Global Research journal of Natural Science and Technology (Grjnst)
Volume: 03 - Issue 1 (2025)
p.ISSN: 2790-7643 e.ISSN: 2790-7651
www.grjnst.com

Multivariate Analysis of Phenotypic Traits to Assess Diversity and Breeding Potential of Rice Hybrids in Pakistan

Muhammad Ijaz
Rice Research Institute, Kala Shah Kaku, Punjab, Pakistan
director.rice@gmail.com
Tahira Bibi*
Rice Research Institute, Kala Shah Kaku, Punjab, Pakistan
tahira167_uaar@yahoo.com
Muhammad Sabar
Rice Research Institute, Kala Shah Kaku, Punjab, Pakistan
muhammadsabar@gmail.com
Rashid Mehmood
PMAS- Arid Agriculture University, Rawalpindi, Punjab, Pakistan
rashid.pbg@uaar.edu.pk

Abstract

Understanding the genetic architecture and phenotypic diversity within elite germplasm is paramount for accelerating rice breeding programs aimed at enhanced productivity and resilience. This study meticulously analyzed 134 rice hybrids evaluated at KSK over one year, across 13 crucial agronomic and quality traits, employing Principal Component Analysis (PCA) and detailed genetic parameter estimation. ANOVA results confirmed highly significant ($P \leq 0.01$ or $P \leq 0.05$) phenotypic variations among the 134 hybrids for all measured traits, including Days to Flowering (DF), Days to Maturity (DM), Plant Height (PH), Yield, Grains per Panicle (GPP), and Thousand Grain Weight (TGW), validating the substantial genetic diversity. Significant block effects ($P \leq 0.01$ or $P \leq 0.05$) were also observed for most traits, underscoring the effectiveness of the experimental design. Trait reliability was high, as indicated by low Coefficients of Variation (CVs), with Yield at 0.51% and DM at 1.56%, ensuring vigorous data quality. PCA effectively summarized the multidimensional data, with the Scree plot demonstrating that the first two principal components, Dim1 and Dim2, cumulatively explained 35.6% of the total phenotypic variance (19.1% and 16.5% respectively), thus capturing the most significant patterns of variation. The PCA biplot revealed that Dim1 was primarily driven by developmental traits, with strong positive loadings for DM and DF. Critically, Yield, GPP, average grain length (AGL), PH, and panicle length (PL) also loaded positively on Dim1, yet their overall vectors indicated an inverse relationship with maturity; earlier-maturing hybrids often exhibited higher yields. Dim2 was predominantly shaped by grain morphology, strongly associating with length width ratio (positive loading) and grain width (negative loading), while head rice recovery (HRR) and elongation ratio (ER) showed limited contribution to these initial components.

The distribution of rice hybrids in the PCA space further elucidated phenotypic clustering and environmental influences. Hybrids such as 'Badbaan-07', 'Dhanak (FM - HBS-2001)', and 'EG-1907' consistently clustered in the early-maturing, high-yielding quadrant. Genetic parameter estimation provided quantitative support for breeding potential. Broad-sense heritability (h^2 (BS) %) was exceptionally high (over 80%) for key traits including GPP (98.52%), TGW (98.36%), Yield (99.84%), HRR (98.98%), PH (91.57%), Tillers per Plant (86.36%), Grain Width (80.99%), Length Width Ratio (94.29%), and ER (94.09%), signifying strong genetic control and high selection efficiency. Genetic Advance (GA) was particularly high for Yield (1396.76) and GPP (74.60), predicting substantial gains through selection. Days to Flowering, Days to Maturity, and AGL also displayed moderate heritability (around 64%), offering reasonable prospects for improvement. In conclusion, this research identified significant genetic diversity and high heritability for critical agronomic and quality traits within the 134 rice hybrids. PCA effectively mapped the complex trait interrelationships and identified promising genotypes based on their multi-trait profiles, especially highlighting the potential of early-maturing, high-yielding accessions. The findings offer invaluable guidance for NARC's breeding program at KSK, facilitating targeted selection of superior germplasm and optimizing breeding strategies for developing high-performing, climate-resilient rice varieties.

Keywords: PCA, Rice, Hybrid Rice, Phenotypic diversity, Multivariate, GxE interaction, Breeding

Introduction

Rice (*Oryza sativa* L.) remains a cornerstone of food security, feeding over half of the global population. With increasing population pressure, climate change, and shrinking agricultural resources, there is an urgent need to enhance rice productivity and adaptability Virmani (1994). Hybrid rice technology has emerged as a pivotal innovation to bridge the yield gap by exploiting heterosis (hybrid vigor), enabling yield gains of 15–20% over conventional inbred varieties (Yuan, 2017). Countries like China, India, and more recently Pakistan have made significant investments in hybrid rice research to address national food security objectives and enhance export competitiveness (Nalley et al., 2016).

Rice (*Oryza sativa* L.) remains an important part of the diet of approximately half of the world's population and serves an important role in agriculture in Pakistan (Shahzadi et al., 2018). After wheat, rice is the second important cereal crop and provides 3.5 percent towards value added in agriculture and 0.7 percent towards GDP of Pakistan (Hanif, 2023). The crop is grown in multiple agro-climatic regions, with Punjab and Sindh being the two main rice growing provinces. As per the Pakistan Bureau of Statistics (PBS), rice cultivation for the fiscal year 2022-23 was about 3.3 million hectares with 8.8 million tons award, which means the average yield is 2.67 tons per hectare. However, yields remain below potential due to biotic and abiotic stresses, suboptimal input management, and genetic limitations (Iqbal et al., 2023). The development and selection of high-performing hybrid rice genotypes is thus essential to improving productivity, particularly in the face of climate variability and resource constraints (Pandit et al., 2016).

