



## Linear Discriminant Function Analysis for the Different Characters of Pigeonpea (*Cajanus cajan* (L.) Millsp) Accessions

Pavan Kumar S. T.\*†, Avinalappa H. H.\*\*, Tuppad G. B.\*\*\*, Adam Kamei\*\*\*\*, Nethravathi K. H.\*\*\*\*\*, Dharmaraj P. S.\*\*\*\*\* and D. Mazumdar\*\*\*\*\*

\*Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur-741 252, India

\*\*Department of Genetics and Plant Breeding, Agricultural Research Station, Gulbarga UAS, Raichur-585 101, Karnataka

\*\*\*Department of Agronomy, UAS, Dharwad-580 005, Karnataka, India

\*\*\*\*Department of Plant Pathology, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur-741 252, India

\*\*\*\*\*Department of Genetics and Plant breeding, UAS, Dharwad-580 005, Karnataka, India

\*\*\*\*\*Agriculture Research Station, Gulbarga, UAS, Raichur- 585 101, Karnataka, India

\*\*\*\*\*Agricultural Statistics, Director of Research, B.C.K.V, Kalyani-741 235, West Bengal, India

†Corresponding author: Pavan Kumar S. T.

Nat. Env. & Poll. Tech.  
Website: www.neptjournal.com

Received: 5-12-2014

Accepted: 21-1-2015

### Key Words:

Pigeonpea

*Cajanus cajan*

K-means clustering

Discriminant function

### ABSTRACT

Variations in different characters of the pigeonpea were studied using 122 accessions under rainfed condition during 2013. The characters considered viz., days to 50 per cent flowering, plant height (cm), days to physiological maturity, pods per plant, pod bearing length (cm), seeds per pod, 100 seed weight (g) and seed yield per plant (g) were used for the analysis. K-means cluster analysis was carried out based on the mean values and classified 122 accessions into four distinct clusters. The cluster-I (14 accessions), cluster-II (40 accessions), cluster-III (17 accessions) and in fourth cluster 51 accessions were distinctly classified. It is found that, days to physiological maturity, pods per plant, days to 50 per cent flowering and pod bearing length (cm) were significant discriminating characters. The stepwise discriminant analysis showed that pods per plant and days to physiological maturity were highly discriminating characters and these two characters were explained by the 98.8 percent variation in the accessions studied and these accessions were classified 97.5 percent correctly in the four clusters.

### INTRODUCTION

Pigeonpea (*Cajanus cajan* (L.) Millsp) is the major pulse crop of India after chickpea, which contributes significantly to the nourishment of a sizeable population. It belongs to the family 'Fabaceae' and is also known as 'Arhar' or 'Tur', generally used in preparing *dal* which is rich in protein (24-25%) and minerals and eaten by majority of the Indian vegetarian population. Pigeonpea is widely grown in the semi-arid tropics, particularly in the Indian subcontinent where it accounts for over 70% of the world's production and coverage (FAO 2007).

It can grow under low fertility and harsh conditions due to its ability to fix atmospheric nitrogen through biological nitrogen fixation up to 40 kg N ha<sup>-1</sup> (Nene 1987) which is useful in maintaining soil health through increasing nitrogen availability and microbial activities. It has also been reported that, pigeonpea root exudates dissolve phosphate containing rocks (e.g. phosphate rocks) to make P available for crop use (Ae et al. 1990). Thus, presently pigeonpea is proved to be a prime pulse crop in Asian continent.

The evaluation of the diversity existing in accessions is essential for understanding and fully utilizing its potential value (Singh et al. 1997). The evolution of pigeonpea through natural and human selection in diverse elevation zones and under different cropping systems has resulted in a wide variety of locally adapted landraces. These landraces have evolved over the years to fit into local cropping patterns and diverse end uses and represent a wide range of patterns of crop diversity. The knowledge of patterns of genetic variation of a crop species in any given region or country is very important for planning future germ plasm collection missions and for efficient utilization of the collected germ plasm in crop improvement programmes. Information on variability in crop plants is of much importance to the plant breeder, since it provides him with the necessary tool to work with (Bennet-Lartey & Ofori 1999) and the success of any breeding programme depends on the availability and use of genetic variation.

