PRINCIPAL COMPONENT ANALYSIS IN F_2 MAPPING POPULATION FOR HIGH GRAIN ZINC CONTENT IN RICE (*ORYZA SATIVA* L.)

**S.K. Singh, Sonali Vijay Habde*, Amrutlal Khaire, Korada Mounika,
Prasanta Kumar Majhi and D.K. Singh**

Department of Genetics and Plant Breeding, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi
(Uttar Pradesh), India.

Abstract

Rice is one of the most important staple food crop in world however white rice it is poor source essential micronutrients. Increasing bioavailable zinc in rice endosperms is aim of breeding efforts in rice bio fortification. Rice mapping population for high grain zinc content consisting of 312 F_2 individuals was evaluated for 15 yield related, grain quality and nutritional traits at Agriculture Research Farm, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi. Data collected was subjected to principal component analysis study to examine pattern of variation exhibited by 15 variables and to identify group of variables based on their relative contribution to total variability by reducing dimensionality of data. Principal component analysis, revealed five PCs having eigen value more than 1.0 contributing 73.84% of cumulative variability. PC1 showed 21.445% while PC2, PC3, PC4 and PC5 exhibited 18.356%, 14.302%, 11.785% and 7.955% variability respectively. 2D and 3D scatter diagrams of principal component analysis demonstrated highly divergent F_2 entries clearly. Based on eigen vectors values, factors loading and biplots traits like total grains per panicle, fertile grains per panicle, grain yield per plant, plant height, kernel length, kernel length to breadth ratio, total tillers per plant, effective tillers per plant, 1000 grain weight and panicle length were major determinant for differentiating genotypes and contributed to major variation in present data set. Hence selection based on this trait combination contributing for major variations will be effective for trait improvement in present mapping population.

Key words: Rice, Principal component analysis, Grain zinc, Biplot

Introduction

Zinc (Zn) and iron (Fe) and are highly essential micronutrients for human health. Zinc (Zn) is one of the essential micronutrients, serves as a co-factor for more than 300 enzymes involved in the metabolism of carbohydrates, lipids, proteins and nucleic acids, so zinc is important micronutrient in growth and development processes of human, other animals as well as plants (Roohani *et al*., 2013). Fe is an important component of hemoglobin. Rice (*Oryza sativa* L.) is the most important food crop of world and source of food for more than half of world's population. It is also a predominant staple food and energy source for Asian population. Although rice is rich in starch, but milled rice is a poor source essential micronutrients such as Iron (Fe) and Zinc (Zn) of micronutrients (Bouis and Welch, 2010), hence population solely depend on rice as staple food suffers from

micronutrient deficiency or hidden hunger. Across the globe more than two billion people, especially children, pregnant and lactating women suffer from Fe and Zn deficiencies (Wessels and Brown, 2012). Fe deficiency causes anemia and reduced growth while zinc deficiency results in diarrhea, pneumonia, stunting and child mortality (Prasad, 2004). Biofortification strategy implemented through plant breeding is a sustainable solution which can have massive impact while dealing with micronutrient malnutrition.

Understanding of genomic regions and genetic basis grain mineral content is the first step of breeding zinc and iron efficient plants and mineral dense rice varieties. F_2 mapping population is the simple and easiest population for any QTL mapping experiment. It is a unique population which multiplies heterozygosity of F_1 and it permits simultaneous identification of additive as well as dominance effect QTLs. Theoretically it represents all

**Author for correspondence :* E-mail: sonali.habde@gmail.com

possible recombinants (subject to population size) with unique recombination in each individual. Spectrum of variation in F_2 is large enough to conduct variability and diversity studies. As large size F_2 population evaluated for multiple traits generates large data set which is difficult to interpret necessitates reduction in the dimensionality of such datasets without loss of information. Multivariate analysis like principal component analysis is important to reduce dimensions of data table in which observations are delineated by different inter-correlated quantitative variables. Reduction is achieved by linear transformation of the original variables into a new set of uncorrelated variables known as principal components (PCs) by capturing and retaining original variations (Jolliffe and Cadima, 2016). It allows visualization of the differences among the individuals and identifies possible groups. It is also useful to structure diversity in segregating mapping populations. It permits examination of pattern of variation and to determine relative contribution of various characters to total variability.