Pakistan's rice industry has significant potential to increase production; however, lack of effective management in dealing with climate variability, inefficient resource use, and overall low input productivity curtail the yield and competitiveness in international markets (Maji & Shaibu, 2012). Effective schematic change can begin with understanding how different regions and seasons affect production attributes through policy-making, resource allocation, and strategic crop improvement. Multiple traditional statistical methods tend to overlook the multifaceted problem of interdependency that exist with regard to rice yield and production (Gaballah et al., 2022).

To address this complexity, Principal Component Analysis (PCA), a multivariate statistical technique, has been widely adopted in agricultural research (Tiwari et al., 2022). PCA reduces the dimensionality of large datasets by transforming correlated variables into a smaller set of uncorrelated components, thereby highlighting the most influential factors contributing to variation in data (Jolliffe & Cadima, 2016). In rice research, PCA has been applied to explore genotype-environment interactions (Wadood et al., 2022), identify yield-contributing traits, and evaluate regional production patterns (Akinwale et al., 2011)

This study employs PCA to analyze morphological and agronomic traits among diverse hybrid rice genotypes under field conditions (Vanisri et al., 2020). The objective is to identify the most informative traits contributing to yield performance and variability, thereby facilitating the development of selection indices for hybrid rice breeding. The findings aim to assist breeders in optimizing selection strategies and accelerating the development of superior hybrid cultivars suited to local agro-ecological conditions.

Materials and Methods

The present study was conducted at the Rice Research Institute (RRI), Kala Shah Kaku, Punjab, Pakistan, during the 2022–2023 growing season to evaluate the genetic diversity and performance of rice hybrids under local agro-climatic conditions. A total of 134 rice hybrids (including 2 checks and 132 test treatments) were received from the National Agricultural Research Council (NARC), Islamabad, and were tested in the field using an augmented block design (ABD). The experiment comprised 11 blocks, with each block including two standard check varieties—GNY-53C1 and Diamond-121C2—to facilitate comparative analysis. Each hybrid entry was evaluated in three replications to ensure accuracy and minimize environmental variation. The sowing was carried out on 10 June 2022 transplanting on 10 July 2022, and harvesting on 28 October 2022.

Standard agronomic practices were followed uniformly throughout the experiment, including land preparation, transplanting, fertilization, irrigation, and pest management, as per regional recommendations. Data of 13 traits were recorded on several important agro-morphological and yield-related traits viz., Days to flowering-DF, Days to maturity-DM (days), Plant height-

PH (cm), Tillers per plant-TPP, Panicle length-PL (cm), Grains per panicle-GPP, Thousand grain weight-TGW (grams), Yield (kg/ha), Head rice recovery-HRR (%), Average grain length-AGL (mm), Grain width (mm), Length width ratio-LW, Elongation ratio-ER (%). These measurements were taken from randomly selected plants in each plot, and standard procedures were used to ensure consistency and reliability in data collection. In addition to the current season's evaluation, some of the hybrids tested in this study were previously included in the National Uniform Yield Trials (NUYT) conducted at multiple locations across Pakistan in the year 2021. The data from NUYT served as a comparative reference for evaluating hybrid stability and adaptability across different environments.

To assess genetic diversity among the tested hybrids, both qualitative and quantitative traits, along with seed morphological characteristics, were studied. This comprehensive evaluation aimed to identify superior genotypes for further development and potential commercial release, while also exploring the extent of phenotypic variation among rice hybrids under the specific agro-ecological conditions of Kala Shah Kaku.

Results and Discussion

The ANOVA results assess the significance of differences among 134 rice hybrids for 13 agro-morphological traits, while accounting for block effects (Table 1). This ANOVA presents a critical statistical analysis, showcasing the results of a two-way Analysis of Variance, adjusted for treatments, conducted for each of the 13 distinct phenotypic traits evaluated across the 134 rice hybrids. This comprehensive analysis serves to statistically ascertain if the observed phenotypic differences among the rice hybrids are significant (Stoker et al., 2020), and equally important, to understand the extent to which the experimental blocks contributed to the variation in trait expression (Talekar et al., 2022).

The analysis reveals that the "Block (ignoring Treatments)" source of variation, with 10 degrees of freedom, had a highly significant impact on the expression of most traits, including plant height, tillers per plant, grains per panicle, thousand grain weight, yield, head rice recovery, average grain length, grain width, length width ratio, and elongation ratio (all significant at $p \leq 0.01$). Days to flowering and days to maturity also showed a significant block effect ($P \leq 0.05$). This pervasive significance across most traits strongly underscores the importance of the blocked experimental design in mitigating environmental heterogeneity and thereby enhancing the precision of the hybrid performance assessment. Only panicle length remained unaffected by block variation, showing a non-significant effect ($P > 0.05$), suggesting its relative stability across varying micro-environmental conditions within the experiment.

Crucially, the "Treatment (eliminating Blocks)" source of variation, with 133 degrees of freedom, provides the statistical validation for the phenotypic differences among the 134 rice hybrids themselves. A highly significant effect ($P \leq 0.01$) was observed for plant height, tillers per plant, grains per panicle, thousand grain weight, yield, grain width, length width ratio, and elongation ratio. Furthermore, days to flowering, days to maturity, panicle length, average grain length, and head rice recovery also showed a significant effect ($P \leq 0.05$) due to treatments. This overwhelmingly significant variation across nearly all measured traits unequivocally confirms that your 134 rice hybrids represent a genetically diverse collection with substantial phenotypic differences in key agronomic and quality attributes. This robust statistical evidence directly supports the visual patterns observed in the PCA biplots, where the spread of hybrids and distinct trait associations are a direct consequence of this underlying genetic and phenotypic variability.

