Journal of Plant Breeding and Crop Science Vol. 4(6), pp. 87-93, 30 March, 2012
Available online at http://www.academicjournals.org/JPBCS
DOI: 10.5897/JPBCS11.093
ISSN 2006-9758 © 2012 Academic Journals

Full Length Research Paper

Application of principal component analysis for rice germplasm characterization and evaluation

Maji A. T. and Shaibu A. A.*

National Cereals Research Institute, Badeggi, P. M. B. 8, Bida, Niger State, Nigeria.

Accepted 19 March, 2012

This study gave empirical evidence on sixteen agro-morphological data that were collected from one hundred and twenty three rice germplasm comprising of *Oryza sativa* and *Oryza glaberrima* lines including checks. The data was collected from thirteen villages in two States in Nigeria and were characterized using ANOVA model. Among the studied traits, high coefficients of variation were observed for number of unfilled grain per head (45.8%), grain weight (29.1%), 1000 grain weight (23.0%), tiller number at three weeks after planting (22.5%), and tiller number at maturity (20.9%). Seven out of the sixteen phenotypic traits measured were statistically significant at (P = 0.001 and P = 0.05), and 7 phenotypic variables also showed significant differences when subjected to univariate statistics at (P = 0.001 and P = 0.05). The association of all morphological traits was estimated by phenotypic correlation coefficient and showed that eight dependent variables were positively related. Cluster analysis using Ward's method classified the 123 populations into seven distinct groupings. A large number of genotypes was placed in cluster 5 (65 genotypes) followed by cluster 1 (20), cluster 4 (14) and cluster 3 (9), cluster 2 (8) and cluster 6 (7). Cluster 6 includes five checks with few *sativa* lines, cluster 5 with large grouping of *sativa* lines with only FARO 56 in that group. Cluster 1 consists of only the *O. glaberrima*. Clusters 2, 3 and 4 consisted of only *O. sativa* groups indicating no association between clustering pattern and eco-geographical distribution of genotypes. The maximum inter-cluster distance was observed between clusters indicating the possibility of high heterosis if individuals from these clusters are cross-bred. Principal component analysis resulted in the first two components with Eigen value greater than 1 accounting for 78% of the total variation. The results of Principal Component Analysis (PCA) were closely in line with those of the cluster analysis. These results can now be used by breeders to develop high yielding rice varieties and new breeding protocols for rice improvement.

Key words: Principal component analysis, germplasm characterization, correlation coefficient, correlation matrix, cluster analysis, genetic variability.

INTRODUCTION

A large number of variables are often measured by plant breeders, some of which may not be of sufficient discriminatory power for germplasm evaluation, characterization, and management. In such case, Principal Component Analysis (PCA) may be used to reveal patterns and eliminate redundancy in data sets (Adams, 1995; Amy and Pritts, 1991) as morphological and physiological variations routinely occur in crop species. Knowledge of the nature, extent and organization of this variation could be useful for genetic improvement of crop species. Until a collection has been properly evaluated and its attributes become known to breeders, it has little practical use. Germplasm evaluation in the broad sense and in the context of genetic resources is the description of the material in a collection that covers the entire range of activities starting from the receipt of the new samples by the curator and growing of these for seed increase, characterization and preliminary evaluation, and also for further evaluation. Rice is considered as an important food crop in Africa and

*Corresponding author. E-mail: a.shaibu@hotmail.com.

millions of people depend on it for their livelihood (Anonymous, 2000). The development of varieties is a continuous process and the success of the plant breeding programme aimed at the evaluation of high yielding, better quality, fertilizer responsive, disease and insect resistance varieties depends upon the selection of suitable plants to be utilized in breeding programme. The effectiveness of selection depends primarily upon the magnitude of genetic variability in the breeding material at hand. After collection of germplasm, there is the need for its systematic evaluation in order to know its various morphological, physiological and developmental characters including some special features, such as stress tolerance, pest and disease resistance. Such systematic and detailed evaluation operations, though expensive and time consuming, are of great value. The principal goal in exploiting useful genes from germplasm collections vary greatly among crops and for different ecological zones within a crop. For example, stable resistance to races of rusts, semi-dwarfness, improved quality, tolerance to drought and wider adaptability in higher yield potential, improved quality and multiple resistance to pests and diseases in rice, resistance to viruses, nematodes and bacterial diseases. Appropriate and most efficient approach should be used for germplasm evaluation and characterization, and while further detailed evaluation is mostly done by the breeders for taking additional information. Hotelling (1933) indicated that principal component analysis (PCA) is an exploratory tool designed by Karl (1901) to identify unknown trends in a multi-dimensional data set. However, in a typical micro-array experiment, the expression of thousands of genes is measured across many conditions such as treatments or time points. Therefore, it becomes impossible to make a visual inspection of the relationship between genes or conditions in such a multi-dimensional matrix. One way to make meaning of this data is to reduce its dimensionality (Hotelling, 1933). Several data decomposition techniques are available for this purpose and multivariate data analysis (Cooley and Lohnes, 1971). PCA is among these techniques that reduced the data into two dimensions (Smith, 2002; Rao, 1964; Raychaudhuri et al., 2000).

