Physical Education, Health and Social Sciences
https://journal-of-social-education.org
E-ISSN: 2958-5996
P-ISSN: 2958-5988

Diversity Assessment of Rice (Oryza sativa L.) Genotypes Based on Quantitative Traits and Multivariate Analysis

[1]Tahira Bibi, [1]Ayesha Bibi, [1]Rana Ahsan Raza Khan, [1]Mudassar Ali, [1]Muhammed Usman Saleem, [1]Hira Sehar, [1]Misbah Riaz, [2]Iurem Shahzadi

[1] Rice Research Institute Kala Shah Kaku
[2] University of Agriculture Faisalabad

Abstract:
Rice (Oryza sativa L.) is an important staple crop and a major contributor to food security and agricultural economies worldwide. This study evaluated the genetic diversity among 25 fine-grain rice genotypes, including advanced breeding lines and commercial checks, using 14 agro-morphological, yield, and grain quality traits under field conditions at the Rice Research Institute, Kala Shah Kaku, Punjab, Pakistan. Significant variability was observed among the genotypes for most traits, indicating substantial genetic diversity. Principal Component Analysis (PCA) revealed that the first six principal components explained 81.27% of the total variation, with PC1 and PC2 accounting for 39.87%. Traits such as panicle length, days to flowering, days to maturity, and paddy yield contributed markedly to genotype differentiation. Cluster analysis grouped the genotypes into distinct clusters, reflecting varying levels of genetic divergence. Highly divergent genotypes identified through Euclidean distance analysis may serve as valuable parents in breeding programs. The results demonstrate the effectiveness of multivariate analysis in characterizing genetic diversity and provide useful information for the selection and development of improved high-yielding rice cultivars.

Keywords: Rice, morphological traits, PCA, Cluster, Multivariate analysis, Phenotypic diversity

Introduction:

Rice (Oryza sativa L.) is a foundational crop for global food security, serving as the daily meal for over two-thirds of the world's population and providing approximately 20% of all human caloric intake (Mohidem et al., 2022). Globally, rice production is heavily centralized, with more than 90% of total world production originating from Asian countries (Mohidem et al., 2022). In recent market cycles, global rice production scaled to 537 million tonnes, out of which roughly 53 million tonnes are traded internationally, meaning that approximately 9.87% of the world's rice production enters global trade networks (Glauber & Mamun, 2025). Within this international trade framework, Pakistan plays a vital strategic role, traditionally positioning itself as a major global exporter (5.8 million tons), which accounts for 11% of the total world rice trade.

Domestically, rice stands as Pakistan's primary agricultural economic asset and is classified as the country's second-largest cash crop, trailing only behind cotton (Khan et al., 2020). It contributes roughly 1.9% of the value-added in agriculture and adds about 0.4% to the national gross domestic product (GDP) (Khan et al., 2023). Cultivated as a premier commercial crop, Pakistan firmly secures its position as the 9th largest rice producer worldwide (Khan et al., 2023). In the 2024–2025 cropping season, Pakistan achieved a production output of 9.4 million metric tons of rice, harvested from a total cultivated area under rice expanding to 3.80 million hectares. This production is concentrated within well-irrigated ecological zones, particularly within the Punjab

province, which is globally celebrated for its premium, long-grain aromatic Basmati varieties, and the Sindh province, which specializes in high-yielding, non-basmati long-grain rice and modern hybrids.

Despite these production milestones, significant bottlenecks remain to achieving long-term food security and maintaining export momentum. Rice cultivation in Pakistan is increasingly constrained by climate change, erratic weather disruptions, skyrocketing input prices, and escalating water scarcity, which heavily pressures traditional, water-intensive puddled transplanting techniques. Meeting the demands of a rapidly growing population while confronting dwindling natural resources highlights an urgent need to improve and deploy high-yielding varieties. A comprehensive understanding of the genetic variation present within available germplasm is an indispensable preliminary step in breeding programs aimed at selecting parents and developing resilient new varieties (Kumbhar et al., 2015). Dissecting rice genotype diversity remains an essential method for identifying and transmitting vital genetic information across generations (Nambara & Nonogaki, 2012).

To systematically navigate genetic variability, plant breeders utilize advanced statistical frameworks. The recent biplot technique provides breeders with a complete visual representation of all aspects of variables by producing a biplot that simultaneously represents both mean efficiency and stability (Yan & Frégeau-Reid, 2018). Multivariate analysis serves as a highly sufficient measurement of the true degree of difference between genotypes. In this context, Principal Component Analysis (PCA) and cluster analysis are both heavily relied upon as powerful multivariate methods to assess variation (Tiwari et al., 2020). Specifically, PCA is deployed to study diversity and to determine commitment to several specific characteristics, isolating the agronomic and morphological traits that contribute most significantly to total phenotypic variation. On the other hand, cluster analysis helps breeders choose unique parental combinations to maximise heterosis by grouping genotypes based on genetic or agronomic characteristics (Shabir et al., 2013).

