



RESEARCH ARTICLE

# Multivariate genetic analyses unveil the complexity of grain yield and attributing traits diversity in *Oryza sativa* L. landraces from North-Eastern India

Anjan Roy<sup>1,2\*</sup>, Sanghamitra Rout<sup>2</sup>, Lakshmi Hijam<sup>1</sup>, Supratim Sadhu<sup>3</sup>, Pavithra S<sup>1</sup>, Abir Ghosh<sup>4</sup>, Moumita Chakraborty<sup>1</sup>, Nandita Sahana<sup>5</sup>, Saurav Singla<sup>6</sup> & Suddhasuchi Das<sup>7</sup>

<sup>1</sup>Department of Genetics and Plant Breeding, Uttar Banga Krishi Viswavidyalaya, Pundibari, Cooch Behar, West Bengal- 736165, India

<sup>2</sup>Department of Genetics and Plant Breeding, MS Swaminathan School of Agriculture, Centurion University of Technology and Management, Paralakhemundi, Odisha-761211, India

<sup>3</sup>School of Agriculture Science, JIS University, Kolkata, West Bengal, India

<sup>4</sup>College of Forestry, Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad, Uttar Pradesh- 211007, India

<sup>5</sup>Department of Agriculture Biochemistry, Uttar Banga Krishi Viswavidyalaya, Pundibari, Cooch Behar, West Bengal- 736165, India

<sup>6</sup>School of Chemical Engineering and Physical Science, Lovely Professional University, Phagwara, Punjab- 144401, India

<sup>7</sup>Krishi Vigyan Kendra, Malda, Uttar Banga Krishi Viswavidyalaya, West Bengal- 732205, India

\*Email: royanjanag328@gmail.com

 OPEN ACCESS

## ARTICLE HISTORY

Received: 13 March 2023

Accepted: 23 December 2023

Available online

Version 1.0 : 19 February 2024

Version 2.0 : 01 April 2024



## Additional information

**Peer review:** Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

**Reprints & permissions information** is available at [https://horizonepublishing.com/journals/index.php/PST/open\\_access\\_policy](https://horizonepublishing.com/journals/index.php/PST/open_access_policy)

**Publisher's Note:** Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Indexing:** Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc  
See [https://horizonepublishing.com/journals/index.php/PST/indexing\\_abstracting](https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting)

**Copyright:** © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

## CITE THIS ARTICLE

Roy A, Rout S, Hijam L, Sadhu S, Pavithra S, Ghosh A, Chakraborty M, Sahana N, Singla S, Das S. Multivariate genetic analyses unveil the complexity of grain yield and attributing traits diversity in *Oryza sativa* L. landraces from North-Eastern India. Plant Science Today. 2024; 11(2): 29–37. <https://doi.org/10.14719/pst.2500>

## Abstract

In the North-Eastern region of India, rice stands as the predominant staple, with diverse cultivars evolving over the past six decades. This study systematically evaluated 20 rice landraces, analyzing eleven variables related to yield and its attributing traits. The aim was to identify promising genotypes for potential breeding programs and to ascertain the minimum number of components essential for explaining the total diversity. Among the eleven principal components (PCs) examined, four PCs exhibited eigenvalues surpassing 1.0, collectively contributing to 80.45% of the total variability in the traits. PC1, which explained 31.19% of the overall variance, was associated with plant height, days to 50% flowering, panicle length, grain breadth, and grain length-to-breadth ratio. Utilizing cluster analysis, the 20 rice landraces were categorized into seven distinct clusters. Maximum inter-cluster divergence was observed between clusters VI and I, as well as clusters VI and V, indicating greater genetic distinctiveness among genotypes in these clusters compared to others. Notably, rice landraces such as Borosolpana, Phougak, Satyaranjan, Kakcheng Phou, Moniram, Kanaklata, and Bahadur were identified as genetically divergent. These genotypes hold promise for generating segregating populations, serving as valuable source materials for targeted yield improvement through meticulous selection, as indicated by inter-cluster distances.

## Keywords

Genetic diversity; grain yield; landraces; multivariate analysis; rice

## Introduction

Rice (*Oryza sativa* L.) stands as a crucial linchpin for global food security, serving as a dietary staple for over half of the world's population. The cultivation of rice has been intricately woven into human history for millennia, leading to the emergence of diverse landraces and locally adapted varieties shaped by natural and farmer-driven selection processes. These landraces serve as invaluable genetic reservoirs, harboring essential traits for sustainable and resilient production systems (1). With the global population expected to reach 8 billion in November 2022 and India poised to surpass Chi-

na as the most populous country in 2023 (World Population Prospects, 2022), there is an urgent imperative to augment rice production to meet escalating demands. Achieving a 50% increase in rice production necessitates a focused effort by breeders to develop cultivars not only characterized by heightened yield potential but also possessing desirable agronomic traits (3). Effectively addressing these challenges requires a profound understanding of the genetic diversity within rice landraces. This paper aims to elucidate the significance of rice landraces as genetic resources and their pivotal role in contemporary agricultural research and development. Additionally, it underscores the critical necessity of harnessing this genetic diversity through the application of multivariate analysis, providing a robust toolkit to dissect the intricate relationships between genotypic and phenotypic variations (4). This analytical approach holds great promise for steering breeding programs towards the creation of improved rice varieties characterized by heightened adaptability, productivity, and resilience in the face of current and future challenges.

A comprehensive investigation, (5) delved into the genetic diversity elucidation of thirty-two elite rice varieties. Utilizing Mahalanobis' D2 statistics and Principal Component Analysis (PCA), the authors discerned pivotal factors influencing genetic variability. Time to maturity, single plant yield, and days to 50 percent flowering emerged as primary contributors to genetic diversity. The varieties were stratified into seven clusters. PCA revealed that the initial three components explicated 77.3 percent of the variability, with the first component, encapsulating single plant yield, panicles per plant, and time to maturity, attributing to 40.6 percent. This research yields crucial insights into the genetic diversity inherent in elite rice varieties.

The principal aim of this study was to delineate the inherent diversity among rice varieties. To achieve this objective, Mahalanobis D2 statistics have been employed as a robust statistical tool to quantify genetic distances between landraces. This methodology proves highly efficacious in categorizing varieties into discrete groups (6, 7). Furthermore, the utilization of PCA complements the insights derived from the D2 analysis by reinforcing the identification of key contributors to the variability between genotypes (8, 9). Through PCA, essential independent variables are succinctly represented while preserving the original variability (10). The amalgamation of Mahalanobis D2 statistics and PCA holds paramount significance in the selection of genetically divergent parents from a pool of 20 rice landraces, a critical aspect for the success of breeding programs.