Multivariate statistical techniques which simultaneously analyze multiple measurements on each individual under investigation, are widely used in analysis of genetic diversity irrespective of whether it is morphological, biochemical or molecular marker-based and subsequently, classification of germplasm collections. Among the multivariate techniques, cluster analysis, PCA, principal co-ordinate analysis (PCoA) and Multi-Dimensional Scaling (MDS) are at present, most commonly employed and appear particularly useful (Mohammadi and Prasanna, 2003). Multivariate analysis has been used frequently for genetic diversity analysis in many crops such as barley (*Hordeum vulgare* L.) (Cross, 1992),Sorghum (*Sorgum bicolor* L. Moench) (Ayana and Bekele, 1999), wheat (*Triticum* spp.) (Hailu et al., 2006), peanut (*Arachis hypogaea* L.) (Upadhyaya et al., 2009) and vineyard peach (*Prunus persica* L. Batsch) (Nikolic et al., 2010) and in rice, (Bharadwaj et al., 2001). The study was undertaken to run a classificatory analysis on the rice genotypes by means of PCA, and descriptive ANOVA statistic, which would enable us to classify the available germplasm into distinct groups on the basis of their genetic diversity. The information, thus obtained, would be helpful to develop an effective rice-breeding programme and as such a quantification of the degree of divergence would be helpful in choosing suitable genotypes and traits of interests for ongoing breeding programmes.

MATERIALS AND METHODS

Plant materials, data collection and experimental design

One hundred and thirty six rice lines including six checks were sampled from different eco-geographical locations of Niger State, in the North Central part of Nigeria including Minna, Kuta, Paiko, Againe, Lapai, Bida, Jina, Mambe, Gonagi, Tufa, Naico and Kotangora, and in Kebbi State, in the North West eco-geographical location of Nigeria including Yawuri, Jega, and Ilubi. The rainfall pattern in the two eco-geographical locations is between three to four months of rain starting from the month of May and sometimes in June and ends in the month of August or September and high temperatures in early part of the year (February to April). The National Cereals Research Institute, Badeggi, Niger State, Nigeria, situated in Latitude 9' 45° N, Longitude 6' 07 E and 70.5 m. a. s. l in the South Guinea Savannah ecological zone of Nigeria, and in Kebbi on Latitude N, Longitude E and m. a. s. l) and investigated during 2009 and 2010. The experiment was conducted in an augmented design with checks replicated 3 times within the blocks. Experimental units in each block comprised of one line of four meters long. Row to row and plant to plant spacing was 20 cm between and within rows respectively. Different morphological and agronomical traits were measured that included tiller number at three weeks (TN-3 wks) after planting, plant height at maturity (PHT-mat) and plant height at vegetative stage (PHT-veg), leaf length (LL), leaf width (LW), ligule length (LGL), tiller number at maturity (TN-mat), days to 50% flowering (DFL50), panicle weight (PW), grain weight (Gwt), panicle length (PL), number of grain per panicle (NGPP), number of unfilled grain per head (NUNFGPH), 1000 grain weight (1000-gwt), grain length (GL), and grain width (GW).

Data analysis

Analysis of variance using descriptive statistics such as mean, standard deviation and coefficient of variation and correlation coefficient for each one of the 16 studied traits were calculated. Clustering of genotypes into similar groups was performed using Ward's hierarchical algorithm based on squared Euclidean distances. For the sixteen 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 pseudo F statistic and the pseudo T2 statistic (Jobson, 1992) were examined to establish the numbers of clusters using SAS version 9.2 software (SAS Institute, 1996, Cary, NC). In order to identify the patterns of morphological variation, (PCA) was conducted.

