International Journal of Scholarly
Research and Reviews
Journal homepage: https://srrjournals.com/ijsrr/
ISSN: 2961-3299 (Online)

(RESEARCH ARTICLE)

Principal component analysis for seven quantitative traits of different rice (*Oryza sativa L.*) genotypes tested at Pawe, northwestern Ethiopia

Gedifew Gebrie Muchie * and Desta Abebe Belete

Ethiopian Institute of Agricultural Research, Pawe Agricultural Research Center, Pawe, Ethiopia.

International Journal of Scholarly Research and Reviews, 2022, 01(01), 009–016

Publication history: Received on 31 May 2022; revised on 07 July 2022; accepted on 09 July 2022

Article DOI: https://doi.org/10.56781/ijsrr.1.1.0022

Abstract

Principal component analysis (PCA) is aimed at reducing the dimensionality to find a smaller number of dimensions (usually 2 or 3) that exhibit most of the variations present in the data helping to identify the relative importance of individual traits on the genotypic diversity of the genotypes. The PCA was computed using seven quantitative traits measured from 321 rice genotypes evaluated using augmented RCBD experimental design with a plot area of 2.5m^2 involving 4 rows per plot. The seeds were drilled in rows with a seed rate of 60kg per hectare. NPS (Nitrogen-Phosphorus-Sulfur) (124 kg per hectare) and Urea (100 kg per hectare) fertilizers were applied. The quantitative traits such as days to 50% heading, days to 85% maturity, plant height, panicle length, number of filled and unfilled grains per panicle, and 1000 seed weight were collected and subjected to the principal component analysis using XLSTAT 5.03 statistical software so as to determine the importance of the measured quantitative traits for the genetic diversity of the tested rice genotypes. The first three principal components (PC1, PC2 and PC3) were identified with a total cumulative variation of 78.90% showing that the genotypes could be grouped at least into three main varied classes. From the observed distribution plot, the tested genotypes were almost uniformly distributed in four quadrants pointing the presence of genetic diversity among the genotypes.

Keywords: Distribution plot; Eigen values; Genetic diversity; Rice; Principal component analysis

1. Introduction

Rice (Oryza sativa L., 2n = 2x = 24) is the second most widely grown cereal crop and the staple food for more than half of the world's population, providing two-thirds of caloric intake for more than three billion people in Asia and one-third of nearly 1.5 billion people in Africa and Latin America [14] and is increasingly important in Ethiopia and named as the "millennium crop" to ensure food security of the country though the crop is relatively a new grain to the country, with its introduction only taking place in the 1970s [4]. It is estimated that the country is endowed with about 30 million hectares of land, of which 5.6 million hectare is categorized as highly suitable for rice production [8] but the current rice production area is only 85,288.87 hectares [7].

In an effort to improve the productivity of rice and increasing its annual production and area coverage, the Ethiopian national rice breeding research program introduced and evaluated germplasms from external sources targeting for their environmental adaptability and agronomic performance regarding with their high grain yield, tolerant for biotic and abiotic stresses and for other quality traits. Measuring the available genetic diversity is of utmost importance for effective evaluation and utilization of germplasms [16] to explore their variability so as to identify desirable agronomic attributes [5].

---

[*] Corresponding author: Gedifew Gebrie Muchie
Ethiopian Institute of Agricultural Research, Pawe Agricultural Research Center, Pawe, Ethiopia.

Patterns of genetic variation among the germplasms are popularly estimated by multivariate analysis statistical approaches such as principal component analysis (PCA). Thus, this research activity was designed to asses and determine the extent and pattern of genetic diversity vested on the pool of rice genotypes based on important agronomic traits using principal component analysis by reducing the dimensionality and getting smaller number of dimensions showing most of the variations present in the measured quantitative data and identifying the importance of individual traits on the genotypic diversity of the genotypes.

