Principal Component Analysis and Genetic Divergence Studies for Yield and Quality Related Attributes of Hybrid Rice

Kuldeep Karoda [a], Yogendra Singh [a*], S.K. Singh [a], Pramod Kumar Prajapati [b] and Vinod Kumar [c]

[a] Department of Plant Breeding and Genetics, JNKVV, Jabalpur, India.
[b] JNKVV, College of Agriculture, Tikamgarh, India.
[c] PC Unit (S& N) JNKVV, Jabalpur, India.

Authors' contributions

This work was carried out in collaboration among all authors. Author KK designed the study, performed the statistical analysis, wrote the protocol, and wrote the first draft of the manuscript. Authors YS, SKS guided for research work and author PKP managed the analyses of the study. Author VK managed the literature searches. All authors read and approved the final manuscript.

Original Research Article

ABSTRACT

This study investigates the genetic divergence and yield and quality-related attributes of hybrid rice (*Oryza sativa L.*) through principal component analysis (PCA) and genetic divergence studies. Conducted during the Kharif 2022 season at the Seed Breeding Farm, Rice Improvement Project, Department of Plant Breeding & Genetics, JNKVV, Jabalpur, the research focused on 98

_____

*Corresponding author: E-mail: yogendrasinghbt@gmail.com;

genotypes, evaluating 30 agronomic attributes based on DUS guidelines. Genetic variability, heritability, genetic advance, correlation, path coefficient, D2 analysis, and PCA were employed to analyze the data.The findings reveal significant genetic diversity among the genotypes, with traits like stem length, fertile spikelet, and panicle weight contributing notably to genetic divergence. PCA identified eight principal components accounting for 81.97% of the total variability, with the first component alone contributing 22.34%. The study emphasizes the importance of understanding genotypic and phenotypic characteristics for effective selection and breeding of diverse genotypes, ultimately enhancing hybrid rice technology and regional adaptability.The research highlights key traits contributing to genetic divergence and provides insights into the genetic architecture of hybrid rice, aiding in the development of high-yield, quality rice varieties. The comprehensive analysis of yield and quality traits offers a valuable resource for breeders aiming to improve hybrid rice cultivars, ensuring food security and nutritional value for the global population.

1. INTRODUCTION

Rice, a key staple food more than 50 percent of worldwide [1] and belongs to the genus Oryza, which includes two cultivated species, *Oryza sativa* and *Oryza glaberrima*, along with 22 wild species. Of these wild species, 14 are diploid with 24 chromosomes, while 8 are tetraploid with 48 chromosomes. Both cultivated species are diploid with 24 chromosomes. Oryza sativa is particularly significant, serving as the primary dietary staple for approximately 3.5 billion people globally. Rice is crucial for global nutrition, providing carbohydrates, protein, vitamins, essential fatty acids, minerals, zinc, and iron [2]. It holds symbolic importance in various cultures and is the only cereal crop that can thrive in submerged conditions. With over 100,000 distinct landraces and improved cultivars, rice exhibits exceptional genetic diversity [3]. In 2022–2023, global rice production reached 511.6 million tons on 164.19 million hectares [4]. India's hybrid rice sector is expanding, with an increase in hybrid rice cultivation and seed production expected. Ensuring genetic purity and developing regionally adaptive hybrids are essential for the success of hybrid rice technology.

