P.ISSN: 0971-8184 II E ISSN: 0976-1926
DOI: https://doi.org/10.56093/ijpgr.v39i3.178885

Indian Journal of Plant Genetic Resources

RESEARCH ARTICLE

Phenotypic Diversity, Correlation and Principal Component Analysis in a Mutagenized M_2F_2 Population of Hybrid Rice (*Oryza sativa*) cv. Arize 6444 Gold

Arjun Kumar Agarwal[1]*, Manigopa Chakraborty[1], Nutan Verma[1], Manoj Kumar Barnwal[2], Sumit Kumar Mishra[1], Shivam Mishra[3], Ankita Mishra[4], Raushan Kumar[5], Radhe Shyam Yadav[6] and Fatima Araish[1]

Abstract

This study evaluated phenotypic variability, frequency distribution, trait associations, and principal components in an M_2F_2 population developed from mutagen-treated seeds of hybrid rice (*Oryza sativa* L.) Arize 6444 Gold. A total of 100 plants were assessed for ten quantitative traits during Kharif 2024. Descriptive statistics revealed substantial variability for key yield-related traits, particularly grain yield, filled grains per panicle, and spikelet fertility, with near-normal frequency distributions indicating good scope for selection. Principal component analysis identified four components explaining 61.46% of total variation, with PC1 largely associated with panicle architecture and grain-related traits. Correlation analysis showed strong positive associations among primary branches, secondary branches, filled grains, and spikelet fertility, while grain yield itself showed weak direct correlations with most component traits owing to their compensatory relationships. Overall, the results identify panicle architecture and grain formation traits as reliable indirect selection criteria for improving yield in early segregating generations of rice.

Keywords: Correlation analysis, M_2F_2 population, Mutagenesis, Principal component analysis, Rice

[1]Department of Genetics and Plant Breeding, Ranchi Agriculture College, BAU, Ranchi, Jharkhand, 834006, India

[2]Department of Plant Pathology, Ranchi Agriculture College, BAU, Ranchi, Jharkhand, 834006, India

[3]Department of Agricultural Statistics and Computer Application, Ranchi Agriculture College, BAU, Ranchi, Jharkhand, 834006, India

[4]Department of Genetics and Plant Breeding, College of Agriculture, OUAT, Bhubaneswar, Odisha, 751003, India

[5]Department of Entomology, Ranchi Agriculture College, BAU, Ranchi, Jharkhand, 834006, India

[6]Department of Agronomy, Ranchi Agriculture College, BAU, Ranchi, Jharkhand, 834006, India

***Author for correspondence:**
arjunagarwal016@gmail.com

Received: 11/05/2026
Revised: 24/09/2026
Accepted: 28/09/2026

How to cite this article: Arjun Kumar Agarwal, Manigopa Chakraborty, Nutan Verma, Manoj Kumar Barnwal, Sumit Kumar Mishra, Shivam Mishra, Ankita Mishra, Raushan Kumar, Radhe Shyam Yadav and Fatima Araish (2026) Phenotypic Diversity, Correlation and Principal Component Analysis in a Mutagenized M_2F_2 Population of Hybrid Rice (*Oryza sativa*) cv. Arize 6444 Gold. *Indian J. Plant Genet. Resour.* 39(3): 54-60.

DOI: https://doi.org/10.56093/ijpgr.v39i3.178885

Introduction

Rice (*Oryza sativa* L.) is a primary staple food for more than half of the global population and plays a central role in ensuring food security, particularly in Asia and developing countries (Bin Rahman and Zhang, 2023). However, increasing demand for rice production is challenged by climate change, resource limitations, and the complex genetic architecture of yield traits (Ray *et al*., 2013). Yield in rice is a polygenic trait governed by multiple interacting components, and conventional breeding, while indispensable, is constrained by the narrow genetic base of elite cultivars and the limited natural variability available for several yield-contributing traits.