2. Material and methods

2.1 Description of the Experimental site

The trial was conducted at Pawe agricultural research center experimental station (Figure 1) during 2019/2020 main cropping season under rain-fed. The experimental station is located in Pawe district of northwestern Ethiopia at about 575 km north west of Addis Ababa at a latitude of 11°19'N and longitude of 36°24'E and at an altitude of 1120 masl. The area is characterized by hot to warm moist conditions with mean minimum and maximum temperature of 16^{o}c and 32^{o}c, respectively and has an average annual rainfall of 1587 mm with five to seven months duration [9].

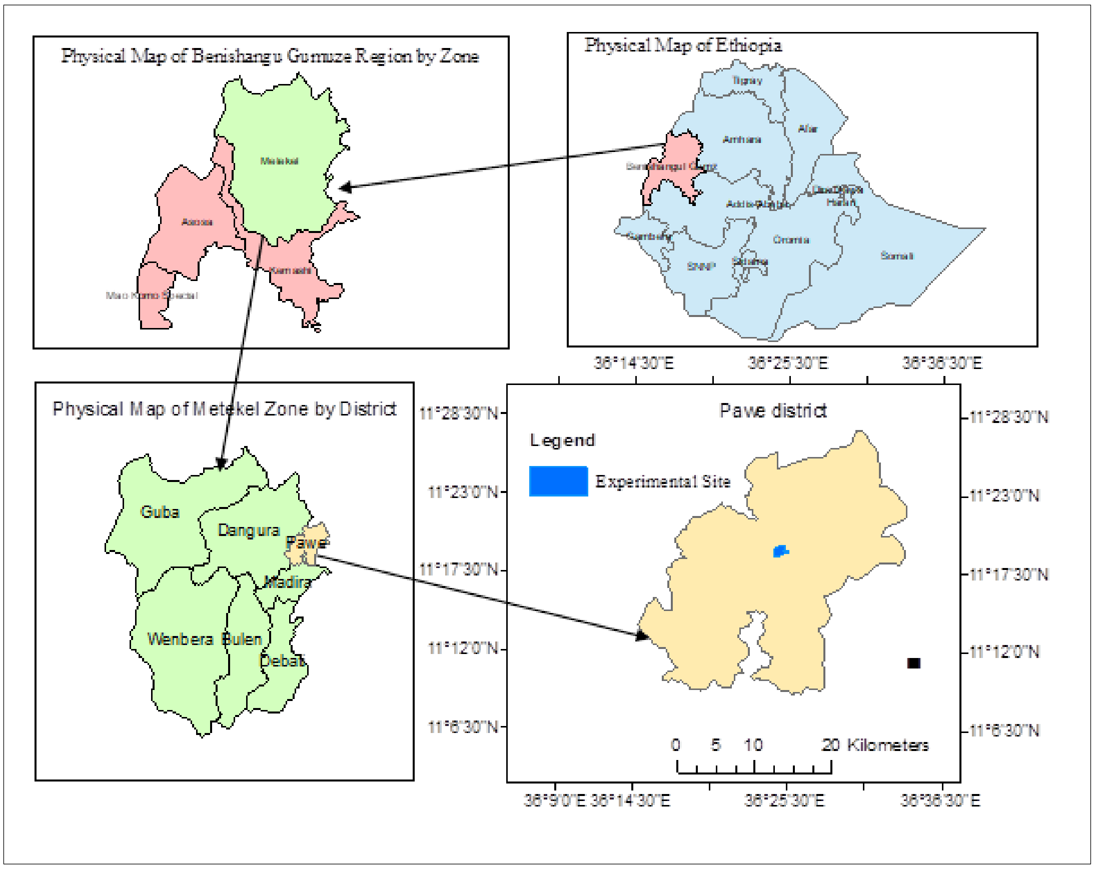

Figure 1 The ArcGIS map projection of the experimental site

2.2 Plant Materials

352 lowland rice genotypes with 4 standard checks (Suplement-1) were used. The genotypes were collected and introduced from China and Japan in collaboration with IRRI and AfricaRice rice research project.

