Original Research Article

Yield Trait Association Studies through Principal Component Analysis in Bi-Parental Population Derived from *indicaltropicaljaponica* Inter Sub-SpecificCross in Rice (*Oryza sativa* L.)

ABSTRACT

In the present investigation, a mapping population consisting of 189 lines derived from an inter sub-specific cross between an *indica* cultivar, ‘Samba Mahsuri’ and a *tropicaljaponica* accession, ‘Cuba 65’ wasinvestigated toassess the trait contribution to variabilityemploying principal component analysis (PCA). The study aimed at determination of inter-relationships among various yield attributing traits *vis-à-vis* individuals of the mapping population. Out of eleven principal components (PC), first four PCs exhibited Eigenvalue more than one accounting to a total of 68.73% cumulative variance among the characters. The first PC accounted for the highest variability (PC1 31.6%) to the total variability, followed by 14.51% (PC2), 12.69% (PC3) and 9.91% (PC4). The highest positive Eigenvalue observed for the tiller number (0.33) and panicle number (0.32) in PC1 indicated their pronounced effect in the overall variation in the population.The analysis revealed the traits that contribute maximum for the variation. PCA also revealed negligible effect of days to fifty percent flowering, grain size and spikelet fertility to the variability in the mapping population. Although all the studied traits except grain size showed significant positive association with grain yield;negative correlation was observed between some of the component traits. Panicle weight, thousand grain weight and grain number showed negative association with tiller number and panicle number.The latter two traits contributed to the maximum variability in the population. The trait associations identified in the present investigation will help in determining the selection criteria to be adopted in the development of high yielding cultivars involving *indica/tropical japonica* populations.

Keywords: Eigen Value, inter sub-specific cross, variability, Cumulative Variance.

INTRODUCTION

Rice (*Oryza sativa* L.) cultivated globally, holds significant importance as a primary cereal crop, and a crucial staple food for one-third of the world's population. Among the leading rice-producing nations, India dedicates 44 million hectares to rice cultivation and yields 121.46 million tonnes (Ministry of Agriculture and Farmers Welfare, Government of India, 2020-21).In recent decades, there has been a significant decrease in the amount of arable land available per person, posing a major challenge for meeting the food needs of a growing population in the future.Therefore, rice production must be boosted through further increases in the maximum attainable yield per unit land area, namely, yield potential. To address the global rice deficit, it will be necessary to utilize traits with high-yield potential to develop high-yield rice production systems.In the last 50 years, there has been a significant improvement in the yield of rice varieties whichwas achieved through the development of semi-dwarf varieties and the utilization of heterosis. These factors have played a crucial role in ensuring global food security [1]. However, it has been observed that there has been a stagnation in the yield potential of semi-dwarf *indica* inbred rice varieties in the tropics since the release of IR8 [2].

Comment [B1]: Can be replaced with earmarked

Insufficient genetic diversity has been identified as a major reason for the yield limitation in rice, a problem that has been present for almost two decades. As yield is a complex quantitative trait controlled by a polygenic gene action, to enhance the yield potential, a broad range of germplasms needs to be utilized to gather beneficial alleles of important genes for each yield-related trait in breeding efforts [3]. The two sub-species exhibit distinct morphological and physiological characteristics that are linked to their adaptation to different latitudes and altitudes [4].Broadening the genetic base of elite *indica*cultivars with introgression of ideal new plant type traits from *tropicaljaponica* germplasm is essential in view of the stagnant yield levels of *indica* cultivars [5]. Because of reproductive isolation between the *indica* and *japonica* subspecies, they carry different favourable alleles of some yield-related genes [6]. It has also been observed that derivatives of *indica*/*japonica*cross have higher yield vigour than either Indica /Indica or Japonica /Japonica.Therefore, one of the strategies being considered to further boost the yield potential of rice involves creating inter-subspecific (indica/japonica) hybrids.Moreover, the evaluation of genetic diversity is important to know the source of genes for a particular trait within the available germplasm.