Table 1. Mean squares from analysis of variance of some agro-morphological traits data collated in rice germplasm collections.

Source of variation	Df	TN-3WKS	PHT-veg	PHT-mat	LL	LW	LGL	TN-mat	DFL50	Pan-wt	Gwt	PanLgt	NGPP	NUNFGPH	1000-gwt	GL	GW
Block	2	16.9	612.8	49.7	0.9	0.02	0.01	0.28	43.4	0.002	0.012	4.33	58.02	14.65	5.88	0.007	0.003
Trt	174	35.1**	123.2	629.1***	42.1	0.04	0.05**	22.9	915.6***	198.9***	0.134**	4.94	292.2	6.31	24.13	0.48***	0.005***
Error	348	6	128.3	90.4	21.6	0.01	0.01	10.9	36.6	0.024	0.012	2.17	384.4	6.28	30.7	0.002	0.001
R^2		0.98	0.93	0.99	0.96	0.98	0.98	0.97	0.99	0.99	0.99	0.96	0.91	0.94	0.91	0.99	0.99
C.V (%)		22.5	16.9	6.2	11.6	13.9	12.4	20.9	4.9	7.7	29.1	5.6	19.8	45.8	23.0	5.1	1.0

TN-3WKS = Tiller number at three weeks after planting; PHT-veg=Plant height at vegetative stage; PHT-mat=Plant height at maturity stage; LL= Leaf length; LW = Leaf width; LGL= Ligule length; TN-mat = Tiller number at maturity; DFL50 = Days to 50% flowering; Pan-wt=Panicle weight; Gwt=Grain weight; PanLgt = Panicle length; NGPP = Number of grain per panicle; NUNFGPH=Number of unfilled grain per head; 1000-gwt=1000 grain weight; GL=Grain length; GW=Grain width. ns = not significant. *, **, *** significant at P = 0.05, 0.01, 0.001; values without asterisks are not significant.

Those PCs with Eigen values greater than one were selected as proposed by Jeffers (1967). Correlations between the original traits and the respective PCs were calculated. Data were processed using statistical program (Statistical analysis System). The principal component analysis was computed using the following equation:

PCA

$$PC1 = \sum_{1}^{p} a_j X_j$$

Where; PC = Principal component, a1j = Linear coefficient – Eigen vectors

RESULTS AND DISCUSSION

The analysis of variance for means indicated that the differences among the genotypes were significant for eight of the sixteen characters studied (Table 1). In this study, we found the significant mean squares results for 8 of the 16 quantitative characters in 136 genotypes of rice germplasm collection. Bharadwaj et al. (2001) and Kotaiah et al. (1986) reported that the significant differences among the rice genotypes indicated the necessity to group them into clusters to identify the divergent groups, and Mehdi and Asghar (1999) also classified the sorghum genotypes in five distinct groups using this method. Among the traits of interest that agreed with Bharadwaj et al. (2001) was days to 50% flowering and panicle length. In rice breeding, the earliness to attain flowering cycle from 21 days after transplanting is a good trait that can be utilized for screening drought tolerance genotype. The analysis of variance for all morphological traits were significant for tiller number at three weeks after planting (TN-3wk), ligule length (LGL), and grain weight (Gwt), at (P = 0.01), and plant height at maturity (PHT-mat), days to 50% flowering (DFL50), panicle weight (PW), grain length (GL) and grain width (GW), at (P = 0.001), (Table 2) among the treatments.