2.3 Experimental Design and Procedures

An Augmented RCBD design was employed with a plot size of 2.5m^{2} (2.5 m x 1m) each with four rows. A spacing between plots and between blocks of 0.5m and 1m was used respectively. An inter-row spacing of 0.25m and a seed rate of 60 kgha^{-1} was used. Inorganic fertilizer (NPS = 124 kgh^{-1}; Urea = 100 kgh^{-1}) was applied. The total NPS fertilizer was applied at planting, whereas the urea was applied with three splits (1/3 at planting, 1/3 at tillering after weeding and 1/3 at the crop's panicle initiation stage).

2.4 Data Collection and Statistical Analysis

Data were collected on plot and plant bases following the appropriate growth stage of the crop for each respective measured trait. 31 genotypes failed to germinate and hence data were not collected on plots where those missed genotypes were planted. Agronomic traits such as days to 50% heading (DH), days to 85% maturity (DM), plant height in cm (PH), panicle length in cm (PL), number of filled grains per panicle (NFG), number of unfilled grains per panicle (NUFG), and 1000 seed weight in gram (TSW) were considered and subjected to principal component analysis using XLSTAT statistical software 5.03 so as to identify the contributing traits to the total genetic variation among the rice genotypes on which the principal components having eigen value greater than 1.00 unit were considered as significant. Bi-plots were drawn separately based on the first two respective PCA's to facilitate visualization of the genotypes by trait associations.

3. Results and discussion

Principal components with an eigen value of less than a unit account for less variance than did the original variable (which had a variance of 1.00 unit), and so are of little use and usually such a component should be eliminated so that fewer components are dealt with [15]. In addition to the value of eigen values, the scree plot which is a simple line segment that shows the fraction of total variance in the data is also a useful visual aid for determining an appropriate number of principal components [13]. Therefore, from the principal component analysis of 321 rice genotypes, the first three principal components were observed with the eigen values greater than 1.00 unit (Table 1) and the scree plot line (Figure 2) after the third principal component was almost flat indicating that the corresponding principal components were accounted for lesser amounts of the cumulative variability.

Table 1 The eigen values of the principal components and their contributions to the variability of 321 rice genotypes

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	2.61	1.83	1.09	0.54	0.41	0.31	0.22
Variability (%)	37.25	26.11	15.54	7.74	5.83	4.38	3.15
Cumulative Variability %	37.25	63.36	78.90	86.64	92.47	96.85	100.00

Where: PC-principal component, %-Percentage

Therefore, based on the eigen values of each principal components and the appearance of the scree plot, it has been confirmed that the quantitative traits with a higher loading on PC1, PC2 and PC3 contributed the most variability observed in the corresponding seven quantitative traits where a great emphasis had been given for those traits having a higher loading on the first three principal components.

The first three principal components (PC1, PC2 and PC3) accounted for a total cumulative variance of 78.90 % where the largest variance was contributed by PC1 (37.25%) with an eigen value of 2.61. These principal components confirmed the presence of genetic variation among the tested rice genotypes. With a similar fashion, [6] reported the first three principal components with eigen values above 1unit and accounted for 65.79% of the total variation among 39 evaluated rice genotypes.

It has been also explained that 64% % of the total variability using the first three principal components among 64 rice genotypes using seven morphological traits [10]. Unlike this report, [4] reported four principal components with eigen values of 6.01, 1.88, 1.77 and 1.54 for PC1, PC2, PC3 and PC4 respectively which were accounted for 80.04% of the total variation, and two principal components (PC1 and PC2) for 56% of the total phenotypic and morphological variation was also reported by [15] while evaluatig a collection of rice genotypes.

Taking the eigen vector values of principal component analysis results for seven quantitative traits of the tested genotypes (Table 2) into consideration, it had been confirmed that PH, PL and TSW contributes positively to PC1 with a contributing coefficient of 0.46, 0.25 and 0.48 respectively whereas DH, DM, NFG and NUFG were a negative contributor. Since PL and TSW were directly related to yield, an attention could be given for these quantitative traits which were contributing positively for the variability of this component. The variability on PC3 was positively contributed by NFG and NUFG but negatively contributed by the other five quantitative traits. [3] also discussed that the variability on PC1 was positively contributed by PH, PL and TSW, and negatively contributed by DH and DM but positively contributed to the variability on PC2. Unlike this report the author reported a positive contribution of DH and DM to the observed variability on PC3.