The first step in PCA is to calculate eigen values, which define the amount of total variation that is displayed on the PC axes. The loading values are standardized in such a way that the sum of square of loadings within a principal component is equal to one. The loadings are viewed as weights defining the contribution of characters in respective principal component. PCs are the result of orthogonal transformation of variables and are independent of each other therefore each PC discloses different properties of the original data. The proportion of variation accounted for by each PC is expressed as the eigenvalue divided by the sum of the eigen values. The eigenvector defines the relation of the PC axes to the original data axes. Several multivariate studies with principal component analysis have been reported in rice recently. (Maji and Shaibu, 2012; Nachimuthu *et al.*, 2014; Govintharaj *et al.*, 2018; Riaz *et al.*, 2018; Madhubabu *et al.*, 2020; Sheela *et al.*, 2020).

Due to its adaptive properties, different variants of principal component analysis has wide applicability in dealing with different data types based analysis like germplasm characterization and evaluation (Maji and Shaibu, 2012), stage specific fluorescence spectrum based nitrogen content analysis (Yang *et al.*, 2016) accurate and quick disease recognition rate (Xiao *et al.*, 2018), GWAS studies for rice architecture (Yano *et al.*, 2019). In this context data collected for 15 variables in present mapping population was subjected to principal component analysis study to examine variation exhibited by all 15 traits under study, to identify group of variables based on their relative contribution to total variability by

reducing dimensionality of data.

Materials and Methods

F_2 mapping population derived from cross of Rajendrakasturi×URG-30 was evaluated in *kharif-2018* at Agriculture Research Farm, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, India. F_2 seeds were sown on Nursery bed in 1st week of June, 2018 and after 21 days seedlings were transplanted in field with spacing of 40cm × 30cm. 312 F_2 individual were raised along with their parents with recommended package of practices of this region. Observations were recorded for variables considered for present study *viz* morphological traits like days to flowering, plant height (cm), total tillers, number of effective tillers; yield related traits *viz* panicle length (cm), total grains per panicle, fertile grains per panicle, spikelet fertility (%), grain yield per plant (grams), 1000 grain weight; grain quality traits like kernel length, kernel breadth, kernel length to breadth ratio; nutritional quality traits like grain iron and zinc content in ppm. Each F_2 plant was harvested separately. After harvesting grain samples were prepared for Zinc and Iron analysis. Grain Zinc (in ppm) and Iron content (in ppm) were analyzed at IRRI South Asia Hub, with XRF machine facility available at Harvest Plus lab, ICRISAT, Hyderabad. Observations recorded for 15 traits were analysed in XLSTAT, 2019 for principal component analysis. Dimensionality of data matrix is reduced by creating several significant principal components having eigen value more than one. The loadings are viewed as weights defining the contribution of characters in respective principal component.

Results and Discussion

Knowledge of existing variability in breeding population is important before its use for effective selections for traits of interest. As yield is complex trait, it has association with several other trait which interact with each other (Nandeshwar *et al.*, 2010). Therefore selection for grain yield alone in early segregating population may not be promising, hence characters combinations with proven significant contribution to total variability will be effective to identify transgressive segregants and recombinant individuals in F_2 population.

Principal Component Analysis is one of the important multivariate analysis which measures the importance and contribution of each component to total variance. It measures independent impact of variables to the total variance whereas each coefficient of proper vectors indicates the degree of contribution of every original variable with which each principal component is

Table 1: Eigen value, variability and cumulative variability contributed by 15 principal components.