Induced mutagenesis offers a direct route to broadening this genetic base by generating novel allelic variation that is otherwise absent or rare in natural germplasm (Mba *et al*., 2010). Physical and chemical mutagens have been used to develop rice genotypes with improved yield potential, quality, and stress tolerance (Ahloowalia *et al*., 2004), and remain one of the few approaches capable of introducing de novo variation independent of existing genetic diversity. However, most mutation breeding efforts in rice to date have been carried out in inbred, homozygous backgrounds, while the application of induced mutagenesis to hybrid backgrounds remains

© IJPGR, 2025. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit https://creativecommons.org/licenses/by-nc-sa/4.0/.

comparatively underexploited despite hybrids offering a genetically heterogeneous and heterozygous platform on which mutagen-induced variation could interact with existing heterosis-related loci in ways not captured by mutagenesis in fixed lines. This represents a notable gap, since combining mutagenesis with hybrid vigour could, in principle, generate transgressive segregants and novel trait combinations unattainable through either approach alone.

Induced mutations become phenotypically expressed in advancing generations, particularly from M_2 onward, as recessive mutations segregate out and recombination further increases variability. The M_2F_2 generation is therefore a critical stage for a mutagenized hybrid population, combining segregation from hybridity with mutation-induced variability, and offering an early window to assess the extent and pattern of genetic variation generated. However, systematic characterization of trait variability, correlation structure, and multivariate trait relationships at this stage specifically in a mutagenized hybrid population has received limited attention in the literature.

Understanding the interrelationships among yield and its component traits at this stage is therefore essential for guiding selection strategy. Correlation analysis is widely used to quantify such associations (Falconer & Mackay, 1996), while multivariate approaches such as principal component analysis (PCA) help resolve complex trait interdependencies, reduce dimensionality, and identify the principal axes of variation driving divergence in a segregating population (Jolliffe & Cadima, 2016). The present study was therefore undertaken to characterize the extent of phenotypic variability, trait associations, and multivariate structure in an M_2F_2 mutagenized hybrid rice population, with the aim of identifying key traits for effective selection and addressing the underexplored potential of mutation breeding in a hybrid genetic background.

Materials and Methods

The experiment was conducted out during the *Kharif* 2023 and 2024 at the Rice Research Farm, Birsa Agricultural University, Kanke, Ranchi. In *Kharif* 2023, one kilogram of seeds from the commercial hybrid Arize 6444 Gold were subjected to treatment with a combination of chemical mutagens, namely colchicine, EMS, and sodium azide. The hybrid Arize 6444 Gold was selected for mutagenesis owing to its wide cultivation and popularity among farmers in the study region, making it agronomically relevant for generating and evaluating mutation-induced variability. The treated seeds (M_1F_1 generation) were grown to maturity and allowed to self-pollinate to obtain M_2F_2 seeds. These M_2F_2 seeds were then sown in *Kharif* 2024 to develop the M_2F_2 population. Observations were recorded on 100 randomly selected plants from a large M_2F_2 population for ten quantitative traits (Table 1), following the Standard Evaluation System for Rice (IRRI, 2013).

All statistical analyses were performed using R statistical software within the RStudio integrated development environment (RStudio version 2024.09.1; Posit Team, 2024). The analyses included descriptive statistics such as mean, range, standard deviation, coefficient of variation, skewness, and excess kurtosis, along with Pearson's correlation coefficients and principal component analysis (PCA). Principal components were considered significant based on the Kaiser criterion (eigenvalue >1) (Kaiser, 1960), and a biplot was generated to illustrate the relationships between traits and genotypes.

Results and Discussion

Trait variability

The M_2F_2 population showed wide phenotypic variation across the traits studied (Table 1). Trait means ranged from 7.28 productive tillers per plant to 144.67 days to first flowering. Frequency distributions for all yield and yield-related traits are shown in Fig. 1. Four traits varied widely: grain yield (2.06–33.24 g/plant), filled grains per panicle (2–108), spikelet fertility (1.87–87.88%), and secondary branches per panicle (7–30). Their standard deviations were correspondingly high (8.29 g/plant, 24.95, 22.07% and 6.07). This spread points to substantial variability in the population, most likely from mutagen-induced segregation combined with recombination from the hybrid background, and gives a good basis for selection.