## Materials and Methods

### Experimental site

The current investigation was conducted at Uttar Banga Krishi Viswavidyalaya situated in Pundibari, Coochbehar, West Bengal, within the terai region of Bengal. The geographical coordinates of the site were 26°19'86" N latitude

and 89°23'53" E longitude, at an elevation of 43 meters above mean sea level.

### Experimental materials

Twenty (20) distinct rice genotypes were procured from the Regional Agricultural Research Station located in Titabar, Jorhat, Assam, for the experimentation (Table 1). A comprehensive list of these genotypes is provided in Table 1. The study was conducted over the course of two seasons in 2017, encompassing both the pre-kharif and kharif periods. The soil characteristics at the experimental site include a sandy loam texture, optimal drainage conditions, and a pH level of 5.74

**Table 1.** List of 20 landraces of rice.

Name of Landraces	Denotation
Phougak	1
Disang	2
Kakcheng Phou	3
Joymati	4
Satyanranjan	5
Chakhao Sempak	6
Luit	7
Mahsuri	8
Borosolpana	9
Haripowa Sali	10
Kanaklata	11
Kushal	12
Ranjit	13
Dhansiri	14
Piolee	15
Diphalu	16
Moniram	17
Bahadur	18
Teti Sali	19
Phourin Nakuppi	20

### Experimental layout

A Randomized Complete Block Design (RCBD) experiment employing three replications was conducted to assess various genotypes for both yield and yield attributing traits. The experimental design comprised 20 treatments, with each plot measuring 2 m by 1.5 m. Plots were spaced at a 50 cm interval, with plant-to-plant spacing set at 15 cm and row-to-row spacing at 20 cm. Standardized amounts of nitrogen (N), phosphorus (P), and potassium (K) fertilizers were applied at rates of 60, 40, and 40 kg per acre, respectively. During land preparation, the entire dose of potash and phosphorus, along with half of the nitrogen, was applied as the basal dose. Subsequently, 22 days post-transplanting, one-fourth of the nitrogen was administered as the first top dressing, followed by the application of the remaining one-fourth nitrogen as the second top dressing 22 days later. Intercultural operations and crop protection measures were implemented as needed throughout the experimental duration.

### Observations recorded

Plant height (PH) was quantified in centimeters, measuring from the base to the apex of the primary panicle before the plants were harvested. Days to 50% flowering (DF) were recorded when half of the population within the experimental plot exhibited panicle emergence. Panicle length (PL) was determined in centimeters, measured from the basal node of selected plants to the tip of the uppermost spikelet. Effective tillers per plant (ETP) were assessed at maturity by counting from five randomly chosen plants, with the total divided by five to yield the effective tillers per plant in each replication. Filled grains per tiller (FGT) were calculated as the average number of fertile grains on 10 randomly selected panicles. Grain dimensions, including length (GL) and breadth (GB), were measured with digital slide calipers on 10 randomly chosen rice grains with husks, and the mean values were computed. The grain length-to-breadth ratio (GLBR) was determined as the averaged ratio of length to breadth for 10 grains. Thousand grain weight (TGW) was assessed by weighing selected grains with an electric balance, and the mean was calculated. Grain yield per plant (GY) was determined by harvesting five randomly selected plants, and the obtained grain amount was divided by five to derive the yield per plant. Harvest index (HI) was calculated and expressed as a percentage for each plot.

### Statistical analysis

The statistical analysis in this study involved aggregating average data from each replication. To ascertain the significance of variance among distinct genotypes (treatments), a randomized block design was implemented (11). Genetic diversity within a pool of 20 landraces was assessed using Mahalanobis D2 statistics, following the methodology outlined in reference 12. The resulting clustering pattern was proposed and visualized through hierarchical clustering (13) utilizing the dendextend package in R software (14) to elucidate relationships among genotypes based on their similarities and differences. Furthermore, principal components with eigenvalues surpassing one were explored using the FactoMineR package (15). This transformation facilitated the conversion of the original set of variables into a new set of uncorrelated variables, denoted as principal components. Biplots, generated using the Factoextra package (16) in R, were employed to visualize relationships between quantitative variables and individual observations, as well as to depict the clustering structure. The resulting dendrogram was integrated into these visualizations. These methodologies collectively offered a comprehensive evaluation of genetic diversity and relationships among the investigated genotypes.

### Results and discussion

The genetic diversity within a population significantly influences genetic advancement. Hence, it is imperative to assess the genetic diversity among a collection of breeding materials, enabling the categorization of genotypes into genetically analogous and dissimilar types. The anticipation lies in the recognition that genotypes displaying sig-

nificant genetic divergence are conducive to recombination breeding, thereby offering a broad array of genetic variations and increased opportunities for the emergence of transgressive segregants (17, 18). Consequently, the assessment of the degree of genetic divergence among the existing rice genotypes has been undertaken.

Utilizing Euclidean genetic distance measurements between potential pairs of test landraces, the research categorized the 20 rice landraces into seven distinct genetic clusters based on their grain yield and attributing traits, as outlined in Table 2. Notably, Cluster IV emerged as the largest cluster, encompassing six landraces (Joymati, Mahsuri, Haripowasali, Kushal, Piolee, and Teti Sali). Subsequently, Cluster II featured four landraces (Disang, Luit, Dhansiri, and Diphalu), and Cluster V comprised three landraces (Satyaranjan, ChakhaoSempak, and Phourin Nakuppi). Clusters I, III, and VII each accommodated two landraces: Phougak and Moniram in Cluster I, Kakcheng Phou and Kanaklata in Cluster III, and Ranjit and Bahadur in Cluster VII, as detailed in Table 2. The clustering was visually represented through a dendrogram, illustrating

**Table 2.** Grouping of 20 landraces of rice on different clusters for yield and yield attributing traits.

Cluster No.	Total no. of genotypes	Name of genotypes
I	2	Phougak, Moniram
II	4	Disang, Luit, Dhansiri, Diphalu
III	2	Kakcheng Phou, Kanaklata
IV	6	Joymati, Mahsuri, Haripowasali, Kushal, Piolee, Teti Sali
V	3	Satyaranjan, ChakhaoSempak, PhourinNakuppi
VI	1	Borosolpana
VII	2	Ranjit, Bahadur

the relationships among the 20 rice landraces concerning both yield and yield attributing traits. Fig. 1 illustrates the outcomes of clustering analyses, revealing the grouping patterns of 20 rice landraces through D2 analysis. Intriguingly, the examination disclosed seven distinct clusters, challenging the conventional assumption that germplasm with a shared origin would inherently belong to the same category. Instead, it underscored that genetic divergence was predominantly influenced by morphological traits rather than geographical provenance alone. Moreover, the application of Mahalanobis D2 statistics validated the clear categorization of the 20 rice landraces into five groups (Fig. 1). Subsequent chi-square testing accentuated highly significant differences ( $p < 0.01$ ) among these clusters, aligning with earlier research (19) that utilized D2 analysis to classify rice genotypes into seven clusters, thereby affirming the robustness of this analytical approach. The observed genetic diversity, leading to the formation of distinct solitary clusters, is presumed to result from geographical barriers impeding gene flow or the effects of intense natural and human selection for unique and adaptive gene complexes. This finding aligns with similar results reported in a previous study (20). Another study on