Table 1: ANOVA Summary (Treatment Adjusted) for 13 Traits in 134 Rice Hybrids

Source	**Df**	**DF**	**DM**	**PH**	**TPP**	**PL**	**GPP**	**TGW**	**Yield**	**HRR**	**AGL**	**Width**	**LW**	**ER**
Block (ignoring Treatments)	10	15.04 *	15.04 *	399.19 **	32.38 **	10.55^{ns}	4864.06 **	3.71 **	1522654.07 **	73.90 **	0.08 *	0.01 *	0.04 **	0.35 **
Treatment (eliminating Blocks)	133	10.36 *	10.36 *	79.09 **	8.51 **	13.38 *	1240.82 **	7.33 **	342886.96 **	52.30 **	0.06 *	0.01 **	0.03 **	0.60 **
Treatment: Check	1	2.23^{ns}	2.23^{ns}	4.55^{ns}	46.55 **	58.91 **	26323.68 **	88.00 **	836.04^{ns}	1087.82 **	0.86 **	0.00^{ns}	0.11 **	16.39 **
Test and Test vs. Check	132	10.42 *	10.42 *	79.66 **	8.22 **	13.04 *	1050.80 **	6.72 **	345478.26 **	44.46 **	0.06 *	0.01 **	0.03 **	0.48 **
Residuals	10	3.73	3.73	8.35	1.45	4.21	19.88	0.10	753.20	0.51	0.02	0.00	0.00	0.03

^{ns} = not significant ($P > 0.05$); * = significant at $P \leq 0.05$; ** = highly significant at $P \leq 0.01$

The table 2 presents the Coefficient of Variation (CV) for each of the 13 traits evaluated across the 134 rice hybrids, offering a crucial measure of the relative variability and reliability of the data for each specific trait. The Coefficient of Variation, expressed as a percentage, indicates the ratio of the standard deviation to the mean, providing a standardized measure of dispersion that allows for comparison across traits with different units or scales. A lower CV generally signifies greater precision and reliability of the experimental data for that trait, while a higher CV indicates greater variability relative to the mean, potentially suggesting less control or inherently more diverse responses. Notably, traits such as Yield (0.51%), Thousand Grain Weight (TGW, 1.24%), Head Rice Recovery (HRR, 1.21%), Days to Maturity (DM, 1.56%), and Elongation Ratio (ER, 1.61%) exhibit very low CV values, indicating high precision and consistency in their measurement across the rice hybrids. Conversely, Tillers per Plant (TPP) shows the highest CV at 10.47%, followed by Panicle Length (PL) at 8.02%, suggesting greater inherent variability among the hybrids for these traits or potentially more environmental influence on their expression, though these values are still within acceptable ranges for field experiments. The remaining traits, including days to flowering (2.06%), plant height (2.52%), grains per panicle (3.03%), average grain length (1.98%), grain width (2.42%), and length width ratio (1.26%), all demonstrate low to moderate CVs, reinforcing the reliability of the experimental data for all traits and ensuring that the observed differences in the ANOVA and PCA are robust.

Table 2: Coefficient of Variation (CV%) for 13 Agro-Morphological Traits in 134 Rice Hybrids

Trait	CV (%)	Trait	CV (%)	Trait	CV (%)
DF	2.06	PL	8.02	HRR	1.21
DM	1.56	GPP	3.03	AGL	1.98
PH	2.52	TGW	1.24	Width	2.42
TPP	10.47	Yield	0.51	LW	1.26
ER	1.61				

This table 3 presents crucial genetic parameters for each of the 13 traits, providing insights into the magnitude of genetic variability, the influence of environmental factors, and the potential for genetic improvement in your 134 rice hybrids. For each trait, the Mean indicates the average performance, while Phenotypic Variance (PV) represents the total observed variation, encompassing both genetic and environmental components. Genotypic variance (GV) quantifies the variation attributable solely to genetic differences among the hybrids, and Environmental variance (EV) accounts for the non-heritable variation due to environmental factors. The genotypic coefficient of variation (GCV) and phenotypic coefficient of variation (PCV) express the genetic and phenotypic variability relative to the mean, respectively; a higher GCV relative to PCV indicates a greater proportion of variation that is genetically controlled. Environmental coefficient of variation (ECV) quantifies the environmental variability relative to the mean. Broad-sense heritability (h^2 (BS) %) is a critical parameter, indicating the proportion of phenotypic variation that is due to genetic factors, ranging from 0 to 100%; high heritability (e.g., GPP, TGW, Yield, HRR, PH, TPP, LW, ER, Width all over 80%) suggests that selection for these traits would be highly effective as genetic factors predominantly determine their expression. Finally, genetic advance (GA) predicts the expected improvement in a trait from one generation of selection, which is highest for GPP (74.60) and Yield (1396.76), indicating substantial potential for improving these traits through breeding. Traits like days to flowering and days to maturity, while having lower GA values, still show moderate heritability (64.16%), suggesting that selection for these developmental traits would also be reasonably effective. Overall, the high heritability estimates combined with substantial genetic advance for many key traits, particularly yield and its components, underscore the significant potential for genetic improvement within this germplasm of rice hybrids.

Table 3: Genetic Variability Parameters for 13 Agro-Morphological Traits in 134 Rice Hybrids

Trait	Mean	PV	GV	EV	GCV (%)	PCV (%)	ECV (%)	h^2 (BS) (%)	GA
DF	93.38	10.40	6.67	3.73	2.77	3.45	2.07	64.16	4.27
DM	123.38	10.40	6.67	3.73	2.09	2.61	1.56	64.16	4.27
PH	115.59	99.02	90.67	8.35	8.24	8.61	2.50	91.57	18.80
TPP	11.44	10.60	9.15	1.45	26.45	28.46	10.51	86.36	5.80
PL	25.09	11.65	7.44	4.21	10.87	13.60	8.18	63.86	4.50
GPP	144.25	1347.15	1327.27	19.88	25.26	25.44	3.09	98.52	74.60
TGW	25.71	6.11	6.01	0.10	9.53	9.61	1.23	98.36	5.02
Yield	5430.4	459908.2	459155.0	753.2	12.48	12.49	0.51	99.84	1396.76
HRR	58.97	50.06	49.55	0.51	11.94	12.00	1.21	98.98	14.45
AGL	6.82	0.05	0.03	0.02	2.69	3.34	1.98	64.97	0.31
Width	2.01	0.01	0.01	0.00	4.99	5.55	2.42	80.99	0.19
LW	3.40	0.03	0.03	0.00	5.11	5.27	1.26	94.29	0.35
ER	10.70	0.50	0.47	0.03	6.40	6.59	1.60	94.09	1.37