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Eigen value	3.217	2.753	2.145	1.768	1.193	0.969	0.802	0.643	0.467	0.423	0.362	0.226	0.026	0.002	0.002
Variability (%)	21.445	18.356	14.302	11.785	7.955	6.461	5.348	4.289	3.111	2.822	2.413	1.510	0.175	0.016	0.013
Cumulative (%)	21.445	39.801	54.103	65.888	73.843	80.304	85.651	89.940	93.050	95.873	98.286	99.796	99.971	99.987	100.0

associated. Higher the value of coefficients, regardless of the sign, the more effective they will be in discriminating between accessions (Nachimuthu *et al.*, 2014).
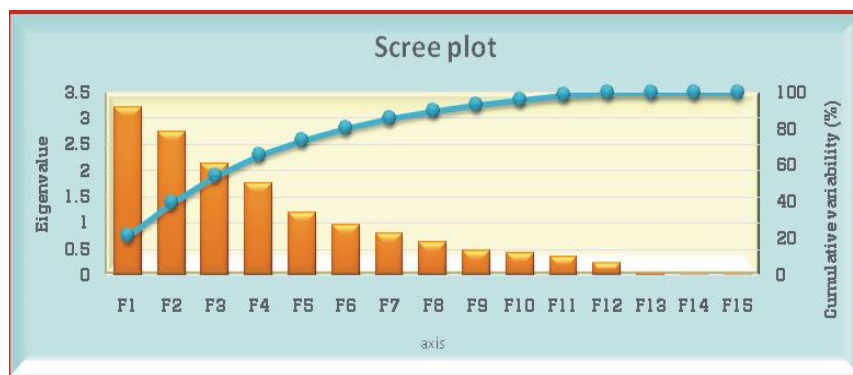
In present study, Principal component analysis was based on correlation matrix, created 15 new variables as principal components (PC1, PC2…, PC15) with 312 entries without changing their relative position. Eigen value, variability and cumulative variability contributed by 15 principal components is presented in table 1. Each principal component was linear combination of 15 attributes of data matrix. As PC's having eigen value more than one are more informative over single variable studied. Five principal component with eigen values more than one contributed 73.843% towards the total variability. Other principal components with eigen value less than one were considered as non significant, indicated that majority of variation has been explained by first five components. Contribution of all characters in first five principal components are presented in table 2. 72.24% of contribution to total variability was explained by first three principal components (Sanni *et al.*, 2012). In one principal component study with 39 rice accessions in rainfed lowland ecology of Nigeria, Gana *et al.*, 2013 recorded 65.4% variability combined by first five

Table 2: Component loading of 15 traits to first five PCs.

Traits	F1	F2	F3	F4	F5
DTF	0.095	-0.009	-0.074	0.463	0.427
PH	0.347	0.217	-0.011	-0.171	0.202
Ti	0.276	0.312	0.381	0.233	-0.071
ET	0.287	0.316	0.381	0.216	-0.059
PL	0.297	0.130	-0.236	-0.313	0.059
G/PL	0.405	-0.139	-0.362	-0.071	-0.107
FG/PL	0.413	-0.175	-0.313	-0.122	-0.170
SF	0.056	-0.145	0.146	-0.191	-0.237
GY/P	0.341	0.324	0.154	-0.170	-0.121
TWt	-0.196	0.375	-0.136	-0.299	0.026
Zinc	-0.144	-0.085	0.351	-0.389	-0.082
Iron	-0.001	-0.205	0.289	-0.142	-0.412
KL	-0.248	0.454	-0.237	-0.037	-0.148
KB	-0.092	0.139	0.111	-0.428	0.540
LB Ratio	-0.202	0.384	-0.284	0.165	-0.405

components. Riaz *et al.*, 2018 reported 97.9% variation contributed by first five components. During assessment of genetic variability for micronutrient content and agromorphological traits in rice, 76.4% of variability was accounted by 6 principal components (Madhubabu *et al.*, 2020). Sheela *et al.*, 2020 reported 72.24% of variability accounted by first four components.

