Journal of Experimental Biology and Agricultural Sciences, June - 2024; Volume – 12(3) page 419 – 425

Journal of Experimental Biology and Agricultural Sciences

http://www.jebas.org

ISSN No. 2320 – 8694

Principal component analysis of morpho-floral traits in *Oryza sativa* × *Oryza longistaminata* advanced backcross lines of rice

Madhu Choudhary*, Ravi P Singh, PK Singh, Jayasudha S

Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh, India

Received – March 24, 2024; Revision – June 07, 2024; Accepted – July 04, 2024
Available Online – July 15, 2024

DOI: http://dx.doi.org/10.18006/2024.12(3).419.425

KEYWORDS

Hybrid rice

Morpho-floral

Principal component analysis

Out-crossing

Stigma exsertion

ABSTRACT

Hybrid rice technology substantially improves the food security of South Asian countries where rice (*Oryza sativa* L.) is a staple food. Several traits contribute to hybrid seed production efficiency, among which stigma exsertion is crucial for enhancing production by facilitating out-crossing pollination. This study evaluated the variation patterns and relative impact of 12 morpho-floral traits on overall variability in advanced backcross lines derived from crosses CRMS 32B cv. *Oryza sativa* and *Oryza longistaminata*. For this study, 290 BC_4F_2 lines were grown during Kharif 2019 in 3 replications using a randomized complete block design (RCBD). Principle component analysis (PCA) was performed on all traits, and the findings revealed 11 principal components (PCs). Out of 11 PCs, the first five displayed eigenvalues exceeding 1, collectively explaining 78.78% of the total variability. PC1, PC2, PC3, PC4, and PC5 contributed 26.36%, 19.94%, 14.22%, 9.81%, and 8.44% of the variation, with eigenvalues of 3.16, 2.39, 1.71, 1.18 and 1.01, respectively. PC1 was predominantly associated with yield-related traits such as panicle length, plant height, grain yield per plant, grains per panicle, and effective tillers per plant. On the other hand, PC2 was mainly associated with outcrossing-related floral traits such as total stigma exsertion percentage, dual stigma exsertion percentage, and single stigma exsertion percentage. However, PC3 and PC4 were associated with both floral and yield-related traits, *i.e.*, days to 50% flowering (DF), days to maturity (DM), plant height (PH), effective tillers per plant (ETPP), spikelet fertility percentage (SFP), grain yield per plant (GYPP) and grains per panicle (GPP). Therefore, PC1, PC2, PC3, and PC4 were major contributors to rice hybrid seed production.

---

* Corresponding author

E-mail: anamikaz0129@gmail.com (Madhu Choudhary)

Peer review under responsibility of Journal of Experimental Biology and Agricultural Sciences.

Production and Hosting by Horizon Publisher India [HPI]
(http://www.horizonpublisherindia.in/).
All rights reserved.

1 Introduction

Rice (*Oryza sativa* L.) is a major staple crop, serving as the primary food source for over half of the global population and fulfilling more than 21% of the world's caloric requirements (Sathe et al. 2021). The world's population is increasing and is estimated to reach 9.7 billion people by 2050; thus, increasing the productivity of major cereal crops such as rice is urgently needed to keep pace with the population surge (Kumar et al. 2022). Although increased rice yields were successfully achieved by integrating semidwarf genes, the limited genetic diversity among breeding lines has decreased the yield of released rice varieties. It is now crucial to expand the genetic diversity of rice varieties to surpass the yield limit and satisfy the rising demand for rice. Hybrid rice technology has emerged as the most feasible and adaptable option to address this situation, providing a yield advantage of 15-20% over conventional varieties (Qian et al. 2016). However, the major challenge hindering the widespread adoption of hybrid rice is the low seed yield ($\leq$2.5 ton ha^{-1}), which consequently leads to high seed costs, and farmers often purchase hybrid seeds at a relatively high price for every cropping season (Xie 2009). Rice is an autogamous plant that prevents it from being naturally out-crossing; therefore, the most efficient strategy for enhancing hybrid seed production is to create male sterile lines with a high out-crossing rate. The out-crossing rate is affected by multiple traits, including both parents' flowering behavior and morphological traits (Marathi and Jena 2015). Among these traits, the stigma exsertion rate is critical for enhancing out-crossing and ensuring efficient hybrid rice seed production (Zhang et al. 2018). It is a genetic trait that varies among male sterile lines and can be improved through targeted breeding efforts with appropriate donor varieties. Wild rice species serve as valuable repositories for revealing novel variations in flowering and morphological traits and could be utilized to increase the genetic background of elite cultivars (Ramos et al. 2016). Furthermore, these wild rice species exhibit a higher out-crossing rate than cultivated rice, ranging from 3.2% to 70.0%, while some wild rice species, such as *Oryza longistaminata* and *O. rufipogon* exhibit out-crossing rates of up to 100% (Prahalada et al. 2021). Incorporating favorable traits from wild species into elite breeding materials through wide hybridization has been a longstanding strategy (Zeliang and Pattanayak 2013). Although backcross introgression is also a feasible method for incorporating favorable traits from wild species, this method primarily involves the genome of recurrent parents with minimal donor segments and offers advantages for accurately estimating novel genes or QTLs and diversifying existing germplasms (Todorovska et al. 2013; Balakrishnan et al. 2016).

