Received: 27 Dec 2022 | Revised: 23 Jan 2024 | Accepted: 17 Mar 2024

Electronic Journal of Plant Breeding

Research Note

Assessment of variation in rice maintainer lines using principal component analysis

B. Edukondalu[1], V. Ram Reddy[1], T. Shobha Rani[1], CH. Aruna Kumari[2] and B. Soundharya[3]

[1]Department of Genetics and Plant Breeding, Agriculture College, PJTSAU, Jagtial, , Telangana -505 529
[2]Department of Crop Physiology, Agriculture College, Jagtial, PJTSAU, Telangana -505 529
[3]Department of Genetics and Plant Breeding, Regional Sugarcane and Rice Research Station, PJTSAU, Rudrur, Nizamabad, Telangana (503 188), India
***E-Mail:** edukondalu0208@gmail.com

Abstract
The aim of this study was to explore the characteristics essential for a maintainer line to effectively complement the A lines in hybrid rice production. The experiment was conducted at the Regional Agricultural Research Station, Jagtial and Telangana, India during *kharif*, 2016 (June-October). A total of 40 genotypes were raised in Randomized Block Design (RBD) with two replications. PCA identified five principal components (PCs) with Eigen values over 1, collectively accounting for approximately 75.50% of the total variance. PC1 predominantly representing yield and related features (number of tillers per plant, panicle length, length-to-breadth ratio, grain yield per plant), while the other PCs corresponded to unique aspects like grain numbers, morphological and quantitative traits. The study also utilized biplot analysis to elucidate the relationships among these traits, revealing significant correlations and interactions crucial for rice breeding. It indicated a negative correlation between 1000 grain weight, kernel breadth, and the number of grains per panicle, while showing positive correlations among traits influencing grain yield. This method also proved assistance in identifying superior genotypes for specific traits, as exemplified by genotypes JMS18B and JMS20B excelling in grain numbers per panicle and genotype B18 standing out in grain yield and other yield-related traits.

Keywords: PCA, Maintainer lines, Hybrid rice, Biplot

Heterosis, also known as hybrid vigour, is the phenomenon of two genetically distinct parents producing an F_1 hybrid that surpasses the parental lines in multiple phenotypic characteristics (Birchler *et al.,* 2015). Harnessing heterosis is an important strategy to improve the yield and broaden the adaptability of rice crop. Rice is an important cereal crop worldwide. Strong hybrid vigour has been demonstrated in rice, leading to a yield increase of approximately 20% in the F_1 hybrid compared with the parental lines (Cheng *et al.,* 2007; Ma and Yuan 2015). Hybrid rice has been commercialized in China since the 1970s, and it has significantly contributed to rice production and food supply domestically and globally (Li and Yuan 2000). Currently, two technologies are widely used in hybrid rice production: the three-line method and the two-line method. However, restorer lines face a severe germplasm limitation, because they must carry specific nuclear *Rf* alleles for a given sterile line. With the discovery of photoperiod-sensitive genic male sterility (PGMS) (Zhang and Yuan 1987) and temperature-sensitive genic male sterility (TGMS) (Chen *et al.,* 1994) and (Zhou *et al.,* 2014), rice breeders developed a two-line hybrid system consisting of a sterile line and a restorer line. In this system, because both PGMS and TGMS alleles are nuclear-recessive and male fertility can be restored by nearly all normal cultivars

(Ding *et al.,* 2012) (Zhou *et al.,* 2012) there is no germplasm limitation for the restorer lines, and nearly all rice accessions can be used as restorer lines. Thus, the two-line hybrid system has facilitated rice hybrid breeding compared with the three-line system (Wu *et al.,* 2016). But the two-line system is generally less stable than the three-line system, because the fertility of PGMS/TGMS is vulnerable to uncontrollable weather fluctuations (Tao *et al.,* 2003). Moreover, reliance on a single source of cytoplasmic male sterility (CMS), specifically the Wild Abortive (WA) system, in all hybrids released in the country raises concerns of genetic vulnerability and a narrow genetic base. Addressing the genetic vulnerability associated with a prolonged reliance on a single source of cytoplasmic male sterility (CMS), such as the Wild Abortive (WA) system, becomes increasingly pertinent in the face of evolving challenges posed by diseases and insect pests. Therefore, diversifying the genetic makeup of maintainer lines is crucial for hybrid seed production. Maintainer lines, responsible for producing seeds for hybrid varieties, play a key role in ensuring the vigor and productivity of these hybrids through their genetic diversity. The characteristics essential for a maintainer line to effectively complement the A line in hybrid rice production include high fertility, cytoplasmic compatibility, biotic stress resistance, and environmental adaptability , for the successful conversion and development of high-performing hybrids.