Induced mutagenesis is well established as a way to widen variability for yield traits beyond what the parent carries (Oladosu *et al*., 2016). Oladosu *et al*., (2014) found significant phenotypic variation for yield and its components among rice mutant lines, with grain-filling traits contributing most to yield differences. Sushmitharaj *et al*., (2019) likewise recorded higher variability and wider trait ranges in M2 rice populations treated with gamma rays, electron beam and EMS than in the untreated control, particularly for grains per panicle and grain yield per plant. Amri-Tiliouine *et al*., (2018) reported the same pattern in gamma-irradiated chickpea, where seed number per plant and seed yield

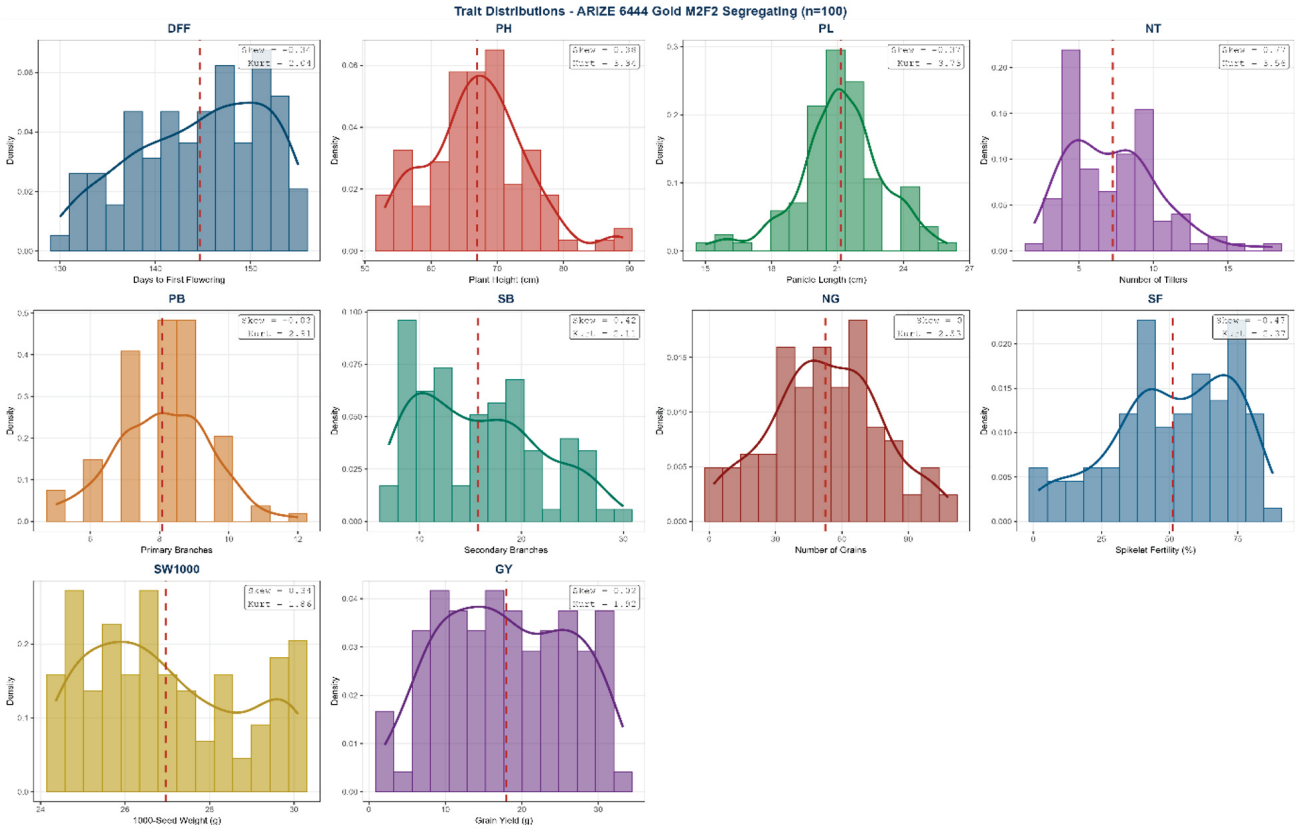


Fig. 1: Frequency distribution of yield-related traits in the M_2F_2 population.