The effectiveness of any breeding program relies on understanding genetic variability to determine appropriate selection strategies for improving targeted traits. Multivariate analysis aids plant breeders in devising these selection approaches (Ravindra et al. 2012). Principal component analysis (PCA) is a widely utilized multivariate statistical tool for compressing, reducing, and transforming data. This approach aims to simplify complex datasets by reducing their dimensionality to a minimum number of components while retaining most of the variance (Rahangdale et al. 2021). Through orthogonal transformation, PCA transforms a set of possibly related variables into a fresh set of linearly unrelated variables, facilitating data simplification and interpretation. PCA is extensively used to assess morphological and physiological traits by analyzing multiple parameters of each individual simultaneously. The eigenvalue of a given principal component signifies the extent of variance in attributes explained by that principal component, making it crucial for trait selection (Singh et al. 2020). Hence, the present research aimed to explore the genetic variations and to identify key factors contributing to the overall variance among selected morpho-floral traits of rice advanced backcross lines using PCA.

2 Materials and Methods

The plant material utilized in this study consisted of 290 BC_4F_2 lines with recurrent parents. These advanced backcross lines (BC_4F_2) were developed through interspecific crossing between the recurrent parent CRMS 32B (maintainer of male sterile line with low stigma exsertion) cv. *O. sativa* and the donor parent *O. longistaminata* (wild rice with high stigma exsertion). The BC_4F_2 lines were evaluated during *Kharif* 2019 at the Research Farm of the Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh. This study site is located in northern India's northern Gangetic Alluvial Plain zone, at latitude 25.18° N and longitude 83.03° E. The material was planted with three replications using a randomized complete block design (RCBD), with each row 3.0 meters long and a 20 × 15 cm spacing. Standard packages and practices were implemented to ensure optimal crop growth and quality. Observations were taken for 12 quantitative traits, which included days to 50% flowering (DF), days to maturity (DM), plant height (PH), effective tillers per plant (ETPP), panicle length (PL), grains per panicle (GPP), spikelet fertility percentage (SFP), grain yield per plant (GYPP), test weight (TW), single stigma exsertion percentage (SSE%), dual stigma exsertion percentage (DSE%) and total stigma exsertion percentage (TSE%). These traits were measured on five randomly selected plants from each line in every replicate. Morphological characterization followed the standard evaluation system (SES) for rice, as the IRRI (2013) outlined. After the data were collected, the mean values for each trait were calculated and subjected to further statistical analysis. PCA was performed to identify key traits that significantly impacted the overall variability, and biplots were generated to visually represent the data effectively. The analysis, including PCA and biplot, was carried out using R software version 4.3.2.