Principal Component Analysis (PCA) was employed to dissect the complex phenotypic variation among the 134 rice hybrids evaluated, effectively reducing the dimensionality of the multi-trait data. The analysis revealed that the first two principal components, Dim1 and Dim2, cumulatively accounted for 35.6% of the total phenotypic variance (19.1% and 16.5% respectively), offering a significant, albeit partial, summary of the inherent trait relationships and hybrid differentiation. Dim1 was primarily driven by traits related to maturity and yield potential, highlighting a crucial trade-off or synergy between development time and productivity. The subsequent interpretation of both variable loadings and individual hybrid scores allowed for the identification of key trait associations and the characterization of distinct phenotypic groups among the diverse rice germplasm under study. This detailed breakdown of variance partitioning and trait relationships, as supported by the accompanying supplementary data (Supplementary Table S1), forms the basis for understanding the breeding value and genetic diversity within this set of rice hybrids.

The Scree plot, titled "Scree plot," serves as a crucial diagnostic tool in Principal Component Analysis, graphically illustrating the proportion of total variance explained by each principal component (Figure 1). This plot is fundamental for determining the optimal number of dimensions to retain for meaningful interpretation of the underlying data structure, particularly for the 13 traits measured across your 134 rice hybrids.

On the x-axis, the plot displays the "Dimensions" (or principal components) from 1 to 10, while the y-axis represents the "Percentage of explained variances." Each bar indicates the variance captured by a specific principal component, and the overlaid line connects these percentages, typically showing a steep decline followed by a more gradual slope.

From this Scree plot, it is evident that the first principal component (Dimension 1) accounts for the largest proportion of variance, at 19.1%. The second principal component (Dimension 2) explains the second largest amount of variance, at 16.5%. This is consistent with the percentages displayed on your biplots. Beyond these first two dimensions, there is a clear "elbow" or inflection point in the curve. Dimension 3 explains 13.8% of the variance, Dimension 4 explains 12%, and then the explained variance drops more gradually for subsequent dimensions: 8.5% for Dim 5, 7.2% for Dim 6, 6.6% for Dim 7, 5.5% for Dim 8, 4.3% for Dim 9, and 3.7% for Dim 10 (Table 5).

The "elbow criterion" is commonly applied to Scree plots to identify the point at which the marginal gain in explained variance significantly diminishes. In this case, the elbow appears to be around Dimension 2 or Dimension 3. While Dim1 and Dim2 together explain 35.6% of the variance, adding Dim3 increases the cumulative explained variance to 49.4% (35.6% + 13.8%). The decision to retain components often balances the desire to explain as much variance as possible with the need for interpretability. Given that your subsequent biplots focus on Dim1 and Dim2, this Scree plot confirms that these first two components capture a substantial portion of the most dominant patterns of variation within your rice hybrid data. Although more dimensions would explain additional variance, the initial rapid drop in explained variance highlights Dim1 and Dim2 as the most informative, representing the primary axes of variation among the 13 traits and 134 rice hybrids (Table 4).

Figure 1: Scree plot illustrating the percentage of total variance explained by each principal component

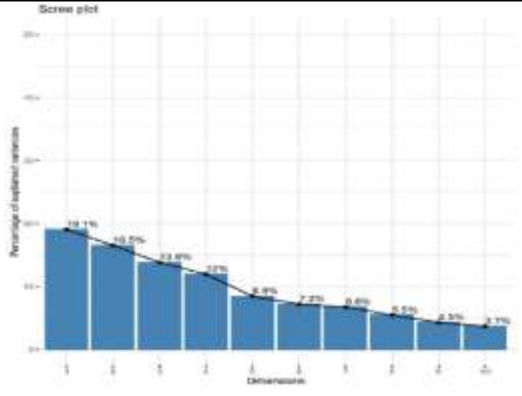


Table 4: Cumulative Percentage of Variance Explained by the First Ten Principal Components in 134 Rice Hybrids

Principal Component	**Cumulative Variance (%)**	**Principal Component**	**Cumulative Variance (%)**
PC1	19.08	PC6	77.07
PC2	35.61	PC7	83.82
PC3	49.44	PC8	89.30
PC4	61.41	PC9	93.61
PC5	69.87	PC10	97.31

Table 5: Principal Component Loadings for 13 Agro-Morphological Traits across 10 Principal Components in 134 Rice Hybrids.

Trait	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Dim.9	Dim.10
DF	0.8295	-0.2944	-0.3789	-0.0347	0.0766	-0.0914	-0.2091	-0.1182	-0.0062	0.0922
DM	0.8295	-0.2944	-0.3789	-0.0347	0.0766	-0.0914	-0.2091	-0.1182	-0.0062	0.0922
PH	0.5612	0.2256	-0.0893	0.5171	-0.1223	-0.0158	0.2342	0.0401	0.1815	-0.4957
TPP	0.0847	-0.0165	0.0936	0.6828	0.2968	-0.2923	-0.0129	0.5281	-0.1622	0.1912
PL	0.1063	0.2255	-0.0159	-0.1134	0.7511	-0.0526	0.4993	-0.2880	-0.1570	0.0023
GPP	0.1946	0.6015	0.2120	0.4530	0.0356	0.1474	-0.0125	-0.2045	0.4010	0.2954
TGW	0.3365	-0.0658	-0.3696	-0.4099	-0.1603	0.2045	0.5259	0.4193	0.1231	0.1215
Yield	0.3678	0.4624	0.0534	0.1866	-0.2634	0.5750	0.0266	-0.0383	-0.4215	0.1133
HRR	-0.2049	0.0639	-0.3293	-0.0711	0.5489	0.5470	-0.3734	0.2335	0.1589	-0.1427
AGL	0.4904	0.2407	0.6390	-0.2944	0.1106	-0.0475	-0.1906	0.1292	-0.1629	-0.1755
width	0.2750	-0.6318	0.6742	0.0500	0.0888	0.2200	0.0493	-0.0037	0.0148	-0.0300
LW	0.0182	0.8382	-0.3084	-0.2385	-0.0268	-0.2746	-0.1712	0.0690	-0.1163	-0.0672
CGL	0.3553	0.3595	0.4934	-0.4982	0.0417	-0.1076	-0.0404	0.1591	0.2388	0.0985