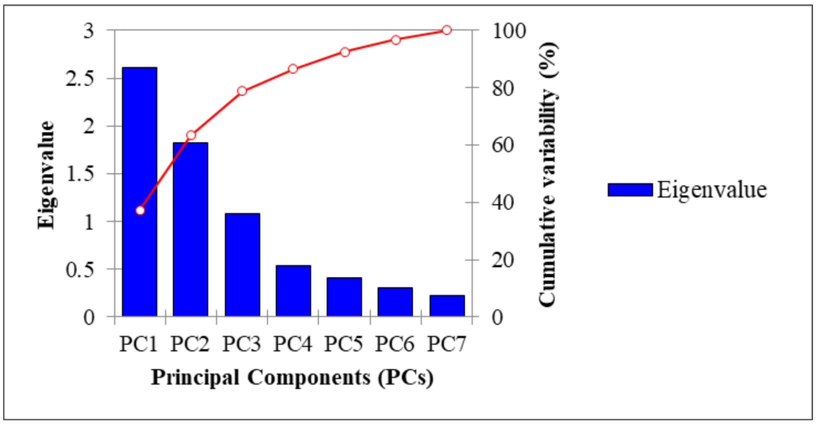

Figure 2 Scree plot for Eigen values and variance explained by the principal component analysis of seven quantitative traits of 321 lowland rice genotypes

Table 2 Principal component analysis and the eigen vectors of seven quantitative traits for 321 lowland rice genotypes

Variables	PC1	PC2	PC3
DH	-0.51	0.12	-0.39
DM	-0.47	0.09	-0.45
PH	0.46	0.26	-0.31
PL	0.25	0.50	-0.37
NFG	-0.13	0.65	0.08
NUFG	-0.07	0.49	0.58
TSW	0.48	-0.05	-0.27

Where: DH-days to heading, DM-days to maturity PH-plant height, PL-panicle length, NFG-number of filled grains per panicle, NUFG-number of unfilled grains per panicle; TSW-1000 seed weight, PC-principal component

The principal component loadings for the quantitative traits of the tested rice genotypes (Table 3), showed that the total variability of PC1 was highly and positively loaded by plant height (0.74), panicle length (0.41) and 1000 seed weight (0.77). The variability of PC2 was highly loaded by number of filled grains per panicle (0.88) and panicle length (0.67) on which the traits contributed highly and positively on its total variability. Number of unfilled grains per panicle (0.60) had a high loading on PC3.

Similarly, [2] reported a positive loading of plant height (0.33) and panicle length (0.37) for the cumulative variance of

PC1(29.72%), days to heading (0.59) for the cumulative variance of PC2 (50.63%) and a negative loading of days to heading (-0.02), plant height (-0.42) and panicle length (-0.35). In contrast with this study the author also recorded a positive loading of days to heading (0.15) and days to maturity (0.21) for PC1 and a negative loading of plant height (-0.24) and panicle length (-0.11) on PC2, and 1000 seed weight on PC1. In addition, [12] also reported a PC1 positively loaded on plant height and grain weight, and a PC2 positively loaded on number of unfilled grains per panicle.

Taking the results of the trait loadings into consideration, a great attention could be given for the genotypes included in PC1 and PC2. But genotypes laid on PC3 may not be considered due to the predicted yield reduction because of a high loading by the number of unfilled grains, the least contribution of number of filled grains and a negative contribution of 1000 seed weight on its total variability (15.54%).

Table 3 Principal component loadings of seven quantitative traits of 321 rice genotypes

Variables	PC1	PC2	PC3
DH	-0.82	0.16	-0.41
DM	-0.76	0.12	-0.47
PH	0.74	0.35	-0.32
PL	0.41	0.67	-0.38
NFG	-0.21	0.88	0.08
NUFG	-0.11	0.66	0.60
TSW	0.77	-0.07	-0.29

Where: DH-days to heading, DM-days to maturity, PH-plant heigh, PL-panicle length, NFG-number of filled grains per panicle, NUFG-number of unfilled grains per panicle, TSW-1000 seed weight, PC-principal component.