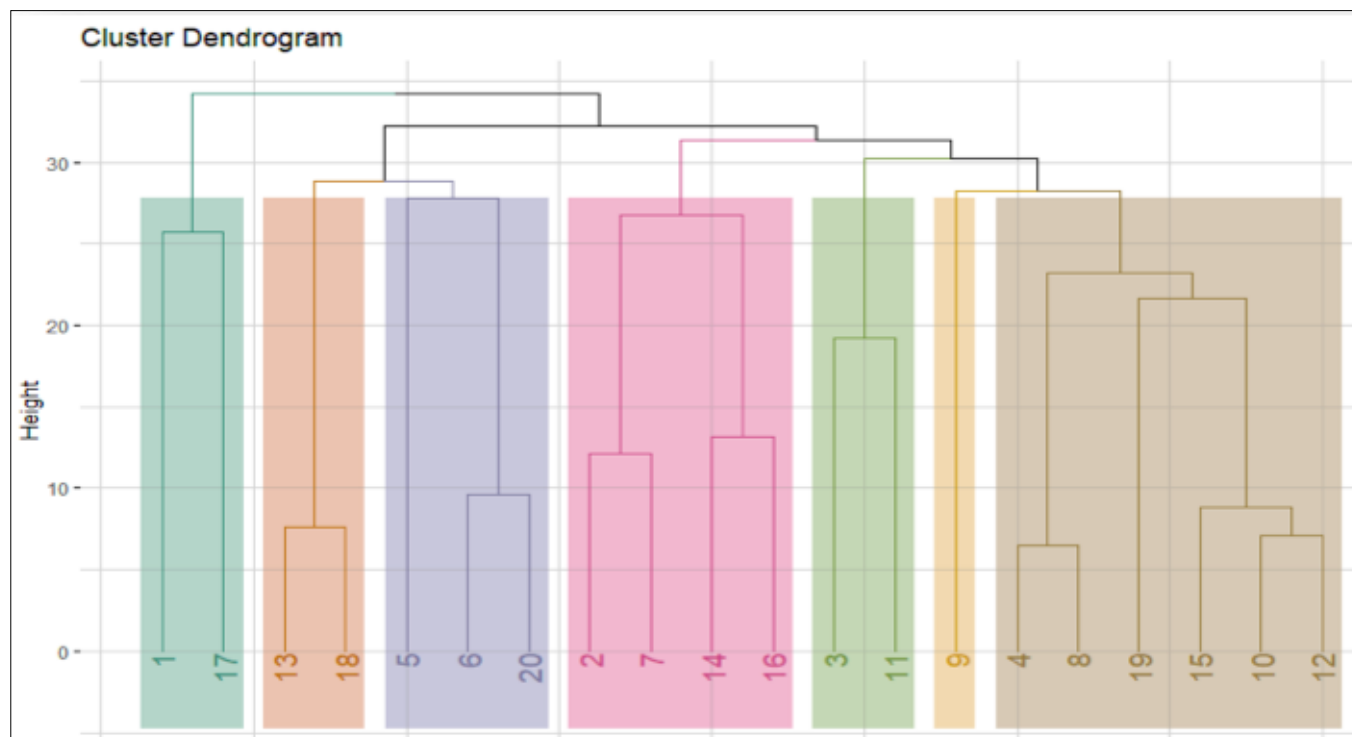


Fig. 1. Dendrogram showing the clustering of 20 landraces of rice for yield and yield attributing trait.

rice diversity analysis (21) allocated twenty-six genotypes into six clusters, with the maximum inter-cluster distance noted between clusters V and VI. Additionally, a previous investigation on rice genotypes (22) identified ten clusters, where cluster I comprised a single genotype, while clusters II and III each encompassed nine genotypes.

In Table 3, it is discernible that Cluster I manifests the highest mean intra-cluster distances for both grain yield and its associated traits, registering a value of 12.87. Closely following are Clusters III and V, exhibiting distances of 9.60, succeeded by Cluster II at 8.97, and Cluster IV at 7.18. Cluster VI, being monogenotypic, records an intra-cluster distance of 0, while Cluster VII displays a distance of 3.81. Notably, Clusters VI and I demonstrate the most substantial inter-cluster distance at 32.33, succeeded by Clusters VI and V (29.18), III and I (28.79), V and I (27.75), VI and III (27.15), and finally, Cluster VII and I with a distance of 26.51. Based on the inter-cluster distances, it is anticipated that hybridization between genotypes from Cluster VI with Cluster V, Cluster III with I, and Cluster V with I could yield promising candidates with enhanced grain yield and other pivotal traits in rice (Table 3). Moreover, incorporating genotypes from more distantly positioned clusters in hybridization programs has the potential to

introduce a broader spectrum of variability, facilitating the generation of superior candidates for grain yield and essential agronomic characteristics. It is recommended to select parents from two clusters with wider inter-cluster distances, as suggested by prior investigators (23, 24), to achieve a robust heterotic impact and foster high variability.

The examination of genetically divergent clusters and the calculation of distances ( $D^2$  values) among selected *Oryza sativa* L. landraces are presented in Table 4. Upon meticulous scrutiny of the distances, it was noted that Borosolpana in cluster VI and Phougak in cluster I exhibited a conspicuously high genotypic distance ( $D^2 = 34.12$ ). Analogous patterns of substantial genetic distances between landraces were discerned in other clusters, such as the case of Borosolpana in cluster VI and Satyaranjan in cluster V ( $D^2 = 32.31$ ). Kakcheng Phou in cluster III and Maniram in cluster I manifested a considerable genotypic distance ( $D^2 = 34.70$ ). Furthermore, Satyaranjan in cluster V and Phougak in cluster I displayed a substantial genotypic distance ( $D^2 = 30.42$ ), and Bahadur in cluster VII and Phougak in cluster I exhibited a noteworthy distance as well ( $D^2 = 30.11$ ) (Table 4). The identification of parental lines from these distinct clusters holds considerable potential for

Table 3. Average intra (diagonal) and inter-cluster (off-diagonal)  $D^2$  values of 20 rice landraces for different yield and yield attributing traits.