The association of all morphological traits was estimated by phenotypic correlation coefficient (Table 2). Among the 16 morphological traits studied, Leaf length (LL), days to 50% flowering (DFL50), plant height at maturity (PHT-mat) had significant and positive correlations with plant height at vegetative (PHT-veg), P = 0.001 and significant correlation with ligule length (LGL), number of grain per panicle (NGPP) and number of unfilled grain per head (NUNFGPH) at P = 0.1 and all other traits were not significant. Leaf length (LL) had significant correlation with leaf width (LW), plant height at maturity (PHT-mat) and number of grain per panicle (NGPP) at P = 0.01, 0.001 and 0.1, and every other traits were not significant. Leaf width (LW) had negative, but significant correlation with plant height at maturity (PHT-mat) at P = 0.01 and every other traits had no significant association. Ligule length (LGL) had positive and significant association with plant height at maturity (PHT-mat) at P = 0.1) and all other traits were not significant. Plant height at maturity (PHT-mat) had positive correlation with days to 50% flowering (DFL50) and number of grain per panicle (NGPP) at P = 0.001 and days to 50% flowering (DFL50) had positive correlation with grain weight (Gwt) at P = 0.1; and grain weight (Gwt), had positive correlation with 1000 grain weight (1000-gwt) at P = 0.1 and panicle length (PL) positively correlated with number of unfilled grain per head (NUNFGPH) at P = 0.1. All other traits were not significant. The non significance of some traits may be due to less contribution to their development or might be due to genetic constitution differences in the breeding materials evaluated. The significance association of some traits are expected. For example, panicle length which positively correlated with grain 1000 grain weight (1000-gwt) is expected as these traits are important parameters that determine the grain yield of rice plant, likewise, plant height at

Table 2. The correlation coefficient of agro-morphological traits evaluated in rice germplasm collections.

Variable	TN-3WKS	PHT-veg	LL	LW	LGL	PHT-mat	TN-mat	DFL50	Pan-wt	Gwt	PanLgt	NGPP	NUNFGPH	1000-gwt	GL	GW
TN-3WKS	-	0.13	0.09	0.20	-0.10	0.08	0.11	0.08	0.09	0.68	-0.00	0.07	-0.02	0.14	-0.12	-0.01
PHT-veg			0.54***	-0.08	0.26*	0.53***	0.05	0.43***	-0.03	0.13	0.18	0.25*	0.24*	-0.20	0.00	0.16
LL				0.28**	0.17	0.39***	0.09	0.16	0.03	0.12	0.10	0.24*	0.12	0.05	0.02	0.00
LW					-0.19	-0.29**	-0.00	-0.10	0.10	0.12	0.03	-0.09	0.13	0.11	0.10	0.07
LGL						0.27*	0.05	0.09	-0.19	0.02	0.22	0.15	-0.02	0.08	0.05	-0.18
PHT-mat							0.13	0.31***	0.08	0.14	0.12	0.35***	0.03	0.02	0.06	-0.11
TN-mat								-0.00	0.05	-0.01	0.05	0.29*	0.12	0.08	0.02	-0.04
DFL50									-0.07	0.22*	0.18	0.19	0.12	0.02	-0.01	0,09
Pan-wt										-0.00	0.05	0.00	0.01	-0.03	-0.02	0.14
Gwt											0.05	0.05	0.09	0.24*	0.03	-0.00
PL												0.20	0.29*	-0.15	-0.12	0.13
NGPP													0.14	0.09	0.03	-0.00
NUNFGPH														-0.18	-0.12	0.11
1000-gwt															0.08	-0.07
GL																-0.02
GW																-

TN-3WKS = Tiller number at three weeks after planting; PHT-veg = Plant height at vegetative stage; PHT-mat = Plant height at maturity stage; LL = Leaf length; LW=Leaf width; LGL = Ligule length; TN-mat=Tiller number at maturity; DFL50=Days to 50 percent flowering; Pan-wt = Panicle weight; Gwt = Grain weight; PanLgt = Panicle length; NGPP = Number of grain per panicle; NUNFGPH = Number of unfilled grain per head; 1000-gwt = 1000 grain weight; GL = Grain length; GW = Grain width. ns = not significant. *, **, *** significant at P = 0.05, 0.01, 0.001; values without asterisks are not significant.

vegetative stage that is highly correlated with days to 50% flowering, number of grain per panicle and number of unfilled fingers per head indicating that rice growth at vegetative stage is also important and crucial for the physiological process that determines the output of its performance in terms of yield and other physiological attributes that contributes to its development. The positive and significant association of these traits will provide plant breeders an understanding to phenotypic traits and their degree of association to be able to plan breeding schemes and managements of plant germplasm. Consequently, there were significant differences for the test of univariate statistics, and out of 16 morphological variables, 7 were statistically significant including plant height at vegetative stage (PHT-veg), leaf length (LL), plant height at maturity (PHT-mat), days to 50% (DFL50), panicle weight (Pan-wt) and number of unfilled grains per head (NUNFGPH) at P = 0.001 and ligule length (LGL) at P = 0.05 respectively (Table 3).