Table 1: Descriptive statistics with Skewness & Kurtosis - Arize 6444 Gold M_2F_2 (n=100)

Trait	Mean	SD	Min	Max	CV	Skew	Kurtosis _Excess
Days to first flowering (DFF)	144.67	6.93	130.00	155.00	4.79	-0.34	2.04
Plant height (PH, cm)	66.99	7.53	53.00	89.00	11.24	0.38	3.34
Number of productive tillers per plant (NT)	21.15	1.98	15.00	26.00	9.37	-0.37	3.73
Panicle length (PL, cm)	7.28	3.12	2.00	18.00	42.82	0.77	3.56
Primary branches per panicle (PB)	8.08	1.40	5.00	12.00	17.30	-0.03	2.91
Secondary branches per panicle (SB)	15.73	6.07	7.00	30.00	38.60	0.42	2.11
Number of filled grains per panicle (NG)	52.52	24.95	2.00	108.00	47.50	0.00	2.53
Spikelet fertility percentage (SF, %)	51.23	22.07	1.87	87.88	43.08	-0.47	2.37
1000-seed weight (1000-SW, g)	26.96	1.81	24.36	30.09	6.72	0.34	1.86
Grain yield per plant (GY, g)	17.94	8.29	2.06	33.24	46.21	0.02	1.92

SD, Standard deviation; Min, Minimum; Max, Maximum; CV, Coefficient of variation; Skew, Skewness.

had the highest genotypic and phenotypic coefficients of variation among all M2 traits. Three traits varied much less: panicle length (15–26 cm; SD = 1.98 cm), primary branches per panicle (5–12; SD = 1.40) and 1000-seed weight (24.36–30.09 g; SD = 1.81 g). Their expression was comparatively uniform across the population. Panicle architecture and grain weight are generally less responsive to induced variation than grain number and yield. Luz *et al*. (2016), screening 340 EMS-derived M3 rice families, found plant height to be the trait most altered by the mutagen, whereas only seven families differed from the control for main panicle length. Sao *et al.,* (2022) similarly reported that gamma-ray-derived rice mutants shifted mainly in flowering duration and plant height, while panicle and grain-weight traits stayed close to the parent. The narrow variation seen here for these traits is therefore consistent with stronger buffering of panicle structure and grain size against

both mutational and environmental perturbation.

Most traits followed a near-normal distribution with skewness values ranging from –0.47 to 0.77 (Table 1). Productive tillers, plant height and secondary branches per panicle were positively skewed, meaning most plants fell below the trait mean and only a few transgressive segregants exceeded it. Days to flowering, panicle length and spikelet fertility were negatively skewed, with most plants above the mean and a short lower tail. Sushmitharaj *et al*., (2019) recorded a similar pattern in M2 rice populations derived from gamma-ray, electron-beam and recurrent EMS treatment and interpreted the direction of skew as a guide to how far each polygenic trait had responded to the mutagen. Most traits were also leptokurtic, meaning values clustered tightly around the mean but with a heavier tail of extreme individuals than a normal curve would give (Table 1). In the same M2 rice populations, Sushmitharaj *et al*., (2019) reported leptokurtic curves for panicle length and grain weight and interpreted the degree of peakedness as reflecting how few genes were controlling the response to selection. Taken together, the moderate to high variability in yield-related traits, combined with near-normal distributions and a leptokurtic tendency, indicates a phenotypically diverse M_2F_2 population with scope for effective selection in later generations consistent with the role of induced mutagenesis in widening variability for polygenic yield traits (Oladosu *et al*., 2016). Because replicated trials were not possible at this generation, genotypic and environmental variance could not be separated, so all variability reported here is phenotypic and not a direct estimate of genetic variance.