Cluster	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V	Cluster VI	Cluster VII
Cluster I	<b>12.87</b>	26.49	28.79	23.84	27.75	32.33	26.51
Cluster II		<b>8.97</b>	23.61	19.91	23.39	26.44	21.00
Cluster III			<b>9.60</b>	21.63	24.56	27.15	24.28
Cluster IV				<b>7.18</b>	21.32	23.08	18.85
Cluster V					<b>9.60</b>	29.18	21.91
Cluster VI						<b>0.00</b>	24.82
Cluster VII							<b>3.81</b>

**Table 4.** Genetically divergent clusters and distance ( $D^2$  value) between the landraces selected for yield and its attributing traits in 20 landraces of rice.

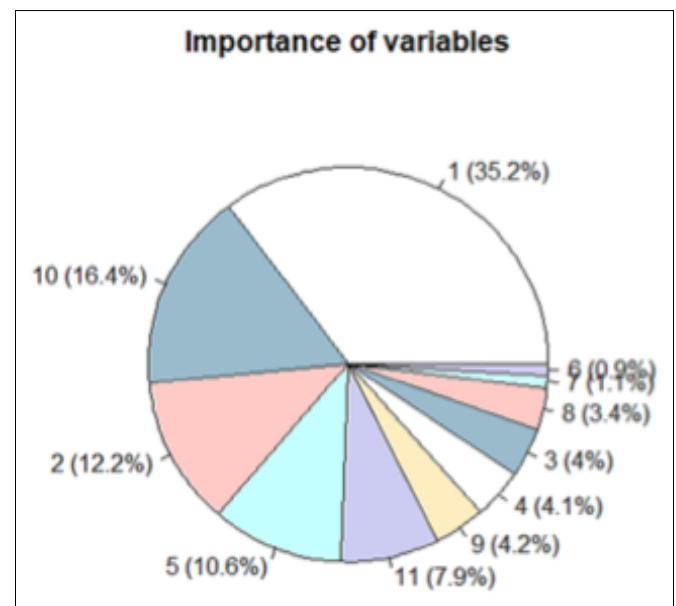
Cluster combination	Inter cluster distance ( $D^2$ value)	Landraces selected from the cluster	Distance between the landraces selected ( $D^2$ value)
Cluster VI and cluster I	32.33	Borosolpana in cluster VI and Phougak in cluster I	34.12
Cluster VI and cluster V	29.18	Borosolpana in cluster VI and Satyaranjan in cluster V	32.31
Cluster III and cluster I	28.79	Kakcheng Phou in cluster III and Maniram in cluster I	34.70
Cluster V and cluster I	27.75	Satyaranjan in cluster V and in Phougak in cluster I	31.97
Cluster VI and cluster III	27.15	Borosolpana in cluster VI and Kanaklata in cluster III	30.42
Cluster VII and cluster I	26.51	Bahadur in cluster VII and Phougak in cluster I	30.11

hybridization programs, as mating genetically divergent parents often leads to enhanced variability and the emergence of transgressive segregations with robust heterotic effects. The magnitude of inter-cluster distances directly corresponds to the level of variability among genotypes between clusters and vice versa. Consequently, a deliberate mating strategy involving genotypes from these clusters has the potential to yield robust hybrids or recombinants with heightened vigor. This observation aligns with previous findings (25) reported in the literature.

### Cluster means

Table 5 illustrates the mean values of eleven yield and yield-related traits within each identified cluster. In Cluster I, comprising two distinct landraces, noteworthy attributes included maximum GB at 2.35 mm, highest GYP at 20.24 g, moderate FGT at 368.50, and TGW at 21.92 g. Additionally, Cluster I exhibited the shortest PH at 123.67 cm and PL at 22.99 cm. Cluster II, consisting of four landraces, exhibited characteristics such as high HI at 52.86, moderate ETP at 10.17, and GL at 7.95 mm, alongside late DF at 76.46 and the lightest TGW at 21.02 g. Cluster III, comprising two landraces, displayed distinctive features including the highest ETP at 11.41, elevated PH at 149.06 cm, increased FGT at 429.67, moderate GYP at 18.97 g, and the lowest GLBR at 3.56. Cluster IV, encompassing six landraces, was characterized by early DF at 98.69, high ETP at 11.41, moderate PH at 134.94 cm, GB at 2.06 mm, GLBR at 3.88, and HI at 39.62. In Cluster V, composed of three landraces, notable traits included the highest GL at 9.04 mm, elevated GLBR at 4.53, high TGW at 24.50 g, moderate DF at 82.39, and PL at 25.78 cm, along with the lowest FGT at 293.05. Cluster VI, with a single landrace, exhibited the tallest PH at 196.68 cm, earliest DF at 106.17, largest PL at 33.33 cm, maximum FGT at 535.67, heaviest TGW at 27.40

g, highest HI at 55.83, and GYP at 14.98 g. Lastly, Cluster VII, consisting of two landraces, exhibited lengthy PL at 28.16 mm and GB at 2.24 mm. However, this cluster also presented the lowest HI at 33.52 and GYP at 14.98 g. The contribution of individual yield and yield-related traits to total divergence is depicted in Fig. 2, where PH contributed the maximum at 35.2%, followed by HI at 16.4%, DF at 12.2%, FGT at 10.6%, and GYP at 7.9%. This information underscores the significance of these traits in genotype selection, with plant height emerging as a primary determinant for divergence and potential utility in parent selection (26, 27).

**Fig. 2.** Contribution of different yield attributing traits towards divergence in 20 landraces of rice. 1= PH (cm), 2= DF, 3=PL (cm), 4=ETP, 5=FGT, 6=GL (mm), 7=GB (mm), 8=GLBR, 9=TGW (g), 10= HI (%) and 11=GYP (g).**Table 5.** Cluster means of 20 landraces of rice for different yield and yield attributing traits.