The utility of these multivariate tools is well-documented in agricultural research. For instance, Nachimuthu et al. (2014) used PCA to assess variation between diverse rice genotypes, studying morphological traits like No. of grains per panicle, grain yield per plant, panicle length, and plant height. They reported high coefficients of variation for both the number of filled grains per panicle and grain yield per plant, successfully defining the key traits driving variation. Similarly, Chakravorty et al. (2013) evaluated the variation in rice genotypes through PCA, showing that the total number of grains per panicle heavily impacted the underlying diversity of the studied genotypes. Furthermore, Sanni et al. (2012) employed cluster analysis to evaluate the variability of 434 rice genotypes across 10 agronomic traits, successfully partitioning the germplasm into seven distinct groups and stating that specific clusters provide valuable guidance for establishing parental genotypes in breeding blocks. While these frameworks have been applied to distinct international populations, it is vital to evaluate how domestic lines adapt specifically to Pakistan's agro-climatic conditions. Therefore, the objective of this research was to estimate the possible diversity between the rice genotypes studied by multivariate analysis, define the contribution of each studied trait to the total variation, and determine the best genotypes for the breeding of new high-yielding cultivars in Pakistan.

Materials and Methods

The field research was carried out during the 2018–2019 cropping season at the experimental farm of the Rice Research Institute (RRI), Kala Shah Kaku, Punjab, Pakistan. A total of 25 fine-grain rice lines were evaluated, including 22 advance lines and three commercial checks: Super Basmati,

Basmati 515, and PK1121 Aromatic. These advance lines were acquired from National Coordinator, National Agricultural Research Council (NARC), Islamabad, Pakistan.

The experiment was laid out using a Randomized Complete Block Design (RCBD) with three replications. Each replicate consisted of plots containing 7 rows, 5 meters in length, with a 22.5 cm plant-plant and row-row spacing. Nursery sowing was initiated on 20th June, 2018, followed by manual transplanting on 20th July, 2018, and harvesting on 28th October, 2018. Standard regional agronomic interventions, encompassing thorough land preparation, precise irrigation, balanced fertilization, and targeted pest management, were applied uniformly across all experimental units.

Data measurements for 14 quantitative traits were recorded from ten randomly selected guarded plants per plot to eliminate edge effects. Phenological traits included days to flowering (DF) and days to maturity (DM). Morpho-yield characteristics comprised plant height (PH, cm), tillers per plant (TPP), panicle length (PL, cm), grains per panicle (GPP), thousand-grain weight (TGW, g), and overall grain yield (Yield, kg/ha). Physical grain quality and seed morphology traits included head rice recovery (HRR, %), average grain length (AGL, mm), grain width (mm), length-width ratio (LW), and elongation ratio (ER), Bursting (%).

The quantitative dataset was subjected to standard descriptive and multivariate statistical analyses. Principal Component Analysis (PCA) was used to minimise dimensionality and identify important characteristics influencing the overall phenotypic variance. Concurrently, a hierarchical Cluster Analysis was applied to group the 25 fine-grain rice lines into distinct clusters, helping pinpoint divergent parental lines for future premium-quality rice breeding programs.

Table 1: List of Rice Genotypes tested at Rice Research Institute, Kala Shah Kaku (RRI, KSK)

Genotypes	Decoded Name	Genotypes	Decoded Name
G1	P-47	G13	NB- 1395
G2	PK 9966	G14	BASMATI 515 (Check)
G3	PK1121 Aromatic (Check)	G15	PK 10324
G4	RRI 3	G16	NB-13122
G5	PKBB- 15-116	G17	BBF-BB- EM-25-7- 4
G6	NS-5	G18	P-35
G7	PK 8892	G19	BR-51
G8	NB-1519	G20	BBF-AP- EM-117- 23-23
G9	BBF-BB- EM-25-7- 2	G21	P-48
G10	PK 10683	G22	SRI-25
G11	PK 9444	G23	PK PB8
G12	SRI-23	G24	BR-1
		G25	Super Basmati (check)

Results and Discussion

The descriptive statistics for fourteen agronomic, yield, grain quality, and milling traits evaluated in 25 rice genotypes are presented in Table 2. The mean days to flowering (DF) and days to maturity (DM) were 109.88 and 141.83 days, respectively, indicating moderate variation among the genotypes, with ranges of 34 days for both traits. Plant height (PH) averaged 112.70 cm and varied from 91.73 to 130.47 cm, reflecting considerable diversity in plant stature. The number of total productive tillers per plant (TPP) ranged from 13.73 to 20.87, with a mean of 17.25, while panicle length (PL) averaged 25.81 cm and varied between 22.20 and 31.30 cm. Grains per panicle (GPP) exhibited substantial variation, ranging from 68.40 to 132.50 grains with a mean of 97.60 grains. Thousand-grain weight (TGW) showed moderate variability, averaging 25.20 g and ranging from 19.01 to 32.44 g. Paddy yield (PY) displayed a wide range from 2524.80 to 4250.57 kg ha^{-1}, with a mean value of 3511.96 kg ha^{-1}, indicating the presence of high-yielding genotypes in the experimental material.