2. MATERIALS AND METHODS

The study titled " Principal component analysis and genetic divergence studies for yield and quality-related attributes of hybrid rice " was conducted during Kharif 2022 at the Seed Breeding Farm, Rice Improvement Project, Department of Plant Breeding & Genetics, JNKVV, Jabalpur. This chapter reviews and presents literature on various aspects, including morphological characterization (following DUS recommendations), genetic variability, heritability, genetic advance, correlation coefficient, path coefficient analysis, D^2 analysis, principal component analysis, combining ability, and heterosis. It also covers biochemical and molecular analysis for genetic purity testing of rice hybrids. Thirty agronomic attributes were assessed to evaluate the genetic diversity and rank the variability of restorer lines, based on DUS guidelines. These attributes included days to fifty percent flowering (DFF), days to maturity (DTM), flag leaf length (FLL) and width (FLW), total tillers per plant (TT/P), productive tillers per plant (PT/P), stem thickness (ST) and length (SL), plant height (PH), number of panicles per plant (Pa/P), panicle length (PL), biological yield per plant (BY/P), panicle weight (PaWt/P), number of fertile spikelets per panicle (Fsp/pa), total spikelets per panicle (TSp/pa), spikelet fertility (SF) and density (SD), thousand seed weight (ThSWt), panicle index (PI), harvest index (HI), grain length (GL) and width (GW), decorticated grain length (DGL) and width (DGW), length to breadth ratio (L/B), hulling percentage (H), milling percentage (M), and head rice recovery percentage (HRR). Observations were recorded on 10 randomly selected plants from each replication and analyzed using mean values and statistical methods, including D^2statistics by [5]. Inters- and intracluster D^2 mean values were estimated following [6], genotypes were grouped using Tocher's method, and PCA analysis was conducted based on methodologies by [7,8]. Understanding a crop's genotypic, phenotypic, and other relevant characteristics is essential for developing a successful breeding system. The amount of heritable variation in crop material determines the effectiveness of selection in breeding. The primary goal of plant breeding is to select, develop, and maintain diverse genotypes, making variability crucial for the selection process and identifying diverse parents.

Table 1. Contribution of different characters towards genetic divergence

S.No	Character	Times ranked 1st	Percentage (%) contribution of traits towards divergence
1	**ST**	3234	60.38 %
2	**FS**	608	11.35 %
3	**PW**	489	9.13 %
4	**1000GW**	308	5.75 %
5	**SS**	198	3.7 %
6	**GYPP**	160	2.99%
7	**PI%**	139	2.6 %
8	**HI%**	72	1.34 %
9	**SWPP**	69	1.29 %
10	**F%**	45	0.84 %
11	**DTF**	13	0.24 %
12	**TNOS**	10	0.19 %
13	**GL (mm)**	5	0.09 %
14	**DGL (MM)**	2	.04 %
15	**HRR%**	2	0.04 %
16	**DGW (mm)**	1	0.02 %
17	**FLL**	1	0.02 %

3. RESULTS AND DISCUSSION

3.1 Assessment of Genetic Divergence

Mahalanobis D^2 analysis, a quantitative method for measuring the relationship between geographic and genetic diversity, evaluates genetic differences across biological populations using a generalized distance [5]. By converting character-based correlated means into uncorrelated means through Tocher's approach, D^2 values are obtained. Genetic divergence between pairs of genotypes is then calculated by summing the squares of the differences between the corresponding uncorrelated values of the two genotypes. This study used seed yield and related variables to measure genetic divergence across 98 genotypes.

Regarding the contribution of individual characters towards genetic divergence, most characteristics played a role except for flag leaf length, stem length, panicle length, plant height, number of effective tillers, panicle weight, biological yield per plant, spikelet density, grain width, hulling percentage, milling percentage, and length-to-breadth ratio for decorticated grain.This finding was not in agreement with findings of [9 and 10] for plant height. The collective percentage contributions to genetic divergence were significant for stem length (60.38%), fertile spikelets (11.35%), panicle weight (9.13%), thousand-grain weight (5.75%), sterile spikelets (3.7%), grain yield per plant (2.99%), panicle index (2.6%), harvest index (1.34%), and stem weight per plant (1.29%).

Similar findings were obtained by [11] for 1000 seed weight and [11 and 12] for fertile spikelets . Some traits, including days to flowering (0.24%), total number of spikelets (0.19%), grain length (0.09%), decorticated grain length (0.04%), head rice recovery percentage (0.04%), decorticated grain width (0.02%), and flag leaf length (0.02%), contributed less than 1% to genetic divergence. Other traits such as flag leaf length, stem length, panicle length, plant height, number of tillers, number of productive tillers per plant, panicle weight, biological yield, spikelet density, grain weight, hulling percentage, milling percentage, and length-to-breadth ratio for decorticated grain did not contribute to divergence.

