Journal of Experimental Agriculture International
16(1): 1-8, 2017; Article no.JEAI.31662
Previously known as American Journal of Experimental Agriculture
ISSN: 2231-0606

SCIENCEDOMAIN *international*
www.sciencedomain.org

Application of Multivariate Analysis to Access Selected Rice Germplasm Phenotypic Diversity

A. A. Shaibu[1,2*] and M. I. Uguru[1]

[1]Department of Crop Science, Faculty of Agriculture, University of Nigeria, P.M.B. 50, Nsukka, Nigeria.
[2]Rice Research Division, National Cereals Research Institute, P.M.B. 8, Badeggi, Niger State, Nigeria.

Authors' contributions

This work was carried out in collaboration with both authors. Author AAS designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Author MIU supervised the research and managed the analyses of the study. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JEAI/2017/31662
Editor(s):
(1) Anita Biesiada, Department of Horticulture, Wroclaw University of Environmental and Life Sciences, Poland.
(2) Chrysanthi Charatsari, Department of Agricultural Economics, School of Agriculture, Aristotle University of Thessaloniki, Greece.
(3) Daniele De Wrachien, State University of Milan, Italy.
Reviewers:
(1) Edward Tsinigo, Innovations for Poverty Action, Ghana.
(2) E. F. El-Hashash, Al-Azhar University, Cairo, Egypt.
(3) Aamir Raina, Aligarh Muslim University, Aligarh, India.
(4) Oluwatoyin Sunday Osekita, Adekunle Ajasin University, Ondo State, Nigeria.
Complete Peer review History: http://www.sciencedomain.org/review-history/18774

Original Research Article

Received 18th January 2017
Accepted 7th April 2017
Published 25th April 2017

ABSTRACT

Without variability, it is not possible to conduct a plant breeding program. Germplasm is hence the critical first step in initiating a breeding program. Diversity in *O. glaberrima* accessions is enormous. It needs to be organized and characterized in order to facilitate its use by plant breeders. Two principal components (PRIN1) and (PRIN2) accounted for most of the variability observed in characters studied. PRIN 1 accounted for 56% of the phenotypic and morphological variation. The PRIN 1 was loaded on plant height, number of panicle, biomass wet weight, panicle wet weight and grain yield traits. PRIN 2 accounted for 23% of the variation. PRIN 2 was loaded on biomass wet weight, biomass dry weight, panicle wet weight, panicle dry weight, and harvest index traits. The

**Corresponding author: E-mail: A.Shaibu@cgiar.org;*

test for univariate statistics described individual variable to explore pattern of response to variation showed strong statistical significantly (P<0.001) on phenotypic differences in all the variables that were measured. These traits studied are the most important contributing to the overall variability. The dendogram produced grouping that defined nine distinct clusters and minimum genetic distance between clusters varies from 0 to 5. All the selected *Oryza glaberrima* accessions and the *Oryza sativa* were distributed across the nine clusters respectively.

Keywords: Principal component analysis; germplasm evaluation; genetic diversity; multivariate.

