Journal of Tropical Crop Science Vol. 9 No. 3, October 2022
www.j-tropical-crops.com

RESEARCH ARTICLE

Morphological and Genetic Diversity Study of Upland Rice Varieties under Rainfed Environment

Bantalem Zeleke[A]*, Fisseha Worede[B]

[A] Amhara Regional Agricultural Research Institute, Adet Agricultural Research Center, Bahir Dar, Ethiopia
[B] Ethiopian Institute of Agricultural Research, Fogera Rice Research and Training Center, Bahir Dar, Ethiopia

*Corresponding author; email: bantalem2006@gmail.com

Abstract

A field experiment was conducted at Fogera Northwest Amhara region to study the morphological traits to variability in 20 upland rice varieties, consisting of nine NERICA and eleven parents. The data were collected from ten randomly selected plants of each plot (plant height, panicle length, culm length, flag-leaf length, number of spikelet per panicle, number of grains per panicle, number of filled grains per panicle, numbers of fertile tillers per plant, yield per plant) and from plot bases (days to heading, days to maturity, grain-filling period, thousand-seed weight, biomass yield, grain yield, and harvest index). The results of the principal component analysis showed that four components account for 76.7% of the total variation, giving a clear idea of the structure underlying the variables analysed. Cluster analysis using un-weighted Pair Group Method using Arithmetic Average linkage (UWPGMA) classified the twenty varieties into five distinct groups. The maximum inter-cluster distances were; recorded 8.05 between cluster I & V, 6.67 between cluster I and IV; and 5.5 between Cluster I and III, indicating that the possibility of high heterosis if individuals from these clusters are cross bred. The results of the principal component analysis were closely in line with those of the cluster analysis. This study has provided useful information, on evaluation of genetic diversity of rice varieties and will indicate the way, how plant breeders screen out large populations and to develop new breeding protocols for rice improvement.

Keywords: *Oryza sativa,* principal component, clustering, component traits, varieties

Introduction

Rice is one of the important food crops in Ethiopia, and its importance has led to the development of the ten-year National Rice Strategy Plan (2020 – 2030). Rice was introduced in the 1970s to Ethiopia where its breeding and other research components are still at the initial stages (Asmelash, 2012). The current rice production area in Ethiopia is 58,000 ha with a total yield of 184,000 tons (EARI, 2015).

A scarcity of water for irrigation, and a breakdown of resistance genes against emerging races of the pathogen due to intensive cultivation have threatened rice production. Upland rice comprises 11% of the total global rice production and is cultivated on about 14 million hectares (Sohrabi et al., 2012). Rice is crop that recently introduced to Ethiopia, but it is increasingly important, indicated by the expansion of its growing area from about 10,000 ha in 2006 to over 50,000 ha in 2018. In addition, the number of growers engaged in rice production has grown year after year (Tadesse, 2020).

NERICA is the product of interspecific hybridization between the cultivated rice species of Africa (*O. glaberrima*) and Asia (*Oryza sativa*). However, some scientists suggested that introgression between the two rice species occurs in the field (Second, 1982). Some experimental studies showed that introgression from *O. glaberrima* to *O. sativa* are possible, although at a low frequency (Yabuno, 1977). Artificial backcrosses produced fertile progenies which resembled the parental phenotypes, indicating that under natural conditions it will be difficult to detect hybrid derivatives (Sano et al., 1980). This means that, for example, plants belonging to *O. glaberrima* can incorporate *O. sativa* genetic material but remain typically *O. glaberrima* to the eye. NERICA varieties have high yield potential and shorter growth cycle compared to traditional rice varieties. Several NERICA varieties have vigorous vegetative growth, which is a useful trait for competitiveness against weeds, and the grains have higher protein content and amino acid balance than most of the imported rice varieties (WARDA, 2008). Continuous evaluation of rice germplasm should be conducted to broaden the genetic base of the species, particularly those that control a particular trait for use in crop improvement. Genetic variability in crop species has potentials to develop new rice varieties that are more resilient to biotic and abiotic stresses (Gana, 2006). Our study was conducted to assess the extent of morphological variation in upland rice varieties, to cluster the rice varieties into relatively homogenous groups, and to identify the major characters contributing to the overall diversity of the varieties.