3 Results and Discussion

Principal component analysis is pivotal for identifying the primary contributors to overall variability. It reduces the dimensionality of large datasets by extracting a small set of major independent variables that encapsulate the original variability without sacrificing its integrity (Sharma et al. 2022). PCA was performed on a set of 12 morpho-floral traits observed in 290 advanced backcross lines of rice, and the findings revealed 11 principal components (PCs). The eigenvalues, percentage of variability, and cumulative percentage are provided in Table 1. Eigenvalues represent the variance explained by each principal component, with higher values indicating more significant contributions to variability. Generally, components with eigenvalues >1 are considered essential because they capture more variance than does a single original variable (Brejda et al. 2000). Out of 11 PCs, five PCs displayed eigenvalues exceeding 1, and these PCs collectively explained 78.78% of the total variability. The first principal component (PC1) explained 26.36% of the overall variation with an eigenvalue of 3.16, whereas PC2, PC3, PC4, and PC5 individually accounted for 19.94%, 14.22%, 9.81%, and 8.44% of the variation with eigenvalues of 2.39, 1.71, 1.18 and 1.01, respectively. The results demonstrated that the first five principal components captured a substantial portion of the total variability (78.78%), effectively summarizing the information in the original set of traits. This reduction in dimensionality while retaining most of the variance proved valuable for identifying key traits that influenced the genetic variation among the rice lines. Similarly, Sheela et al. (2020) reported four principal components that explained 72.24% of the total observed variability in rice germplasm.

The analysis of factor loadings revealed that the phenotypic traits having the most significant impact on variation displayed high positive loadings across various principal components (Table 1). In PC1, traits such as PL, PH, GYPP, GPP, and ETPP exhibited positive loadings, indicating that these traits were positively correlated and contributed significantly to the variation explained by PC1. Conversely, the remaining traits displayed negative loadings in PC1, suggesting an inverse relationship with this principal component. Thus, lines with high values of PC1 tended to have longer panicles, taller plants, greater grain yield, more grains per panicle, and more effective tillers. In PC2, the TSE%, DSE%, SSE%, GYPP, and PL traits had more positive loadings than did the other traits, suggesting that these traits, particularly those related to stigma exsertion, were significant contributors to the

Table 1 Eigenvalues, variability percentages, cumulative percentages, and factor loadings of different morph-floral traits across principal components

Principal Components	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
Eigenvalues	3.163	2.393	1.707	1.177	1.013	0.889	0.494	0.456	0.262	0.234	0.211
Variability Percentage	26.362	19.940	14.224	9.808	8.445	7.405	4.121	3.799	2.187	1.953	1.758
Cumulative Percentage	26.362	46.302	60.526	70.333	78.778	86.183	90.304	94.102	96.289	98.242	100.000
Component matrix	Factor loadings										
DF	-0.556	-0.070	0.666	-0.088	-0.086	-0.320	-0.032	0.006	-0.201	-0.280	0.066
DM	-0.562	-0.104	0.624	-0.023	0.001	-0.400	0.083	-0.020	0.222	0.256	-0.031
PH	0.669	0.203	0.072	-0.415	-0.032	-0.249	-0.129	0.491	0.005	0.003	-0.111
ETPP	0.469	0.245	0.327	0.723	-0.048	-0.026	-0.013	-0.021	0.091	-0.124	-0.250
PL	0.755	0.301	0.107	-0.342	0.023	-0.133	-0.003	-0.261	0.280	-0.172	0.131
GPP	0.564	0.256	0.499	-0.383	-0.064	0.207	0.137	-0.260	-0.222	0.134	-0.145
SFP	-0.255	-0.113	0.587	-0.104	0.098	0.696	0.024	0.210	0.150	-0.063	0.041
GYPP	0.641	0.343	0.356	0.435	0.156	-0.030	-0.076	0.095	-0.116	0.139	0.289
TW	0.005	-0.163	0.008	-0.059	0.976	-0.088	-0.009	-0.038	-0.031	-0.032	-0.073
SSE%	-0.483	0.756	-0.009	-0.068	0.037	0.078	-0.416	-0.084	0.009	0.035	-0.036
DSE%	-0.256	0.785	-0.158	0.020	0.075	-0.044	0.516	0.128	-0.002	-0.046	0.022
TSE%	-0.455	0.881	-0.076	-0.040	0.059	0.036	-0.070	-0.004	0.005	0.005	-0.016

PC- principal component; DF-days to 50% flowering (days); DM-days to maturity (days); PH-plant height (cm); ETPP-effective tillers per plant (no.); PL-panicle length (cm); GPP-grains per panicle (no.); SFP-spikelet fertility%; GYPP-grain yield per plant (g); TW-test weight (g); SSE%- single stigma exsertion%; DSE%- dual stigma exsertion%; TSE%- total stigma exsertion%