The correlation circle (Figure 3) formed with the narrow angles by the vector lengths of days to heading (DH) and days to maturity (DM) confirmed that these quantitative traits were positively correlated with each other while correlating the first two principal components with the respective quantitative traits. The right angles formed by the vector length of number of unfilled grains (NUFG) and plant height (PH) showed that these two traits were highly unrelated to each other.

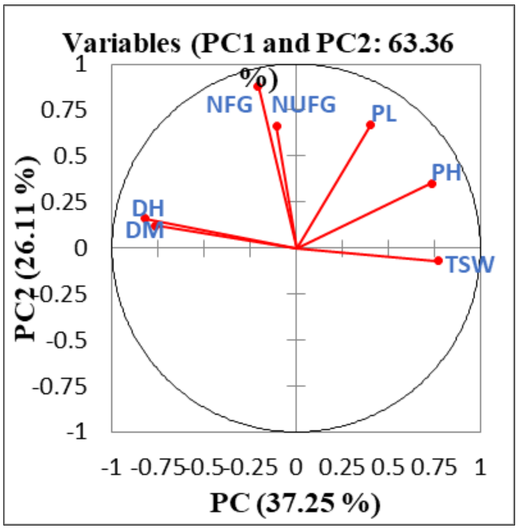

Figure 3 A correlation circle showing a correlation between seven quantitative traits and the principal components of 321 rice genotypes

Out of the seven quantitative traits evaluated among the genotypes, it has been clearly showed that days to 50% heading (DH), days to 85% maturity (DM), plant height (PH) and 1000 seed weight (TSW) brought a higher contribution to the total variability of PC1(25.76%, 22.33%, 20.72% and 22.81% respectively) whereas panicle length (PL), number of filled grains per panicle (NFG) but number of unfilled grains per panicle (NUFG) showed a lower contribution (Table 4) to its total variability.

Number of filled grains per panicle (NFG) had a highest contribution (42.26%) to the total variability of PC2 whereas 1000 seed weight (TSW) was the least contributed trait (0.27%) to this principal component. The total variability percentage of PC3 was highly contributed by number of unfilled grains per panicle (33.37%) and days to maturity (20.48%). [11] It has been also reported that plant height showed a higher contribution for the total variability of PC1 (23.35 %). Unlike this study the author reported a negative contribution of panicle length and number of filled grains per a panicle to PC2 while studding the genetic diversity of 36 upland rice genotypes in southwestern Ethiopia.

Table 4 Contribution of the quantitative traits to each principal component (%) for 321 rice genotypes

Variables	PC1	PC2	PC3
DH	25.76	1.40	15.12
DM	22.33	0.80	20.48
PH	20.72	6.73	9.49
PL	6.30	24.73	13.44
NFG	1.64	42.26	0.65
NUFG	0.44	23.81	33.37
TSW	22.81	0.27	7.45

Where: DH-days to heading, DM-days to maturity, PH-plant height, PL-panicle length, NFG-number of filled grains per panicle, NUFG-number of unfilled grains per panicle, TSW-thousand seed weight in gram, PC-principal component

Table 5 The squared cosines of the variables evaluated from 321 rice genotypes

Variables	PC1	PC2	PC3
DH	0.67	0.03	0.16
DM	0.58	0.01	0.22
PH	0.54	0.12	0.10
PL	0.16	0.45	0.15
NFG	0.04	0.77	0.01
NUFG	0.01	0.44	0.36
TSW	0.59	0.01	0.08

Where: DH-days to heading, DM-days to maturity, PH-plant height, PL-panicle length, NFG-number of filled grains per panicle, NUFG-number of unfilled grains per panicle, TSW-1000 seed weight and PC-principal component