Principal component analysis

Principal component analysis identified four components with eigenvalues greater than one (Fig. 2), together explaining 61.46% of the total variation, with individual and cumulative contributions of each component (Table 2). PC1 was dominated by panicle productivity traits filled grains per panicle, spikelet fertility, primary branches, productive tillers and secondary branches, while days to flowering, grain yield and 1000-seed weight contributed negligibly. PC2 was shaped mainly by panicle branching traits (secondary and primary branches), set against negative loadings for flowering duration and spikelet fertility, reflecting a trade-off between branching complexity and grain-filling efficiency. PC3 contrasted vegetative traits (productive tillers, plant height, days to flowering) against grain yield, spikelet fertility and filled grains,

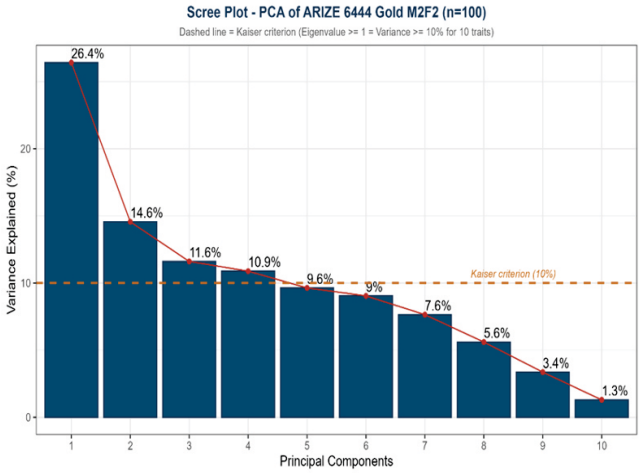


Fig. 2: Scree Plot depicting the eigenvalues of principal components and cumulative variance explained.

Table 2: Principal components analysis of morpho-agronomic traits

Variables	PC1	PC2	PC3	PC4
Days to first flowering (DFF)	0.023	-0.399	0.303	0.457
Plant height (PH, cm)	0.296	-0.049	0.366	-0.201
Panicle length (PL, cm)	0.222	-0.046	0.131	0.01
Number of productive tillers per plant (NT)	0.35	-0.286	0.385	-0.138
Primary branches per panicle (PB)	0.394	0.475	0.086	-0.126
Secondary branches per panicle (SB)	0.303	0.579	0.119	0.065
Number of filled grains per panicle (NG)	0.521	-0.154	-0.304	0.168
Spikelet fertility percentage (SF, %)	0.466	-0.292	-0.331	0.17
1000-SW	-0.014	0.268	-0.238	0.673
Grain yield (GY)	0.071	-0.113	-0.574	-0.448
Eigenvalues	2.641	1.456	1.16	1.088
Percentage of variance	26.415	14.557	11.603	10.881
Cumulative percentage of variance	26.415	40.972	52.575	61.456

indicating that vegetative vigour was largely decoupled from realized yield. PC4 was driven by 1000-seed weight and flowering duration against a negative loading for grain yield, reflecting the classic grain size–grain number trade-off. The PCA biplot depicting the relationship among traits and the distribution of genotypes across the first two principal components (Fig. 3). From the biplot, promising genotypes can be selected based on higher positive loadings of grain yield, panicle length and spikelet fertility.

Similar component structures have been reported in other rice populations developed through mutation and pedigree breeding. Sushmitharaj *et al*., (2020), analysing gamma-ray, electron-beam and EMS-derived mutant families of rice, likewise reported filled grains per panicle and spikelet number as the dominant contributors to PC1, with panicle length, grain number and 100-grain weight structuring the subsequent components. Riaz *et al*., (2018), evaluating advanced uniform rice lines through multivariate analysis, similarly found yield-contributing traits loading most heavily on the early components, supporting panicle- and grain-related traits as the primary axis of variation in segregating and advanced breeding material. Tejaswini *et al*. (2018), characterizing rice F5 families using PCA, reported a comparable pattern in which grain and panicle traits accounted for the bulk of variance across the first few components, while grain size and phenology traits structured the later ones consistent with the PC4 pattern observed here. This convergence across independently derived populations supports panicle productivity as the primary axis of trait variation, and underscores that yield improvement depends on multi-trait selection rather than any single component.