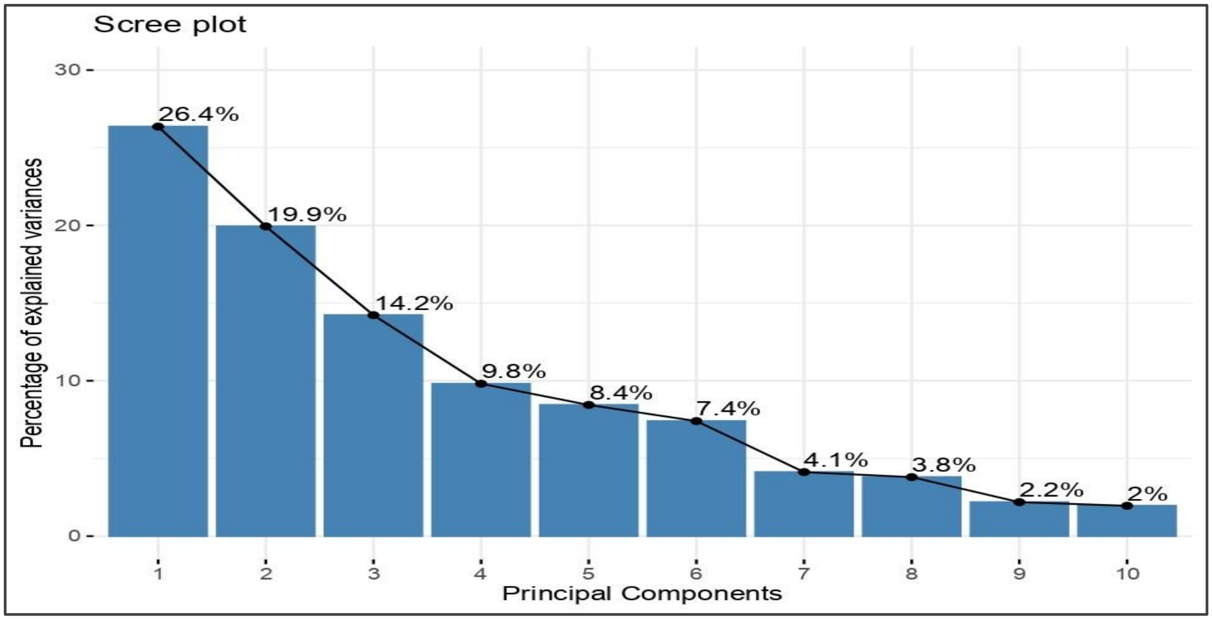

Figure 1 Scree plot visually representing the distribution of variance among principal components.

variability explained by PC2. The DF, DM, SFP, and GPP traits showed high positive loadings in PC3, indicating that these traits were important for the variability captured by PC3. On the other hand, the ETPP, GYPP, and TW traits showed substantial positive loadings in PC4 and PC5, respectively. This suggests that the ETPP and GYPP traits were key contributors to the variability explained by PC4, whereas TW was the primary contributor to the variability explained by PC5. These results demonstrated the complex interplay of various morpho-floral traits contributing to the overall genetic variability among the rice advanced backcross lines. The scree plot in Figure 1 illustrates the variance distribution across principal components, with PC1 displaying the highest variability (26.36%), followed by a decreasing trend in variability for the subsequent PCs. This indicates that PC1 captured the most significant variation in the dataset, followed by diminishing contributions from the subsequent principal components.

The interpretation of the rotated component matrix demonstrated that every principal component is distinctly loaded with various morphological and floral characteristics. The contributions of 12 traits to the first five PCs are detailed in Table 2. PC1 was primarily influenced by yield-related morphological traits, notably PL (17.998%), PH (14.163%), and GYPP (12.970%). On the other hand, PC2 was primarily contributed by outcrossing-related floral traits, particularly the stigma exsertion traits TSE% (32.402%), DSE% (25.739%), and SSE% (23.884%). PC3 contributed to both floral and yield-related traits, *viz.*, DF (26.019%), DM (22.791%), SFP (20.164%), and GPP (14.562%).