Materials and Methods

The study was undertaken at Fogera National Rice Research and Training Center, Ethiopia. The experimental site is located at 11°58'N latitude, 37°41'E longitude and at an elevation of 1810 m above sea level. Based on ten years' average meteorological data, the annual rainfall is 1300 mm, and the mean annual minimum and maximum temperatures are 11.5° C and 27.9°C, respectively. The soil type is Vertisol with a pH of 5.90.

Experimental Materials

Experimental materials were 20 upland rice varieties released by Ethiopian research centers from 1998-2016 (Table 1). Eight of the varieties were NERICA (New Rice for Africa) types initially developed for the upland ecosystems of Africa Rice. Genetic transformation of NERICA, interspecific hybrid rice between *O. glaberrima* and *O. sativa*, developed and mediated by Agrobacterium tumefaciens (Ishizaki and Kumashiro, 2008).

Experimental Design and Layout

A field experiment was conducted using 20 released upland rice varieties from Fogera National Rice Research and Training Center during the 2017 main cropping season. Randomized complete block design with three replications in 14m x 39.5m total area was used. Each experimental plot had a total area of 6m2 (1.5m x 4m) and six rows at 0.25 m interval, while the distance between plots and between blocks were 0.5m and 1m, respectively. Seeds have been sown in rows with manual drilling at a rate of 60 kg.ha^{-1}. Fertilizer application was at a rate of 60.5 kg NPS and 125 kg urea per hectare (Zeleke, 2017). All NPS were applied at planting whereas urea application was in three splits at planting, tillering, and panicle initiation stage.

Data Collection

Observation and data recording for the traits were based on the standard evaluation system for rice (IRRI, 2013). The data were collected from ten randomly selected plants of each plot for traits treated on plant basis like plant height (cm), panicle length (cm), number of panicles per plant, culm length (CL), flag leaf length (FL), number of total grains per panicle, number of filled grains per panicle, number of fertile tillers per plant, yield per plant (g). Days to heading, days to maturity, grain-filling period, thousand-seed weight (g), biological yield (t. ha^{-1}), harvest index (HI), and grain yield (t.ha^{-1}) were taken on plot basis; the four central rows were considered. Grain yield was adjusted at 14% moisture level.

Statistical Analysis

Correlation coefficient analysis was done using Statistical Analysis System (SAS) version 9.4 Computer software program following SAS statement (syntax) for randomized complete block design (SAS, 2013). Path coefficient analysis was done using MS-Excel.

Principal Component Analysis

The principal component analysis was carried out using Statistical Analysis System Version 9.4 (SAS, 2013). The principal component analysis used to identify few characters that plays prominent role in classifying the variation existing in the varieties / genotypes (Nachimuthu et al., 2014).

Cluster Analysis

Divergence analysis is performed using Mahalanobis (1936) D2 distance to classify the diverse genotypes for hybridization purpose. Clustering was conducted using the proc-cluster command of SAS version 9.4 software (SAS, 2013) with using un-weighted Pair Group Method using Arithmetic Average linkage (UWPGMA). The values of pseudo F statistic (PSF) and Hotellin's (1931) pseudo T2 statistic were used for defining optimum number of clusters.

Results and Discussion

Cluster Analysis

Clustering analysis is a multivariate statistical analysis technique involving partitioning a set of varieties into groups so that varieties within a group are more similar and varieties in different groups are more dissimilar. Cluster Analysis depending upon the

Table 1. List of upland rice varieties from Fogera National Rice Research and Training Center used in the experiment