Principal Component Analysis (PCA) is a widely recognized and powerful multivariate statistical method that compresses, reduces and transforms data.PCA reduces data with a large number of correlated variables into a much smaller set of new variables through a linear combination of the original variables capturing most of the variation present in the data [7]. It is used in data analysis to identify the minimum number of components that can explain the maximum variability from the total variability [8]. PCA is also employed to rank genotypes based on their principal component scores. These components are typically derived from either the correlation matrix or the covariance matrix. The eigenvalue of a specific principal component represents the amount of variation in traits that this principal component explains, which is of great importance in breeding programs. Given the significance of PCA, the present investigation was conducted on mapping population derived from the intersub-specific cross between 'Samba Mahsuri' an elite *indica* cultivar and 'Cuba 65' a *tropical japonica* accession,to identify the quantitative traits responsible for variations among the lines in the population.

MATERIALS AND METHODS

The experiment was carried out using a bi-parental mapping population consisting of 189 lines derived from the cross between 'Samba Mahsuri',an elite*indica* cultivar, widely grown in the southern states ofIndia and 'Cuba65', a *tropicaljaponica* accession identified to be a novel source with thick and strong culm, heavy panicles and high yield [5,9,10]. The experimental material was sown in a raised bed nursery and 25-day-old seedlings were transplanted to the main field.The experiment was conducted in an augmented block design (ABD) with seven blocks. The mapping population along with the four checks viz., 'Samba Mahsuri'-P1, 'Cuba65'-P2, 'Swarna', an elite high-yielding *indica* cultivar and 'Lima', an *indica* accession from rice 3K genome with strong culm, [10, 11]was planted at the research farm, ICAR-Indian Institute of Rice Research, (ICAR-IIRR) Hyderabad during *kharif*-2023 with a spacing of 20 x 15 cm. Lines of the mapping population were planted only once while the checks were replicated in each block. Recommended agronomic practices were followed for good crop maintenance.Observations on 11 yield contributing traits viz., days to 50% flowering (DFF), plant height (PH) in centimetres, panicle length (PL) in centimetres, tiller number (TN), number of panicle bearing tillers or panicle number (PN), grain number per panicle (GN), grain yield (GY)per plant in grams, panicle weight (PW) in grams, thousand grain weight or test weight (TW) in grams, spikelet fertility in % (SF) and grain size in terms of length/breadth ratio (LBR) were recorded.PCA was used to identify the traits with a maximum contribution to the variation in the population. The mean data of 11 yield and yield-related traits was used for principal component analysis and the scree plot and biplots were constructed using R Studio (version 4.3.2.).

RESULTS AND DISCUSSION

PCA is a well-known method of dimension reduction that can be used to reduce a large set of variables to a small set that still contains most of the information in the large set. The result of the PCA explained the genetic diversity of the genotypes. In the present PCA, a total of 11 principal components (PCs) were extracted, equivalent to the number of traits studied and it revealed the four most informative PCs with eigenvalues more than one which accounted for 68.73% cumulative variance (Table 1). However, more than 50 % of the variance in the population was explained by the first three major PCs (PC1 31.6%; PC2-14.51%; and PC3-12.69%). Similar to our present findings, first three PCs captured maximum variability in a set of introgression lines derived from multi-parent crossing involving marker assisted forward breeding [7] and in a diverse collection of *indica* and *tropical japonica* genotypes[12].

Table 1. Eigen values, Percentage of variation and Cumulative percentage for principal components in mapping population from the cross of 'Samba Mahsuri' and 'Cuba 65'.

	Eigen Value	Variance%	Cumulative Variance%
PC 1	3.47	31.6	31.6
PC 2	1.59	14.51	46.12
PC 3	1.39	12.69	58.81
PC 4	1.09	9.91	68.73
PC 5	0.87	7.93	76.67
PC 6	0.74	6.76	83.43
PC 7	0.59	5.36	88.79
PC 8	0.49	4.54	93.33
PC 9	0.42	3.83	97.17
PC 10	0.18	1.7	98.87
PC 11	0.12	1.12	100

The contribution of 11 yield traits to the principal components is presented in Table 1. The tiller number (0.33), panicle number (0.32) and length/breadth ratio (0.26) showed positive loading in PC1 while other traits showed negative loadings. In PC2, the parameters viz., test weight (0.01), and length/breadth ratio (0.06) showed positive loading and the remaining factors showed negative loadings. In PC3, traits like test weight (0.57), spikelet fertility (0.42), grain yield (0.26), panicle weight (0.12) and tiller number (0.06) showed positive loading whereas, the remaining traits showed negative loadings. In PC4 test weight (0.24), panicle length (0.25), panicle number (0.15), tiller number (0.11) and plant height (0.02) showed positive loading whereas, the remaining traits showed negative loadings (Table 2). These traits are largely engaged in the variance and they also carry most of the variability. In another study, the first three PCs together contributed 68.69 percent to the total variability and that the number of productive tillers per plant showed positive loading in PC1 [13]. Hence, the selection of traits with high variability will be rewarding for future breeding programs.