Principal Component Analysis (PCA) is a multivariate statistical tool, initially pioneered by Pearson in 1901 and further developed independently by Hotelling in 1933. The primary objective of PCA is to identify the minimum number of components that can explain the maximum variability within the total variability, as highlighted in studies by Anderson (1972) and Morrison (1982). Additionally, PCA facilitates the ranking of genotypes based on their PC scores. By utilizing only a few components, PCA allows each sample to be represented by a relatively small number of values, in contrast to the potentially thousands of variables, a technique elaborated by Ringner, 2008. PCA transforms a set of possibly correlated variables into a set of linearly uncorrelated variables. PCA can be employed to identify germplasm characterization features, visualize individuals' differences and relationships, and assess their contribution to total variation (Singh *et al.,* 2016 and Martínez-Calvo *et al.,* 2008) and This method is particularly valuable for screening large number of genetic resources by a large number of descriptor variables (Beiragi *et al.,* 2001) and Golbashy *et al.,* 2010). In the current study, this method was employed to characterize rice maintainer lines to identify promising ones for exploitation in hybrid rice breeding.

The experiment was conducted at the Regional Agricultural Research Station, Jagtial, Telangana, India during *kharif*, 2016 (June–October). A total of 40 genotypes (**Table 1**) were laid in Randomized Block Design (RBD) with two replications and a spacing of 20×15 cm^2 . Twenty-eight days old seedlings were transplanted in the main field and all the necessary package of practices were followed to raise a healthy crop. Observations were recorded on

Table 1. List of rice genotypes (maintainer lines) utilized for the study

S.No.	Genotype	Source	S.No.	Genotype	Source
1	B1	RARS, Jagtial	21	B24	RARS, Jagtial
2	B2	RARS, Jagtial	22	B35	RARS, Jagtial
3	B3	RARS, Jagtial	23	B42	RARS, Jagtial
4	B4	RARS, Jagtial	24	B86	RARS, Jagtial
5	B5	RARS, Jagtial	25	B88	RARS, Jagtial
6	B6	RARS, Jagtial	26	JMS11B	RARS, Jagtial
7	B7	RARS, Jagtial	27	JMS13B	RARS, Jagtial
8	B8	RARS, Jagtial	28	JMS14B	RARS, Jagtial
93	B9	RARS, Jagtial	29	JMS17B	RARS, Jagtial
10	B11	RARS, Jagtial	30	JMS18B	RARS, Jagtial
11	B12	RARS, Jagtial	31	JMS19B	RARS, Jagtial
12	B13	RARS, Jagtial	32	JMS20B	RARS, Jagtial
13	B15	RARS, Jagtial	33	JMS21B	RARS, Jagtial
14	B16	RARS, Jagtial	34	CMS11B	IRRI, Philippines
15	B17	RARS, Jagtial	35	CMS14B	IRRI, Philippines
16	B18	RARS, Jagtial	36	CMS23B	IRRI, Philippines
17	B19	RARS, Jagtial	37	CMS46B	IRRI, Philippines
18	B20	RARS, Jagtial	38	CMS52B	IRRI, Philippines
19	B22	RARS, Jagtial	39	CMS59B	IRRI, Philippines
20	B23	RARS, Jagtial	40	CMS64B	IRRI, Philippines

https://doi.org/10.37992/2024.1501.024

yield and yield attributing characters and quality traits on five randomly selected competitive plants for each entry in each replication for 15 characters *viz.,* days to 50% flowering (DFF), days to maturity (DM), plant height (PH), number of tillers plant-1 (NTPP), panicle length (PL), number of grains panicle-1(NGPP), kernel length (KL), kernel breadth (KB) L/B ratio (L/B), 1000 grain weight (TW), hulling percentage (HL), milling percentage (ML), head rice recovery percentage (HRR), grain yield plant-1 (GYP), bran oil percentage (BO). PCA was performed using the "FactoMineR" package in R software version 4.1.3 (Husson *et al.,* 2020).