In PC4, the traits ETPP (44.402%), GYPP (16.082%), PH (14.637%), and GPP (12.455%) were prominent contributors. Therefore, PC1, PC2, PC3, and PC4 were key contributors to rice hybrid seed production. PC5 was mainly driven by the TW (94.083%) trait, and similar findings were reported by Bhargava et al. (2023). Overall, the results suggested that yield-related morphological traits played a significant role in PC1, highlighting their potential for increased productivity. Conversely, outcrossing-related floral traits were the major contributors to PC2, reflecting their role in reproductive success and potential for genetic diversity. PC3 and PC4 were influenced by a combination of flowering time-related and panicle characteristics, indicating their importance in determining the developmental and yield-related aspects of the plants. PC5 was influenced mainly by TW, indicating its role in determining grain size and weight. These results are consistent with Riaz et al. (2018), who reported that PC1 and PC2 were predominantly associated with yield-contributing traits, while flowering time-related traits, such as DF and DM, were clustered within the PC3 component.

The biplot is a powerful graphical tool that integrates information from both variables and observations into a single plot, offering a visual representation of the relationships between variables (traits) and observations (advanced backcross lines) in the reduced-dimensional space defined by the principal components. In the biplot, the length and direction of the vectors representing variables indicate the magnitude and direction of their contributions to the principal components, respectively. The clustering of variables in the biplot revealed the patterns of correlation or covariance among traits, while the positioning of observations relative to variable vectors provides insights into the performance of individual lines across various traits. A biplot between PC1 and PC2 was created using different morpho-floral

Table 2 Contribution of morpho-floral traits to the variation observed in the first five principal components

Traits	PC1	PC2	PC3	PC4	PC5
DF	9.784	0.203	26.019	0.651	0.726
DM	10.000	0.450	22.791	0.047	0.000
PH	14.163	1.730	0.304	14.637	0.104
ETPP	6.965	2.511	6.276	44.402	0.229
PL	17.998	3.790	0.666	9.949	0.051
GPP	10.057	2.737	14.562	12.455	0.403
SFP	2.050	0.535	20.164	0.925	0.953
GYPP	12.970	4.908	7.415	16.082	2.415
TW	0.001	1.111	0.003	0.292	94.083
SSE%	7.377	23.884	0.005	0.394	0.134
DSE%	2.079	25.739	1.459	0.034	0.557
TSE%	6.556	32.402	0.336	0.133	0.344

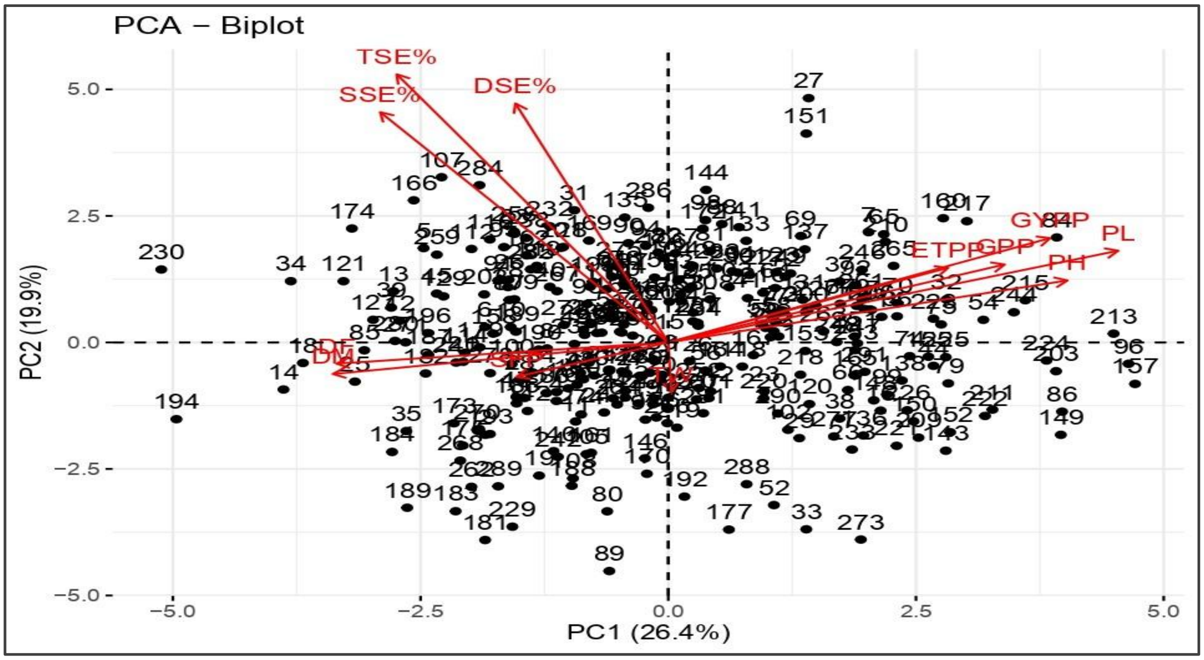

Figure 2 Biplot between PC1 and PC2 using different morpho-floral traits of rice advanced lines