Table 2. Contribution of principal components to variation in mapping population from the cross of Samba Mahsuri and Cuba 65.

Trait	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10	PC 11
DFF	-0.16	-0.20	-0.51	-0.28	0.46	-0.04	-0.49	-0.24	0.22	0.07	-0.08
PH	-0.37	-0.08	-0.03	0.02	-0.49	-0.14	-0.48	-0.08	-0.45	-0.35	0.03
PL	-0.37	-0.09	-0.09	0.25	-0.39	0.07	-0.10	0.35	0.66	0.19	0.00
TN	0.33	-0.53	0.06	0.11	-0.19	-0.14	0.01	-0.11	0.03	0.01	-0.71
PN	0.32	-0.56	-0.01	0.15	-0.09	-0.14	-0.02	-0.17	0.08	0.03	0.69
PW	-0.43	-0.19	0.12	-0.20	-0.11	0.01	0.24	-0.31	-0.24	0.69	0.00
GY	-0.16	-0.48	0.26	-0.07	0.29	0.53	-0.08	0.47	-0.21	-0.09	0.01
GN	-0.35	-0.23	-0.34	-0.11	0.01	-0.03	0.64	-0.09	0.07	-0.50	0.00
SF	0.00	-0.05	0.42	-0.71	-0.07	-0.42	-0.03	0.19	0.26	-0.11	0.04
TW	-0.24	0.01	0.57	0.24	0.20	0.13	-0.09	-0.55	0.31	-0.25	-0.02
LBR	0.26	0.06	-0.10	-0.42	-0.42	0.66	-0.02	-0.27	0.14	-0.07	0.02

DFF- Days to 50% flowering, PH- Plant height (cm), PL- Panicle length(cm), TN- Tiller number, PN- Panicle number, PW- Panicle weight(g), GY- Grain yield (g), GN- Grain number per panicle, SF- Spikelet fertility (%), TW- Test weight (g), LBR- Length/Breadth ratio (grain size).

Scree plot elucidated the variation percentage between Eigenvalues and the Principal components (Figure 1A). In this study, PC1 showed 31.6 per cent variability with Eigenvalue of 3.47. From the graph, it is clear that the maximum variation was observed in PC1 in comparison to other PCs. The genotypes selected from PC1 would be useful in breeding programs for the improvement of the traits contributing maximum variability (Figure 1B). The biplot diagrams give the picture of interaction among the traits and also the genotypes performing better for the traits (Figure 1C). The vector length of each trait depicts the contribution of the trait to total divergence, longer vector length denotes more contribution of concerned traits (Figure 1D). The distribution and nature of diversity for genotypes and yield attributing quantitative traits are described in the biplot diagram between PC1 and PC2. The trait tiller number and panicle number showed maximum vector length indicating its contribution to the total divergence followed by length/breadth ratio. The angle between the trait vectors indicates the direction of association between the traits. An angle <90º indicates a positive correlation, an angle equal to 90º indicates no correlation and an angle >90º indicates negative correlation. All the studied traits except grain size were found to have positive association with grain yield, however, negative correlation was observed between some of the component traits. Out of 11 traits studied, PW, PL, GN, PH, TW, DFF and SF showed a positive correlation with grain yield per plantamong which PW, PL, GN and PH contributed significantly to the variability while TW, DFF and SF contributed very less to the total divergence. Further TN and PN contributed significantly to the variabilitydisplayinga positive correlation with GY but negative correlation with PW, PL, GN and PH. Therefore, selection for GY based on GN, PW and PL would be rewarding in improving GY.The genotypes that are present close to a trait vector of the same quadrant would perform best for those traits. Thus, the present study can be utilized to identify the variability contributing parameters and to select the lines to utilize them as donors for the improvement of traits in future breeding programs.