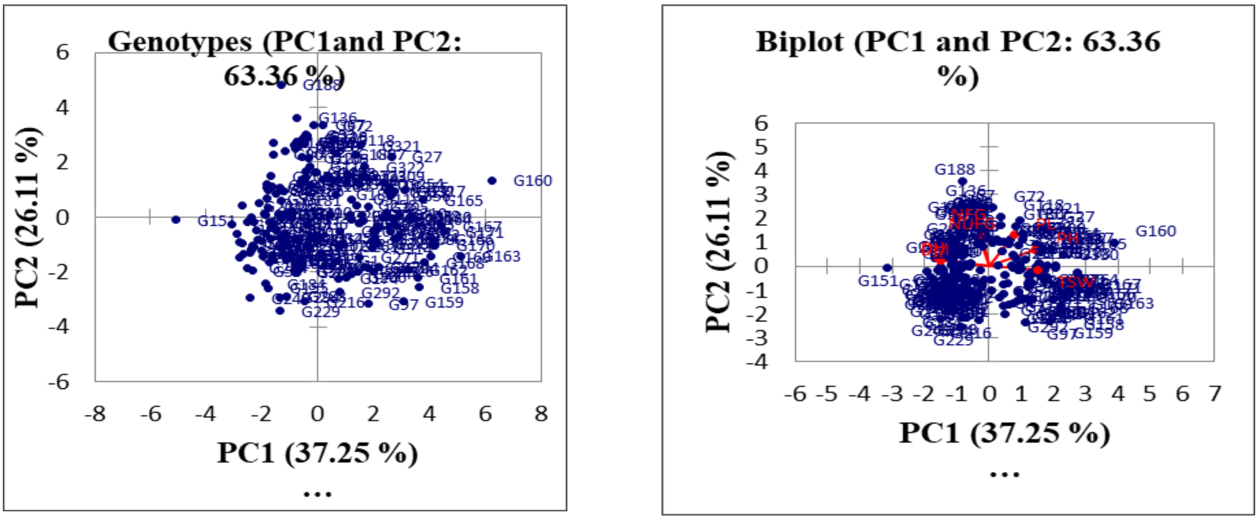

Figure 4 Distribution plot of 321 lowland rice genotypes across the first two components (A) and the corresponding seven measured quantitative traits (biplot) across the first two principal components (B) based on PCA

To clearly identify that which quantitative trait is well linked with each principal component, the squared cosine of each trait was considered (Table 5). The greater the squared cosine, the greater the link with the corresponding principal components [17]. Hence, days to heading (DH), days to maturity (DM), plant height (PH) and 1000 seed weight (TSW) were well linked with PC1 whereas panicle length (PL), number of filled grains per a panicle (NFG) and number of unfilled grains per a panicle (NUFG) were highly linked with PC2, and PC3 had relatively a higher linkage with days to maturity (DM) and number of unfilled grains per a panicle (NUFG).

The distribution pattern of 321 rice genotypes explained by seven different quantitative traits (Figure 4-A and 4-B) confirmed that the genotypes were distributed in different coordinates, which clearly showed the presence of a genetic diversity among the tested rice genotypes.

4. Conclusion

The statistical result that the tested rice genotypes were grouped into different quadrants while computing the distribution plot confirms the presence of a genetic diversity among the genotypes and the cumulative variability of the first three significant principal components (PC1, PC2 and PC3) is mainly contributed by the number of filled grains per a central panicle. The total variability of PC1 is highly contributed by Days to 50% heading, days to 85% maturity, plant height, and 1000 seed weight whereas the total variability of PC2 is contributed by number of filled grains a central panicle, panicle length, and number of unfilled grains per a central panicle. It is also identified that the total variability percentage of PC3 is highly contributed by number of unfilled grains per a central panicle and days to 85% maturity while predicting the presence of genetic diversity among the tested rice genotypes. Generally, the total variability of each significant principal components contributed by different quantitative traits predicts the presence of a genetic diversity among the tested rice genotypes. Thus, it is recommended that an attention should be given for those genotypes with a greater value of number of filled grains per panicle, 1000 seed weight but with a lower number of unfilled grains per panicle while selecting genotypes for improving the desired traits of the genotypes to be implemented in different rice breeding research program.