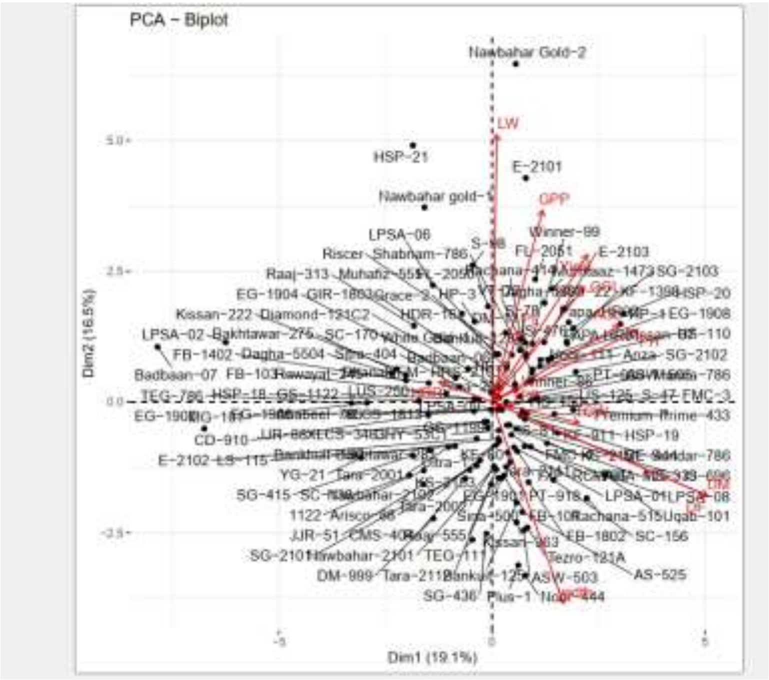
The various rice hybrids are plotted as labeled points, their positions indicating their overall phenotypic similarity, with closer points signifying more resemblance (Figure 2). The arrows represent the traits, their length indicating their influence on the principal components, and their direction revealing correlations. A longer arrow denotes a trait that is strongly associated with the displayed components, thus playing a more significant role in defining the phenotypic space. The angle between any two trait arrows approximates their correlation, with small angles indicating positive correlation, large angles (close to 180 degrees) indicating negative correlation, and angles near 90 degrees suggesting low or no correlation. A close examination of the trait vectors reveals distinct groupings and relationships. Days to maturity (DM) and days to flowering (DF) are prominent, pointing strongly in the positive direction of Dim1. Their long arrows confirm their high contribution to this dimension and their strong positive correlation, reflecting that later-flowering hybrids invariably lead to later-maturing ones. This establishes Dim1 as a critical axis primarily driven by the developmental duration of the rice hybrids.

In the same general direction as DM and DF, but with slightly differing angles, are several key yield and growth-related traits: Yield (kg/ha), Grains per Panicle (GPP), Average Grain Length (AGL), Plant Height (PH), and Panicle Length (PL). These traits are positively correlated among themselves, forming a cluster that points towards the positive side of Dim1. This implies that hybrids with higher values for these yield components and morphological traits also tend to have longer days to flowering and maturity, or conversely, earlier maturing hybrids might sacrifice some of these growth parameters. Tillers per Plant (TPP) and Thousand Grain Weight (TGW) are also present in this general quadrant, indicating positive contributions to Dim1, though TPP shows a slight positive influence on Dim2 as well. This cluster as a whole suggests that Dim1 also captures the overall productivity and vegetative vigor, with a nuanced relationship to maturity period.

Moving towards the top of the biplot, Length Width Ratio (LW) stands out, loading strongly in the positive direction of Dim2. This indicates that Dim2 is significantly influenced by grain shape, specifically the slender vs. bold characteristic. Conversely, Grain Width ('width') loads negatively on Dim2, directly opposing LW, which confirms that wider grains are associated with lower length-width ratios. These two traits define the second principal component, illustrating variation in grain morphology. Head Rice Recovery (HRR) and Elongation Ratio (ER) are situated closer to the origin, with relatively short arrows. This positioning suggests that these traits contribute less to the variation captured by Dim1 and Dim2 compared to others, or their variability is more effectively explained by subsequent principal components not depicted in this two-dimensional space. Therefore, conclusions regarding these traits based solely on this biplot should be made with caution.