PCA is carried out to reduce the dimensionality of multivariate data by lowering the number of traits responsible to the maximum percentage of total variation of the experimental data in to new variables (components) so as to maximize the information in first few components (Stauffer *et al.,* 1985; Abdi and Williams, 2010). In the present study, PCA analysis of fifteen traits revealed that only five principal components (PCs) with eigenvalues greater than 1 were significant. They accounted for the majority of the variability, representing approximately 75.50% of the variance among the traits studied in 40 maintainer lines. The variations accounted for by PC1, PC2, PC3, PC4, and PC5 were 21.08%, 17.78%, 14.44%, 11.7%, and 10.5%, respectively (**Table 2**). Consequently, these five PCs were prioritized for further discussion in this study.

Scree plot obtained by drawing a graph between Eigen values and principal component numbers (**Fig.1**) explained the percentage of variance associated with each PCs. PC1 showed 21.08% variability with Eigen value 3.16. Similar results were reported by Nalajala *et al.,* 2023. In determining significant traits, an Eigen vector coefficient with an absolute magnitude of 0.3 or higher was used as a threshold, as reported by Badu-Apraku *et al*., 2006. Traits with a coefficient greater than 0.3 were deemed important due to their substantial effect. Conversely, traits with a coefficient value less than 0.2 were considered to have negligible impact on overall variation, as indicated by Laude and Carena 2015 and Sharifi *et al.,* 2018. **Table 3** shows the eigenvectors (loadings) for the first five principal components. PC1 had most of the yield and its related features like the

Table 2. Eigen values, %of variance, and cumulative % of variance for 15 PC components

Traits	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
Eigen value (Root)	3.16	2.67	2.17	1.76	1.58	0.98	0.8	0.62	0.44	0.36	0.18	0.14	0.1	0.04	0.01
Variance (%)	21.08	17.78	14.44	11.7	10.5	6.53	5.36	4.15	2.96	2.41	1.18	0.96	0.65	0.25	0.04
Cumulative variance (%)	21.08	38.86	53.3	65	75.5	82.04	87.39	91.54	94.51	96.92	98.09	99.06	99.71	99.96	100

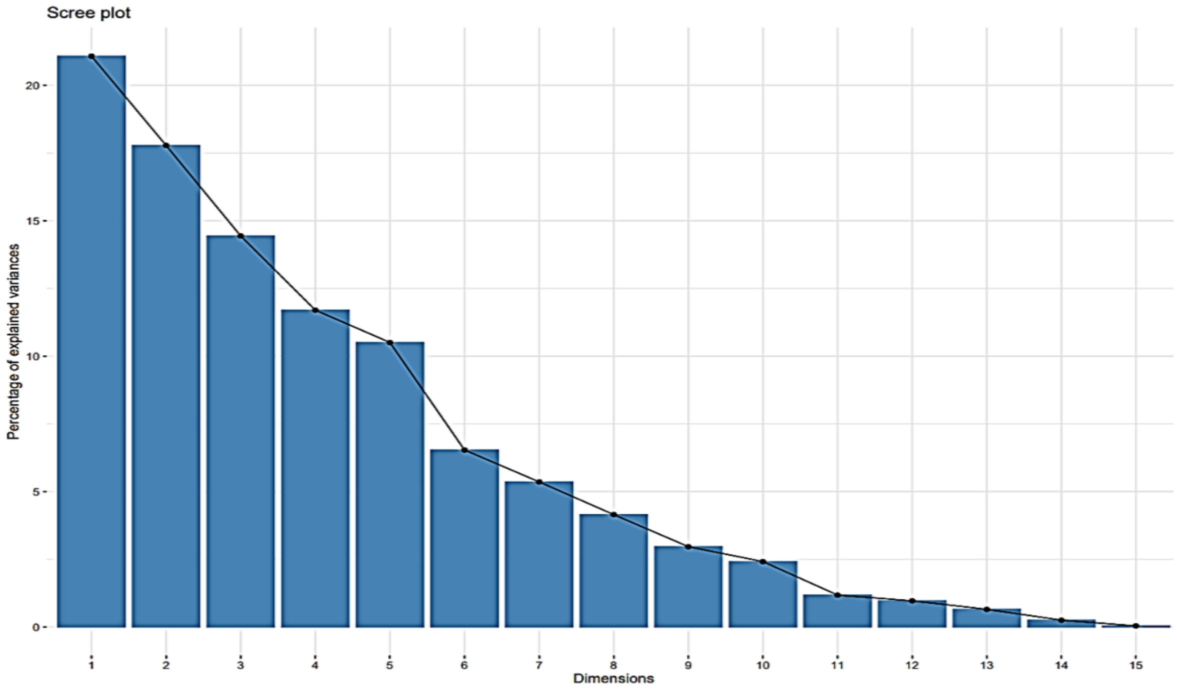

Fig. 1. Scree plot showing the variation between their Eigen value and the number of principal components