Cluster	PH (cm)	DF	PL (cm)	ETP	FGT	GL(mm)	GB (mm)	GLBR	TGW (g)	HI (%)	GYP (g)
Cluster I	123.67	81.09	22.99	9.67	368.50	8.17	2.35	3.72	21.92	50.95	20.24
Cluster II	130.57	76.46	25.74	10.17	327.88	7.95	2.00	4.05	21.02	52.86	16.53
Cluster III	149.06	78.08	25.06	11.41	429.67	7.59	2.20	3.56	21.73	38.50	18.97
Cluster IV	134.94	98.69	26.10	11.31	402.47	7.87	2.06	3.88	21.56	39.62	19.60
Cluster V	144.65	82.39	25.78	9.22	293.05	9.04	2.00	4.53	24.50	34.42	17.11
Cluster VI	196.68	106.17	33.33	7.50	535.67	8.30	1.92	4.34	27.40	55.83	23.15
Cluster VII	129.36	89.92	28.16	10.75	336.67	7.88	2.24	3.79	22.93	33.52	14.98
Contribution%	35.20	12.20	4.00	4.10	10.60	0.90	1.10	3.40	4.20	16.40	7.90

PH= Plant height (cm), DF=Days to 50% flowering, PL=Panicle length (cm), ETP= effective tillers plant<sup>-1</sup>, FGT= filled grains tiller<sup>-1</sup>, GL= grain length (mm), GB= grain breadth (mm), GLBR=Grain l:b ratio, TGW= 1000 grain weight (g), HI= harvest index (%) and GYP= grain yield plant<sup>-1</sup> (g).

### Principal component analysis (PCA)

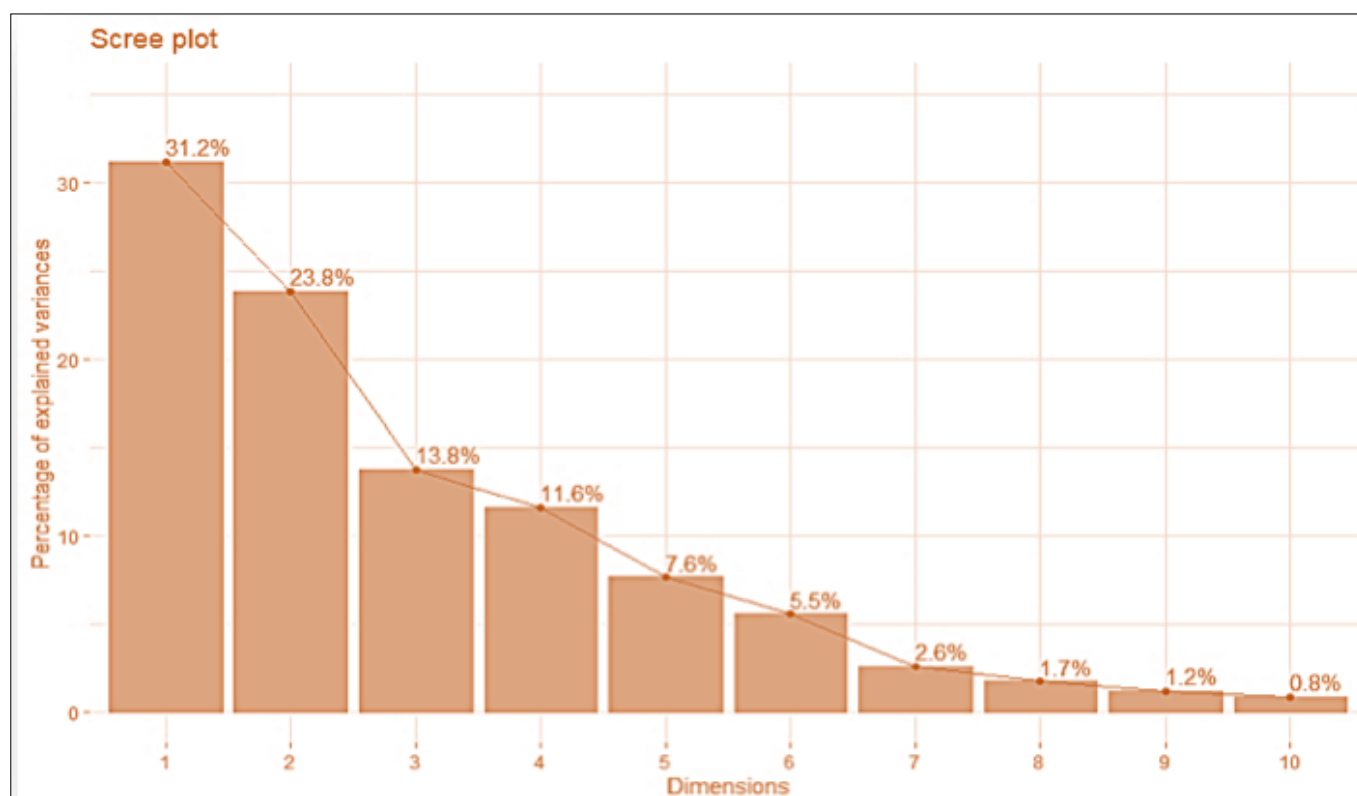
PCA serves as a multivariate statistical method aimed at selecting a reduced set of components from the PCs, effectively eliminating less critical information. This process involves arranging genotypes based on PC scores, prioritizing and elucidating the major variability within the overall variance (28, 29). The determination of the number of retained factors is guided by eigenvalues. Table 6 presents the eigenvalues, percentage of variance, and cumulative variance for quality-related traits across 20 landraces. Our investigation identified eleven PCs, and the significance was attributed to the first four PCs with eigenvalues exceeding one. The remaining seven PCs offered limited additional insights and only partially accounted for the observed variation. In this study, PCA highlighted four principal components, namely PC1, PC2, PC3, and PC4, each possessing eigenvalues greater than one (3.43, 2.63, 1.51, and 1.27, respectively). Collectively, these four components explained approximately 80.45% of the total variation. To gain a more nuanced understanding, emphasis was placed on the variation associated with these four PCs (Table 6). Notably, characters in the first principal component with absolute values closer to unity exerted a more pronounced influence on grouping than those with smaller absolute values nearing zero.

**Table 6.** Principal component analysis eigen values, variance and cumulative variance of yield and attributing traits in rice landraces.

	PC1	PC2	PC3	PC4
Eigen Value	3.43	2.63	1.51	1.27
Variance (%)	31.19	23.86	13.75	11.63
Cumulative (%)	31.19	55.06	68.81	80.45

Following the extraction of PC4, a semi-curve line was delineated, exhibiting a trend toward linearity with minimal fluctuations in each PC. The scree plot graph vividly illustrates that PC1 encapsulates the highest proportion of variance compared to the other PCs (Fig. 3). Consequently, favoring the lines derived from PC1 for inclusion in genetic improvement programs holds considerable advantages. These findings are consistent with the outcomes of prior investigations by (30), elucidating the distribution of variance across principal components. In a study involving forty rice genotypes, (31) reported that PC1 accounted for 33% of the variability, while PC2, PC3, and PC4 contributed 14.3%, 11.4%, and 9%, respectively. Similarly, a separate investigation (32) demonstrated that the first two principal components explained 23% of the variation in sixteen agro-morphological parameters across twenty-three rice germplasm lines, with eigenvalues exceeding 1. In an earlier study analyzing 31 rice germplasm lines, (33) found that the first five components with eigenvalues greater than 1 collectively explained 82.90% of the overall variances. Moreover, research on rice landraces (34) unveiled that, among thirteen principal components, five were statistically significant, contributing significantly to the total variance and amounting to 84.67% in cumulative proportion of variance. A prior investigation in Pearl millet, involving forty germplasm, revealed that the first six principal components, each with an eigenvalue surpassing one, collectively contributed to 78.29% of the observed variability (35).