Among grain quality traits, average grain length (AGL) ranged from 6.85 to 9.25 mm with a mean of 7.86 mm, whereas grain breadth averaged 1.75 mm and varied from 1.59 to 2.00 mm. The length-to-breadth (L/B) ratio had a mean value of 4.50, ranging from 3.42 to 5.40, suggesting diversity in grain shape characteristics. Elongation ratio (ER) exhibited the highest skewness (-3.43) and kurtosis (12.52), indicating the presence of extreme values and a highly non-normal distribution. The mean ER was 1.95, with values ranging from 0.32 to 2.32.

Cooking quality trait bursting percentage averaged 12.68%, ranging from 6% to 22%, while head rice recovery (HRR) had a mean value of 56.32% and varied between 53.00% and 61.00%. Most traits exhibited relatively low skewness and kurtosis values, indicating approximately normal distributions. However, DF and DM showed negative skewness, suggesting a greater concentration of genotypes with higher flowering and maturity durations. Overall, the observed variability across agronomic, yield, and grain quality traits indicates substantial genetic diversity among the evaluated rice genotypes, providing opportunities for selection and genetic improvement in breeding programs.

Table 2: Descriptive statistics for the estimated variables in 25 rice genotypes

Variable (vars)	Mean	SD	Median	Trimmed	MAD	Min	Max	Range	Skew	Kurtosis	SE
DF	109.88	8.70	112.00	110.95	6.92	86.67	120.67	34.00	-1.07	0.33	1.74
DM	141.83	8.66	144.00	142.87	6.92	118.67	152.67	34.00	-1.08	0.36	1.73
PH	112.70	8.91	112.80	112.66	8.60	91.73	130.47	38.73	-0.01	-0.08	1.78
TPP	17.25	1.88	17.73	17.30	1.19	13.73	20.87	7.13	-0.57	-0.48	0.38
PL	25.81	2.01	25.33	25.66	1.43	22.20	31.30	9.10	0.84	0.50	0.40
GPP	97.60	18.23	98.60	97.44	18.68	68.40	132.50	64.10	0.10	-1.11	3.65
TGW	25.20	3.53	25.08	25.03	3.09	19.01	32.44	13.43	0.37	-0.48	0.71
PY	3511.96	460.33	3533.61	3530.31	373.56	2524.80	4250.57	1725.77	-0.49	-0.47	92.07

AGL	7.86	0.64	7.68	7.81	0.52	6.85	9.25	2.40	0.67	-0.51	0.13
Breath	1.75	0.12	1.71	1.74	0.12	1.59	2.00	0.41	0.57	-0.66	0.02
LB Ratio	4.50	0.44	4.62	4.51	0.28	3.42	5.40	1.98	-0.44	-0.18	0.09
ER	1.95	0.37	2.05	2.01	0.13	0.32	2.32	2.00	-3.43	12.52	0.07
Bursting	12.68	4.71	14.00	12.43	5.93	6.00	22.00	16.00	0.33	-0.99	0.94
HRR	56.32	2.16	56.00	56.21	2.22	53.00	61.00	8.00	0.43	-0.66	0.43

DF = Days to flowering; DM = Days to maturity; PH = Plant height; TPP = Tillers per plant; PL = Panicle length; GPP = Grains per panicle; TGW = Thousand-grain weight; PY = Paddy yield; AGL = Average grain length; L/B Ratio = Length-to-breadth ratio; ER = Elongation ratio; HRR = Head rice recovery

To reduce the dimensionality of the dataset while preserving maximum variance, a Principal Component Analysis (PCA) was performed on the evaluated quantitative traits across the 25 rice genotypes. A scree plot was generated by plotting the eigenvalues against their corresponding principal components to determine the optimal number of components to retain for structural interpretation. The scree plot displays a sharp, steep decline in eigenvalues from Principal Component 1 (PC1) to Principal Component 2 (PC2), followed by a distinct inflection point—or "elbow"—at PC3. Beyond PC3, the curve transitions into a gradual, linear decline, indicating that subsequent components contribute progressively less to explaining the total experimental variance. According to the elbow criterion, components prior to the primary inflection point capture the most meaningful biological variation. Therefore, the first two (or three) principal components should be retained for further multivariate analysis (Greenacre et al., 2022). These leading components account for the majority of the total dataset variability, effectively simplifying the complex trait interactions into a lower-dimensional space without substantial loss of information (Figure 1).