1. INTRODUCTION

Estimation of relatedness in crop species is one of the most appropriate way for parental selection and hybridization while selection of the parents is essential to enhance the potential yield increase [1]. Cluster and PCA analysis have been used in the germplasm identification and parental selection, tracing the pathway to evolution of crop species diversification and origin [2,3,4,5]. In PCA analysis, the two first principal components in most studies contributed for high variation percentage will be a useful method to find the clusters [6]. Many procedures have been used in studying of diversity in crop species, but cluster analysis have been used extensively for exploring both phenotypic and genetic diversity and grouping of plant materials with respect to the relationship of family based on their genetic material [5]. [5], reported that PCA and other techniques can be appropriately used for grouping crop genotypes into their distinct group while cluster analysis has been described an appropriate technique for determining relationships [7]. A huge number of variable are regularly measured by plant breeders of which some may not be of sufficient discriminatory power for germplasm evaluation and characterization of traits [8]. In such a situation, principal component analysis (PCA) may offer a useful insight in data set to eliminate redundancy. [9], describe that in any crop, germplasm resource not only serves as a valuable source of useful genes but also provides scope for building up a basic population of wide genetic variability. Bringing improvement over existing crop varieties is a continuous process in plant breeding. However, to achieve this objective, the breeder has to identify diverse parents having high genetic variability for combining desirable characters. Therefore, knowledge of sound genetic diversity is essential for undertaking any breeding program targeting crop improvement. [9], demonstrated in their study that geographical diversity was considered as a measure of genetic diversity but recently it is observed that genetic materials from same eco geographic origin also possess diverse genetic makeup. Thus, it is not uncommon that the genetic materials of different eco geographic origin possess similar genetic architecture. The usefulness of multivariate analysis for the study of morphologically complex individual and the degree of divergence between biological populations has been shown in different fields of research. Multivariate statistical techniques have been widely used in analysis of genetic diversity irrespective of whether it is morphological, biochemical or molecular marker-based and subsequent, classification of germplasm collections. It has been well documented in literature that principal component analysis (PCA) and cluster analysis had been shown to be very useful in selecting genotypes for breeding program that meet the objective of a plant breeder [5]. PCA has been reported to reveal patterns and eliminate redundancy in data sets [10] while morphological and physiological variations routinely occur in crop species. Cluster analysis is commonly used to study genetic diversity and for forming principal subclass for grouping accessions with similar characteristic into one homogenous category. Clustering is also used to summarize information on relationships between objects by grouping similar units so that the relationship may be easily understood for scientific knowledge and advancement. [11] described that multivariate analysis has been used frequently for genetic diversity analysis in many crops such as barley (*Hordeum vulgare* L.) sorghum (*Sorghum bicolor* L. Moench), wheat (*Triticum* spp.), peanut (*Arachis hypogaea* L.) [12], vineyard peach (*Prunus persica* L. Batsch), rice [13], and finger millet [14]. The main objective of any plant scientist is to identify an optimum number of plant characters which are sufficient to elucidate the maximum variability in field crop. This study was undertaken to run a classification analysis on selected *Oryza glaberrima* accessions and Asian rice varieties by means of descriptive statistic and to understand the association of various characters, PCA and cluster analysis which would enable breeders to classify the

available germplasm into distinct groups on the basis of their genetic diversity. PCA and cluster analysis allowed a natural grouping of genotypes [15,16]. However, results work in wheat showed that cluster analysis based on PCA is a more precise indicator of differences among wheat genotypes [17,18,19]. However, the results obtained would be valuable in developing an effective breeding.

2. MATERIALS AND METHODS

2.1 Plant Materials, Data Collection and Experimental Design

The study was conducted to characterize 30 rice germplasm accessions consisted of 22 selected *Oryza glaberrima* and 8 *Oryza sativa* lines obtained from AfricaRice gene bank were evaluated under lowland conditions at AfricaRice lowland breeding station, at the International Institute of Tropical Agriculture, Ibadan, Nigeria. IITA is located at (Latitude7^{o} 3 "N and Longitude 3^{o} 45 "E in the forest-savannah agro ecology). These lines were evaluated in replicated trials in alpha lattice design during the dry season of 2014. The 21 day old seedlings were transplanted into a well levelled puddled field of 1.2 m^{2} plot at spacing of 20 cm between rows and hills consisting two row plots of 3 m length. Ten morphological traits were measured including days to flowering, plant height, number of tiller, number of panicle, biomass wet weight, biomass dry weight, panicle wet weight, panicle dry weight, harvest index and grain yield.

2.2 Data Analysis

Univariate statistics, standard deviation and probability level for each one of the 10 characters were calculated. Clustering of genotypes into similar groups was performed using Ward's hierarchical algorithm based on squared Euclidean distances. For the ten agro-morphological characters, the data were standardized to have a mean of zero and a variance of one prior to squared Euclidean distance calculation.

The principal component analysis method explained by [20], was followed in the extraction of the components and determined of the percentage variability explained by each component, so identify the patterns of morphological variation. The univariate descriptive statistics, cluster and principal component analysis were performed using the SAS version 9.2 software [21] of ten morphological traits in thirty rice genotypes.