traits, as depicted in Figure 2. The results indicated that traits showing positive correlations were clustered together, while those with negative correlations were positioned on opposite sides of the plot origin. The PL, GYPP, and PH traits had greater vector lengths (distance from the origin) in PC1, while the TSE%, DSE%, and SSE% traits had greater vector lengths in PC2, indicating a good representation of these traits compared to other traits. Figure 2 also illustrates the distribution and diversity of both variables in the advanced backcross lines. Overall, PC1 and PC2 accounted for 46.30% of the total variability. Among the traits, TSE% exhibited the greatest vector length, suggesting its substantial contribution to the overall divergence, followed by SSE%, DSE%, PL, GYPP, PH, and GPP. Advanced backcross lines adjacent to the trait vector within the same quadrant were more likely to exhibit favorable performance for the corresponding traits. Lines 27, 151, 107, 284, 213, 84, 230, 157, 96, 194, and 89 significantly contributed to the overall diversity, owing to their high PC scores and considerable distance from the origin of axes. Therefore, rigorous selection methods can be devised to expedite the enhancement of the aforementioned morpho-floral traits in advanced backcross populations of rice. Similarly, Rahangdale et al. (2021) identified superior genotypes for yield and quality traits by focusing on high PC scores in their PCA.

Conclusion

PCA was employed to assess the contribution of different morpho-floral traits to overall variability. The analysis revealed that the first five PCs collectively accounted for an average of 78.78% of the total variability. The important traits (PL, PH, GYPP, TSE%, DSE%, and SSE%) grouped within PC1 and PC2 helped explain maximum variability and displayed a tendency to cooccur consistently. Therefore, this study provides valuable insights into identifying parameters contributing to variability and selecting suitable lines for breeding, facilitating crop improvement for morpho-floral traits. Furthermore, the investigated advanced backcross population exhibited notable segregation and diversity in morphological and floral characteristics, suggesting its potential for mapping genomic regions associated with yield and out-crossing (stigma exsertion) traits.

Acknowledgment

The authors would like to convey their gratitude to Dr. R.L. Verma, Crop Improvement Division, National Rice Research Institute, Cuttack, Odisha, for providing the essential materials to accomplish this research work. Furthermore, the authors wish to express their gratitude to the Department of Science and Technology (DST), Ministry of Science and Technology, Government of India, for providing partial funding support through a DST INSPIRE fellowship to the first author for pursuing a doctoral degree program at Banaras Hindu University, Varanasi.

Declaration of competing interest

The authors declare that they have no competing interests related to this study.

References

Balakrishnan, D., Subrahmanyam, D., Badri, J., Raju, A. K., Rao, Y. V., et al. (2016). Genotype × environment interactions of yield traits in backcross introgression lines derived from *Oryza sativa* cv. Swarna/*Oryza nivara*. *Frontiers in Plant Science*, *7*, 1530.

Bhargava, K., Sreedhar, M., Lakshmi, V. J., Rathod, R., Parimala, K., & Vanisri, S. (2023). Genetic analysis of backcross derived lines for yield and yield attributing traits in rice (*Oryza sativa* L.). *Theoretical Biology Forum*, *12*(2), 51-59.

Brejda, J. J., Moorman, T. B., Karlen, D. L., & Dao, T. H. (2000). Identification of regional soil quality factors and indicators. I. Central and Southern High Plains. *Soil Science Society of America Journal*, *64*, 2115-2124.

IRRI. (2013). *Standard evaluation system (SES) for rice* (5th ed.). IRRI, Philippines.

Kumar, L., Chhogyel, N., Gopalakrishnan, T., Hasan, M. K., Jayasinghe, S. L., et al. (2022). Climate change and future of agri-food production. In *Future Foods: Global Trends, Opportunities, and Sustainability Challenges* (pp. 49-79). Cambridge, MA, USA: Academic Press. DOI: https://doi.org/10.1016/B978-0-323-91001-9.00009-8.