3.2 Principal Component Analysis

Principal component analysis (PCA) evaluates the relevance and contribution of each component to the total variance, with each coefficient of proper vectors reflecting the degree of contribution of each original variable related to each principal component [13]. This mathematical process converts potentially linked variables into a smaller number of uncorrelated variables known as principal components. The first principal component accounts for as much variability in the data as possible, and each subsequent component accounts for as much of the remaining variability as possible. PCA aims to identify the fewest components that explain the most variability [14,15] and rank genotypes based on PC scores. Principal components are generally calculated using a are measured in various units and scaling effects occur,

Table 2. Distribution of rice hybrid in different clusters

Cluster No.	No. of genotype	Genotypes
1	83	IR 64,Kranti × WGL 32100,IR 64 × Pusa tarangini ,NPT 10-24 × WGL 32100,97 x NPT 29 ,56Ax NPT 13-01,3A X MTU 1081 × NPT 13-01,3A X MTU 1081 × PS-4, 3A X JNPT 514-22, 3A X AVT 1 IME 3535,3A X AVT-1 IME-3536, 3A X IVT- 1 IME-3664, 3A X Pusa tarangini ,97A x NPT 65, 3A X MTU 1010 ×25B × NPT 101, 3A X JNPT 23-1, R-671 × WGL 32183, NPT 10-24 × WGL 32183, Kranti ×WGL 32183,31A X JNPT 1059-10,25A X JNPT 1059-9 25A X Sonam,32A X MTU 1081 × NPT 111, 32A X JNPT 1059-10, 32A X AVT-1 IME-3531,Sonam ,Pusa tarangini ,Swarnyapriya, R 712, R 671, NPT-29, NPT 13-01, NPT-25,NPT 65,NPT 10-24, NPT 10-65 , JNPT 1058, JNPT 23-01, JNPT 514-22, JNPT 1059-9, JNPT 1059-10, MTU 11320, P 3123, MC -13, RTCNP -28, WGL 32100,MTU 1081 X NPT 111, MTU 1010 ×25B × NPT 101, AVT- 1 IME- 3531, AVT -1 IME- 3535, AVT -1 IME- 3536, IVT IME-3624,JR 206,3B,99B, 22B, 25B,31B,56B, WGL 32183, WGL 14, MTU 1010×PS-5, MTU 1081 X PS-4, MTU 1081 X NPT 13-01, 56A X NPT 23, 56A XJNPT 1058, 56A X WGL- 14, 56A X RTCNP-28, 97A X JR 206, 97A X P 3123, 99A X NPT- 25, 99A X JNPT- 1058, 99A X MTU 1010×PS-5, 99A X MTU 11320, 99A X MC-13,Kranti ,IR 64, 22A X Pusa tarangini , 22A X Swarnyapriya ,29A X R 712, 29 A X P 3123, 31A X MTU 1081 × PS-4, JRH 19
2	9	Kranti × JNPT 1058,Kranti × WGL -14, Kranti × NPT 26-1, Shyamla x IR 64,32A X MTU 1081 × NPT 13-01,NPT 23,NPT 26-1,JNPT-58,NPT 14-7
3	1	32B
4	1	Shyamla
5	6	NPT 10-24 × R 712 , NPT 10-65 × JNPT 58,3A X NPT 13-01,3A X Swarnyapriya,29B,97B
6	2	R 671 × NPT 29,31A X MTU 1081 × NPT 111
7	2	25A X NPT 26-1, 25A X NPT 14-7

the composition of derived components might be influenced, making it desirable to standardize the variables. Table 4 shows the results of PCA for yield and quality contributing variables of fertility restorer rice lines studied. For rice genotypes, PCA was performed using yield and attributing components. Eight principal components (PCs) showed more than 1.00 Eigen values, accounting for approximately 81.97% of the total cumulative variability across all characteristics tested. PC 1 accounted for 22.34% of the variability, similar results showed by [16,17,18] with maximum variability in PC 1, while PC 2, PC 3, PC 4, PC 5, PC 6, PC 7, and PC 8 accounted for 14.81%, 11.51%, 9.71%, 7.85%, 6.10%, 5.25%, and 4.38% of the variability, respectively, across the genotypes for the variables under investigation.