No	Varieties	Year of release	Yield (t/ha)		Breeder/maintainer
			On- farm	On- station	
1	Adet [WAB450-1-B-P-462-HB]	2014	2.4	4.2	ADARC/ARARI
2	Andassa [AD-012]	2007	2.5	3.8	ADARC/ARARI
3	Chewaqa [YIN lu20]	2013	3.3	4.2	BARC/OARI
4	FOFIFA-3737	2010	5.0	6.5	GOPARC/SORPARI
5	Fogera-1	2016	3.2-3.9	4.2-5.0	FNRRTC
6	Getachew [AD-01]	2007	2.1	3.0	ADARC/ARARI
7	Hidassie [WAB515-B-16A1-2]	2012	2.2-3.2	3.0-4.2	ADARC
8	Kokit [IRAT-209]	2000	2.8	3.6	ADARC/ARARI
9	NERICA-2	2007	3.5	5.5	GOPARC/SORPARI
10	NERICA-3 [WAB-450-IB-P-2B-HB]	2006	2.9	4.5	PARC/EIAR
11	NERICA-4 [WAB-450-IB-P-9/1]	2006	3.0	4.8	PARC/EIAR
12	NERICA-6	2011	5.5	6.3	GOPARC/SORPARI
13	NERICA-10	2013			Only registered
14	NERICA-12[WAB880-1-38-20-17-P1-HB]	2013	2.3-3.4	3.5-4.1	ADARC
15	NERICA-13 [Maytsebri-1]	2014	3.3	3.8	MARC/TARI
16	NERICA-14	2010	5.0	6.2	GOPARC/SORPARI
17	NERICA-15	2011	5.0	6.2	GOPARC/SORPARI
18	Pawe-1 [M-55]	1998	2.0	3.0	PARC
19	Suprica-1 [WAB-450]	2006	2.3	5.1	PARC/EIAR
20	Tana [AD-048]	2007	2.4	4.4	ADARC/ARARI

range of diversity, 20 genotypes were grouped into five clusters (Table 2, Figure 1). Significant differences among the genotypes for all traits suggested the presence of variation among the genotypes for all traits under study. The distribution pattern revealed the maximum number of genotypes (11 genotypes) in cluster I whilst cluster II and V included the minimum number of genotypes (1 genotype). Clusters III and IV included 4 and 3 genotypes, respectively. The limited number of genotypes in the clusters was probably due to the high correlation among most of the traits and the duplication effect of the traits included in this study. The inter-cluster distances in all cases were greater than the intra-cluster distances, suggesting wider diversity among the genotypes of the distant groups (Table 3).

The highest inter cluster distances were 8.05 between clusters I & V, 6.67 between cluster I and IV; and 5.5 between cluster I and III. The lowest inter cluster distances were 0.27 between cluster II and III; 1.17 between cluster III and IV; and 1.36 between cluster IV and V. The results of inter cluster analysis revealed that the varieties had wide genetic divergence with

Table 2. Distribution of 20 upland rice varieties over five clusters based on 16 quantitative traits

Cluster	Number of varieties	Proportion (%)	Name of rice varieties
I	11	55	Fogera-1, Adet, Fofifa 3737, Hidasse, NERICA-2, NERICA-3, NERICA-4, NERICA-6, NERICA-12, NERICA-13, Superica-1
II	1	5	NERICA-10
III	4	20	Andassa, Chewaqa, Getachew, Tana
IV	3	15	NERICA-14, NERICA-15, Kokit
V	1	5	Pawe

Table 3. Intra- and Inter-cluster distance (D^2) values among five clusters in 20 upland rice varieties.

Cluster	I	II	III	IV	V
I	1.52				
II	5.23				
III	5.5	0.27	2.53		
IV	6.67	1.44	1.17	3.6	
V	8.05	2.82	2.55	1.36	