The distribution of the 134 rice hybrids across this two-dimensional space further refines our understanding. Rice hybrids positioned far to the right, such as 'Nawbahar Gold-2', 'E-2101', 'HSP-20', and 'Nawbahar gold-1', are characterized by longer Days to Maturity and Days to Flowering, aligning with the direction of those trait vectors. Conversely, hybrids found prominently on the far left, like 'Badbaan-07', 'Dhanak(FM - HBS-2001)', 'EG-1907', and 'Kissan-222', tend to exhibit earlier maturity and flowering. It is notable that some of these early hybrids are also positioned towards the upper-left or lower-left quadrants, suggesting various combinations of other traits. For example, some early types might have higher LW, while others have higher grain width. Hybrids located in the upper portion of the plot, like 'HSP-21', 'Nawbahar Gold-2', and 'E-2101', generally possess higher Length Width Ratio (LW). Conversely, those in the lower part, such as 'DM-999', 'JJR-51', 'Naram-2002', and 'SG-436', are characterized by greater Grain Width ('width'). A large number of rice hybrids are clustered around the origin, indicating that their trait profiles are somewhat average for these first two components, or their unique variations are explained by higher principal components. This central clustering also suggests a degree of genetic homogeneity for the traits most strongly explained by Dim1 and Dim2, among a large portion of the germplasm. The presence of outliers at the periphery of the plot, such as 'Nawbahar Gold-2' or 'Badbaan-07', highlights unique combinations of traits, making them interesting for targeted breeding. Overall, this biplot provides a powerful visual summary of the phenotypic landscape of your 134 rice hybrids, enabling the identification of key trait associations and the characterization of distinct phenotypic groups based on their comprehensive trait profiles.

Figure 2. PCA biplot illustrating the relationship between rice hybrids and agro-morphological traits based on the first two principal components.



This PCA biplot, titled "Variables - PCA," serves as a visual representation of the relationships among various biological attributes and their contributions to the principal components derived from the dataset (Figure 3). The arrows extending from the origin on the biplot represent individual variables. The length of each arrow signifies the strength of the variable's contribution to the principal components: longer arrows indicate that a variable is well-represented by the components displayed. The direction of the arrows is equally important; variables with arrows pointing in similar directions are positively correlated, suggesting they move in tandem. Conversely, variables with arrows pointing in opposite directions are negatively correlated. Variables whose arrows are nearly perpendicular to each other exhibit little to no correlation.

A crucial aspect of this biplot is the color scale, which represents the `cos2` value, or the squared cosine, indicating the quality of representation of each variable on a given principal component (or set of principal components, as in a 2D biplot showing Dim1 and Dim2). For each variable, its `cos2` value ranges from 0 to 1. Warm colors, such as red and orange, correspond to higher `cos2` values (e.g., 0.6 and above), signifying that the variable is exceptionally well-represented by Dim1 and Dim2. This implies that a substantial proportion of the variability in that variable is captured by these two dimensions. In contrast, cooler colors like green and blue denote lower cos2 values (e.2.g., 0.2 and below), suggesting that the variable is not adequately represented by these two dimensions, and a significant portion of its variability might be explained by higher-order principal components not depicted in this 2D plot.

Examining specific variables with high `cos2` values, indicated by orange and red hues, we find **Days to Maturity (DM)** and **Days to Flowering (DF).** These two variables are strongly aligned with the *positive direction* of Dim1 and exhibit high `cos2` values, indicating their excellent representation by these two components, particularly Dim1. Their close proximity and similar direction suggest a strong positive correlation between them: rice hybrids that flower early also tend to mature early, and vice-versa. This is a very common biological relationship in crop development.

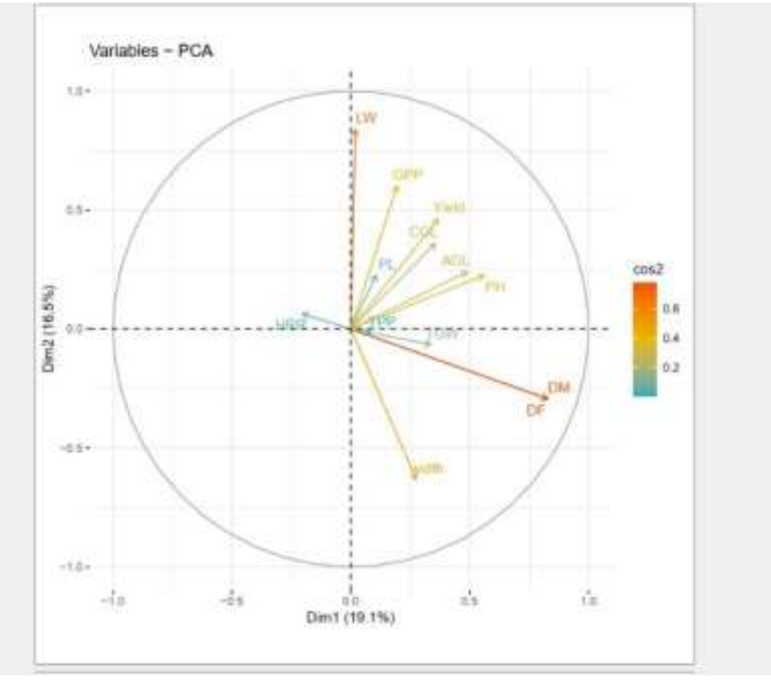
In stark contrast, a distinct cluster comprising Yield, GPP (Grain per panicles), AGL(mm) (Average grain length), and PH Plant height (cm) is observed in the *positive quadrant of Dim1 and the positive quadrant of Dim2*, though more strongly correlated with Dim1. These variables also exhibit good `cos2` values (yellow to orange), signifying their relatively strong representation. Their close proximity indicates strong positive correlations among themselves. This suggests that higher values for GPP, CGL, AGL, and PH are associated with higher Yield. What is critical here, with the correct understanding of DM and DF, is their *opposite direction* to DM and DF along Dim1. Since DM- days to maturity (days) and DF (days to flowering) point positively along Dim1, and Yield, GPP, etc., also have positive components along Dim1 but are somewhat *opposed* in the overall vector direction from DM/DF, this suggests that *earlier maturity (lower DM/DF values)* might be positively associated with *higher yield and related growth traits* (GPP, CGL, AGL, PH). This implies that faster-developing hybrids might also be higher yielding.