The PCA was used to eliminate redundancy in the data set. Two principal components (PRIN 1) and PRIN 2) accounted for most of the variability observed among the rice germplasm collections from different locations (Table 4). PRIN 1 accounted for 55% of the morphological variation in the rice germplasm collection and was loaded on plant height at vegetative stage (PHT-veg), leaf length (LL), ligule length (LGL), plant height at maturity stage (PHT-mat), days to 50% flowering (DFL50), grain weight (Gwt) and number of grains per panicle (NGPP). PRIN 2 accounted for 23% of the variation and was loaded on total number of suckers at three weeks after planting (TN-3WKS), leaf width (LW), grain weight (Gwt) and number of unfilled grains (NUNFGPH), while the remaining variables had weak or no discriminatory power. Thus, the most important descriptors were those associated with PRIN 1 and 2. Similarity indices and pattern of relationships among the *O. sativa* and *O. glaberrima* from clusters and principal component analysis are useful in evaluating the potential breeding value of the checks, sativa and glaberrima lines. The checks included in this analysis are FAROs 44, 46, 52, 55, 56, and 57. The phenogram produced grouping that defined

Table 3. Univariate test statistics of morphological variables identified by clusters.

Variable	Total STD	Pooled STD	Between STD	R-square	R-square (1-RSq)	F-Value	Prob.
TN-3WKS	5.8834	5.8198	1.7504	0.0781	0.0847	1.38	0.2207
PHT-veg	11.8614	7.9864	9.5582	0.5729	1.3413	21.84	0.0001***
LL	6.6800	5.9579	3.5597	0.2505	0.3343	5.44	0.0001***
LW	0.2081	0.2006	0.0780	0.1241	0.1417	2.31	0.0309
LGL	0.2100	0.1937	0.0997	0.1987	0.2479	4.04	0.0005*
PHT-mat	23.5706	12.6319	21.4320	0.7294	2.6956	43.90	0.0001***
TN-mat	5.0793	5.1469	0.9766	0.0326	0/0337	0.55	0.7955
DFL50	29.9405	12.5888	29.1008	0.8334	5.0038	81.49	0.0001***
Pan-wt	14.6808	0.2194	15.6283	0.9899	47.9672	77.9	0.0001***
Gwt	0.3744	0.3682	0.1187	0.0887	0.0974	1.59	0.1465
PanLgt	2.1191	2.0324	0.8241	0.1334	0.1540	2.59	0.0196
NGPP	17.0690	12.6075	12.6689	0.4860	0.9455	15.40	0.0001***
NUNFGPH	2.8703	2.8289	0.8901	0.0849	0.0927	1.51	0.1708
1000-gwt	4.8267	4.6840	1.7255	0.1127	0.1271	2.07	0.0524
GL	1.0060	1.0270	0.1443	0.0181	0.0185	0.30	0.9521
GW	4.3087	4.3879	0.6945	0.0229	0.0235	0.38	0.9113

. *, *** significant at P = 0.05, 0.001; values without asterisks are not significant.

Table 4. Eigenvector ("Weight") and Eigen value ("Load") of the correlation matrix and their contribution to total variation of rice germplasm collections.

Descriptor variables	Prin 1	Prin 2
TN-3WKS	0.17	0.36
PHT-veg	0.45	-0.05
LL	0.34	0.08
LW	-0.08	0.38
LGL	0.25	-0.13
PHT-mat	0.43	-0.11
TN-mat	0.12	0.07
DFL50	0.37	-0.13
Pan-wt	-0.01	0.13
Gwt	0.23	0.31
PL	0.18	0.12
NGPP	0.30	-0.08
NUNFGPH	0.15	0.21
1000-gwt	0.09	0.14
GL	0.02	-0.47
GW	0.03	-0.47
Eigen value	2.924	2.049
Proportion (%)	0.55	0.23

six distinct clusters and minimum genetic distance between clusters varies from 0.5 to 2.4. Cluster 1 composed of *O.glaberrima* diverged at 0.5, 0.8 and 2.4 level. One of the checks (FARO 44) was found in the same cluster grouping with other sativa lines in cluster 4 (Figure 1). FAROs 46, 55, 52 and 57 were grouped into cluster 6 with only few sativa lines while cluster 5 comprised of FARO 56 with large groups of sativa lines. Clusters 2, 3 and 4 consisted of the sativa lines with genetic distance diverged at 0.3 to 0.5 level (Figure 1). Germplasm evaluation and characterization is a routine endeavour for plant breeders, and application PCA tool, cluster and multivariate statistical analysis provide a useful means for estimating morphological diversity within and between germplasm collections. These tools are useful for the evaluation of potential breeding value and