The first principal component (PC1) contributed maximum to total variability (21.445%). The most important traits in first component (PC1) were fertile grains per panicle, grains per panicle, plant height and grain yield per plant followed by panicle length, effective tillers and total tillers influenced effectively, days to flowering and spikelet fertility also showed positive loading and remaining traits showed negative loadings. Second component (PC2) accounted 18.356% of total variation was primarily contributed by kernel length, LB ratio, 1000 grain weight, grain yield per plant, effective tillers and total tillers while grains per panicle, fertile grains per panicle, spikelet fertility, grain zinc content, grain iron content showed negative loading for PC2. Principal component 3 (PC3) contributed 14.302% variation for which characters like total tillers, effective tillers, grain zinc content, grain iron content, grain yield per plant, spikelet fertility and kernel breadth showed positive loading on PC3 where as grains per panicle, fertile grains per panicle, kernel length, panicle length showed negative loading. Component PC4 explained 11.785% of variation for which days to flowering contributed more and least contributed by tillers, effective tillers and LB ratio whereas remaining traits showed negative loading. PC5 accounted 7.955% variability which was mainly influenced by kernel breadth. Similar findings have been reported by Nachimuthu *et al.*, 2014. Mahendran *et al.*, 2015 identified plant height, days to flowering, panicle length as main contributor to PC1, however effective tillers account for PC2 and kernel length, kernel breadth and LB ratio contributed more in PC3 which is different from our findings. Madhubabu *et al.*, 2020 analyzed results which are similar to our present findings in which PC1 was loaded positively with days to flowering and negative with zinc content, PC2 showed positive loadings with grain length, grain width and test weight, PC3 was mainly
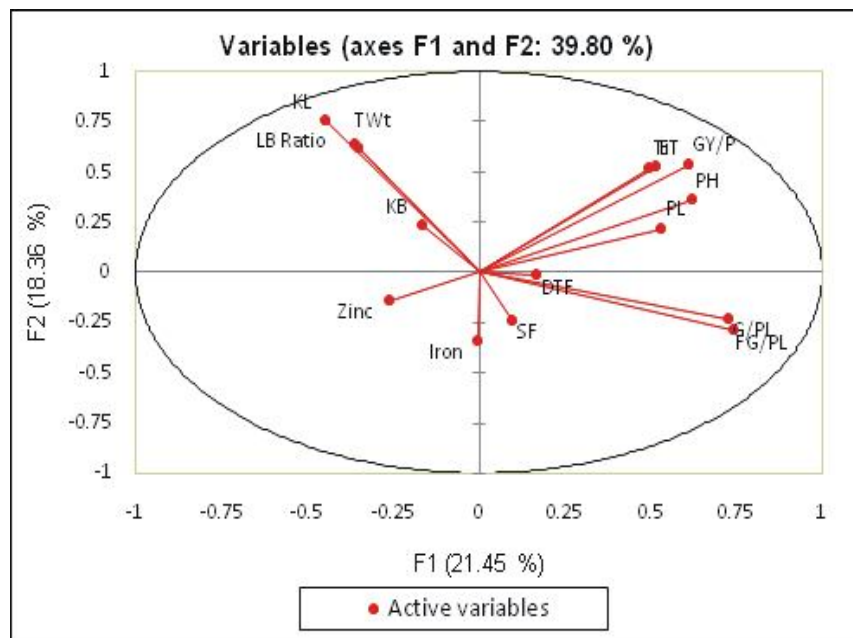
Fig. 1: Scree plot showing Eigen value variation in 15 Principal components.

contributed by total tillers and effective tillers per plant.

Based on loading values, exact picture of component traits contributing to maximum variability was obtained. According to which contribution of yield related was more in first component; relative contribution of grain quality traits was more in PC2, total tillers and effective tillers, grain zinc and iron content contributed more in PC3. Relation of PC1 and PC3 with yield factors, PC2, PC4 and PC5 with grain quality traits has been highlighted by Gour *et al*., 2017. In another study by Riaz *et al*., 2018 they reported association of PC1 and PC2 with yield traits and PC3 with days to flowering and maturity. Graph between eigen values and principal component as Scree plot is given in fig. 1. which explained percentage of variation by each PC and its eigen value. As depicted in graph majority of variations are contributed by first five PCs.