Compliance with ethical standards

Acknowledgments

A gratitude will be given for the researchers who are working as a rice breeder at Pawe agricultural research center. I wish to acknowledge a technical staff of the center for their help in collecting the phenotypic data and managing the overall activities of this trial. My gratitude also goes to Ethiopian institute of agricultural research for its full financial support.

Disclosure of conflict of interest

No conflict of interest among the authors.

References

[1] Mbe OJ, Saraswathi R, Jeyaprakash PJI. Valuation of Genetic Diversity for Quantotatove traits in Rice (Oryza sativa L) Genotypes. *Nigerian Agricultural Journal*. 2019; 50(1): 200-207.

[2] Astewel Takele. Determinants of rice production and marketing in low producer farmers: the Case of Fogera district, north-western Ethiopia. *International journal of environment, agriculture and biotechnology*. 2017; 2(5): 2534-2545.

[3] CSA. The Federal Democratic Republic of Ethiopian Central Statistical Agency Agricultural Sample Survey 2020/2021. *Statistical report*, The Federal Democratic Republic of Ethiopian Central Statistical Agency, Addis Ababa. 2021.

[4] Chakma SP, Huq H, Mahmud F, Husna A. Genetic Diversity Analysis in Rice (Oryza sativa L.). *Bangladesh J. Pl. Breed. Genet*. 2012; 25(1): 31-39.

[5] Shaibu AA, Uguru MI. Application of Multivariate Analysis to Access Selected Rice Germplasm Phenotypic Diversity. *Journal of Experimental Agriculture International*. 2017; 16(1): 1-8.

[6] Atsedemariyam T, Wassu M, Alemayehu A. Genetic Diversity Analysis of Upland Rice (Oryza Sativa L.) Genotypes in Benshangul Gumuz, Ethiopia. *International Journal of Research Studies in Agricultural Sciences (IJRSAS)*. 2019; 5(10): 22-28.

[7] Dawit A, John T. The emerging importance of rice as a strategic crop in Ethiopia. 2020

[8] Mudasir H, Zahoor A, Sher A. Breeding Strategies for Improving Rice—A Review. *Agricultural Sciences*. 2015; 6(5): 467-478.

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

[10] Bhattarai U, Subudhi PK. Genetic Diversity, Population Structure and Marker-Trait Association for Drought Tolerance in US Rice Germplasm. *Plants*. 2019; 8(12): 530

[11] Gedifew G, Tsige G. Morphological Characterization and Evaluation of Sorghum [Sorghum bicolor (L.) Moench] Landraces in Benishangul Gumuz, North-western Ethiopia. *Greener Journal of Agricultural Sciences*. 2019; 9(1): 37-56.

[12] Altaye T, Wosene G, Abush T. Genetic diversity study on upland rice (Oryza sativa L.) genotypes based on morphological traits in southwestern Ethiopia. Asian journal of crop science. 2019; 11(1): 17-24.

[13] Alamir A, Tiegist D, Fisseha W. Genetic divergence analyses of lowland rice genotypes in North Western Ethiopia. *African Journal of Plant Science*. 2020; 14(4), 165-171.

[14] Mahendran R, Veerabadhiran P, Robin S, Raveendran M. Principal component analysis of rice germplasm accessions under high temprature stress. *International Journal of Agricultural Science and Research (IJASR).* 2015; 5(3): 355-360.

[15] Girma BT, Kitil MA, Banje DG, Biru HM, Serbessa TB. Genetic Variability Study of Yield and Yield Related Traits in Rice (Oryza sativa L.) Genotypes. *Adv Crop Sci Tech*. 2018; (6): 381.

[16] Syafii M, Cartika, Ruswandi. Multivariate Analysis of Genetic Diversity among some Maize Genotypes under Maize-Albizia Cropping System in Indonesia. *Asian Journal of Crop Science*. 2015; 7(4): 244-255.