Correlation analysis

Correlation coefficients among all yield and yield-related traits are presented in Table 3. Panicle-architecture traits formed the strongest interrelated cluster: plant height, panicle length, primary and secondary branches, filled grains, and spikelet fertility were all positively and significantly associated with one another, consistent with reports that panicle branching intensity and panicle length jointly determine grain-filling capacity in rice (Sushmitharaj *et al*., 2020; Jasmin

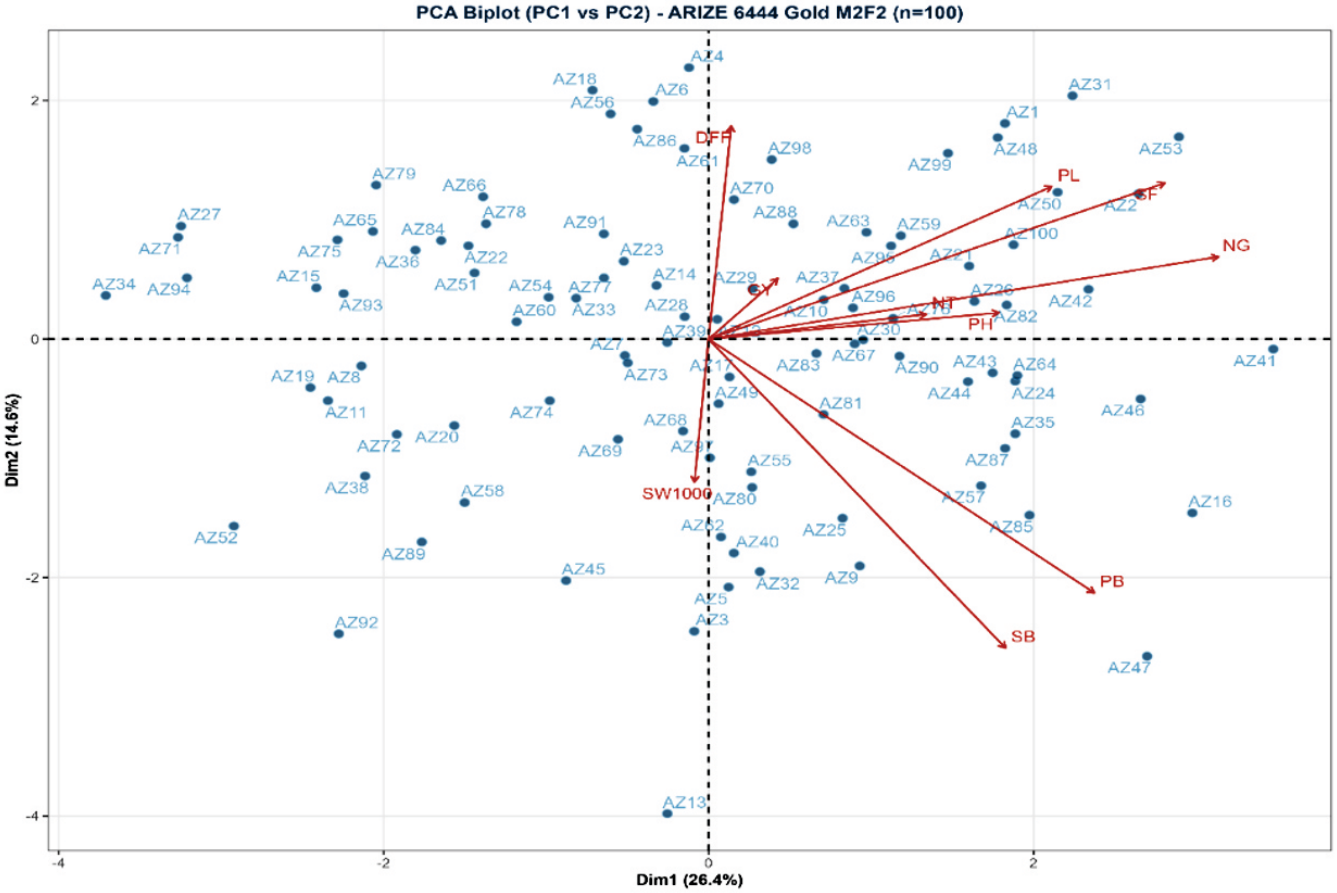


Fig. 3: PCA biplot showing the distribution of genotypes and the contribution of traits based on the first two principal components.