Figure 1: Scree plot diagram of Eigenvalues constructed on studied traits

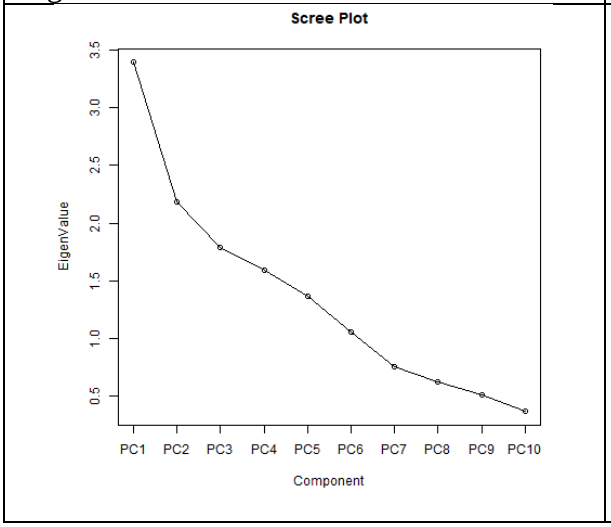
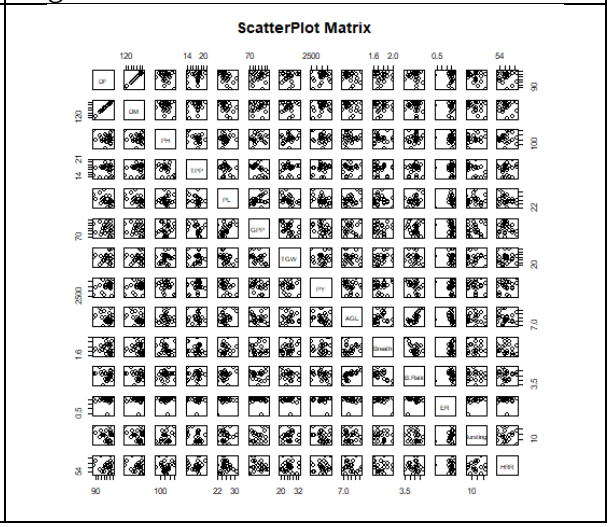


Figure 2: Scatter plot diagram of Eigenvalues constructed on studied traits

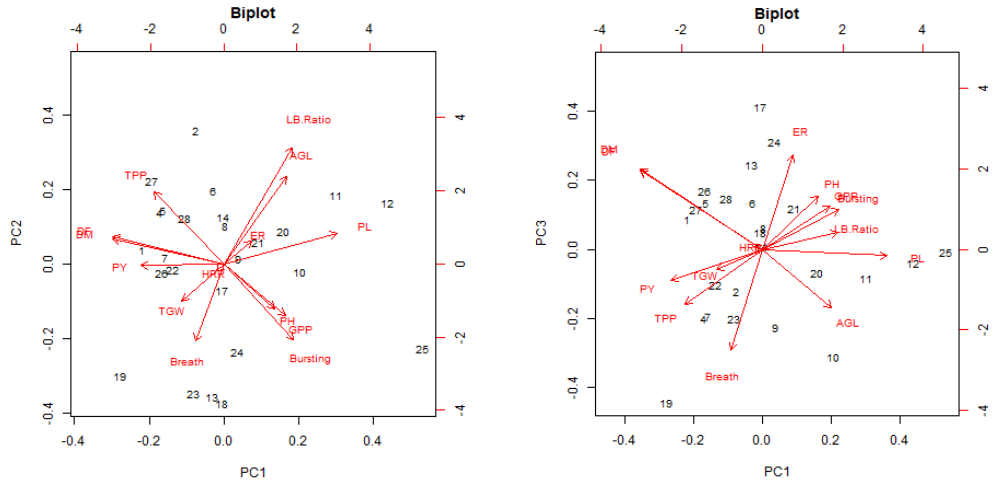


To investigate the multi-directional relationships and underlying distribution patterns among the evaluated agro-morphological traits, a pairwise Scatterplot Matrix (SPLOM) was constructed. This visual matrix provides a comprehensive, simultaneous overview of the bivariate relationships across all studied quantitative variables for the rice genotypes. The diagonal elements of the matrix represent the individual traits (including labels such as DF, DM, PH, TPP, PL, GPP, TGW, Grain Length, Grain Breadth, L/B ratio, and Yield). The off-diagonal panels display the individual bivariate scatterplots for every possible trait combination, allowing for immediate identification of linear, non-linear, or clustered distributions. A preliminary assessment of the scatterplot matrix reveals distinct patterns of association:Highly prominent, tight linear distributions are visible between days to maturity and morphological traits (e.g., the strong positive linear relationship in the upper-left quadrant between Days to Flowering [DF] and Days to Maturity [DM]), signifying tightly linked genetic or physiological mechanisms. Yield and yield-contributing components (such as GPP and TGW) display more dispersed, multi-directional scatter patterns against vegetative traits, suggesting a complex, polygenic network of influence rather than simple linear dependencies.The uniform distribution of data points across the majority of the bivariate plots indicates a lack of severe anomalies or isolated clustering, confirming the statistical reliability of the phenotypic data for subsequent multivariate modeling (Figure 2).

To simultaneously evaluate the relationships among the 14 agro-morphological traits and project the multivariate distribution of the 25 rice lines, PCA biplots were constructed. The analysis is visualized across two distinct projections: PC1 versus PC2, and PC1 versus PC3. These plots combine both the vector loadings of the traits (indicated by red arrows) and the individual scores of the genotypes (indicated by numbers 1 to 25). The cosine of the angle between any two trait vectors approximates their correlation; acute angles indicate positive correlations, obtuse angles indicate negative correlations, and right angles (90 degree) signify independence. A strong positive correlation is evident among grain quality traits such as LB.Ratio and AGL, which share a tight acute angle in the upper-right quadrant. Similarly, yield components like PH (Plant Height) and GPP (Grains Per Panicle) form a closely aligned cluster in the lower-right quadrant, pointing to a coordinated inheritance or physiological link. Conversely, traits like DF (Days to Flowering) and PY (Paddy yield) are positioned directly opposite to the right-quadrant clusters along the PC1 axis, indicating a strong negative correlation with traits like PL (Panicle Length) (Figure 3). The same results were reported in research article by Ta et al., 2018.