In biplot analysis the cosine of the angle between any two vectors representing variables is the coefficient of correlation between those variables. Also the cosine of the angle between any vector representing a variable and the axis representing a given PC is the coefficient of correlation between those two variables (Jolliffe and Cadima, 2016). In present study biplot between PC1 and PC2 was plotted by using variability of all 15 variables under study were represented by arrow. Sign difference in loading values of variables with respect to PC1 and PC2 is clearly visible in biplot given in fig. 2. First quadrant comprised of 5 yield trait related variables which suggest high correlation among them and they are also strongly correlated with PC1. Second quadrant contained 4 grain quality related traits which are positively correlated with PC2 and negatively correlated with PC1. Third quadrant mainly composed of grain nutritional quality traits *viz* grain zinc and iron content both with negative correlation with PC1 as well as PC2. Fourth quadrant involved flowering and panicle related traits in positive correlation with PC1 and negative correlation with PC2. Variables negatively correlated with each other are positioned on opposite quadrants.

The quality of represented variables on biplot is expressed by its square cosine value, so that those variables with higher value are positioned near to circumference and are important for first two PCs. In our study as presented in table 3 square cosine value of grains per panicle, fertile grains per panicle, plant height, grain yield per plant and panicle length are higher in PC1., kernel length, kernel length to breadth ratio, 1000 grain weight are higher in PC2., total tillers, effective tillers per plant in PC3 while grain zinc and iron content have high cosine values in PC4 and PC5 respectively. Relative length of vector from origin representing each variable is comparable to its magnitude of squared cosine value in respective principal component so that larger vector has higher contribution to the variations.

Based on above all findings of principal component analysis *viz* eigen vectors values, factors loading and biplots of PC1 and PC2, traits like total grains per panicle, fertile grains per panicle, grain yield per plant, plant

Fig. 2: Biplot of 15 variables across first two Principal components.

Table 3: Square cosines of the variables.

Traits	F1	F2	F3	F4	F5
DTF	0.029	0.000	0.012	0.379	0.217
PH	0.387	0.129	0.000	0.052	0.049
Ti	0.245	0.269	0.311	0.096	0.006
ET	0.266	0.276	0.312	0.083	0.004
PL	0.283	0.046	0.120	0.173	0.004
G/PL	0.528	0.053	0.282	0.009	0.014
FG/PL	0.549	0.084	0.210	0.026	0.034
SF	0.010	0.058	0.046	0.064	0.067
GY/P	0.374	0.288	0.051	0.051	0.017
TWt	0.123	0.387	0.039	0.158	0.001
Zinc	0.067	0.020	0.264	0.267	0.008
Iron	0.000	0.116	0.179	0.036	0.202
KL	0.197	0.568	0.120	0.002	0.026
KB	0.027	0.054	0.026	0.323	0.347
LB Ratio	0.132	0.405	0.173	0.048	0.196

height, kernel length, kernel length to breadth ratio, total tillers per plant, effective tillers per plant, 1000 grain weight and panicle length were major determinant for differentiating genotypes and contributed to major variation in present data set. All these traits have great influence on phenotypes; hence selection based on these trait combinations will be effective for improvement of present mapping population for high yield with good grain quality. Days to flowering, total tillers, panicle length, kernel length, kernel breadth, Iron and zinc content were considered as the most contributing traits to total variations by Madhubabu *et al*., 2020.