Table 3. Principal component analysis for 15 yield and its attributing traits

Traits	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
DF	0.21	-0.04	0.53	-0.26	0.21	-0.11	0.05	-0.04	0.06	-0.05	0.11	-0.07	0.72	0.1	-0.06
DM	0.14	0.02	0.5	-0.34	0.15	-0.16	-0.08	-0.31	0.2	-0.16	-0.18	0.02	-0.6	-0.03	0.05
PHT	0.11	-0.22	0.4	0.4	-0.14	0.18	0.19	0.06	-0.32	-0.27	0.51	-0.21	-0.2	-0.15	0.04
NTP	0.46	-0.11	0	0.03	-0.33	0.03	-0.35	0.07	-0.13	0.08	-0.03	-0.06	-0.09	0.71	-0.06
PL	0.3	-0.29	0.03	0.15	0	-0.06	0.7	0.17	0.05	0	-0.41	0.32	-0.04	0.09	-0.07
NGP	0.16	0.45	0.16	0.02	-0.09	0.04	0.28	-0.18	-0.2	0.73	0.02	-0.2	-0.07	-0.07	0
KL	0.2	-0.25	-0.06	0.35	0.48	0.07	-0.19	-0.24	0.11	0.28	0.1	0.22	0.03	0.02	0.54
KB	-0.29	-0.42	0.17	0.02	-0.19	0.07	-0.13	-0.28	0.07	0.36	0.17	0.41	0	-0.03	-0.49
LB	0.32	0.13	-0.14	0.27	0.52	0.02	-0.13	0	0.02	-0.07	-0.01	-0.14	-0.08	-0.09	-0.67
GYP	0.44	-0.1	0.01	0.05	-0.36	-0.13	-0.35	0.04	-0.1	0.02	-0.26	0.08	0.14	-0.64	0.03
TGW	-0.06	-0.54	-0.17	-0.05	-0.03	-0.08	0.1	-0.27	0.1	0.12	-0.19	-0.72	0.03	-0.03	-0.01
HL	0.27	-0.06	-0.09	-0.31	-0.07	0.64	0.07	0.22	0.52	0.09	0.23	-0.05	-0.06	-0.13	0.01
ML	0.22	0.23	-0.27	0.11	-0.29	-0.06	0.21	-0.68	0.23	-0.31	0.21	0.09	0.13	0.04	-0.02
HRR	-0.11	0.14	0.22	0.46	-0.19	-0.36	-0.07	0.27	0.66	0.11	0.04	-0.14	-0.01	0.03	0
BO	-0.21	0.16	0.28	0.33	-0.07	0.59	-0.11	-0.2	0.03	-0.15	-0.54	-0.07	0.15	0.07	0

number of tillers plant^{-1}, panicle length, length-to-breadth (L/B) ratio, and grain yield plant^{-1}. PC2 was associated with only one factor, the number of grains panicle^{-1}. PC3 is linked to morphological traits such as the days to 50% flowering, days to maturity, and plant height. PC4 was connected to quantitative traits like plant height, kernel length, head rice recovery, and bran oil content. PC5 was related to quantitative traits like kernel length and the L/B ratio.

In a biplot, vectors extending from the origin to each trait marker are used to visualize the relationships among different traits. As described by Yan and Rajcan, 2002, when a biplot accounts for a significant portion of the total variation, the correlation coefficient between any two traits can be estimated by the cosine of the angle between their vectors. Therefore, r = cos180° = -1, cos0° = 1, and cos90° = 0. The most significant variation in the biplot (**Fig. 2**) is attributed to traits such as grain yield per plant, number of tillers per plant, 1000 grain weight, kernel breadth, and number of grains per panicle, as indicated by the lengths of their respective vectors. These interrelationships are particularly crucial in rice breeding. The biplot reveals several key relationships. Yan and Rajcan 2002 noted that traits are positively correlated if the angle between their vectors is less than 90° (acute angles), negatively correlated if more than 90° (obtuse angles), and independent if exactly 90° (near perpendicular vectors). In the present study, a significant negative correlation between 1000 grain weight, kernel breadth, and the number of grains per panicle, was evidenced by the large obtuse angles between their vectors. A positive correlation between the number of tillers per plant and panicle length, both closely related to grain yield per plant was indicated by the acute angles. Additionally, the biplot suggested positive associations among grain yield per plant, hulling percentage, L/B ratio, days to maturity, and days to 50% flowering. These insights are integral to understanding trait interactions in rice breeding. Similar findings were reported by Yan and Rajcan 2002.