Genotypes located in close proximity to specific trait vectors exhibit higher-than-average values for those traits. For instance, in the PC1 vs. PC2 plot, Genotypes 11, 12, and 20 are strongly associated with longer panicle lengths (PL). Genotype 25 stands out as an extreme outlier on the far right of both plots, showing a highly pronounced expression for traits loading heavily on positive PC1, while being sharply divergent from genotypes located on the left hemisphere (such as Genotypes 19, 27, and 2). Genotypes clustered near the origin (e.g., Genotype 14, 8, and 21) represent individuals with average performance characteristics across the majority of the evaluated parameters (Figure 3).

Figure 3: Biplot distribution of 25 rice genotypes and studied traits depending on principal component axes PC1 and PC2



Principal Component Analysis (PCA) was performed to identify the major components contributing to total variation among the genotypes. The PCA extracted fourteen principal components (PCs), corresponding to the fourteen original variables, with eigenvalues ranging from 3.3936 to 0.0004 (Table 4). The first principal component (PC1) exhibited the highest eigenvalue (3.3936) and explained **24.24%** of the total variation. The second principal component (PC2) accounted for **15.62%** of the total variation with an eigenvalue of **2.1875.** Together, PC1 and PC2 explained **39.87%** of the total variability present among the rice genotypes. The third principal component (PC3) contributed **12.75%** of the total variance, increasing the cumulative variance explained to **52.62%.** Among the fourteen principal components, the first six components had eigenvalues greater than one (3.3936, 2.1875, 1.7852, 1.5916, 1.3608, and 1.0594), indicating their significant contribution to the overall variability according to the Kaiser criterion. These six principal components collectively explained **81.27%** of the total variation, suggesting that most of the genetic diversity among the rice genotypes was captured within these components. The remaining principal components (PC7–PC14) individually contributed less than 6% of the total variance and collectively accounted for only **18.73%** of the variation. Therefore, their contribution to genotype differentiation was relatively minor (Table 3).

Table 3. Principal Component Analysis (PCA) of 25 Rice Lines Based on 14 Traits

Statistics	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10	PC 11	PC 12	PC 13	PC 14
Eigenvalue	3.3936	2.1875	1.7852	1.5916	1.3608	1.0594	0.7524	0.6243	0.5100	0.3663	0.2394	0.1278	0.0011	0.0004
Standard Deviation	1.8422	1.4790	1.3361	1.2616	1.1665	1.0293	0.8674	0.7901	0.7142	0.6052	0.4893	0.3575	0.0329	0.0211
Proportion of Variance (%)	24.24	15.62	12.75	11.37	9.72	7.57	5.37	4.46	3.64	2.62	1.71	0.91	0.01	0.00

Cumulative Variance (%)	24.24	39.87	52.62	63.99	73.70	81.27	86.65	91.11	94.75	97.37	99.08	99.99	100.00	100.00

The eigenvectors (factor loadings) of the fourteen principal components indicate the relative contribution of each trait to the total variability among the 25 rice genotypes. Traits with higher absolute loading values (positive or negative) contribute more strongly to a particular principal component and are therefore important in differentiating the genotypes. PC1 accounted for **24.24%** of the total variation and was primarily associated with **panicle length (PL, 0.4208), days to maturity (DM, -0.4123), days to flowering (DF, -0.4118), paddy yield (PY, -0.3077), total productive tillers per plant (TPP, -0.2594), bursting (0.2593)**, and **length-to-breadth ratio (0.2553).** The positive loadings of panicle length and grain quality traits coupled with negative loadings of maturity-related traits suggest that PC1 mainly represents variation related to yield and maturity characteristics. PC2 explained 15.62% of the total variation and was characterized by high positive contributions from length-to-breadth ratio (0.5362), average grain length (0.4032), and total productive tillers per plant (0.3324). In contrast, breadth (-0.3568) and bursting (-0.3487) contributed negatively. This component predominantly reflects grain dimension and grain quality attributes. PC3 contributed **12.75%** of the total variation and was mainly influenced by **elongation ratio (ER, 0.4392), days to maturity (0.3740), days to flowering (0.3670)**, and **average grain length (-0.2705).** The component therefore represents variation associated with cooking quality and maturity-related traits. PC4 accounted for **11.37%** of the total variation and showed strong positive loading for **thousand-grain weight (TGW, 0.6510)** and **plant height (PH, 0.4499),** while **grains per panicle (GPP, -0.4617)** contributed negatively. This component mainly reflects grain weight and plant architecture traits. PC5 explained **9.72%** of the total variation and was largely influenced by **head rice recovery (HRR, -0.5340), elongation ratio (ER, -0.3712), average grain length (AGL, 0.2874), days to flowering (DF, 0.2883)**, and **days to maturity (DM, 0.2857).** This component is associated primarily with grain quality and milling characteristics. PC6 accounted for **7.57%** of the variation and was dominated by **head rice recovery (HRR, -0.6244)** and **elongation ratio (ER, 0.4720).** Therefore, this component reflects important cooking and milling quality parameters (Table 4).