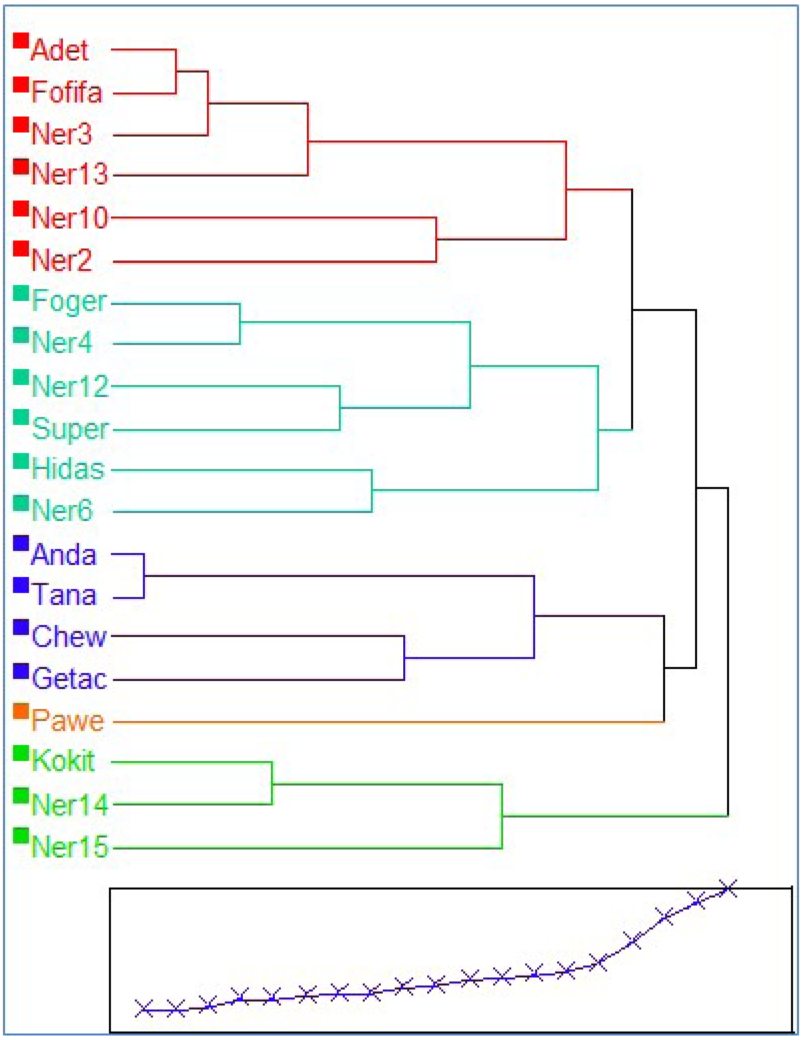

Figure 1. Cluster diagram showing the average intra-and inter cluster distances (D= $\sqrt{D^2}$ values) among 20 upland rice varieties according to Ward's method.

each other indicating existence of high probability for recombination.

Mean performance of the different clusters for the different morphological traits (Table 4) reflected that all early flowering (1st, 50%, and 80% flowering) varieties were grouped into cluster II, whereas cluster V included all late flowering genotypes. Short duration (108.88 days) genotypes were grouped into cluster I, whereas cluster V included long (126.0 days) duration varieties indicating the maximum contribution of this character towards the divergence between clusters I and V. Again all high yielding genotypes with high number of total tillers per hill, high number of effective tillers per hill, maximum panicle length, and high number of filled grain per panicle were grouped into cluster I, whereas cluster V included low yielding genotypes with less number of total tiller per hill, less number of effective tillers per hill, less panicle length, less number of filled grain per panicle indicating the maximum contribution of these traits towards the divergence between cluster I and V. Cluster I was divergent from cluster III mainly due to total tiller per hill, effective tiller per hill, and number of filled grain per panicle indicating maximum contribution of these traits towards the divergence. Cluster II was divergent from cluster IV mainly due to day to heading, day to maturity, grain filling period, Number of grains per panicle, plant height, number of fertile grains per panicle, plant height, yield per plant, culm length, flag leaf length, number of fertile grains per panicle, and harvest index indicating the maximum contribution of these traits towards the divergence. Cluster III was divergent from cluster V mainly due to days to heading , days to maturity, grain filling period, plant height, panicle length, culm length, number of grains per panicle, number of fertile grains per panicle, thousand seed weight, yield per plant, biomass yield, grain yield and harvest index indicating the maximum contribution of these traits towards the divergence.

Principal Component Analysis (PCA)

The principal component analysis used to identify few characters that plays prominent role in classifying the variation existing in the varieties. The principal component analysis resulted in a total of 16 principal component axes (PCAs), equal to the total number of traits used in the analysis. Nonetheless, only four important PCAs, with an eigenvalue (5.13, 3.49, 2.17 and 1.48) greater than one, were retained (Table 6). Data presented in Table 5 demonstrated that an increase in the number of components was associated with a decrease in eigenvalues. The Eigen values are used to determine how many factors to retain and the sum is usually equal to the number of variables. The first step in PCA was to calculate Eigen values, which all together explained total variability that is displayed on the principal component (PC) axes. The principal components (PCs) with eigenvalue > 1.0 are used as criteria to determine the number of PCs. According to the results, the four PCAs with eigen values of 5.13, 3.49, 2.17, and 1.48, respectively, all together explained 76.66% of the total variability, leaving the remaining 23.34% in the last twelve principal components (Table 5). According to Worede et al. (2014), in their analysis a total of 17 principal components which equal to the number of traits were extracted. The result of the first three principal components with eigen values greater than 1 were 73.5 % of the total variance.