Table 3. Cluster means the value of different character

Characters	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V	Cluster VI	Cluster VII
DTF	101.47	106.63	98	108.67	101.28	99.83	103.17
FLL	33.31	39.98	35.03	30.62	32.49	38.56	43.43
FLW	1.38	1.61	0.99	1.03	1.29	1.53	1.5
SL	73.48	82.33	70.1	62.27	76.5	77.88	82.63
PL	24.78	26.06	23.75	23.33	24.06	27.47	29.51
PH	98.26	108.39	93.85	85.6	100.56	105.35	112.13
ST	6.21	9.11	3.71	3.74	3.44	9.18	7.31
NOT	9.58	8.97	9.67	9	9.37	12.91	7
NOETPP	9.22	8.75	9	8	8.74	12.48	6.67
SWPP	26.27	42.56	20.93	14.65	30.48	70.16	22.29
PW	32.15	33.74	21.03	18.4	27.24	90.21	16
BY	58.41	76.29	41.97	33.05	57.72	160.36	38.29
SS	36.36	51.44	56.33	55.67	67.11	28	256.17
FS	214.13	270.04	124.67	122	174.83	264	143.83
TNOS	250.49	321.48	181	177.67	241.94	292	400
F%	84.96	84.54	68.88	68.67	73.16	91.07	34.58
SD	10.11	12.73	7.62	7.62	9.96	10.64	13.53
GL (mm)	8.88	9.35	10	10	8.58	10.08	10.17
GW (mm)	2.61	2.72	2.39	2.49	2.79	2.91	3.16
DGL (MM)	6.35	6.48	6.71	7	6.41	6.98	6.42
DGW (mm)	2.1	2.2	2	1.57	2.1	2.31	2.02
HRR%	62.56	65.35	58.75	72.07	61.75	72.93	59.46
1000GW	20.73	24.98	23.5	24.08	21.53	27.11	20.79
HI%	39.97	39.62	33.72	52.84	42.79	19.4	65.59
PI%	75.06	98.09	67.27	94.92	89.96	34.1	158.59
H%	78.81	74.93	80.43	73	79.86	80.05	70.26
M%	75.48	72.85	77.52	74.59	74.61	79.53	67.66
LBR	3.09	3	3.36	4.48	3.18	3.06	3.18
GYPP	23.14	28.9	14.15	17.46	22.27	31.07	25.76

Table 4. Interpretation of rotated component matrix for 29 yield attributing and quality-related traits

Principal Component (PC)	Dominated by	Contribution
PC 1	Stem length (0.543), Plant height (0.598), Stem weight per plant (0.711), Panicle weight (0.711), Biological yield (0.790), Fertile spikelets per panicle (0.661), Total number of spikelets (0.657), Spikelet density (0.527), Grain width (0.548), Head rice recovery percentage (0.525), Grain yield per plant (0.694)	High
PC 2	Number of effective tillers per plant (0.659), Number of tillers per plant (0.687), Panicle weight (0.570), Biological yield (0.517)	Moderate
PC 3	Hulling percentage (0.600)	Single
PC 4	Stem length (0.552), Plant height (0.607), Grain length (0.674), 1000 grain weight (0.572)	High
PC 5	Fertile spikelets (0.507), Fertility percentage (0.717)	Moderate