Table 4. Eigenvectors (Factor Loadings) of the First six Principal Components for 14 Traits

Trait	**PC 1**	**PC 2**	**PC 3**	**PC 4**	**PC 5**	**PC 6**	**PC 7**	**PC 8**	**PC 9**	**PC 10**	**PC 11**	**PC 12**	**PC 13**	**PC 14**
DF	-0.4118	0.1248	0.3670	-0.0342	0.2883	-0.0175	0.0333	-0.0997	0.1301	-0.1955	0.0817	0.1575	-0.0202	0.7053
DM	-0.4123	0.1136	0.3740	-0.0376	0.2857	-0.0199	0.0382	-0.0928	0.1190	-0.1800	0.0975	0.1610	0.0305	-0.7077
PH	0.1892	-0.2074	0.2477	0.4499	0.2031	-0.2869	0.0042	0.0625	-0.4727	-0.4024	-0.3696	-0.0891	-0.0098	-0.0033
TPP	-0.2594	0.3324	-0.2542	-0.0425	-0.2199	0.1388	0.2936	0.4630	-0.1787	-0.5105	0.1594	-0.2637	0.0053	-0.0031
PL	0.42	0.13	-	-	0.07	0.22	0.19	-	-	-	0.30	0.55	-	0.00

	08	95	0.0298	0.2107	89	15	19	0.2471	0.3873	0.2325	73	37	0.0015	33
GPP	0.2277	-0.2394	0.2006	-0.4617	0.2651	-0.3095	0.0441	0.0712	-0.0823	-0.0135	0.4612	-0.4944	-0.0124	0.0084
TGW	-0.1555	-0.1716	-0.0939	0.6510	0.0580	0.0372	0.2053	-0.0538	-0.1202	0.2638	0.6156	-0.0292	0.0052	0.0105
PY	-0.3077	-0.0083	-0.1445	-0.0810	-0.2293	-0.1743	-0.7477	-0.1469	-0.3874	-0.1045	0.2240	0.0541	-0.0089	0.0013
AGL	0.2336	0.4032	-0.2705	0.2095	0.2874	-0.1216	-0.1929	-0.2316	0.3021	-0.2155	0.0816	-0.1684	-0.5512	-0.0312
Breadth	-0.1053	-0.3568	-0.4700	-0.0309	0.2573	0.1633	0.0279	-0.4045	0.1739	-0.3590	-0.0670	-0.1502	0.4429	0.0075
L/B Ratio	0.2553	0.5362	0.0792	0.1721	0.0726	-0.2135	-0.1752	0.0667	0.1037	0.0628	0.0990	-0.0480	0.7057	0.0218
ER	0.1035	0.1075	0.4392	0.0927	-0.3712	0.4720	-0.0395	-0.4486	-0.0391	-0.0990	0.0189	-0.4487	0.0041	-0.0029
Bursting	0.2593	-0.3487	0.1876	0.1490	-0.1904	0.1461	-0.3001	0.3636	0.4653	-0.3850	0.2463	0.2197	0.0149	0.0001
HRR	-0.0284	-0.0333	0.0107	-0.0013	-0.5340	-0.6244	0.3330	-0.3427	0.2049	-0.1866	0.0592	0.1197	0.0022	-0.0008

DF = Days to flowering; DM = Days to maturity; PH = Plant height; TPP = Tillers per plant; PL = Panicle length; GPP = Grains per panicle; TGW = Thousand-grain weight; PY = Paddy yield; AGL = Average grain length; L/B Ratio = Length-to-breadth ratio; ER = Elongation ratio; HRR = Head rice recovery.

A pairwise distance matrix was constructed using the squared Euclidean distance method based on the evaluated agro-morphological traits (Table 5). The dissimilarity values provide a quantitative measure of genetic diversity, serving as a critical foundation for identifying divergent parents for future hybridization programs. The highest dissimilarity values are observed between specific distantly related pairs, notably between Genotype G16 and Genotype G11 (with a distance value of 8.48), followed closely by pairs such as G22 and G16 (8.09) and G16 and G10 (8.02). Genotypes displaying such high distance values are phenotypically highly divergent and represent excellent candidates for heterosis breeding, as crosses between deeply divergent parents often yield superior segregants. Conversely, the lowest non-zero dissimilarity values indicate high phenotypic similarity i.e; include the pairs G4 and G23 (2.56) and G13 and G7 (3.01). These closely aligned values suggest that these genotypes share highly similar agro-morphological profiles and may possess similar genetic backgrounds. Genotypes (G16) consistently maintain higher average dissimilarity values against almost all other lines in the matrix, indicating a unique

or outlier phenotypic signature within this collection of genotypes, matching its distinct positioning observed in the multivariate cluster models (Table 5).