Analysis of Eleven Key Yield-Related Components in Twenty Rice Landraces: Insights from Multivariate Genetic Assessment. The PC1, explaining 31.2% of the total variability, underscores discriminatory traits such as PH (0.467), PL (0.473), GLBR (0.418), and DF (0.339). This suggests their



**Fig. 3.** Scree plot showing the clustering of 20 landraces of rice for yield and its attributing traits.

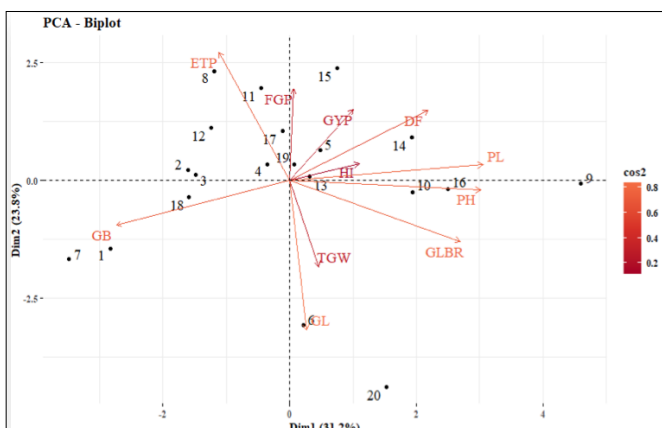
pivotal role in delineating the inherent quality potential of each landrace, leading to distinct cluster formations. Similarly, the second PC2, contributing 23.8% to the overall variation, notably influences ETP (0.481) and FGT (0.343) (Table 7). These findings underscore the richness of trait variants within rice landraces, implying substantial prospects for genetic advancements in the investigated traits.

**Table 7.** Eleven principal components along with their factor's loadings for yield and its attributing traits of 20 rice landraces.

Traits	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
PH (cm)	0.467	-0.037	-0.223	0.135	-0.211	0.023	-0.138	-0.665	-0.291	-0.343	-0.056
DF	0.339	0.262	-0.167	0.271	0.443	-0.345	-0.110	0.383	0.152	-0.465	-0.048
PL (cm)	0.473	0.059	-0.010	0.217	-0.281	-0.183	-0.304	0.037	0.386	0.586	0.165
ETP	-0.172	0.481	0.129	0.300	0.195	0.109	0.377	-0.484	0.451	0.039	-0.004
FGT	0.010	0.343	-0.441	-0.006	-0.505	0.502	0.094	0.329	0.155	-0.193	-0.053
GL (mm)	0.041	-0.562	-0.072	0.027	0.198	0.375	-0.082	-0.057	0.474	-0.253	0.445
GB (mm)	-0.423	-0.168	-0.377	0.040	0.005	-0.163	-0.445	-0.168	0.322	0.009	-0.541
GLBR	0.418	-0.230	0.329	-0.028	0.101	0.338	0.188	0.102	0.146	0.044	-0.686
TGW (g)	0.072	-0.326	-0.583	0.173	0.102	-0.195	0.621	0.029	-0.083	0.275	-0.040
HI (%)	0.171	0.062	-0.045	-0.769	-0.131	-0.363	0.219	-0.121	0.370	-0.159	0.005
GYP (g)	0.156	0.267	-0.337	-0.383	0.557	0.364	-0.229	-0.100	-0.149	0.341	0.033

**PH**= Plant height (cm), **DF**=Days to 50% flowering, **PL**=Panicle length (cm), **ETP**= effective tillers plant<sup>-1</sup>, **FGT**= filled grains tiller<sup>-1</sup>, **GL**= grain length (mm), **GB**= grain breadth (mm), **GLBR**=Grain l:b ratio, **TGW**= 1000 grain weight (g), **HI**= harvest index (%) and **GYP**= grain yield per plant (g).

The PC1 and PC2 were utilized to construct a biplot, facilitating an examination of the interrelationships among twenty rice landraces based on yield and yield-related traits in the present study (Fig. 4). The covariate effects of biplot-based correlations among traits elucidated 55.06% of the total variation, representing a robust approximation for understanding the impact of traits on yield and their inter-similarities. Notably, characters such as DF, PL, PH, GLBR, GL, GB, and ETP exhibited longer vector lengths, indicating a more substantial influence on the variation in a particular dimension. Conversely, traits like FGT, GYP, HI, and TGW displayed shorter vector lengths, suggesting a limited contribution to the observed variation. Piolee, Satyaranjan, and Dhansiri formed a distinctive group in the first quadrant of the biplot's right upper corner, exhibiting positive values for both PCs. Traits such as GYP, FGT, and DF were positioned in the same quadrant, as illustrated in Fig. 4. Previous research on South Indian rice landraces identified positive correlations between PC1 and PC2 for