Variables with moderate `cos2` values, appearing in green or yellow, include PL (likely Plant Length or Height), which shows some correlation with the "Yield" cluster, albeit with a shorter arrow and slightly cooler color. LW (Grain length width ratio) is primarily aligned with the positive direction of Dim2 and slightly negative on Dim1, possessing a moderate `cos2`. Similarly, 'width' (grain width) is mainly aligned with the negative direction of Dim2 and positive on Dim1, also with a moderate `cos2`. These variables contribute to the variation but are not as strongly explained by the first two dimensions as DM, DF, or the yield-related cluster.

Finally, variables with low cos2 values, depicted in blue or light green, such as HRR (Head Rice Recovery (%) and TPP (Tillers per Plant), are positioned close to the origin. Their cooler colors indicate that they are poorly represented by Dim1 and Dim2, suggesting that their variation is more likely explained by higher principal components, or they exhibit less overall variance compared to other attributes in the dataset. While close to the origin, they appear somewhat negatively correlated with each other, being on opposite sides of the central point.

This visual representation can validate existing hypotheses regarding trait correlations, such as the strong positive correlation between days to flowering and days to maturity. Furthermore, it crucially reveals an inverse relationship: early flowering and early maturing hybrids (low DM/DF) appear to be associated with higher yields and improved growth parameters (GPP, CGL, AGL, PH). This finding is particularly significant for breeding programs aiming for early-maturing, high-yielding varieties. The analysis can also inform trait selection and data reduction; highly correlated variables like Yield, GPP, CGL, AGL, and PH provide similar information, suggesting that a representative trait from this group could be chosen for future studies, thereby reducing the number of variables. Conversely, variables poorly represented by Dim1 and Dim2 (low `cos2`) might be less critical for explaining the major variations in the rice hybrids within these two dimensions, although their importance in other dimensions or their inherent lower variability should not be overlooked. This component (positive side) would be high yielding and *early* maturing (as earlier maturity is negatively correlated with the positive direction of Dim1 where yield is). Dim2, with variables like LW and 'width' loading on it, might represent a "Plant Structure/Morphology Component," differentiating hybrids based on their specific structural or morphological characteristics, somewhat independent of the yield and maturity factors. While not shown in this specific biplot, if individual rice hybrids were also plotted, one could observe how they cluster based on these principal components, identifying groups of hybrids with similar trait profiles and highlighting those exhibiting high values for particular traits, for example, high-yielding and early-maturing hybrids.

Figure 3. PCA biplot of variables showing relationships among rice hybrid attributes



This "Individuals - PCA" graph, titled "Individuals - PCA," serves as a direct complement to the "Variables - PCA" biplot, visually representing the positions of your 134 rice hybrids within the very same two-dimensional principal component space (Figure 4). The axes remain consistent: Dim1, explaining 19.1% of the variance, and Dim2, accounting for an additional 16.5%. This means that these two dimensions collectively capture 35.6% of the total variability observed among your rice hybrids. Just as for the variables, the color scale on this graph signifies the `cos2` value, here indicating the quality of representation for each specific rice hybrid on the factor map. Warmer colors, such as orange and red, correspond to higher `cos2` values, signifying that the hybrid's overall characteristics are well-explained by these two principal components. Conversely, cooler colors like blue and green denote lower values, suggesting that the hybrid's profile is better captured by other, higher-order principal components not depicted in this two-dimensional view.

The core principle for interpreting this plot is straightforward: rice hybrids located in close proximity on the graph exhibit similar overall trait profiles, based on the variables included in your PCA. Conversely, rice hybrids positioned far apart are dissimilar. Furthermore, a rice hybrid's placement relative to the origin and to the previously interpreted variable vectors provides crucial insights into its specific trait combination.

Upon observation, a majority of the 134 rice hybrids appear to cluster around the origin of the plot, particularly towards the center of the first quadrant (positive Dim1, positive Dim2). This suggests that a substantial proportion of your hybrids share relatively similar overall characteristics, especially concerning the variations explained by Dim1 and Dim2. However, the presence of distinct outliers, or rice hybrids located further from this main cluster, indicates the existence of unique trait combinations within your germplasm.

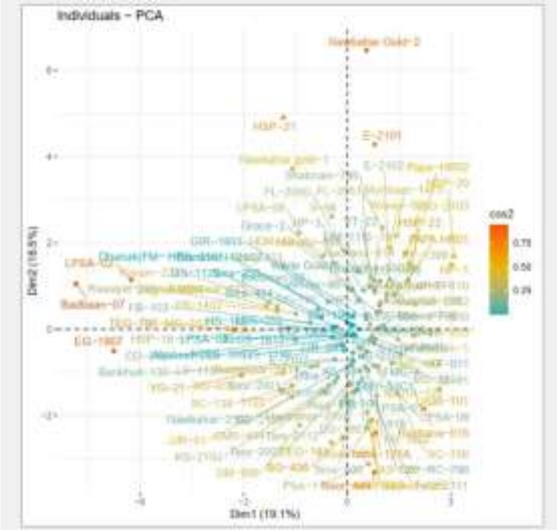
To truly comprehend the distribution of these rice hybrids, it is essential to recall the insights gleaned from the variables plot, especially the corrected understanding of DM as Days to Maturity and DF as Days to Flowering. We previously established that DM and DF load strongly in the positive direction of Dim1, while Yield, GPP, CGL, AGL, and PH also possess positive components along Dim1 but their overall vector direction was somewhat opposed to that of DM and DF. This led to the significant conclusion that earlier maturity (lower DM/DF values) is positively associated with higher yield and its related growth traits. Therefore, rice hybrids positioned prominently to the far left (negative side) along Dim1 are likely to be characterized by higher yield and earlier maturity. Conversely, rice hybrids located far to the right (positive side) along Dim1 would tend to be lower yielding and later maturing. Regarding Dim2, recalling that LW (grain length Width ratio) loaded positively and 'grain width' loaded negatively, rice hybrids positioned high on the positive side of Dim2 would tend to have higher LW values, while those positioned low on the negative side of Dim2 would likely exhibit higher 'width' values.