Table 5: Matrix of dissimilarity according to Euclidean square based on studied traits in rice genotypes

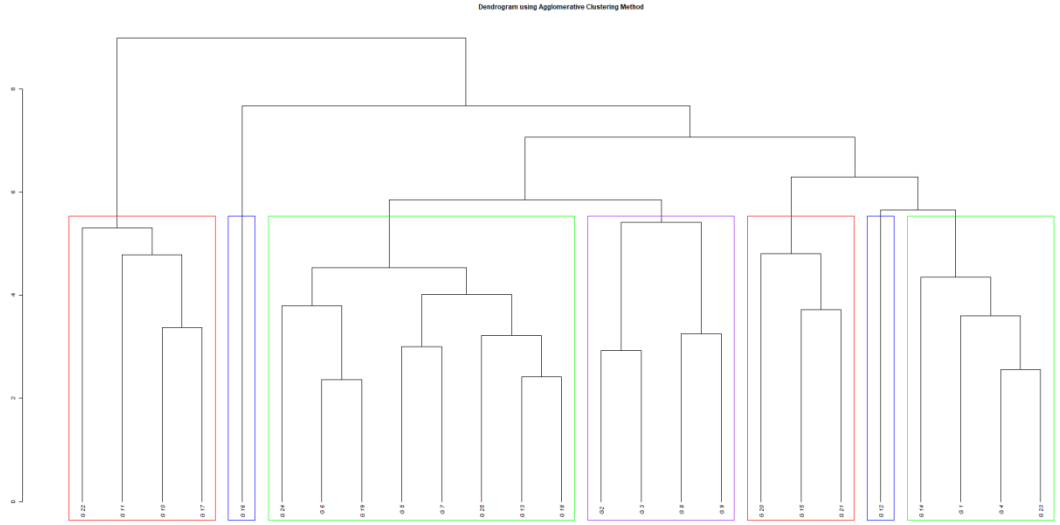
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22	G23	G24	G25
G1	0																								
G2	4.88	0																							
G3	3.53	2.93	0																						
G4	3.27	3.20	2.87	0																					
G5	5.09	4.41	5.23	3.82	0																				
G6	4.39	4.56	4.26	4.25	4.36	0																			
G7	4.53	3.81	4.40	3.64	3.01	3.98	0																		
G8	5.04	4.91	5.15	4.96	5.83	5.02	5.62	0																	
G9	5.52	5.41	5.39	5.75	5.61	4.54	5.45	3.25	0																
G10	6.71	4.66	5.54	5.52	4.84	5.65	3.93	6.32	4.92	0															
G11	7.22	5.89	6.62	6.24	5.26	6.94	4.61	6.20	5.43	3.86	0														
G12	5.52	6.39	5.83	5.65	6.37	6.28	5.50	6.73	6.57	6.59	7.69	0													
G13	3.43	3.68	4.03	3.56	3.55	4.06	3.01	4.38	4.09	4.61	5.40	5.33	0												
G14	4.22	5.66	5.78	4.05	4.72	5.72	4.59	5.02	5.74	6.54	6.24	4.73	4.29	0											
G15	5.41	6.26	5.57	4.85	4.93	4.23	4.45	5.54	5.36	5.81	6.33	4.44	5.24	4.78	0										
G16	6.51	6.96	5.69	6.45	6.44	6.18	6.40	7.00	6.98	8.02	8.48	7.06	6.28	7.67	6.02	0									
G17	5.42	4.35	4.81	4.39	4.13	3.70	3.67	5.21	3.63	3.37	4.78	6.18	3.38	5.32	4.84	7.22	0								
G18	3.63	4.42	4.33	3.63	3.47	4.53	2.82	4.96	4.19	3.70	4.49	5.03	2.42	3.73	4.56	6.63	3.34	0							
G19	4.11	4.36	4.47	3.94	3.85	2.36	3.01	4.39	4.51	5.23	6.42	5.77	3.22	4.81	4.06	5.59	3.90	3.67	0						
G20	5.52	6.56	5.78	5.70	6.33	3.99	4.66	5.72	5.54	6.34	7.29	6.19	5.15	6.29	3.93	6.12	5.03	5.25	3.30	0					
G21	5.84	6.67	6.58	4.65	4.46	5.62	4.33	6.27	6.36	5.91	6.10	6.03	4.99	4.39	3.72	7.01	4.85	4.27	4.52	4.81	0				
G22	7.56	7.69	7.74	7.51	7.10	7.43	6.60	6.09	4.23	5.30	4.84	6.64	5.80	6.25	6.17	8.99	4.91	5.08	7.01	7.02	6.36	0			
G23	3.60	4.65	3.72	2.56	5.22	4.32	4.81	5.63	5.72	6.12	7.40	5.23	4.13	4.35	4.90	6.95	4.18	4.16	4.39	5.36	4.83	7.20	0		
G24	4.58	5.00	5.84	4.57	3.50	3.80	4.07	5.39	5.75	6.45	7.24	7.06	3.86	4.76	5.91	7.29	5.05	4.43	2.91	5.91	5.57	8.21	5.34	0	
G25	4.22	4.61	5.07	3.72	3.18	4.35	4.02	5.20	5.13	5.19	6.77	5.83	2.83	4.64	5.26	6.62	4.28	3.22	3.30	5.74	4.49	6.98	4.20	3.25	0

The dendrogram (Figure 4), constructed using the squared Euclidean distance matrix, successfully groups the genotypes into distinct clusters highlighted by colored bounding boxes.