Table 3: Correlation matrix of yield and component traits of M_2F_2 population of rice hybrid

	DFF	PH	PL	NT	PB	SB	NG	SF	1000-SW	GY
DFF	1	-0.015	0.146	0.072	-0.14	-0.082	0.058	0.065	-0.017	-0.078
PH		1	0.347**	0.165	0.248*	0.097	0.201*	0.207*	-0.014	-0.002
PL			1	0.142	0.2*	0.083	0.353**	0.319**	-0.174	-0.028
NT				1	0.104	0.158	0.19	0.155	-0.009	0.064
PB					1	0.618**	0.372**	0.216*	0.001	0.024
SB						1	0.259**	0.073	0.106	-0.053
NG							1	0.846**	0.031	0.132
SF								1	-0.001	0.109
1000-SW									1	-0.064
GY										1

DFF, Days to first flowering; PH, Plant height; PL, Panicle length; NT, Number of productive tillers per plant; PB, Primary branches per panicle; NG, Number of filled grains per panicle; SF, Spikelet fertility percentage; 1000-SW, 1000-Seed weight; GY, Grain yield.
* and ** are significant at 5% and 1% probability levels, respectively.

et al., 2024). The strongest single association between filled grains and spikelet fertility confirms fertility as the immediate determinant of grain number, a pattern also seen in segregating rice populations by Yuvaraja *et al*., (2019).

By contrast, grain yield itself showed weak and non-significant correlations with days to flowering, productive tillers, 1000-seed weight, and spikelet fertility. Yuvaraja *et al*., (2019), evaluating an F2 segregating rice population, similarly found single-plant yield poorly associated with several component traits at this early generation, attributing it to the fact that component-trait relationships are not yet stabilized while segregation is still ongoing. This is expected rather than anomalous, since rice yield components are known to compensate for one another rather than act additively: a reduction in tiller or panicle number is routinely offset by an increase in spikelets per panicle or grain filling, so tiller number alone does not track final yield. Grain number and grain weight similarly trade off against each other, so an increase in sink size is often accompanied by a decline in individual grain weight, weakening any direct link between either component and total yield on its own (Liu *et al*., 2024). Because yield is the net outcome of several such compensating, polygenic components acting together and because dominance and epistatic effects are still unresolved in an M_2F_2 population, no single component is expected to correlate strongly with yield at this generation. This is why panicle architecture and grain-filling traits, rather than tiller number, grain weight, or fertility in isolation, represent the more reliable indirect selection criteria in this material.

Conclusion

Induced mutagenesis had generated substantial variation for grain yield and its component traits. Statistical analysis consistently identified panicle architecture, particularly filled grains per panicle, spikelet fertility, and primary and secondary branching, as the traits most strongly associated with yield expression and as the principal axes of variation in the mutagenized population. This convergence across independent analyses indicates that the mutagenic treatment successfully broadened the genetic base for these traits, providing the segregating variability necessary for effective phenotypic selection at this generation. The findings thus support panicle architecture as the primary breeding target for developing high-yielding mutant lines from this population, and the identified variability provides a foundation for further genotypic characterization as selection progresses toward fixed, high-performing lines.

Acknowledgement

The authors gratefully acknowledge Birsa Agricultural University for providing the necessary facilities, experimental materials, and institutional support for conducting the present research work.

References

Ahloowalia BS, M Maluszynski and K Nichterlein (2004) Global impact of mutation-derived varieties. *Euphytica*. 135(2): 187-204.

Amri-Tiliouine W, M Laouar, A Abdelguerfi, J Jankowicz-Cieslak, L Jankuloski and BJ Till (2018) Genetic variability induced by gamma rays and preliminary results of low-cost TILLING on M2 generation of chickpea (*Cicer arietinum* L.). *Front. Plant Sci*. 9: 1568.