Biplots are also valuable for aiding in the selection of genotypes based on various traits. As illustrated in **Fig. 2 and 3**, it is evident that genotypes JMS18B and JMS20B excelled in the number of grains per panicle. Genotypes B42 and B3 were outstanding in terms of 1000 grain weight and kernel breadth, while genotype 16 was superior in grain yield per plant, number of tillers per plant, and panicle length. **Fig. 3** highlights the selection process focusing on grain yield per plant, where genotype B18 demonstrated above-average yield during the culling of genotypes.

These findings align with those of Yan and Rajcan 2002. Their approach involved drawing a line through the biplot origin and the seed yield marker, followed by a perpendicular line to the seed yield line from the biplot origin. According to biplot analysis theory, if the biplot accurately represents the data, genotypes on the same side of the perpendicular line as the seed yield are expected to have above-average yields. Conversely, those on the opposite side are likely to yield below average.

Overall, the PCA and biplot analyses have proven to be robust tools in simplifying complex multivariate data and uncovering critical relationships and variations among traits. These findings are particularly valuable in rice breeding, offering a more understanding of trait interactions and assisting in the efficient selection of superior genotypes. The insights gained from this study align with previous research programs.

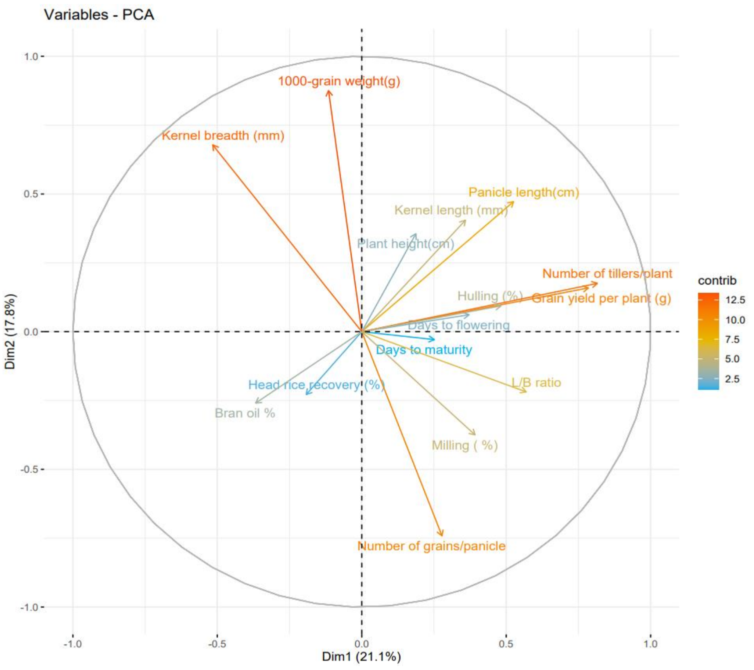

Fig. 2. Loading plot showing contribution of variables towards the first two PCs

Days to 50% flowering (DFF), days to maturity (DM), plant height (PH), number of tillers plant-1 (NTPP), panicle length (PL), number of grains panicle-1(NGPP), kernel length (KL), kernel breadth (KB) L/B ratio (L/B), 1000 grain weight (TW), hulling percentage (HL), milling percentage (ML), head rice recovery percentage (HRR), grain yield plant-1 (GYP), bran oil percentage (BO)

