the cluster (Tejaswini *et al.,* 2016). In this study, intra-cluster distances between clusters is less than inter-cluster distance suggesting that it is a good clustering algorithm. Cluster 1 with 5 individuals is far away from cluster 2 (4.49) followed by cluster 4 (3.55) and closer to cluster

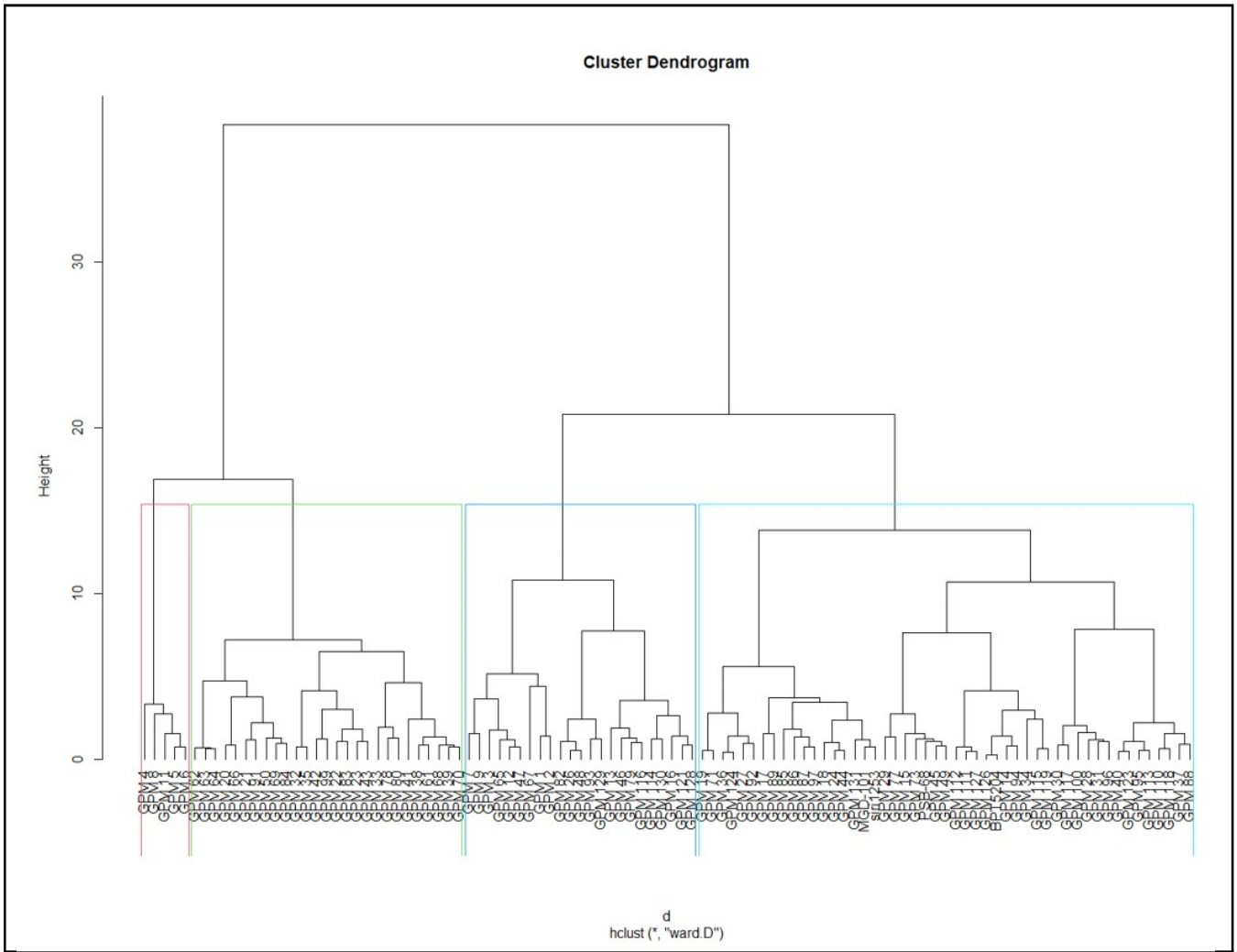

Fig. 2: Cluster dendrogram of 100 germplasm lines and 4 controls evaluated for five yield and yield related traits in rice