Observing specific groupings of rice hybrids, several appear prominently in the far left quadrants (negative Dim1). Examples include 'Dhanak(FM - HBS-2001)', 'Rawayat-04H-B590', 'Badbaan-07', 'EG-1907', and 'Kissan-22'. These rice hybrids are likely defined by earlier flowering and maturity and, concurrently, higher yield, GPP, CGL, AGL, and PH. Their placement strongly suggests they embody a highly desirable combination of traits indicative of robust productivity and accelerated development. The brighter colors (orange/red `cos2` values) for some of these, like 'Badbaan-07', 'EG-1907', and 'Dhanak(FM -

HBS-2001)', further confirm that their position on this two-dimensional map accurately represents their comprehensive trait profiles. In contrast, some rice hybrids are situated far to the right (positive Dim1), with 'Nawbahar Gold-2' and 'E-2101' being prominent examples in the upper right quadrant. These rice hybrids would likely be later maturing (higher DM/DF) and potentially exhibit lower values for yield and associated growth traits. Their distinct position signifies a different phenotypic strategy compared to the earlier-maturing, higher-yielding types. Additionally, hybrids positioned in the upper half (positive Dim2), such as 'Nawbahar Gold-2', 'HSP-21', and 'E-2101', would tend to possess higher LW values, while those in the lower half (negative Dim2), like 'DM-999', 'JJR-51', 'Nawbahar-2103', and 'Taram-2002', would tend to have higher 'width' values. The substantial cluster of rice hybrids located around the origin indicates that for these varieties, the first two principal components do not explain as much of their individual variability; their trait profiles might be more intricate, or their variations are better captured by higher-order principal components not visualized on this 2D plot. Their consistently lower `cos2` values (green/blue) further underscore that these two dimensions are not the primary drivers of their phenotypic variation.

Ultimately, this "Individuals - PCA" plot enables the visual identification of specific rice hybrids that display particularly desirable combinations of traits, such as high yield coupled with early maturity, exemplified by hybrids like 'Dhanak', 'Badbaan', and 'EG-1907'. These hybrids emerge as strong candidates for further breeding efforts or direct selection. Furthermore, the spread of the rice hybrids across the plot provides insight into the extent of genetic diversity within your 134 accessions, where a wide dispersion suggests considerable phenotypic variation. The clustering patterns, if any, can also hint at common parentage, shared genetic backgrounds, or similar adaptive strategies among groups of hybrids.

Fig 4: PCA biplot showing the distribution and quality of representation of 134 rice hybrids based on their phenotypic traits in the first two principal components.



Crucially, this plot facilitates targeted selection; if your breeding objective is, for instance, to develop early-maturing, high-yielding rice, your focus for selection would naturally gravitate towards the hybrids positioned in the bottom-left to mid-left region of the plot. In essence, by integrating the understanding from the "Variables - PCA" plot—which variables drive the variation and their inter-correlations with the "Individuals - PCA" plot—which rice hybrids exhibit which combinations of traits. Therefore, we acquire a powerful and comprehensive understanding of the phenotypic landscape of your 134 rice hybrids, thereby enabling informed decisions for your breeding program.

References

- Akinwale, M., et al. (2011). Heritability and correlation coefficient analysis for yield and its components in rice (*Oryza sativa* L.). *African Journal of Plant Science*, 5(3), 207–212.
- Gaballah, M. M., et al. (2022). Assessment of genetic parameters and gene action associated with heterosis for enhancing yield characters in novel hybrid rice parental lines. *Plants*, 11(3), 266.
- Hanif, M. (2023). Economic governance in Pakistan. *SSRN*.
- Iqbal, J., et al. (2023). Sustainable rice production under biotic and abiotic stress challenges. *Sustainable Agriculture in the Era of the OMICs Revolution*, 241–268.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.
- Maji, A., & Shaibu, A. (2012). Application of principal component analysis for rice germplasm characterization and evaluation. *Journal of Plant Breeding and Crop Science*, 4(6), 87–93.
- Nalley, L., et al. (2016). Quantifying the agronomic and economic performance of hybrid and conventional rice varieties. *Agronomy Journal*, 108(4), 1514–1523.
- Pandit, E., et al. (2016). Principal component analysis for selection of heterotic hybrids for yield component traits in mid and mid-early duration rice genotypes. *ORYZA – An International Journal on Rice*, 53(3), 255–261.
- Shahzadi, N., et al. (2018). Rice in Pakistan: present scenario, trade, problems and prospects. *International Journal of Agricultural Statistics Science*, 14, 1–6.
- Stoker, P., et al. (2020). Analysis of variance (ANOVA). *Basic Quantitative Research Methods for Urban Planners*, 197–219.

- Talekar, S., et al. (2022). Genetic diversity using principal component analysis and hierarchical cluster analysis in rice. *International Journal of Plant Sciences*, 17(2), 191–196.
- Tiwari, S., et al. (2022). Principal component analysis and genetic divergence studies for yield and quality-related attributes of rice restorer lines. *Indian Journal of Genetics and Plant Breeding*, 82(1), 94–98.
- Vanisri, S., et al. (2020). Selection criteria for agronomic traits suitable for cold season through cluster and principal component analysis in rice (*Oryza sativa* L.). *Journal of Pharmacognosy and Phytochemistry*, 9(4), 1129–1137.
- Virmani, S. (1994). Prospects of hybrid rice. *Hybrid Rice Technology: New Developments and Future Prospects*, 7.
- Wadood, S. A., et al. (2022). Rice authentication: An overview of different analytical techniques combined with multivariate analysis. *Journal of Food Composition and Analysis*, 112, 104677.
- Yuan, L. (2017). Progress in super-hybrid rice breeding. *The Crop Journal*, 5(2), 100–102.