Previously, the percentage of variation for the first three PCs (PC1, PC2, PC3) together accounted for 72.9 per cent of variability and the number of productive tillers (0.1482) and flag leaf length (0.1482) showed positive loading in PC1[14]. Another study reported thatPC1 has the contribution from Days to 50% flowering, leaf length, plant height, panicle length, days to maturity and number of filled grains for 28.46% of the total variability[15]. Further, PC 1 and PC 2 together explained 61.2% of the total variation and PC 1 was related to plant height, biomass and grain yield while, the PC 2 was related to total tillers and productive tillers[16]. Additionally, PC1 added highly (31.848%) to the total variability with significant loading of days to 50% flowering (0.497), days to maturity (0.484) and ear bearing tillers (0.359) which were positively correlated[13]. These findings were similar to our studies. The various principal components capture diverse aspects of variability, and the significant eigen values derived from our analysis provide valuable insights into the distinct contributions of different variables, aiding in a better understanding of the genetic foundation of yield traits in rice. This understanding is crucial for developing targeted breeding strategies aimed at yield improvement. The principal component analysis revealed distinct patterns and relationships among the variables.

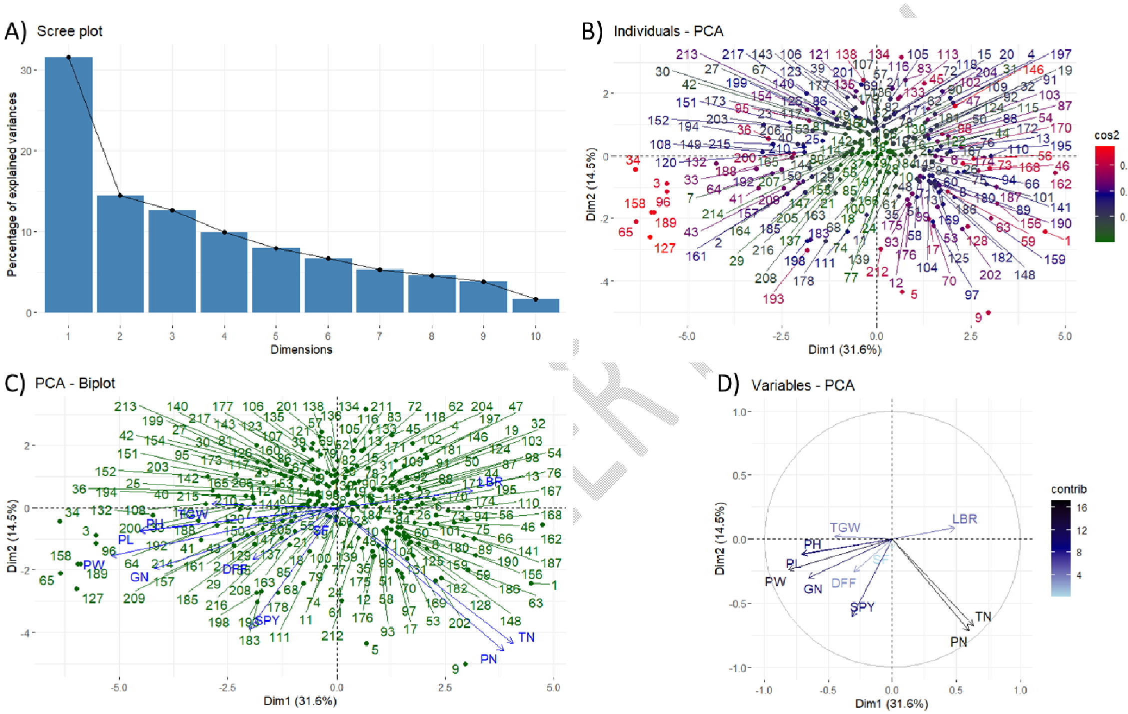

Figure 1 Principal Component Analysis for yield traits among bi-parental population derived from the inter sub-specific cross of Samba Mahsuri *(indica)* and Cuba 65 *(tropical japonica)*. A. contribution of individual lines B. PCA biplot depicting contribution of 11 traits to the total variation in the 189 lines C. Variables PCA plot and D. Scree plot for Percentage of explained variance of 10 PC components. DFF- Days to 50% flowering, PH- Plant height (cm), PL- Panicle length (cm), TN- Tiller number, PN- Panicle number, PW- Panicle weight (g), SPY (GY)- Grain yield (g), GN- Grain number per panicle, SF- Spikelet fertility (%), TGW (TW)- Test weight (g), LBR- Length/Breadth ratio (grain size).