Bin Rahman ANMR and J Zhang (2023) Trends in rice research; 2030 and beyond. *Food Energy Secur.* 12(2): e390.

Falconer DS and TFC Mackay (1996) *Introduction to Quantitative Genetics*, 4th edn. Addison Wesley Longman, Harlow.

Jasmin SA, PH Prasanth, S Ramchander, PD Kumar, N Devasena, R Naveenkumar, S Thankappan and NBJ Kingsly (2024) Assessment of variability parameters and diversity of panicle architectural traits associated with yield in rice (*Oryza sativa* L.). *Plant Sci. Today.* 11(1): 109-118.

Jolliffe IT and J Cadima (2016) Principal component analysis; a review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci*. 374(2065): 20150202.

Kaiser HF (1960) The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* 20(1): 141-151.

Liu K, K Zhang, Y Zhang, J Cui, Z Li, J Huang, S Li, J Zhang, S Deng, Y Zhang, J Huang, L Ren, Y Chu, H Zhao and H Chen (2024) Optimizing the total spikelets increased grain yield in rice. *Agronomy.* 14(1): 152.

Luz VK, SFS Silveira, GM Fonseca, EL Groli, RG Figueiredo, D Baretta, MM Kopp, AM Magalhães Júnior, LC Maia and AC Oliveira (2016) Identification of variability for agronomically important traits in rice mutant families. *Bragantia.* 75(1): 41-50.

Mba C, R Afza, S Bado and SM Jain (2010) Induced mutagenesis in plants using physical and chemical agents. In: MR Davey and P Anthony (eds) *Plant cell culture: essential methods*. Wiley, Chichester, pp 111-130.

Oladosu Y, MY Rafii, N Abdullah, M Abdul Malek, HA Rahim, G Hussin, M Abdul Latif and I Kareem (2014) Genetic variability and selection criteria in rice mutant lines as revealed by quantitative traits. *Sci. World J.* 2014: 190531.

Oladosu Y, MY Rafii, N Abdullah, G Hussin, A Ramli, HA Rahim, G Miah and M Usman (2016) Principle and application of plant mutagenesis in crop improvement; a review. *Biotechnol. Equip.* 30(1): 1-16.

Posit Team (2024) RStudio; integrated development environment for R. Posit Software, PBC, Boston, MA. https://www.posit.co/

R Core Team (2024) R; a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ray DK, ND Mueller, PC West and JA Foley (2013) Yield trends are insufficient to double global crop production by 2050. *PLoS One.* 8(6): e66428.

Riaz A, T Bibi, T Raza, M Sabar and M Akhter (2018) Evaluation of rice (*Oryza sativa* L.) advance uniform lines using multivariate analysis. *Afr. J. Plant Sci.* 12(11): 284-289.

Sao R, PK Sahu, RS Patel, BK Das, L Jankuloski and D Sharma (2022) Genetic improvement in plant architecture, maturity duration and agronomic traits of three traditional rice landraces through gamma ray-based induced mutagenesis. *Plants.* 11(24): 3448.

Sushmitharaj DV, P Arunachalam, C Vanniarajan, J Souframanien and E Subramanian (2019) Comparative efficiency of irradiation and its recurrent mutation on induced variability and population characteristics in rice (*Oryza sativa* L.). *Electron. J. Plant Breed.* 10(2): 542-551.

Sushmitharaj DV, P Arunachalam, C Vanniarajan, J Souframanien and E Subramanian (2020) Multivariate analysis in rice (*Oryza sativa* L.) mutant families from Anna (R) 4 cultivar. *Indian J. Agric. Res.* 54(6): 769-774.

Tejaswini KLY, S Manukonda, R Kumar, PVR Rao, M Lal Ahamed and SK Raju (2018) Application of principal component analysis for rice F5 families characterization and evaluation. *Emergent Life Sci. Res.* 4(1): 72-84.

Yuvaraja A, P Gokulvigneswaran, S Suriya Praba and V Nirubana (2019) An assessment of early generation genetic variability and correlation for yield and quality traits in rice (*Oryza sativa* L.). *Oryza.* 56(4).