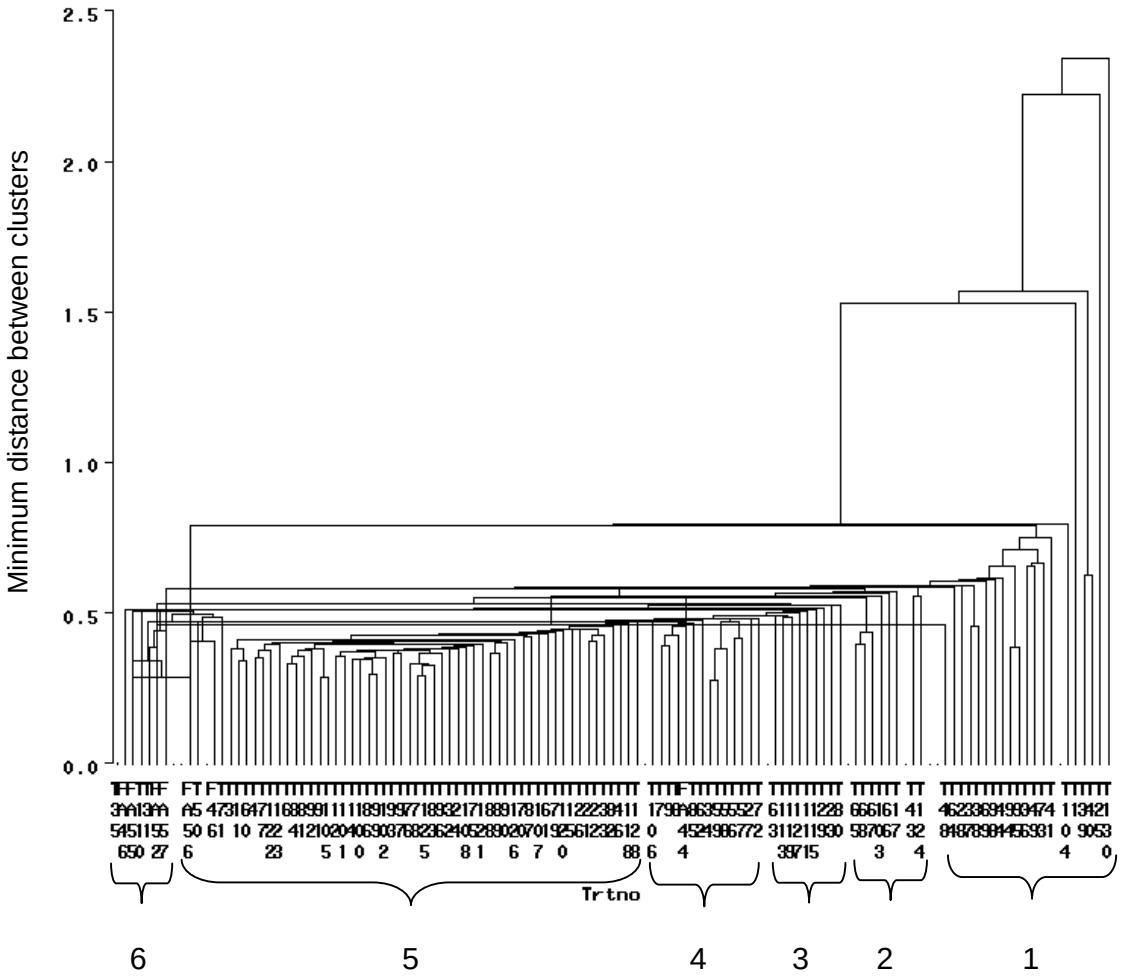

Figure 1. Clustering pattern and dendrogram of the 123 interspecific sativa and *O. glaberrima* lines using simple matching coefficient. Each line is named with a prefix number.

were used to detect significant differences between germplasm and magnitude of deviation among crop species.