Table 5. List of selected genotypes in each PC

| PC1	PC2	PC3	PC4	PC5
Kranti × JNPT 1058	R 671 × NPT 29	Kranti × JNPT 1058	Kranti × JNPT 1058	Kranti × WGL -14
Kranti × WGL 32100	Shyamla x IR 64	Kranti × WGL 32100	Kranti × NPT 26-1	97 Ax NPT 29
Kranti ×WGL 32183	3A X NPT 13-01	Kranti × WGL -14	R 671 × NPT 29	56Ax NPT 13-01
Kranti × WGL -14	3A X MTU 1081 × PS-4	IR 64 × Pusa tarangini	Shyamla x IR 64	97A x NPT 65
Kranti × NPT 26-1	3A X AVT-1 IME-3536	NPT 10-24 × WGL 32183	25A X NPT 26-1	3A X IVT- 1 IME-3624
IR 64 × Pusa tarangini	3A X Pusa tarangini	NPT 10-24 × R 712	25A X NPT 14-7	25A X JNPT 1059-9
NPT 10-24 × WGL 32183	31A X MTU 1081 × NPT 111	R 671 × NPT 29	22A X Pusa tarangini	32A X MTU 1081 × NPT 111
NPT 10-24 × R 712	29A X R 712	97A x NPT 65	22A X Swarnyapriya	99A X JNPT- 1058
R 671 × NPT 29	29 A X P 3123	29 A X P 3123	56A X NPT 23	99A X MTU 1010×PS-5
Shyamla x IR 64	31A X MTU 1081 × PS-4	Kranti	56A X RTCNP-28	99A X MC-13
3A X NPT 13-01	32A X AVT-1 IME-3531	Pusa tarangini	99A X NPT- 25	Pusa tarangini
3A X MTU 1081 × NPT 13-01	56A X JNPT 1058	Swarnyapriya	99A X MTU 1010×PS-5	Swarnyapriya
3A X MTU 1010 ×25B × NPT 101	56A X WGL- 14	NPT -29	R671	NPT 65
3A X JNPT 23-1	56A X RTCNP-28	NPT13-01	NPT 23	NPT 26-1
3A X AVT-1 IME-3536	97A X JR 206	NPT 25	NPT 26-1	NPT 10-24
3A X IVT- 1 IME-3624	R671	NPT 65	NPT 10-24	NPT 10-65
31A X MTU 1081 × NPT 111	JNPT 1058	NPT 10-24	JNPT 514-22	JNPT 1058
25A X NPT 26-1	MPT 1010×PS-5	NPT 10-65	MPT 1010×PS-5	JNPT 1059-10
25A X JNPT 1059-9	MTU 1081 X PS-4	JNPT 23-01	MTU 1081 X PS-4	MTU 1081 X PS-4
29 A X P 3123	JR 206	JNPT 514-22	MTU 1081 X NPT 13-01	JNPT 58
31A X MTU 1081 × PS-4	25B	JNPT 1059-9	JNPT 58	MTU 1081 X NPT 111
32A X MTU 1081 × NPT 13-01	31B	JNPT 1059-10	IVT IME-3624	AVT- 1 IME- 3531
32A X MTU 1081 × NPT 111	32B	MTU 11320	32B	AVT -1 IME- 3535
32A X JNPT 1059-10	56B	WGL 32183	97B	AVT -1 IME- 3536
32A X AVT-1 IME-3531	97B	WGL 14		NPT 14-7
56A X JNPT 1058	99B	MTU 1081 X NPT 13-01		56B
56A X WGL- 14	22B	MTU 1081 X NPT 111		JRH 19
97A X JR 206		MTU 1010 ×25B × NPT 101		
99A X JNPT- 1058		IVT IME-3624		
99A X MC-13		3B		
JNPT 1059-10		NPT 10-24 × WGL 32100		
JNPT 58		R-671 × WGL 32183		
		NPT 10-65 × JNPT 58		
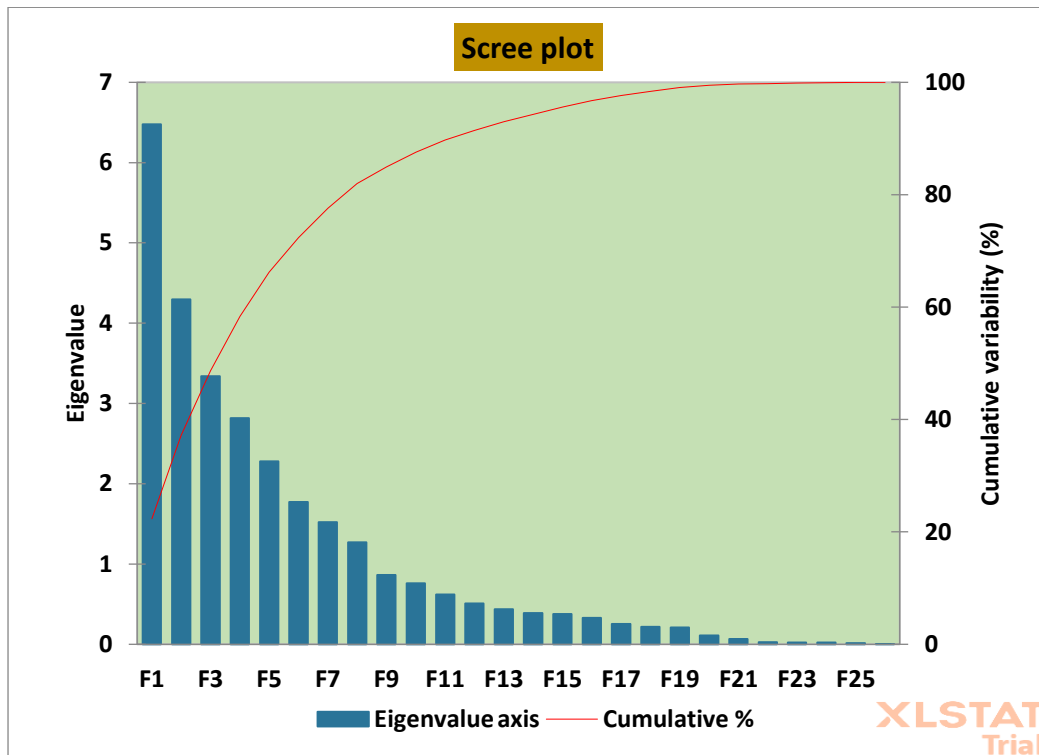
Fig. 1. Scree plot of PCA