The correlation between a variable and a principal component (PC) is used as the coordinates of the variable on the PC. The representation of variables differs from the plot of the observations because observations are represented by their projections, but the variables are represented by their correlations (Abdi and Williams, 2010). Spatial distribution of all 312 F₂ individuals for 15 variables and across first two PCs is presented in fig. 3.
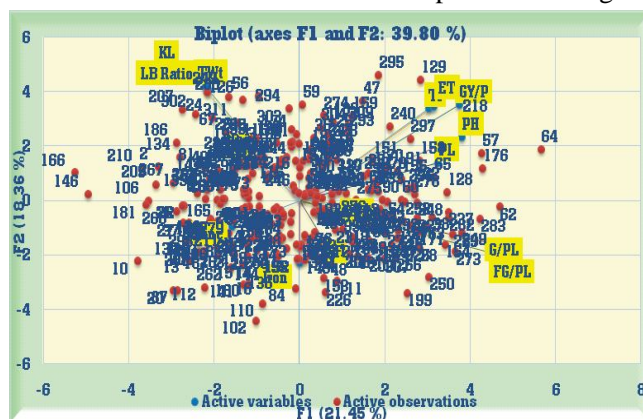
Fig. 3: Biplot of individuals and variables across first two Principal components.
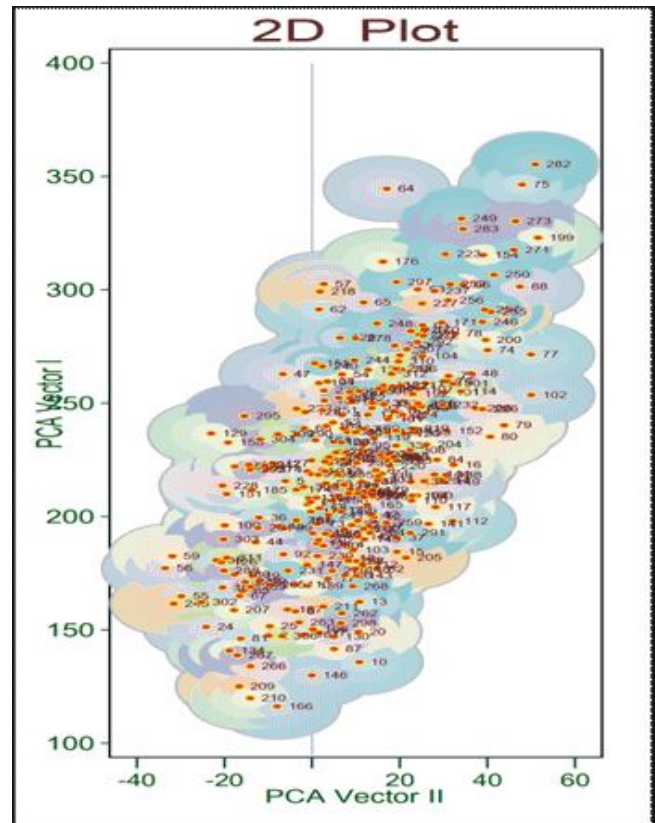
Fig. 4: Two dimensional scatter plot of PC1 and PC2.
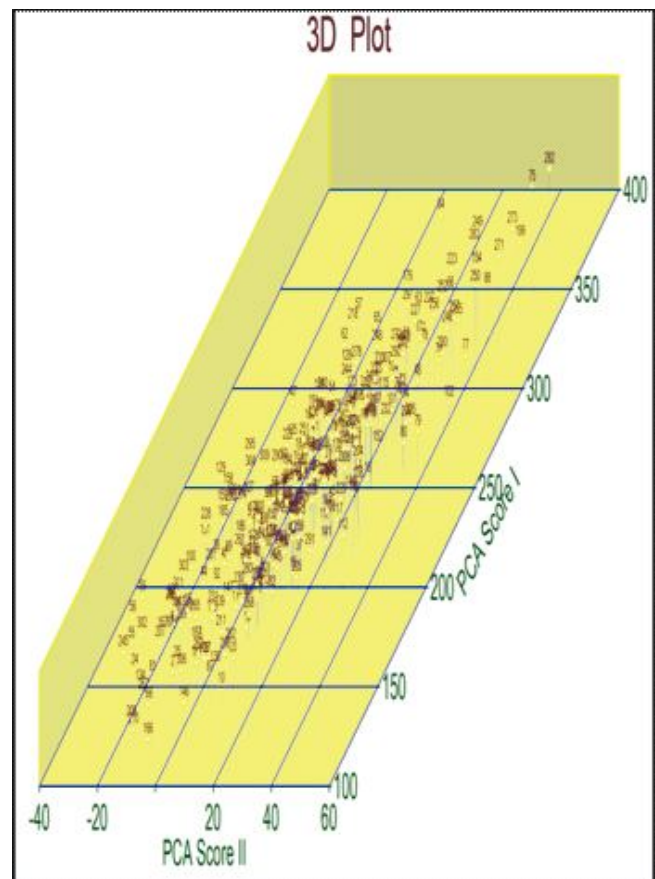
Fig. 5: Three dimensional scatter plot of PC1 and PC2.