Marathi, B., & Jena, K. K. (2015). Floral traits to enhance out-crossing for higher hybrid seed production in rice: present status and future prospects. *Euphytica*, *201*, 1-14.

Prahalada, G. D., Marathi, B., Vinarao, R., Kim, S. R., Diocton, R., 4th, Ramos, J., & Jena, K. K. (2021). QTL Mapping of a Novel Genomic Region Associated with High Out-Crossing Rate Derived from *Oryza longistaminata* and Development of New CMS Lines in Rice, *O. sativa* L. *Rice (New York, N.Y.)*, *14*(1), 80. https://doi.org/10.1186/s12284-021-00521-9.

Qian, Q., Guo, L. B, Smith, S. M., & Li, J. Y. (2016). Breeding high-yield superior quality hybrid super rice by rational design. *National Science Review*, *3*(3), 283-294.

Rahangdale, S., Singh, Y., Upadhyay, P. K., & Koutu, G. K. (2021). Principal component analysis of JNPT lines of rice for the important traits responsible for yield and quality. *Indian Journal of Genetics and Plant Breeding*, *81*(01), 127-131.

Ramos, J. M., Furuta, T., Uehara, T. K., Chihiro, N., Angeles-Shim, R. B., et al. (2016). Development of chromosome segment substitution lines (CSSLs) of *Oryza longistaminata* A. Chev. & Röhr in the background of the elite japonica rice cultivar, Taichung 65 and their evaluation for yield traits. *Euphytica*, *210*, 151-163.

Ravindra, B. V., Sherya, K., Dangi, K. S., Usharani, G., & Shankar, A. S. (2012). Correlation and path analysis studies in popular rice hybrids of India. *International Journal of Scientific and Research Publications*, *2*, 1-5.

Riaz, A., Bibi, T., Raza, Q., Sabar, M., & Akhter, M. (2018). Evaluation of rice (*Oryza sativa* L.) advance uniform lines using multivariate analysis. *African Journal of Plant Science*, *12*(11), 284-289.

Sathe, A. P., Kumar, A., Mandlik, R., Raturi, G., Yadav, H., et al. (2021). Role of silicon in elevating resistance against sheath blight and blast diseases in rice (*Oryza sativa* L.). *Plant Physiology and Biochemistry*, *166*, 128-139.

Sharma, N., Bandyopadhyay, B. B., Chand, S., Pandey, P. K., Baskheti, D. C., Malik, A., & Chaudhary, R. (2022). Determining selection criteria in finger millet (*Eleusine coracana*) genotypes using multivariate analysis. *Indian Journal of Agricultural Sciences*, *92*(6), 763-768.

Sheela, K. R. V. S., Robin, S., & Manonmani, S. (2020). Principal component analysis for grain quality characters in rice germplasm. *Electronic Journal of Plant Breeding*, *11*(01), 127-131.

Singh, P., Jain, P. K., & Tiwari, A. (2020). Principal component analysis approach for yield attributing traits in chilli (*Capsicum annum* L.) genotypes. *Chemical Science Review and Letters*, *9*(33), 87-91.

Todorovska, E., Hadjiivanova, B., Bozhanova, V., Dechev, D., Muhovski, Y., Panchev, I., & Ivanova, A. (2013). Molecular and phenotypic characterization of advanced backcross lines derived from interspecific hybridization of durum wheat. *Biotechnology & Biotechnological Equipment*, *27*(3), 3760-3771.

Xie, F. (2009). Priorities of IRRI hybrid rice breeding. In F. Xie & B. Hardy (Eds.), *Accelerating hybrid rice development* (pp. 49-62). International Rice Research Institute, Los Baños.

Zeliang, P. K., & Pattanayak, A. (2013). Wide Hybridization in the Genus Oryza: Aspects and Prospects. *Indian Journal of Hill Farming*, *26*(2), 71-77.

Zhang, K., Zhang, Y., Wu, W., Zhan, X., BakrAnis, G., et al. (2018). qSE7 is a major quantitative trait locus (QTL) influencing stigma exsertion rate in rice (*Oryza sativa* L.). *Scientific Reports*, *8*(1), 14523.