3. RESULTS AND DISCUSSION

3.1 Univariate Statistics

The test for univariate statistics described individual variable to explore pattern of response to variation showed strong statistical significantly ($P<0.001$) on phenotypic differences in all the variables that were measured (Table 1). This indicates that, the quantitative traits may be modified variously by the environmental conditions and each contributing such a small amount of phenotype that their individual effects cannot be detected by Mendelian methods but by only statistical methods. High variability observed for most of the traits indicated the possibility of improvement of these traits by direct selection. These traits should be given prominence in rice improvement.

3.2 Cluster Analysis

Similarity indices and pattern of relationships among the *O. glaberrima* and *O. sativa* from clusters and principal component analysis are useful in evaluating the potential breeding value of both the checks and *O. glaberrima* lines. The checks - FARO 52, TOX 4004, NERICA-L25, NERCA-U4, IR77298-14.1-2-B-10, IR74371-1-1, CG 14 and Apo were distributed across six clusters. In Fig. 1, the dendogram produced grouping that defined nine distinct clusters and minimum genetic distance between clusters varies from 0 to 5.

The nine, first and second clusters consisted of one (TOG5396), two (TOG7442 and TOG7996) and five (TOG6617, TOG6519, TOG13647, TOG12349 and TOG7400) genotypes belongs to *Oryza glaberrima*, respectively. The fifth cluster composed of three genotypes belongs to *O. sativa* check i.e., CG14, IR77298 and IR74371. Similarly, the seventh cluster comprised of three genotypes, but one genotype (FARO52) belongs to *O. sativa* check and two genotypes (TOMOYANG and RAM133) belong to *Oryza glaberrima.* While, the third, fourth, sixth and eighth clusters composed of four genotypes. The four genotypes belong to *Oryza glaberrima* viz., TOG13429, TOG6335, TOG6379 and TOG6408 in third cluster. During the fourth and sixth clusters, one genotype belong to *O. sativa* check

(NERICA–L25 and TOX4004, respectively) and three genotypes belongs to *Oryza glaberrima* i.e., (TOG6399, TOG9119 and TOG6410) and (A2-123, TOG6311 and TOG6520), respectively. However, the NERICA–U4 and Apo genotypes belongs to *O. sativa* check and the IRAT109 and TOG5534 genotypes belongs to *Oryza glaberrima* at the eighth cluster.

The results of cluster analysis suggested that there is genetic diversity among these genotypes for studied traits. The standalone of one of the selected *O. glaberrima* accession may have some underlying unique features not found in other groups (Fig. 1). This work has demonstrated the level of similarity between and within the selected lines. Hybridization among these genotypes provided more possibility to having more phenotypic diversity and could be used in breeding programs to achieve maximum heterosis and improvement of studied traits in rice.

Table 1. Univariate test statistics of some phenotypic characters identified by principal component analysis clusters

Variable	Total standard deviation	Pooled standard deviation	Between standard deviation	R-square	F value	Pr > F
Days to flowering (days)	347.8	124.3	370.4	0.99	1743.2	<.0001
Plant height (cm)	14.1	11.9	8.5	0.32	10.1	<.0001
Number of tillers (#)	44.2	8.4	46.3	0.96	598.4	<.0001
Number of panicle (#)	121.9	28.9	126.4	0.94	373.0	<.0001
Biomass wet weight (gm^{-2})	114.8	91.6	76.5	0.39	13.7	<.0001
Biomass dry weight (gm^{-2})	188.8	139.0	139.6	0.48	19.8	<.0001
Panicle wet weight (gm^{-2})	60.89	52.8	34.4	0.28	8.35	<.0001
Panicle dry weight (gm^{-2})	57.3	54.4	22.9	0.14	3.5	<.0001
Harvest index (%)	63.3	35.9	56.0	0.69	47.6	<.0001
Grain yield (t ha^{-1})	195.3	79.2	191.1	0.84	114.1	<.0001