All F₂ plants were widely scattered across different quadrants. Each F_2 individual plant is presented as perpendicular projections on variable vector, indicating major traits in each quadrant which are key determinant for distributing F_2 plant. Consideration of biplot of genotype and variables together is important to indicate association between different variable and promising entries in every quadrant display. Individuals distributed near origin have traits values close to mean. This biplot laid basic criteria for selecting recombinants and transgressive segregants identification from present population. Based on PCA scores of F_2 individuals obtained using the first two PCA scores *i.e.,* PC I and PC II two dimensional and three dimensional scatter diagrams are shown in fig. 4 and fig. 5 respectively. Both scatter diagrams provided visual aid for widely scattered F_2 entries and maximum portion of F_2 variability accounted by first two principal components. Similar results for grouping of large number of segregating population across genotype by variable biplot has been presented by Rahimi *et al.*, 2013 and Govintharaj *et al.*, 2018.

Conclusion

In the context of global nutritional security, focus of conventional plant breeding has been shifted more towards grain quality and nutritional aspects. Multivariate analysis using correlation matrix of yield related, grain quality and nutritional traits are gaining equal importance. Present multivariate analysis study has successfully reduced dimension of data set of 15 variables to 5 principal components explaining 73.84% of cumulative variability. It has explained the pattern of usable variation in present mapping population. Present assessment of segregating population for grain quality and nutritional traits formed basis for identification of traits influencing phenotypes and undertaking selection based on principal variables *viz* total grains per panicle, fertile grains per panicle, grain yield per plant, plant height, kernel length, kernel length to breadth ratio, total tillers per plant, effective tillers per plant, 1000 grain weight and panicle length will be effective for yield improvement. Individual by variable biplot analysis will assist in early generation single plant selections and identification of good recombinants based on grouping of F₂ plants.

Acknowledgement

Abbreviation

PH-plant height, Ti-total tillers, ET-effective tillers, PL-panicle length, G/PL-total grains per panicle, FG/PL-Fertile grains per panicle, SF-Spikelet fertility, GY/P-Grain yield per plant, Twt-1000 grain weight, Zinc- grain Zinc content in ppm, Iron-Grain Fe content in ppm, KL-kernel length, KB-kernel breadth, LB ratio- length to breadth ratio.

References

Abdi, H. and L.J. Williams (2010). Principal Component Analysis. *Wiley interdisciplinary reviews: computational statistics*, **2(4):** 433-459.

Bouis, H.E. and R.M. Welch (2010). Biofortification- a sustainable agricultural strategy for reducing micronutrient malnutrition in the global south. *Crop Science*, **50:** S20-32.

Gana, A.S., S.Z. Shaba and E.K. Tsado (2013). Principal component analysis of morphological traits in thirty-nine accessions of rice (*Oryza sativa* L.) grown in a rainfed lowland ecology of Nigeria. *Journal of plant breeding and crop science*, **4(6):** 120-126.

Gour, L., S.B. Maurya, G.K. Koutu, S.K. Singh, S.S. Shukla and D.K. Mishra (2017). Characterization of rice (*Oryza sativa* L.) genotypes using principal component analysis including scree plot & rotated component matrix. *International Journal of Chemical Studies*, **5(4):** 975-983.