A promising method for cultivar identification and classification is offered by the numerical taxonomic approach. However, some other techniques of multivariate analysis

have been applied widely to discriminate and classify cultivars and clones of several plant species. For instance, the linear discriminant function approach is one tool that has been applied to discriminate and classify crop germ plasm (Barone et al. 1996). It has been used to study variation in seed morphology of 31 accessions of lima bean (Asante et al. 2006).

Keeping in view the importance of pigeonpea, the present experiments were conducted to study the cluster analysis for sorting out the accessions obtained from the different locality.

## MATERIALS AND METHODS

The 122 pigeonpea accessions were obtained from IIPR, Kanpur and experiment was laid out in augmented design under rainfed situation at Agricultural Research Station, Gulbarga, Karnataka, India, during 2012. The observations were recorded for the characters viz., days to 50 per cent flowering, plant height (cm), days to physiological maturity, pods per plant, pod bearing length (cm), seeds per pod, 100 seed weight (g) and seed yield per plant (g). The statistical analysis was carried out on K-means clustering and discriminate function analysis.

**Discriminant function analysis (Da):** The discriminant function analysis involves the determination of a linear equation like regression, that will predict which group the case belongs to. The form of the equation or function is (Fisher 1936):

$$DF = v_1X_1 + v_2X_2 + \dots + v_iX_i + a$$

Where, D = discriminate function

v = the discriminant coefficient or weight for that variable

X = respondent's score for that variable

a = a constant

i = the number of predictor variables

This function is similar to a regression equation or function. The v's are unstandardized discriminant coefficients analogous to the b's in the regression equation. These v's maximize the distance between the means of the criterion (dependent) variable and the standardized discriminant coefficients can also be used like beta weight in regression. Good predictors tend to have larger weights to maximize the distance between the categories, i.e. come up with an equation that has strong discriminatory power between groups. After using an existing set of data to calculate the discriminant function and classify cases, any new cases can then be classified. The number of discriminant functions is one less the number of groups.

**K-means clustering:** It is an algorithm to classify or to

Table 1: Number of cases included in each of the cluster (Cluster size).

Number of cases in each cluster	
Cluster 1	14
Cluster 2	40
Cluster 3	17
Cluster 4	51
Valid	122
Missing	0

Table 2: Cluster centers (average/mean) of different characters in 122 pigeonpea (*Cajanus cajan* (L.) Millsp) accessions.

Sl. No.	Characters	Cluster			
		1	2	3	4
1	Days to 50% flowering	94.29	92.43	138.98	73.96
2	Plant height (cm)	140.8	135.1	147.95	113.37
3	Physiological maturity	153.5	154.7	203.95	121.69
4	Pods per plant	183.64	123.68	98.93	71.43
5	Pods length (cm)	23.74	15.96	11.64	25.24
6	Seeds per pod	3.5	3.33	3.04	3.53
7	Seed weight (g)	9.64	9.6	12.44	8.43
8	Seed yield (g)	41.73	27.97	22.63	14.74

group the objects, based on attributes or features into K (positive integer number) number of groups. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus, the purpose of K-mean clustering is to classify the data. The goal is to partition the  $n$  entities into  $k$  sets  $S_p$ ,  $i=1, 2, \dots, k$  in order to minimize the within-cluster sum of squares (WCSS), defined as:

$$\sum_{j=1}^k \sum_{i=1}^n \|x_i^j - c_j\|^2$$

Where, the term  $\|x_i^j - c_j\|$  provides the distance between an entity point and the cluster's centroid.

The cluster centroid is the point with coordinates equal to the average values of the variables for the observations in that cluster. If the cluster centroid are not close to each other then classification is correct with no error.