Cluster I (Red box) contains a distinct group of 4 genotypes (including G22, G11, G10, and G17). These genotypes branch off early from the rest of the germplasm at a high linkage distance, indicating a unique phenotypic baseline compared to the remaining groups. Cluster II (Green Box) represents one of the largest clusters, encompassing approximately 7-8 closely related genotypes (including G24, G5, G15, G8, G7, G25, G13, and G16). The internal nodes of this cluster show very short vertical branch lengths (e.g., between G7 and G25, or G13 and G16), reflecting a high degree of phenotypic similarity and minimum Euclidean distance among these individuals.

Cluster III (Purple Box) contains 4 genotypes clustered together, indicating intermediate similarity values that separate them from the adjacent larger groups. A highly unified group (Cluster IV) containing 3 genotypes (G20, G18, and G21) that share tightly linked branching patterns. Cluster V (Green Box) forms the final major cluster on the right hemisphere of the tree, bringing together 4-5 genotypes (such as G14, G1, G4, and G23) under a single nested branch structure. Notably, isolated singletons or highly independent lines (such as the narrow blue vertical boxes flanking the larger clusters, e.g., near G19 and G12) fail to cluster tightly with the larger groups until much higher linkage thresholds. This highlights their highly divergent nature within the evaluated population.

Figure 4: Dendrogram of 25 rice genotypes based on studied traits



The rice genotypes exhibited substantial genetic variability for key agronomic traits. PCA and cluster analyses effectively differentiated the genotypes and identified promising parental lines for breeding programs. These findings provide valuable information for the development of improved rice varieties.

References

Chakravorty, A., Ghosh, P. D., & Sahu, P. K. (2013). Multivariate analysis of phenotypic diversity in advanced breeding lines of rice (Oryza sativa L.). Journal of Rice Research, 1(2), 112–118.

Glauber, J. W., & Mamun, A. (2025). Global rice market: Current outlook and future prospects. Intl Food Policy Res Inst.

Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. Nature Reviews Methods Primers, 2(1), 100.

Khan, N., Fahad, S., Naushad, M., & Faisal, S. (2020). Economics of Agriculture in the World. Available at SSRN 3603274.

Khan, S. U., Akhter, M., & Haider, Z. (2023). Genetic diversity assessment of selected rice (Oryza sativa L.) genotypes using morphological markers. Applied Ecology and Environmental Research, 21(3), 2525–2538.

Kumbhar, A. M., Khan, M. A., & Ahmed, M. (2015). Evaluation of genetic variability and selection criteria for yield traits in elite rice genotypes. Pakistan Journal of Agricultural Sciences, 52(3), 611–617.

Mohidem, N. A., Hashim, N., Shamsudin, R., & Che Man, H. (2022). Rice for food security: Revisiting its production, diversity, rice milling process and nutrient content. Agriculture, 12(6), 741. https://doi.org/10.3390/agriculture12060741

Nachimuthu, V. V., Robin, S., Sudhakar, D., Rajeswari, S., & Subramanian, K. (2014). Genotypic characterization of rice germplasm lines based on principal component analysis for yield and quality traits. Indian Journal of Science and Technology, 7(12), 2034–2041.

Nambara, E., & Nonogaki, H. (2012). Seed biology and the transmission of genetic information in crop species. Plant Science, 188, 1–3.

Sanni, K. A., Fawole, I., Ghegah, A., & Misra, R. C. (2012). Cluster analysis of agronomic and botanical traits in upland rice genotypes. Journal of Crop Improvement, 26(4), 512–526.

Shabir, H., Malik, T. A., & Khan, A. S. (2013). Application of cluster analysis for parent selection in local rice breeding programs. Journal of Agricultural Research, 51(2), 123–131.

Ta, K. N., Khong, N. G., Ha, T. L., Nguyen, D. T., Mai, D. C., Hoang, T. G., et al. (2018). A genome-wide association study using a Vietnamese landrace panel of rice (Oryza sativa) reveals new QTLs controlling panicle morphological traits. BMC Plant Biology, 18, 282.

Tiwari, D. N., Tripathi, S. R., & Joshi, B. R. (2020). Principal component and cluster analysis for genetic diversity in exotic rice genotypes under irrigated ecosystems. Journal of Agronomy and Crop Science, 206(3), 345–354.

Yan, W., & Frégeau-Reid, J. (2018). Genotype by environment interaction and biplot analysis in plant breeding and variety evaluation. Frontiers in Plant Science, 9, 703.