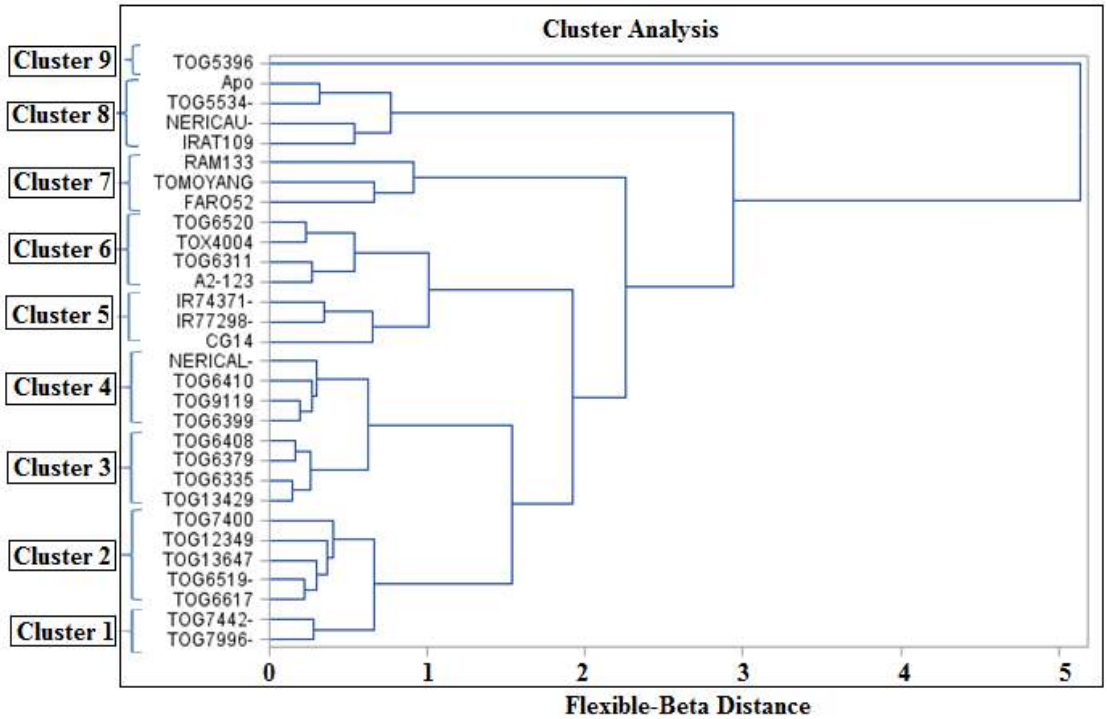

Fig. 1. Clustering pattern and dendrogram of seleceted *O. glaberrima* and *O. sativa* lines

3.3 Principal Component Analysis

The ten phenotypic traits measured, ten principal component axes were revealed from PCA analysis. Two principal components (PRIN1) and (PRIN2) accounted for most of the variability observed in traits measured (Table 2). The PRIN 1 accounted for 56% of the total phenotypic and morphological variation. The PRIN 1 was loaded on plant height, number of panicle, biomass wet weight, panicle wet weight and grain yield traits. The PRIN 2 accounted for 23% of the total variation. The PRIN 2 was loaded on biomass wet weight, biomass dry weight, panicle wet weight, panicle dry weight and harvest index traits. The most important descriptions were those associated with grain yield based on the analysis of natural variability in phenotypic traits measured in rice (Table 2). Days to flowering, number of tillers, biomass dry weight and harvest index traits negatively contributed in the first principal component axes (PRIN1), and only days to flowering negatively contributed to the total variation in the second principal component axes (PRIN2).

In general, we are interested in keeping only those principal components (PRIN1 and PRIN2) whose eigenvalues are greater than 1. The eigenvalues for PRIN1 and PRIN2 were 5.65 and 2.31, respectively. Components with an eigenvalue of less than 1 account for less variance than did the original variable (which had a variance of 1), and so are of little use. Hence, the point of principal components analysis is to redistribute the variance in the correlation matrix (using the method of eigenvalue decomposition) to redistribute the variance to first components extracted (Table 2). The scree plot graphs showed the eigenvalue against the number of component (Fig. 2). The values in the two columns of the Table 2 showed two principal component axes. From the third component, it showed that the line is almost flat (Fig. 1), meaning that each successive component is accounting for smaller and smaller amounts of the total variance.