3 (3.01). Cluster 2 with twenty seven individuals is far away from cluster 3 with an inter cluster distance of 4.64 and close to cluster 4 (3.90). While, cluster 3 with twenty seven individuals is close to cluster 4 (3.08). From this study crossing genotypes of clusters 2 and cluster 3 followed by cluster 1 and cluster 2 may result in desirable novel recombinants in rice. The individuals grouped in a particular cluster were closely related with other individuals within the cluster in comparison with individuals of other clusters. So, it can be understood that the individuals within a cluster were genetically less diverse while they are highly different from the individuals of other clusters. Similar results were reported by Khare *et al.* (2014) and Tejaswini *et al.* (2016).

Cluster means of characters studied:

The cluster means for days to 50% flowering ranged from 86.61 days (cluster 3) to 105.52 days (cluster 2); panicle length ranged from 19.34 cm (cluster 4) to 24.00 cm (cluster 1); panicles per square meter ranged from 269.30 (cluster 3) to 307.89 (cluster 2); 1000-grain weight ranged from 18.26 (cluster 2) to 25.00 (cluster 1); and grain yield ranged from 1562.68 kg/ha (cluster 3) to 5665 kg/ha (cluster 1). Wider range of cluster means values among the studied germplasm lines indicated the presence of genetic variation among the germplasm lines (Table 4). It was discovered that none of the clusters included at least one genotype that had all of the desirable traits, ruling out the idea of selecting one genotype for immediate usage. Therefore to judiciously incorporate all of the desirable features, hybridization between selected genotypes from divergent clusters is required.

Conclusion:

In the present study two principal components with more than 1 eigen value was obtained based on 5 phenotypic traits contributed for 55.60% of the total variance. The traits days to 50% flowering and panicles per square meter contributed to both PC1 and PC2 while panicle length and grain yield contributed only to PC1. Thus these traits can be effectively used to discriminate the individuals in the population. From cluster analysis maximum inter-cluster distance was observed between clusters 2 and cluster 3 followed by cluster 1 and cluster 2. So the genotypes selected from these clusters can be used for selecting genetically diverse parents in hybridization to achieve a useful output from a well framed breeding programme to develop novel high yielding varieties and to exploit heterosis.

REFERENCES

Balol, G. and Patil, M.S. (2014). Biological characterization and detection of Groundnut bud necrosis virus (GBNV) in different parts of tomato. *J. Pure Appl. Microbiol.,* **8** (1) : 749-752.

Balol, G., Channakeshava, C. and Patil, M.S. (2021). Molecular characterization of Candidatus phytoplasma aurantifolia isolates infecting chickpea (*Cicer arietinum*) in Dharwad, Karnataka. *Legume Res.,* **44** (7) : 854-858.

Beena, R., Veena, V., Jaslam, M..PK., Nithya, N. and Adarsh, V.S. (2021). Germplasm innovation for high-temperature tolerance from traditional rice accessions of Kerala using genetic variability, genetic advance, path coefficient analysis and principal component analysis. *J. Crop Sci. Biotech.,* **24** (5) : 555-566.

Christina, G.R., Thirumurugan, T., Jeyaprakash, P. and Rajanbabu, V. (2021). Principal component analysis of yield and yield related traits in rice (*Oryza sativa* L.) landraces. *Electr. J. Plant Breed.,* **12** (3) : 907-911.

Jeevanapriya, P., Saraswathi, R., Thiruvengadam, V., Surendar, K.K. (2019). Assessment of genetic diversity in new restorer lines of hybrid rice. *Int. J. Curr. Microbiol. Appl. Sci.,* **8** (7) : 530-536.

Jolliffie, I.T. (1986). Principal component analysis. Springer, New York. Mahalanobis PC 1936. On the generalized distance in statistics. *Proc. Nat. Inst. Sci.,* **2**: 49-55.

Juliano, B.O. (1993). *Rice in human nutrition*. FAO Food and Nutrition Series No. 21, Rome, Italy 162.