Characters with higher loadings (positive or negative) contribute more to the variability and they are the ones that most differentiate the varieties. Hence, the data presented in Table 5 and graphically shown in Figure 2 showed the most contributing characters in the first principal component were yield per plant,

Table 4. Vegetative growth parameter of five clusters of 20 upland rice varieties

Cluster	DH	DM	GFP	PH	PL	CL	FL
I	91.5	145.1	53.6	83.6	18.6	64.9	21.5
II	79.7	146.0	66.3	70.7	18.2	52.8	18.1
III	105.4	146.6	41.2	103.7	19.5	84.2	18.9
IV	88.1	123.1	35.0	77.0	18.4	58.5	20.0
V	122.7	160.0	37.3	71.4	12.1	58.6	21.0

Note: DH=days to heading, DM=days to maturity, GFP=grain filling period, H=plant height, PL= panicle length, CL= culm length, FL= flag leaf length

Table 5. Generative growth parameter of five clusters of 20 upland rice varieties for generative parameters.

Cluster	NSP	NGP	NFGP	NFTP	TSW	YP	BY	GY	HI
I	10.5	93.1	84.6	7.9	28.3	15.4	8.1	4.4	54.9
II	9.1	80.8	58.1	8.1	26.0	9.0	6.6	3.3	50.5
III	9.7	82.5	73.6	7.8	27.6	12.7	10.3	4.6	45.3
IV	9.1	58.1	53.7	6.6	27.0	7.3	5.7	3.2	55.9
V	9.8	62.7	57.0	7.4	35.8	9.0	10.0	2.8	27.5

Note: NFTP=number of productive tillers per plant, NSP=number of spikes per panicle, NGP=number of grains per panicle, NFGP= number of fertile grains per panicle, TSW= thousand seed weight, YP= yield per plant, BY= biomass yield, GY= grain yield and HI=harvest index.

number of filled grains per panicle, number of grains per panicle and grain yield,; whereas in the second PC were days to heading, harvest index, biomass yield and grain-filling period; in the third PC were panicle length, plant height, culm length, thousand-seed weight and date of maturity; and in the fourth PC were flag leaf length, number of spike per panicle and number of fertile tillers per plant were the major contributing characters for variability to those principal components.

The factor loadings refer to the coefficients in each principal component, or the correlation between the component and the variables. A high correlation between PC1 and a variable indicates that the variable is associated with the direction of the maximum amount of variation in the data set. The components and their contributions to the variables are graphically shown in Figures 1 and 2. The present study confirmed that the upland rice varieties showed wide variations in the character studied and it suggests ample opportunities for conservation of the germplasm for future utilization. Similar findings of grouping rice genotypes by principal component analysis were reported (Worede et al., 2014).

The principal component analysis was carried out using Statistical Analysis System Version 9.4 (SAS, 2013). The results of the PCA explained the genetic diversity of the rice varieties (Table 6). 'Proper values' measure the importance and contribution of each component 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 associated. The higher the coefficients, regardless of the direction (positive or negative), the more effective they will be in discriminating between accessions. There are no standard tests to prove the significance of proper values and coefficients. In this study, the criterion corroborated by Guei et al. (2005) was chosen, which suggested that the first four principal components are often the most important in reflecting the variation patterns among genotypes, and the traits associated with these are more useful in differentiating genotypes. According to this criterion, the first four components account for 76.7% of the total variation (Table 6), giving a clear idea of the structure underlying the variables analysed. This result was

Table 6. Principal components (PCs) for morphological traits of 20 upland rice varieties under rainfed environment