Table 2. Eigenvector ("Weight") and Eigen value ("Load") of the correlation matrix and their contribution to total variation in rice

Descriptor variables	PRIN 1	PRIN 2
Days to flowering (days)	-0.39	-0.01
Plant height (cm)	0.23	0.01
Number of tiller (#)	-0.41	0.04
Number of panicle (m^{-2})	0.40	0.03
Biomass wet weight (gm^{-2})	0.28	0.27
Biomass dry weight (gm^{-2})	-0.28	0.43
Panicle wet weight (gm^{-2})	0.18	0.53
Panicle dry weight (gm^{-2})	0.01	0.60
harvest index (m^{-2})	-0.35	0.30
Grain yield (gm^{-2})	0.39	0.11
Eigen value	5.65	2.31
Proportion	0.56	0.23

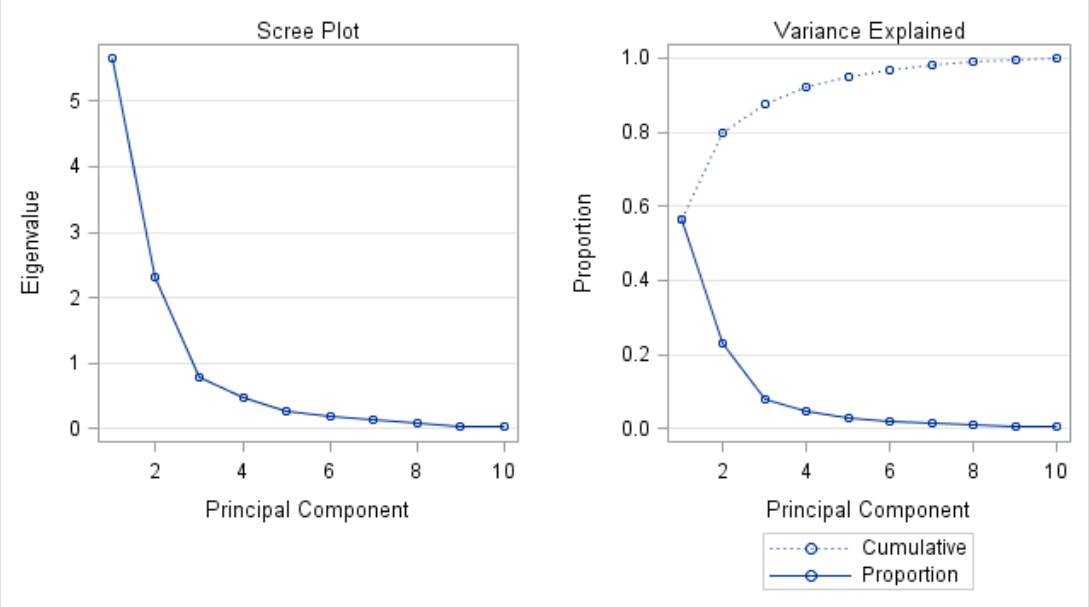

Fig. 2. Scree plot and variance explained by the principal component analysis

The evaluation of germplasm frequently comprises measurement of phenotypic characters or agronomic traits of interests, such as resistance to pests and diseases and tolerance to physiological stresses that are influenced by environment which they grow. The data measuring different characters are most or often required by plant breeders. According to [22] and [23], breeders continually working and looking for variation in germplasm for use in crop improvement. Therefore, continuous evaluation of germplasm should be done to broaden the genetic base of the species and identification of additional genes of interest or alternative source that control a particular attribute or trait for use in crop improvement. As demonstrated in this study, genetic variability in crop species should be exploited further to identify traits of interest so as to develop new rice varieties with high stability to resist or tolerate adverse environments and biotic conditions demonstrated by [24]. [25] studied the genetic relationship of some rice varieties and observed that origin, habitat and breeding background contributed to variation in the rice population. Therefore this study is conducted to explain variation among the selected *Oryza glaberrima* accessions *Oryza sativa* varieties and identify traits that contribute to variability in this population and for their possible exploitation in breeding programs. However, consistent effort is therefore required to evaluate rice germplasm collection and breeding rice adaptable to harsh African environment high grain yield.