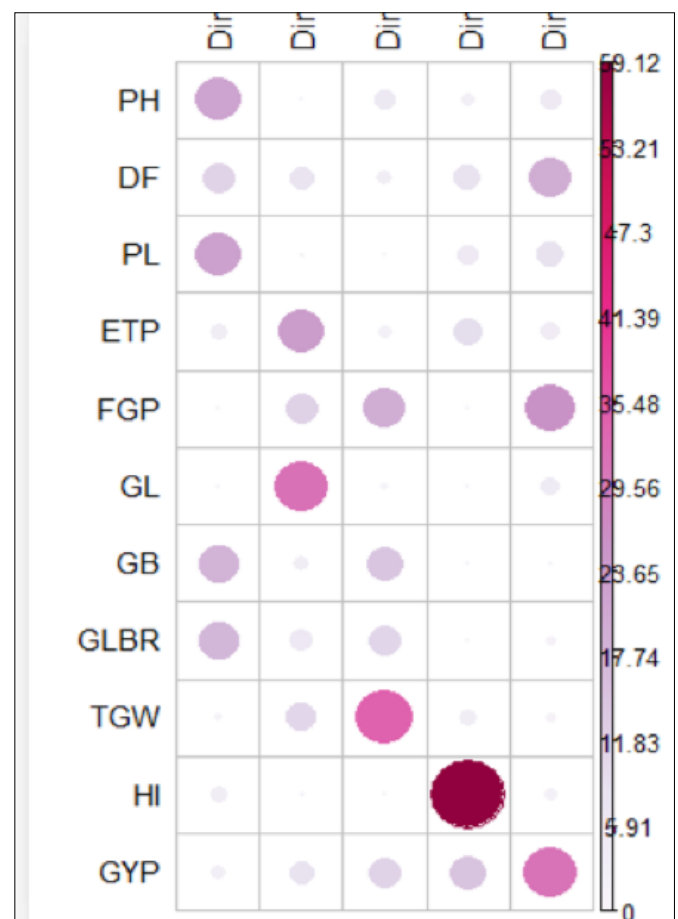


**Fig. 4.** Biplot among landraces for yield and its attributing traits.

traits such as spikelet fertility, kernel length, and length-to-breadth ratio. Similarly, correlations were observed for eight traits, including days to fifty percent flowering, plant height, flag leaf width, panicle length, number of grains, spikelet fertility, length-to-breadth ratio, and single plant yield. The grouping of traits within distinct principal components suggests potential priority selection in breeding

programs, emphasizing their tendency to co-segregate (36, 37).

In the presented study, a comprehensive examination of the contribution of various yield attributing traits to PCs was conducted, as illustrated in the corrplot (Fig. 5).



**Fig. 5.** Contribution of yield attributing traits to significant PC.

Notably, traits such as PH, DF, PL, GLBR, and GB were found to be significantly associated with PC1, explaining 31.19% of the total variance. PC2, contributing to 23.86% of the overall variance, was primarily influenced by traits including ETP, FGT, GL, and TGW. Conversely, PC3, explaining 13.75% of the total variation, predominantly correlated with GLBR and TGW. PC4, contributing 11.63% to the total variation, exhibited a primary association with ETP and HI (Fig. 5). A parallel observation in aromatic rice (38) identified 24 genotypes and specific traits, such as Days to 50% Flowering, Plant Height, and Grain Width, displaying positive values in the biplot between PC1 and PC2. A separate investigation on Indonesian local rice germplasm revealed that PC1 was predominantly influenced by Productive Tiller Number, accounting for 32.54% of the total variability. Additionally, Culm Length and Plant Height collectively contributed 22.1% to the overall variability. Furthermore, PC3, driven by Panicle Length, Culm Diameter, and Flag Leaf Width, explained 9.93% of the overall variation. Notably, the top two PCs collectively elucidated 54.65% of the total variation. Moreover, in-depth research demonstrated that PCA of traits such as Panicle Exertion, Flag Leaf Blade Width, and Panicle Length generated five PCs, collectively contributing 80.00% towards the observed diversity (39).

## Conclusion

The analysis of genetic diversity among rice landraces provides comprehensive insights into improving grain yield and morphological traits. Clustering, based on genetic distance, reveals seven distinct clusters that challenge assumptions about shared landraces, highlighting the influence of morphological traits on geographical origin. Statistical tests and Mahalanobis D2 analysis confirm significant differences among clusters, emphasizing opportunities for hybridization between genetically divergent landraces like Borosolpana, Phougak, Satyaranjan, Kakcheng Phou, Maniram, Satyaranjan, Kanaklata, and Bahadur to create superior segregating populations. Traits such as PH, HI, and DF markedly contribute to divergence among clusters. PCA of twenty rice landraces identifies four pivotal components (PC1 to PC4) explaining about 80.45% of the total variability. PC1, comprising 31.2% variability, highlights crucial traits like PH, PL, GLBR, and DF. Biplot analysis showcases distinct groups among landraces based on yield traits, emphasizing trait correlations with specific PCs. This insight directs targeted genetic improvement efforts, particularly focusing on PC1 to select lines with superior genetic potential for quality traits and yield.

## Acknowledgements

Financial assistance and technical assistance by the Department of Genetics and Plant Breeding, UBKV and Directorate of Research, UBKV is duly acknowledged.

## Authors contributions

The research was conceptualized and designed LH. AR conducted the field experiments. SS, PS and SS evaluated the data. AR, SR and AG drafted the manuscript. LH, MC, NS, and SD edited the manuscript. The findings of the final manuscript were discussed, read and approved by all the authors.

## Compliance with ethical standards

**Conflict of interest:** The authors claim they are not aware of any conflicts of interest that would have seemed to have an impact on the work they described in this manuscript.

**Ethical issues:** None.

## References

1. Azeez MA, Adubi AO, Durodola FA. Landraces and crop genetic improvement. In *Rediscovery of Landraces as a Resource for the Future*. IntechOpen. 2018 Sep 12. <http://dx.doi.org/10.5772/Intechopen.75944>
2. United Nations, Department of economic and social welfares. *World Population Prospectus*; 2022.
3. Akinwale MG, Gregorio G, Nwilene F, Akinyele BO, Ogunbayo SA, Odiyi AC. Heritability and correlation coefficient analysis for yield and its components in rice (*Oryza sativa* L.). *Afr J Plant sci*. 2011 Mar 31;5(3):207-12. <https://doi.org/10.3923/ijpb.2011.224.234>
4. Chang TT, Bardenas EA. The morphology and varietal characteristics of the rice plant. *Int Rice Res Inst*. 1965.
5. Francis N, Packiaraj D. Genetic uniformity of varieties and an assessment on the diversity among the elite varieties of rice (*Oryza sativa* L.). *Electron J Plant Breed*. 2020;11(3):896-900. <https://doi.org/10.37992/2020.1103.146>
6. Hoque A, Begum SN, Robin AH, Hassan L. Partitioning of rice (*Oryza sativa* L.) genotypes based on morphometric diversity. *Am J Exp Agric*. 2015;7:242-50. <https://doi.org/10.9734/AJEA/2015/15687>
7. Fowler C, Mooney PR. *Shattering: Food, politics and the loss of genetic diversity*. University of Arizona Press; 1990.
8. Mahajan RK, Mehan DK. Principal component analysis in rice. *Crop Improvement*. 1980;7(2):83-87.
9. Sheela KS, Robin S, Manonmani S. Principal component analysis for grain quality characters in rice germplasm. *Elect J Plant Breed*. 2020 Apr 6;11(1):127-31. <https://doi.org/10.37992/2020.1101.023>
10. Jolliffe IT. *Principal component analysis for special types of data*. Springer New York. 2002;p. 338-73. [https://doi.org/10.1007/0-387-22440-8\\_13](https://doi.org/10.1007/0-387-22440-8_13)
11. Panse VG, Sukhatme PV. *Statistical methods for Agricultural worker*. Indian Council of Agricultural Research. New Delhi; 1969.
12. Rao CR. *Advance statistical methods in biometrical research edition I*. John Willey and Sons, New York. Rapeseed cultivars. *J Appl Biol Sci*. 1952;2(3):35-39.
13. Ward. Hierarchical grouping to optimize an objective function. *J Am Stat Asso*. 1963;48:236-44. <https://dx.doi.org/10.1080/01621459.1963.105008>
14. Galili T. *Dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering*. *Bioinformatics*; 2015. <https://doi.org/10.1093/bioinformatics/btv428>
15. Le S, Josse J, Husson F. *FactoMine R: An R package for multivariate analysis*. *Journal of Statistical Software*. 2008;25(1):1-18. <https://doi.org/10.18637/jss.v025.i01>