Characters	PCA1	PCA2	PCA3	PCA4
BY	0.27	0.32	-0.06	-0.12
CL	0.22	0.32	0.37	-0.06
DH	0.10	0.45	-0.03	0.20
DM	0.28	0.19	-0.31	0.08
FL	0.12	-0.20	-0.03	0.60
GFP	0.18	-0.29	-0.27	-0.13
GY	0.35	-0.06	0.10	-0.28
HI	0.01	-0.40	0.20	-0.08
NSP	0.20	-0.07	0.08	0.56
NGP	0.38	-0.19	-0.03	0.10
NFGP	0.39	-0.15	-0.04	0.07
NFTP	0.24	-0.09	-0.24	-0.36
PH	0.23	0.27	0.43	-0.04
TSW	0.05	0.27	-0.36	0.04
YP	0.39	-0.12	-0.08	-0.05
Eigen value	5.13	3.49	2.17	1.48
Proportion (%)	32.05	21.83	13.53	9.25
Cumulative (%)	32.05	53.88	67.41	76.66

Note: Vector loading and percentage of explained variation in the first four principal components. PCA=principal component axis, BY=biomass yield, CL= culm length, DH=days to heading, DM= days to maturity, FL= flag leaf length, GFP=grain filling period, GY= grain yield, HI=harvest, NFTP= number of fertile tillers per plant, NSP=number of spikelet per panicle, NGP= number of grains per panicle, NFGP= number of fertile grains per panicle, PH= plant height, PL = panicle length, TSW= a thousand seed weight, YP= yield per plant.

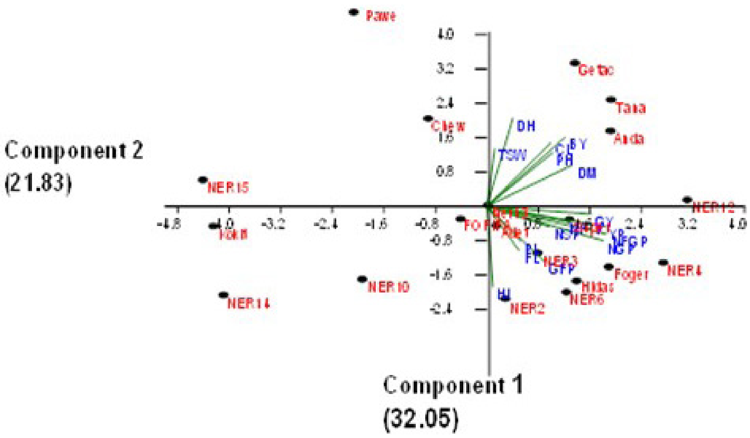

Figure 2. Scatter plot of principal component analysis of 20 rice varieties between their Eigen values and the number of principal components

also supported by the screen test (Figure 2) indicating that the consideration of the first four components which account for 76.7% of the total variation. The first component differentiated those varieties that flowered and mature earlier in the season and identified mainly phonological variables presenting negative contributions. Guei et al. (2005) and Sanni et al. (2010) reported similar results for days to 50% flowering and days to maturity of rice. The second principal component differentiated the high yielding varieties (Table 6). The third principal component differentiated the varieties with good architecture. The fourth principal component accounted for more than 9% of the total variance and associated with flag leaf length, number of fertile tillers per plant, and number of spikelets per panicle. These findings agree with Caldo et al. (1996), who reported that maturity, heading, plant height, culm length, leaf length, and tillering ability were the major factors contributing to the variation of parental lines of modern Philippine rice cultivars.

Conclusion

The matrix of principal components is the product of the eigenvector matrix with the matrix of independent variables. The first principal component accounts for the largest per cent of the total data variation. The second principal component accounts the second largest per cent of the total data variation. The rice varieties showed considerable variability for most of the characteristics that could be used for crop improvement. Cluster analysis based on D^2 values and principal component analysis suggested that 20 upland rice varieties could be grouped into various clusters. Such groupings are useful to breeders in identifying possible varieties that may be used as a parent in a future breeding program. This study has provided useful information on the genetic diversity of 20 upland rice varieties, which will provide directions for plant breeders to screen large populations for potential breeding stock.