CONCLUSION

The phenotypic value of each trait measures the importance and contribution of each component to the total variance. The tiller number, panicle number, test weight, grain yield and panicle weight are the chief contributors to genetic divergence in the mapping population. Grain size, days to fifty percent flowering and spikelet fertility have no significant contribution to the genetic variability in the population, hence selection for the aforesaid traits would be ineffective. Thus, the present study can be utilized to identify the parameters coming together in different principal components and contributing to variability and also to select the genotypes to utilize them for the improvement of yield traits in future breeding programs.

REFERENCES

1. Tester M & Langridge P. Breeding technologies to increase crop production in a changing world. Science. 2010;12;327(5967):818-22.

2. Peng S, Cassman KG, Virmani SS, Sheehy J, Khush GS. Yield potential trends of tropical rice since the release of IR8 and the challenge of increasing rice yield potential. Crop Science. 1999;39(6):1552-9.

3. McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, Burke JM, Charest D. Cloutier S, Cole G, Dempewolf H. Feeding the future. Nature. 2013;499(7456):23-4.

4. Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohtsubo E. Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Molecular Biology and Evolution. 2003;20(1):67-75.

5. Jyothi B, Divya B, Rao LS, Bhavani PL, Revathi P, Rao PR, Rachana B, Padmavathi G, Kumar JA, Gireesh C, Anantha MS. New plant type trait characterization and development of core set among *indica* and *tropicaljaponica* genotypes of rice. Plant Genetic Resources. 2018;16(6):504-12.

6. Li X, Chen Z, Zhang G, Lu H, Qin P, Qi M, Yu Y, Jiao B, Zhao X, Gao Q, Wang H. Analysis of genetic architecture and favorable allele usage of agronomic traits in a large collection of Chinese rice accessions. Science China Life Sciences. 2020;63:1688-702.

7. Varanasi YV, Isetty SR, Revadi P, Balakrishnan D, Hajira S, Prasad MS, Laha GS, Perraju P, Singh UM, Singh VK, Kumar A. Molecular and Morphological Characterization of Introgression Lines with Resistance to Bacterial Leaf Blight and Blast in Rice. Plants. 2023;12(16):3012.

8. Anderson, TW. An Introduction to Multivariate Analysis. Wiley Eastern Pvt. Ltd. New Delhi, 1972.

9. Bagudam R, Eswari KB, Badri J, Devi GL, JaiVidhya LR, Bhavani P, Swathi G, Bharath M, Rao PR, Kumar JA, Fiyaz RA. Morphological and molecular characterization of new plant type core set for yield and culm strength traits in rice (*OryzasativaL.*). Journal of Plant Biochemistry and Biotechnology. 2021;30:233-42.

10. Kamala (2024) Molecular mapping of genomic regions governing culm strength in rice *(Oryza sativa L.)*. Dissertation, Malla Reddy University

11. Kavya Z (2024) Morphological, biochemical and molecular characterization of strong culm lines in rice. Dissertation, Acharya N. G. Ranga agricultural University

12. Deepika K, Lavuri K, Rathod S, Yeshala CM, Jukanti AK, Reddy SN, Lv SR, Badri J. Multivariate analysis of geographically diverse rice germplasm for genetic improvement of yield, dormancy and shattering-related traits. Plant Genetic Resources. 2021;19(2):144-52.

13. Kumari BK, Kumar BR, Jyothula DP, Rao NM. Diversity analysis in rice breeding lines for yield and its components using principal component analysis. Journal of Pharmacognosy and Phytochemistry. 2021;10(1):905-9.

14. Christina GR, Thirumurugan T, Jeyaprakash P, Rajanbabu V. Principal component analysis of yield and yield related traits in rice (*OryzasativaL*.) landraces. Electronic Journal of Plant Breeding. 2021;12(3):907-11.

15. Nachimuthu VV, Robin S, Sudhakar D, Raveendran M, Rajeswari S, Manonmani S. Evaluation of rice genetic diversity and variability in a population panel by principal component analysis. Indian journal of science and technology. 2014;7(10):1555-62.

16. Worede F, Sreewongchai T, Phumichai C,Sripichitt P. Multivariate analysis of genetic diversity among some rice genotypes using morpho-agronomic traits. Journal of plant sciences. 2014; *9*(1):14-24.