Govintharaj, P., S. Manonmani and S. Robin (2018). Variability and genetic diversity study in an advanced segregating population of rice with bacterial blight resistance genes introgressed. *Ciência e Agrotecnologia*, **42(3):** 291-296.

Jolliffe, I.T. and J. Cadima (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **374(2065):** 20150202.

Madhubabu, P., R. Surendra, K. Suman, M. Chiranjeevi, R.A. Fiyaz, D.S. Rao, U. Chaitanya, L.V.S. Rao, V.R. Babu and C.N. Neeraja (2020). Assessment of genetic variability for micronutrient content and agro-morphological traits in rice (*Oryza sativa* L.). *Indian J. Genet*, **80(2):** 130-139.

Mahendran, R., P. Veerabadhiran, S. Robin and M. Raveendran (2015). Principal component analysis of rice germplasm accessions under high temperature stress. *International Journal of Agricultural Science and Research (IJASR)*, **5(3):** 355-359.

Maji, A.T. and A.A. Shaibu (2012). Application of principal component analysis for rice germplasm characterization and evaluation. *Journal of Plant Breeding and Crop Science*, **4(6):** 87-93.

Nachimuthu, V.V., S. Robin, D. Sudhakar, M. Raveendran, S. Rajeswari and S. Manonmani. (2014). Evaluation of Rice Genetic Diversity and Variability in a Population Panel by Principal Component Analysis. *Indian J. Sci. Technol*, **7(10):** 1555-1562.

Nandeshwar, B.C., S. Pal, B.K. Senapati and D.K. De (2010). Genetic variability and character association among biometrical traits in F2 generation of some Rice crosses. *Electronic Journal of Plant Breeding*, **1(4):** 758-763.

Prasad, A.S. (2004). Zinc deficiency: Its characterization and treatment. *Met Ions Biol Syst*. **41**: 103-137.

Rahimi, M., H. Dehghani, B. Rabiei and A.R. Tarang (2013). Evaluation of rice segregating population based on drought tolerance criteria and biplot analysis. *International journal of agriculture and crop sciences*, **5(3):** 194.

Riaz, A., T. Bibi, Q. Raza, M. Sabar and M. Akhter (2018). Evaluation of rice (*Oryza sativa* L.) advance uniform lines using multivariate analysis. *African Journal of Plant Science*, **12(11):** 284-289.

Roohani, N., R. Hurrell, R. Kelishadi and R. Schulin (2013). Zinc and its importance for human health: An integrative review. *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences*, **18(2):** 144-157.

Sanni, K. A., I. Fawole, S.A. Ogunbayo, D.D. Tia, E.A. Somado, K. Futakuchi, M. Sie, F.E. Nwilene and R.G. Guei (2012). Multivariate analysis of diversity of landrace rice germplasm. *Crop Science*, **52(2):** 494-504.

Sheela, K.R.V., S. Robin and S. Manonmani (2020). Principal component analysis for grain quality characters in rice germplasm. *Electronic Journal of Plant Breeding*, **11(1):** 127-131.

Wessells, K.R. and K.H. Brown (2012). Estimating the global prevalence of zinc deficiency: Results based on zinc availability in national food supplies and the prevalence of stunting. *PLoS One*, **7:** e50568.

Xiao, M., Y. Ma, Z. Feng, Z. Deng, S. Hou, L. Shu and Z. Lu (2018). Rice blast recognition based on principal component analysis and neural network. *Computers and electronics in agriculture*, **154:** 482-490.

Yang, J., W. Gong, S. Shi, L. Du, J. Sun and S.L. Song (2016). Estimation of nitrogen content based on fluorescence spectrum and principal component analysis in paddy rice. *Plant, Soil and Environment*, **62(4):** 178-183.

Yano, K., Y. Morinaka, F. Wang, P. Huang, S. Takehara, T. Hirai and S. Yoshida (2019). GWAS with principal component analysis identifies a gene comprehensively controlling rice architecture. *Proceedings of the National Academy of Sciences*, **116(42):** 21262-21267.