4. CONCLUSION

The research highlights key traits contributing to genetic divergence and provides insights into the genetic architecture of hybrid rice, aiding in the development of high-yield, quality rice varieties. The comprehensive analysis of yield and quality traits offers a valuable resource for breeders aiming to improve hybrid rice cultivars, ensuring food security and nutritional value for the global population.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Khush, Gurdev S. What it will take to feed 5.0 billion rice consumers in 2030. Plant molecular biology. 2005;59:1-6.

2. Babu V. Ravindra, et al. Genetic variability studies for qualitative and quantitative traits in popular rice *(Oryza sativa* L.) hybrids of India. International Journal of Scientific and Research Publications. 2012;2(6):1-5.

3. Samal R, Roy PS, Sahoo A, Kar MK, Patra BC, Marndi BC, Gundimeda JNR. Morphological and molecular dissection of wild rices from eastern India suggests distinct speciation between O. rufipogon and O. nivara populations. Scientific reports. 2018;8(1):2773.

4. In 2022–2023, global rice production reached 511.6 million tons on 164.19 million hectareslations. Scientific reports. 8(1):2773.

5. Mahalanobis PC. On the generalized distance in statistics. Proc. Nat. Inst. Sci. India B. 1936;2:49-55.

6. Singh RK, Chaudhary BD. Biometrical methods in quantitative genetic analysis. Kalyani Publishers, New Delhi. 1977; 1-303.

7. Massay WF. Principal components regression in exploratory statistical research. Journal Am. Stat. Assoc. 1965;60:234-246.

8. Jolliffe Ian T. Principal component analysis: A beginner's guide—I. Introduction and application. Weather. 1990;45(10):375-382.

9. Kushwaha N, Kant R, Kumar R, Nilanjaya, Chhaya R, Sinha N, Singh D and Mohanty TA. Study of genetic divergence in lowland rice genotypes of Bihar. International Journal of Current Microbiology and Applied Sciences. 2020;9(2):88-94.

10. Bhushan B, Kumar B, Sudan SK, Singh P, Razdan AK. Nature and magnitude of genetic diversity among locally adapted rice (*Oryza sativa* L.) genotypes. Journal of Pharmacognosy and Phytochemistry. 2019;8(3):2526-2530.

11. Archana RS, Rani MS, Vardhan KMV and Fareeda G. Genetic diversity studies among rice (*Oryza sativa* L.) genotypes for grain yield, yield components and nutritional traits in rice. International Journal of Chemical Studies. 2018;6(6):134-137.

12. Tiwari S, Singh Y, Upadhyay PK, Koutu GK. Principal component analysis and genetic divergence studies for yield and quality-related attributes of rice restorer lines. Indian J. Genet. Plant Breed. 2022;82(1):94-98.

13. Sanni KA, Fawole I, Ogunbayo SA, Tia DD, Somado EA, Futakuchi K, Guei RG. Multivariate analysis of diversity of landrace rice germplasm. Crop Science. 2012;52(2):494-504.

14. Anderson TW. An introduction to multivariate analysis. Wiley Eastern Pvt. Ltd., New Delhi; 1972.

15. Morrison DF. Multivariate statistical methods: 2d Ed. McGraw-Hill; 1976.

16. Singh SK, Habde SV, Khaire A, Mounika K, Majhi PK and Singh DK. Principal component analysis in f2 mapping population for high grain zinc content in rice (*Oryza sativa* L.). Plant Archives. 2020;20(2):6935-6941.

17. Tiruneh A, Gebrselassie W, Tesfaye A. Genetic diversity study on upland rice (*Oryza sativa* L.) genotypes based on morphological traits in south western Ethiopia. Asian Journal of Crop Science. 2019;11:17-24.

18. Gour L, Koutu GK, Singh SK, Patel DD, Shrivastava A, Singh Y. Genetic variability, correlation and path analyses for selection in elite breeding materials of rice (*Oryza sativa* L.) genotypes in Madhya Pradesh. The Pharma Innovation Journal. 2017;6(11):693-696.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/122168