4. CONCLUSION

Variation in plant phenotypic traits among rice accessions is a common phenomenon. Without variability, it is not possible to conduct a plant breeding program. Germplasm is hence the critical first step in initiating a breeding program. Diversity in *O. glaberrima* accessions is enormous. It needs to be organized and characterized in order to facilitate its use by plant breeders. There is need to understand how this tremendous variation originates and the manner in which it is organized or classified. We therefore need to know the sources to which breeders may go to find *O. glaberrima* germplasm with enormous variation to initiate their breeding programs, and the rationale for selecting materials for breeding and improving rice adaptable to African environment. Crop production occurs in a dynamic environment with every growing season, numerous factors that influence their performance, while some factors are within control; many are not. The weather, genetics, soil environment conditions represent broad categories of factors that influence crop adaptation and their usefulness in crop improvement must be deal with on a continual basis. Moreover, introduction of statistical tools has bought significant change in identifying crop adaptable to certain regions while PCA and cluster analysis have also paves the wave to improve rice productivity on the basis on genetic background and pattern of relatedness.

In the context of identifying the pattern of relatedness, the hierarchical approach of cluster analysis and principal component analysis found to be quite suitable. Each cluster thus represents similarity. The following conclusion is drawn in the present study in the context of identifying the pattern relatedness, cluster analysis and principal component method can be applied as demonstrated in this study.

Thus, this study revealed that there is variation among the rice genotypes studied and the first two PCA axes explained and contributed to the majority of variation. However, results showed that cluster analysis based on PCA is a more precise indicator, and can be a useful tool for the selection of the most efficient genotype.

Thus, as demonstrated in this study, members of this group have been reported suitable for breeding programs aimed at improving the yield of crop species [26,27,28,29,30,31]. Similar to the findings by [32] who reported that cluster analysis can be useful for finding high yielding wheat. Considering the ten main components, the first two components explained 79% of total variations in data. PCA and cluster analysis allowed a natural grouping of the rice genotypes and appropriate use of techniques for genotypes grouping have been reported [15,16]. However, results showed that cluster analysis based on PCA is a more precise indicator and can be a useful tool for the selection of the most efficient genotype.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Islam MR. Genetic diversity in irrigated rice. Pak J Biol Sci. 2004;2:226-229.
2. Bhatt GM. Multivariate analysis approach to selection of parents for hybridization

aiming at yield component in self pollination crops. Aus J Agric Rec. 1970; 21:1-7.
3. Carves BF, Smith EL, England HO. Regression and cluster analysis of environmental responses of hybrid and pure line winter wheat cultivars. Crop Sci. 1987;27:659-664.
4. Eivazi AR, Naghavi MR, Hajheidari M, Pirseyedi SM, Ghaffari MR, Mohammadi SA, Majidi I, Salekdeh GH, Mardi M. Assessing wheat (*Triticum aestivum* L.) genetic diversity using quality traits, amplified fragment length polymorphisms, simple sequence repeats and proteome analysis. Ann Appl Biol. 2007;152:81-91.
5. Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants-salient statistical tools and considerations. Crop Sci. 2003;43:1235-1248.
6. Fotokian M, Shahnejat bushehri A, Taleie A. Cluster analysis based on PCA in rice genotypes. Paper presented at the 6rd international conference of Statistics, University of Tarbiat Modares, Iran, 26-28 August; 2002.
7. Mellingers JS. Measures of genetic similarity and genetic distance. Studies in genetics. VII Univ Tex Publ. 1972;27(13): 145-153.
8. Ulaganathan V, Nirmalakumari A. Finger millet germplasm characterization and evaluation using principal component analysis. Sabrao Journal of Breeding and genetics. 2015;47(2):79-88.
9. Shaibu AA, Ogburia MN. Structural organization and morphological diversity of somaclonal variation in triploid plantain (*Musa* Sp. AAB group) landraces. The Nigerian Journal of Research and Production vol. 2002;1:31-37.
10. Adams MW. An estimate of homogeneity in crop plants with special reference to genetic vulnerability in dry season. *Phseolus vulgaris*. Euphytica. 1995;26: 665-679.
11. Maji AT, Shaibu AA. Application of principal component analysis for rice germplasm characterization and evaluation. J. Plant Breeding and Crop Sci. 2012;4:87-93.
12. Upadhyaya HD, Gowda CLL, Pundir RPS, Reddy VG, Singh S. Development of core subset of finger millet germplasm using geographical origin and data on 14 quantitative traits. Gen. Res. Crop Evo. 2006;53:679-685.
13. Gana AS, Shaba SZ, Tsado EK. Principal component analysis of morphological traits in thirty-nine accessions of rice (*Oryza sativa* L.) grown in a rainfed lowland ecology of Nigeria. J. Plant Breed. Crop Sci. 2013;5:120-126.
14. Dagnachew Lule, Kassahun Tesfaye, Masresha Fetene, Santie De Villiers. Multivariate analysis for quantitative traits in finger millet (*Eleusine coracana* subsp. *coracana*) population collected from Eastern and Southeastern Africa: Detection for patterns of genetic diversity. Int. J. Agric. Res. 2012b;7(6):303-314.
15. Bauer I, Drinic SM, Drinic G, Micic DI. Assessing temporal changes in genetic diversity of maize hybrids using RAPD markers. Cereal Res Commun. 2007;35: 1563–1571.
16. Kraic F, Mocák J, Roháčik T, Sokolovičová J. Chemometric characterization and classification of new wheat genotypes. Nova Biotechnol. 2009;9:101-106.
17. Reif JC, Zhang P, Dreisigacker S, Warburton ML, van Ginkel M, Hoisington D, Bohn M, Melchinger AE. Wheat genetic diversity trends during domestication and breeding. Theor Appl Genet. 2005;110: 859–864.
18. Fu YB. Impact of plant breeding on genetic diversity of agricultural crops: Searching for molecular evidence. Plant Genet Resour. 2006;4:71–78.
19. White J, Law JR, MacKay I, Chalmers KJ, Smith JSC, Kilian A, Powell W. The genetic diversity of UK, US, Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. Theor. Appl. Genet. 2008;116:439–453.
20. Harman HH, Modern factor analysis. 3rd Ed. University of Chicago Press, Chicago, 376, 1976.
21. SAS Institute. SAS/STAT software: Users Guide, Version 9.2. SAS Inst., Cary, NC; 1996.
22. Ng NQ, Padulosi S. Constraints in the accessibility and use of germplasm collection. In Biotechnology enhances research on tropical crops in Africa edited by Thottappily, L. M. Monti D. R. Mohan Raj. A. W. Moore IITA Ibadan; 1992.
23. Peters JP, Williams JT. Towards better use of gene bank with special reference to information. FAO/IBPGR Plant genetic resources Newsletter, 1984;60:22-31.
24. Gana AS. Variability studies of the response of rice varieties to biotic and