16. Kassambara A, Mundt F. Factoextra: Extract and visualize the results of multivariate data analyses. R Package Version 1.0.7.; 2020. <https://CRAN.R-project.org/package=factoextra>
17. Zaman MR, Paul DNR, Kabir MS, Mahbub MAA, Bhuiya MAA. Assessment of character contribution to the divergence for some rice varieties. *Asian J of Plant Sci.* 2005;4:388-91. <https://doi.org/10.3923/ajps.2005.388.391>
18. Saxesena RR, Lal GM, Yadav PS, Vishwakarma MK. Diversity analysis and identification of promising lines for hybridization in field pea (*Pisum sativum* L.). *The Bioscan.* 2013 Nov 21;8(4):1437-40.
19. Tripathi A, Kumar S, Singh MK, Kumar A, Karnwal MK. Phenotypic assessment of rice (*Oryza sativa* L.) genotypes for genetic variability and varietal diversity under direct seeded condition. *J Appl Nat Sci.* 2017;9(1):6-8. <https://doi.org/10.31018/jans.v9i1.1138>
20. Beevi AH, Venkatesan M. Genetic divergence studies in rice genotypes under saline condition. *Int J Current Adv Res.* 2015;4:6-8.
21. Lakshmi M, Shanmuganathan M, Jeyaprakash P, Ramesh T. Genetic variability and diversity analysis in selected rice (*Oryza sativa* L.) varieties. *Electron J of Plant Breed.* 2022;13(3):959-66. <https://doi.org/10.37992/2022.1303.124>
22. Lavanya1 K, Suman K, Fiyaz AR, Chiranjeevi M, Surender R, Krishna Satya A, Sudhakar P, Subba Rao LV. Phenotypic assessment of rice landraces for genetic variability and diversity studies under heat stress. *Oryza.* 2022;59(1):31-38. DOI <https://doi.org/10.35709/ory.2022.59.1.4>
23. Mishra LK, Sarawgi AK, Mishra RK. Genetic diversity for morphological and quality traits in rice (*Oryza sativa* L.). *Adv Plant Sci.* 2003;16(1):287-93.
24. Chaturvedi HP, Maurya DM. Genetic divergence analysis in rice (*Oryza sativa* L.). *Adv Plant Sci.* 2005;18(1):349-53.
25. Zewdu Z, Abebe T, Mitiku T, Worede F, Dessie A, Berie A, Atnaf M. Performance evaluation and yield stability of upland rice (*Oryza sativa* L.) varieties in Ethiopia. *Cogent Food & Agriculture.* 2020;6(1):1842679. <https://doi.org/10.1080/23311932.2020.1842679>
26. Nayak AR, Chaudhury D, Reddy JN. Genetic divergence in scented rice. *Oryza.* 2004;41(384):79-82.
27. Ovung CY, Lal GM, Rai PK. Studies on genetic diversity in Rice (*Oryza sativa* L.) *Journal of Agricultural Technology.* 2012;8(3):1059-65.
28. Anderson TW. An introduction to multivariate analysis. Wiley Eastern Pvt. Ltd. New Delhi. 1972; p. 512.
29. Morrison DE. Multivariate statistical methods. 2nd ed. 4th Print, McGraw Hill Kogakusta Ltd; 1978.
30. Sathya Sheela KRV, Robin S, Manonmani S. Principal component analysis for grain quality characters in rice germplasm. *Electron J Plant Breed.* 2019;11(1):127-31. <https://doi.org/10.37992/2020.1101.023>
31. Ashfaq M, Khan AS, Khan SHU, Ahmad R. Association of various morphological traits with yield and genetic divergence in rice (*Oryza Sativa* L.). *Int J Agri and Biol.* 2012;14:55-62.
32. Maji AT, Shaibu AA. Application of principal component analysis for rice germplasm characterization and evaluation. *J Plant Breed Crop Sci.* 2012;4(6):87-93. <https://doi.org/10.5897/JPCS11.093>
33. Ahmed MSU, Khalequzzaman MMD, Bashar K, Shamsuddin AKM. Agro-morphological, physico-chemical and molecular characterization of rice germplasm with similar names of Bangladesh. *Rice Sci.* 2016;23(4):211-18. <https://doi.org/10.1016/j.rsci.2016.06.004>
34. Dhakal A, Pokhrel A, Sharma S, Poudel A. Multivariate analysis of phenotypic diversity of rice (*Oryza sativa* L.) landraces from Lamjung and Tanahun districts of Nepal. *Int J of Agron.* 2020;1-8. <https://doi.org/10.1155/2020/8867961>
35. Singh OV, Gowthami R, Singh K, Shekhawat N. Genetic divergence studies in pearl millet germplasm based on principal component analysis. *Int J Curr Microbiol App Sci.* 2018;7(06):522-27. doi: <https://doi.org/10.20546/ijcmas.2018.706.059>
36. Shanmugam A, Suresh R, Ramanathan A, Anandhi P, Sassikumar S. Unravelling genetic diversity of South Indian rice landraces based on yield and its components. *Elect J Plant Breed.* 2023;14(1):160-69. <https://doi.org/10.37992/2023.1401.007>
37. Chakravorty A, Ghosh PD, Sahu PK. Multivariate analysis of phenotypic diversity of landraces of rice of West Bengal. *Ame J Exp Agri.* 2013;3(1):110. <https://doi.org/10.9734/AJEA/2013/2303>
38. Shoba D, Vijayan R, Robin S, Manivannan N, Iyanar K, Arunachalam P, Nadarajan N, Pillai MA, Geetha S. Assessment of genetic diversity in aromatic rice (*Oryza sativa* L.) germplasm using PCA and cluster analysis. *Electron J Plant Breed.* 2019 Oct 1;10(3):1095-104. <https://doi.org/10.5958/0975-928X.2019.00140.6>
39. Mulsanti IW, Risliawati A, Yunan N. Agro-morphological characterization based genetic diversity of Indonesian local rice germplasm. *Earth Env Sci.* 2021;948. <https://doi.org/10.1088/1755-1315/948/1/012004>