References

Asmelash, Y. (2012). Farmers' Evaluation of Upland Rice Varieties in Fogera District, South Gondar, Ethiopia. *Asian Journal of Agriculture and Rural Development* **2**, 604 – 608**.**

Caldo, R. A., Sebastian, L.S., and Hernandez, J.E. (1996). Morphology-based genetic diversity analysis of ancestral lines of rice in Philippine rice cultivars. *Philippines Journal of Crop Science* **21,** 86–92.

EARI. (2015). "A Host of Achievements in Improving Plant and Animal Productivity in the Five Years of the Growth & Transformation Plan (GTP)". The Ethiopian Agricultural Research Institute (EARI).

Gana, A.S., Shaba, S.Z., and Tsado, E.K. (2006). Principle 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* **5**, 120-126.

Guei, R.G., Sanni, K.A., Abamu, F.J., and Fawole, I. (2005). Genetic diversity of rice (*Oryza sativa* L.). *Journal of Agronomy* **5,** 17–28.

Hotelling, H. (1931). The generalization of Student's ratio. *Annals of Mathematical Statistics* **2**, 360–378.

Ishizaki, T., and Kumashiro, T. (2008). Genetic transformation of NERICA, interspecific hybrid rice between *Oryza glaberrima* and *Oryza sativa* mediated by *Agrobacterium tumefaciens*. *Plant Cell Report* **27**, 319–327.

Mahalanobis, P.C. (1936). On the generalized distance in Statistics. *Journal of Genetics* **41**, 159-193.

Nachimuthu, V. V., Robin, S., Sudhakar, D., Rajeswari, S., Raveendran, M., Subramanian, K. S., Tannidi, S., and Pandian, B. A. (2014). Genotypic variation for micronutrient content in traditional and improved rice lines and its role in biofortification programme. *Indian Journal of Science and Technology* **7**, 1418–1429.

Sanni, K.A., Fawole, I., Ogunbayo, D., Tia, E.A., Somado, K., Futakuchi, M., Sié, F.E., Nwilene and Guei, R.G. (2010). Multivariate analysis of diversity of landrace rice germplasm. *In* “Innovation and Partnerships to Realize Africa’s Rice Potential”. Second Africa Rice Congress, Bamako, Mali. pp. 12-15.

Sano, Y., Chu, Y.E., and Oka, H.I. (1980). Genetic studies of speciation in cultivated rice. 2. Character variations in backcross derivatives between *Oryza sativa* and *O. glaberrima*: M-V linkage and key characters. *Japanese Journal of Genetics* **55**, 19–39.

SAS Institute Inc. (2013). SAS® 9.4 Statements: Reference. SAS Institute Inc. North Carolina

Second, G. (1982). Origin of the genic diversity of cultivated rice (*Oryza spp.*): study of the polymorphism scored at 40 isozyme loci. *Japanese Journal of Genetics* **57**, 25–57.

Sohrabi, M., Rafii, M.Y., Hanafi, M.M., Akmar, A. S. N., and Latif, M.A. (2012). Genetic diversity of upland rice germplasm in Malaysia based on quantitative traits. *The Scientific World Journal*, **9**.

Tadesse T. (2020). History, current status and future directions of rice research in Ethiopia. Fogera national rice research and training center. *Journal of Emerging Technologies and Innovative Research* **7**, 2.

WARDA (West Africa Rice Development Association). (2008). “Rice Interspecific Hybridization Project: Research Highlights”.Bouaké, Côted’Ivoire, p. 36.

Worede, F., Sreewongchai, T., Phumichai, C., and Sripichitt, P. (2014). Multivariate analysis of genetic diversity among some rice genotypes using morpho-agronomic traits. *Journal of Plant Sciences* **9**,14-24.

Yabuno, T. (1977). Genetic studies on the interspecific cytoplasm subsitution lines of japonica varieties of *Oryza sativa L*. and *O. glaberrima* Stued. *Euphytica* **26**, 451–463.

Zeleke, B. (2017). “Genetic Divergence and Genetic Gain for Yield and Yield Related Traits in Upland Rice Varieties at Fogera, Northwest Ethiopia. [Thesis]. Bahir Dar University, College of Agriculture and Environmental Sciences, Bahir Dar, Ethiopia. 74 pp.