abiotic stresses. Unpublished Ph.D Thesis, University of Ilorin; 2006.
25. Awopetu JA, Gana, AS. A Numerical Analysis of Genetic relationship within rice accessions. Nig. J. Gene X. 1997;11:1-8.
26. Nersting LG, Andersen SB, von Bothmer R, Gullord M, Jorgensen RB. Morphological and molecular diversity of Nordic oat through one hundred years of breeding. Euphytica. 2006;50:327–337.
27. Saleem U, Khaliq I, Tariq M, Rafique M. Phenotypic and genotypic correlation coefficients between yield and yield components in wheat. J Agricul Res. 2006; 44: 1-8.
28. Figliuolo G, Mazzeo M, Greco I. Temporal variation of diversity in Italian durum wheat germplasm. Genet Resour Crop Evol. 2007;54:615–626.
29. Hysing S, Sall T, Nybom H, Liljeroth E, Merker A, Orford S, Koebner RMD. Temporal diversity changes among 198 Nordic bread wheat landraces and cultivars detected by retrotransposon-based S-SAP analysis. Plant Genet Resour Charact Util. 2008;6:113–125.
30. Mantegazza R, Biloni M, Grassi F, Basso B, Lu BR, Cai XX, Sala F, Spada A. Temporal trends of variation in Italian rice germplasm over the past two centuries revealed by AFLP and SSR markers. Crop Sci. 2008;48:1832–1840.
31. Aghaee M, Mohammadi R, Nabovati. Agromorphological characterization of durum wheat accessions using pattern analysis. Aust J Crop Sci. 2010;4:505-514.
32. Ali Y, Atta BM, Akhter J, Monneveux P, Lateef Z. Genetic variability, association and diversity studies in wheat (*Triticum aestivum* L.) germplasm. Pak J Bot 40: 2087-2097. Anand IJ, Murrty BR (1968) Genetic divergence and hybrid performance in linseed. Ind J Genet plant Breed. 2008;28:178-185.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/18774