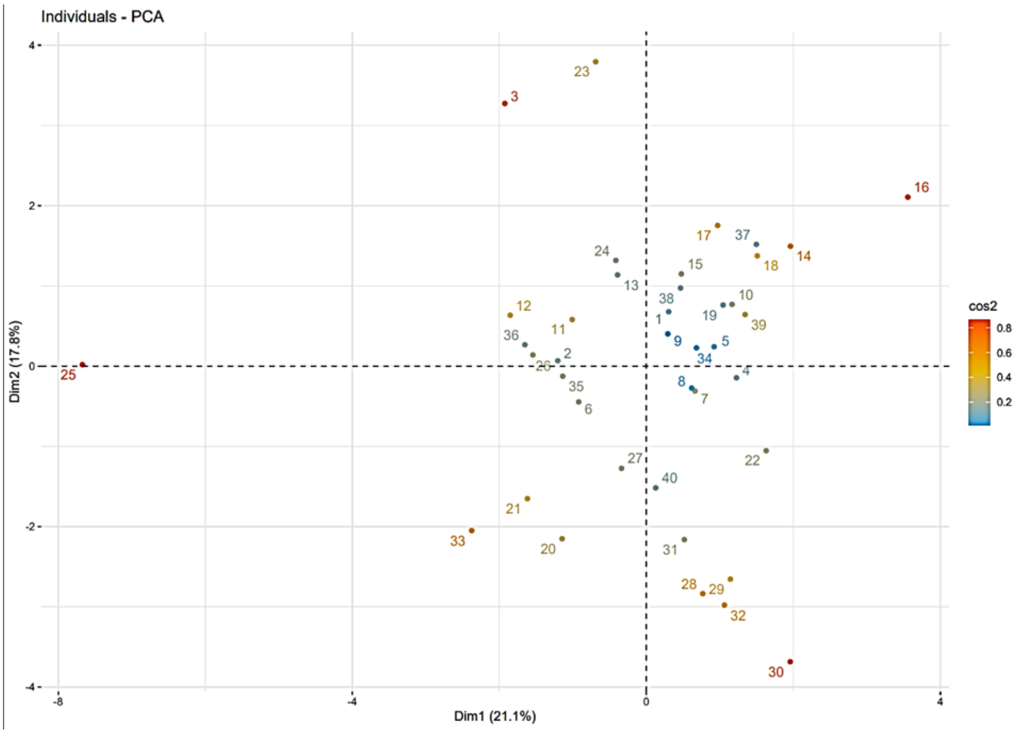

Fig. 3. Biplot showing the distribution of genotypes and contribution of all the haracters in first two principal components

ACKNOWLEDGMENTS

Professor Jayashankar Telangana State Agricultural University (PJTSAU) provided significant materials and funding for this study, which the authors gratefully acknowledge.

REFERENCES

Abdi, H. and Williams, L.J. 2010. Principal component analysis. Wiley interdisciplinary reviews: computational statistics, **2** (4): 433-459. [Cross Ref]

Anderson, T.W. 1972. An introduction to multivariate analysis. Wiley Eastern Pvt. Ltd., New Delhi.

Badu-Apraku, B., Menkir, A., Fakorede, M.A.B., Fontem Lum, A. and Obeng-Antwi, K. 2006. Multivariate analysis of the genetic diversity of forty-seven Striga resistant tropical early maturing maize inbred lines [*Zea mays* L.; Cote d'Ivoire]. Maydica (Italy) **51**(3): 551-559.

Beiragi, M.A., Ebrahimi, M., Mostafavi, K., Golbashy, M. and Khorasani, S.K. 2011. A study of morphological basis of corn (*Zea mays* L.) yield under drought stress condition using correlation and path coefficient analysis. *Journal of Cereals and Oilseeds,* **2**(2): 32-37.

Birchler, J.A. 2015. Heterosis: the genetic basis of hybrid vigour. *Nature plants*, **1**(3): 1-2. [Cross Ref]

Chen, L.B., Zhou, G.Q. and Huang, Y.X. 1994. Influence of temperature and photoperiod on the fertility and physiology of Annong S-1 and Hengnong S-1 of rice. *Acta Botanica Sinica*, **36**: 119-123.

Cheng, S.H., Zhuang, J.Y., Fan, Y.Y., Du, J.H. and Cao, L.Y. 2007. Progress in research and development on hybrid rice: a super-domesticate in China. *Annals of botany*, **100**(5): 959-966. [Cross Ref]

Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., Xu, C., Li, X., Xiao, J. and Zhang, Q. 2012. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proceedings of the National Academy of Sciences,* **109**(7): 2654-2659. [Cross Ref]

Golbashy, M., Ebrahimi, M., Khorasani, K.S. and Choucan, R. 2010. Evaluation of drought tolerance of some corn (*Zea mays* L.) hybrids in Iran. *African Journal of Agricultural Research,* **5**(19): 2714–2719.

Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology,* **24**(6): 417. [Cross Ref]

Husson, F., Josse, J., Le, S. and Maintainer, J.M. 2020. Package "FactoMineR" title multivariate xploratory data analysis and data mining. R Foundation for Statistical Computing: Vienna, Austria.

Laude, T.P. and Carena, M.J. 2015. Genetic diversity and heterotic grouping of tropical and temperate maize populations adapted to the northern US Corn Belt. *Euphytica,* **204**(3): 661-677. [Cross Ref]

Li, J. and Yuan, L. 2000. Hybrid rice: genetics, breeding, and seed production. *Plant breeding reviews*, **17:** 15-158. [Cross Ref]

Ma, L., Li, J. and Deng, X. 2015. A construct for controlling plant pollen fertility and use thereof. Patent no CN103261419B, 10.

Martinez-Calvo, J., Gisbert, A.D., Alamar, M.C., Hernandorena, R., Romero, C., Llacer, G. and Badenes, M.L. 2008. Study of a germplasm collection of loquat (*Eriobotrya japonica* Lindl.) by multivariate analysis. *Genetic Resources and Crop Evolution,* **55:** 695–703. [Cross Ref]

Morrison, D.E. 1982. Multivariate statistical methods (2nd Edn.). McGraw Hill Kogakusta Ltd.,

Nalajala, S., Muniyandi, S.J. and Manjunath, P. 2023. Principal component analysis of quantitative and qualitative traits in sixty mung bean (*Vigna radiata* L. Wilczek) genotypes. *International Journal of Bio-resource and Stress Management***, 14**(4): 643-651. [Cross Ref]

Pearson, K. 1901. Principal components analysis. The London, Edinburgh, and Dublin Philosophical *Magazine and Journal of Science*, **6**(2): 559. [Cross Ref]

Ringner, M. 2008. What is principal component analysis? *Nature Biotechnology,* **26**(3): 303–304. [Cross Ref]

Sharifi, P., Astereki, H. and Pouresmael, M. 2018. Evaluation of variations in chickpea (*Cicer arietinum* L.) yield and yield components by multivariate technique. *Annals of Agrarian Science,* **16**(2): 136-142. [Cross Ref]

Singh, S., Prakash, A., Chakraborty, N.R., Wheeler, C., Agarwal, P.K. and Ghosh, A. 2016. Trait selection by path and principal component analysis in Jatropha curcas for enhanced oil yield. *Industrial Crops and Products,* **86**: 173–179. [Cross Ref]

Stauffer, D.F., Garton, E.O. and Steinhorst, R.K. 1985. A comparison of principal components from real and random data Ecology **66**(6): 1693-1698. [Cross Ref]

Tao, A.L., Zeng, H.L., Zhang, Y.M., Xie, G.S., Qin, F.L., Zheng, Y.L. and Zhang, D.P. 2003. Genetic analysis of the low critical sterility temperature point in photoperiod-thermo sensitive genic male sterile rice. Yi Chuan xue bao= Acta Genetica Sinica, **30**(1): 40-48.

Wu, J., Deng, Q.Y., Yuan, D.Y. and Qi, S.W. 2016. Progress of super hybrid rice research in China. *Chin Sci Bull,* **61**(35): 3787-3796. [Cross Ref]

Yan, W. and Rajcan, I., 2002. Biplot analysis of test sites and trait relations of soybean in Ontario. *Crop science,* **42**(1): 11-20. [Cross Ref]

Zhang, Z. and Yuan, S. 1987. The influence of photoperiod on pollen fertility change of Hubei photoperiod-sensitive genic male sterile rice. *Chin J Rice Sci*.,1: 137-143.

Zhou, H., Liu, Q., Li, J., Jiang, D., Zhou, L., Wu, P., Lu, S., Li, F., Zhu, L., Liu, Z. and Chen, L. 2012. Photoperiod- and thermo-sensitive genic male sterility in rice are caused by a point mutation in a novel noncoding RNA that produces a small RNA. *Cell research*, **22**(4): 649-660. [Cross Ref]

Zhou, H., Zhou, M., Yang, Y., Li, J., Zhu, L., Jiang, D., Dong, J., Liu, Q., Gu, L., Zhou, L. and Feng, M., 2014. RNase ZS1 processes Ub L40 mRNAs and controls thermo sensitive genic male sterility in rice. *Nature Communications,* **5**(1): 4884. [Cross Ref]