However, this study suggest the need for breeders to evaluate germplasm with appropriate tools in order to characterize and evaluate accurately breeding germplasm materials into their distinct groupings and progression on their utility in the rice breeding programme in the development of improved varieties with better genetic base and more so, it is a tool for determining parental contribution and relatedness among progenies. FARO 44 found within the sativa lines and FAROs, 46, 52 and 57 also found within the same clusters will be useful in several backcrosses in rice breeding programme while the *O. glaberrima* can be used for the development of new lines for other important traits of interest such as drought and disease resistance. The variables that are strongly associated in the same group may share some underlying biological relationship, and these associations are often useful for generating hypothesis for better understanding of behaviour of complex traits that would allow breeders to maximize their knowledge.

REFERENCES

Adams MW (1995). An estimate of homogeneity in crop plants with special reference to genetic vulnerability in dry season. *Phseolus vulgaris*. Ephytica, 26: 665-679.

Amy EL, Pritts MP (1991). Application of principal component analysis to horticultural research. Hortscience, 26(4): 334-338.

Anonymous (2000). Economic survey (2000-2001). Government of Pakistan, Economic advisory wing, Finance Division, Islamabad, pp. 16-17.

Ayana A, Bekele E (1999). Multivariate analysis of morphological variation in sorghum (*Sorghum bicolor* L. Moench) germplasm from Ethiopia and Eritrea. Genet. Resource. Crop Evol., 46: 273-284.

Bharadwaj Ch, Tara SC, Subramanyam D (2001). Evaluation of different classifactory analysis methods in some rice (*Oryza sativa* L.) collections. Ind. J. Agric. Sci., 71(2): 123-125.

Cooley WW, Lohnes PR (1971). Multivariate Data Analysis. John Wiley & Sons, Inc., New York,

Cross RJ (1992). A proposed revision of the IBPGR barley descriptor list. Theor. Appl. Genet., 84: 501-507.

Hailu F, Merker A, Singh H, Belay G, Johansson E (2006). Multivariate

analysis of diversity of tetraploid wheat germplasm from Ethiopia. Genet. Resource. Crop Evol., 54: 83-97.
Hotelling H (1933). Analysis of a complex of statistical variable into principal components.; J. Educ. Psych., 24: 417-441.
Jeffers JNR (1967). Two case studies in the application of principal component analysis. Appl. Stat., 16: 225-236.
Jobson JD (1992). Applied multivariate data analysis. Volume II: Categorical and Multivariate Methods. New York: Springer Verlag, USA.
Kotaiah KC, Panduranga C, Rao C, Sreerama N, Reddi N, Sarma, YRB (1986). Mahalanobis's D2 and metroglyph analyses in mid-duration genotypes of rice. Ind. J. Agric. Sci., 56(3): 151-156.
Mehdi SS, Asghar MJ (1999). Genetic variability, heritability and genetic advance for seedling traits in sorghum genotypes. Sci. Int. (Lahore), 11(1): 113-116.
Mohammadi SA, Prasanna BM (2003). Analysis of genetic diversity in crop plants-salient statistical tools and considerations. Crop Sci., 43: 1235-1248.
Nikolic D, Rakonjac V, Milatovic D, Fotiric M (2010). Multivariate analysis of vineyard peach (*Prunus persica* L. Batsch.) germplasm collection. Euphytica, 171: 227-234.
Rao CR (1964). The Use and Interpretation of Principal Component Analysis in Applied Research. Sankhya A, 26: 329 -358.
Raychaudhuri S, Stuart JM, Altman RB (2000). Principal components analysis to summarize microarray experiments: Application to sporulation time series. Pacific Symposium on Biocomputing.
SAS Institute (1996), SAS/STAT software: Users guide, version 9.2. SAS Inst., Cary, NC.
Smith LI (2002). A tutorial on Principal Components Analysis; http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.
Upadhyaya HD, Reddy LJ, Dwivedi SL, Gowda CLL, Singh S (2009). Phenotypic diversity in cold-tolerant peanut (*Arachis hypogaea* L.) germplasm. Euphytica, 165: 279-291.