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 J. Agric*., **12** (2) : 40-51.

Kolekar, R.M., Sujay, V. and Shirasawa, K. (2016). QTL mapping for late leaf spot and rust resistance using an improved genetic map and extensive phenotypic data on a recombinant inbred line population in peanut (*Arachis hypogaea* L.). *Euphytica,* **209** : 147–156.

Kumari, B.K., Kumar, B.R., Jyothula, D. and Rao, N.M. (2021). Diversity analysis in rice breeding lines for yield and its components using principal component analysis. *J. Pharmacog. Phytochem.,* **10** (1) : 905-909.

Massay, W.F. (1965). Principal components regression in exploratory statistical research. *J. Amer. Stat. Assoc.,* **60** : 234-246.

Nachimuthu, V.V., Robin, S., Sudhakar, D., Raveendran, M.,

Rajeswari, S. and Manonmani, S. (2014). Evaluation of rice genetic diversity and variability in a population panel by principal component analysis. *Indian J. Sci. Tech.,* **7** (10) : 1555-1562.

Padmavathi, P.V. (2012). Genetics and stability of promising CMS and restorer lines for yield and quality traits in rice Ph.D. Thesis, Acharya N. G. Ranga Agricultural University, Hyderabad (A.P.) India.

Pathak, H., Tripathi, R., Jambhulkar, N.N., Bisen, J.P and Panda, B.B. (2020). Eco-regional-based rice farming for enhancing productivity, profitability and sustainability. NRRI Research Bulletin No. 22, ICAR-National Rice Research Institute, Cuttack 753006, Odisha, India. 6-30pp.

Pathak, H., Kumar, M., Molla, K.A. and Chakraborty, K. (2021). Abiotic stresses in rice production: impacts and management. *Oryza*, **58** (4) : 103-125.

Pathak, S.K., Lavanya, G.R., Babu, G.S. and Srivastava, N. (2018). Evaluation of rice germplasm for genetic diversity on yield characters by principal component analysis. *The Pharma Innovation J.,* **7** (4) : 661-664.

Rahangdale, S., Singh, Y., Upadhyay, P. and Koutu, G. (2021), Principal component analysis of JNPT lines of rice for the important traits responsible for yield and quality. *Indian J. Genet.,* **81** (1) : 127-131.

Rathna Priya, T.S., Eliazer Nelson, A.R., Ravichandran, K. and Antony, U. (2019). Nutritional and functional properties of coloured rice varieties of South India: A review. *J. Ethnic Foods*, **6** (1) : 1-1.

Saxena, K.B., Kumar, R.V., Saxena, R.K. (2012). Identification of dominant and recessive genes for resistance to *Fusarium* wilt in pigeonpea and their implication in breeding hybrids. *Euphytica,* **188** : 221–227.

Sharma, R., Rao, V.P., Upadhyaya, H.D., Reddy, V.G. and Thakur, R.P. (2010). Resistance to grain mold and downy mildew in a mini-core collection of sorghum germplasm. *https://doi.org/10.1094/PDIS-94-4-0439.*

Sundaresha, S., Sreevathsa, R., Balol, G.B., Keshavareddy, G., Rangaswamy, K.T. and Udayakumar, M. (2012). A simple, novel and high efficiency sap inoculation method to screen for tobacco streak virus. *Physiol. Mol. Bio. Plants*, **18** (4) : 365-369.

Suneetha, K. (2018). Principal component analysis for agro-morphological and quality characters in germplasm of rice. *Int. J. Advanced Bio Res.,* **8** (2) : 1-5.

Talekar, S.C., Lohithaswa, H.C. and Viswanatha, K.P. (2017). Identification of resistant sources and DNA markers linked to genomic region conferring dry root rot resistance in chickpea (*Cicer arietinum* L.). *Plant Breed.,* **136** : 161-166.

Tejaswini, K.L., Manukonda, S., Rao, P.R., Kumar, B.R., Mohammad, L.A. and Raju, S.K. (2016), Cluster analysis studies in rice (*Oryza sativa* L.) using wards minimum variance method. *J. Agric. Crop Res.,* **4** (9) : 129-139.