## RESULTS AND DISCUSSION

K-means non-hierarchical clustering performed based on the mean values of the 8 different attributing characters of 122 accessions of pigeonpea and it is classified into 4 different clusters comprising of 14 accessions in the first cluster, 40 accessions in second cluster, 17 accessions in third cluster and in fourth cluster 51 accessions (Table 1). The centres for days to 50 per cent flowering revealed that cluster-I (94.29), cluster-II (92.43), cluster-III (138.98) and cluster-IV (73.96) found in all the accessions. For plant height cluster-I (140.80cm), cluster-II (135.10 cm), cluster-III

Table 3: Test of significance of different clusters with characters in 122 pigeonpea (*Cajanus cajan*) accessions.

Characters	Cluster		Error		F	Significance
	Mean Square	df	Mean Square	df		
Days to 50% flowering	18051.533	3	160.553	118	112.433	< 0.01
Plant height (cm)	7320.285	3	169.187	118	43.267	< 0.01
Physiological maturity	30245.412	3	247.028	118	122.437	< 0.01
Pods per plant	52874.051	3	431.183	118	122.626	< 0.01
Pods length (cm)	1145.481	3	90.454	118	12.664	< 0.01
Seeds per pod	1.149	3	0.356	118	3.229	0.025
Seed weight (g)	68.474	3	2.621	118	26.123	< 0.01
Seed yield (g)	3124.871	3	53.756	118	58.131	< 0.01

Table 4: Variation explained from the three canonical functions:

Function	Eigen value	% of Variance	Cumulative %	Canonical Correlation
1	6.144 <sup>a</sup>	71.2	71.2	0.927
2	2.384 <sup>a</sup>	27.6	98.8	0.839
3	0.106 <sup>a</sup>	1.2	100	0.31

a. First 3 canonical discriminant functions were used in the analysis.

Table 5: Significance test of discriminant functions.

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 3	0.037	384.463	12	< 0.01
2 through 3	0.267	154.408	6	< 0.01
3	0.904	11.791	2	0.003

(147.95 cm) and cluster-IV (113.37 cm) similarly, for the character physiological maturity was cluster-I (153.50), cluster-II (154.70), cluster-III (203.95) and cluster-IV (121.69). The centres of pods per plant were 183.64, 123.68, 98.93 and 71.43 in cluster-I, cluster-II, cluster-III and cluster-IV respectively.

For pod bearing length was cluster-I (23.74 cm), cluster-II (15.96 cm), cluster-III (11.64 cm) and cluster-IV (25.24 cm) and centres for seeds per pod as 3.50, 3.33, 3.04 and 3.53 were found in cluster-I, cluster-II, cluster-III and cluster-IV respectively. In the same way for 100 seeds, weight was cluster-I (9.64g), cluster-II (9.60g), cluster-III (12.44g), and cluster-IV (8.43g) and the mean values of seed yield were 41.73g, 27.97g, 22.63g and 14.74g in cluster-I, cluster-II, cluster-III and cluster-IV respectively. The centres are provided in Table 2. All clusters were found significant with respect to characters under study (Table 3). The stepwise discriminant analysis was carried out to find the significant characters for accessions and showed that, pods per plant and days to physiological maturity found to be highly discriminating characters and these two characters

exhibited 98.8 per cent variation in all the accessions (Table 4). The significance of the characters, pods per plant and days to physiological maturity, days to 50 flowering and, pod bearing length were tested with Wilks Lambda and these four characters were highly significant to enter the model in the first step (Tables 5 and 6). First discriminant function and third discriminant function and the second and third discriminant functions combinedly explained the variation among the significant characters (Table 5). The canonical correlation was found as high correlation for first (0.927), second (0.839) and less for third (0.310) discriminant functions (Table 7).

To predict the discriminant score, the model;

$$DF1 = 0.200 * DTF + 0.687 * DPM + 0.788 * PPP - 0.053 * PBL$$

$$DF2 = 0.603 * DTF + 0.332 * DPM - 0.771 * PPP + 0.536 * PBL$$

$$DF3 = 0.451 * DTF - 0.202 * DPM + 0.055 * PPP + 0.922 * PBL$$

Where, DF1, DF2, DF3 are the discriminant functions 1, 2 and 3, DTF-days to 50% flowering, DPM-days to physiological maturity, PPP-pods per plant, PBL-pod bearing length could be utilised (Table 8).

The classification matrix which summarizes the predictive ability of discriminatory functions and each accession, was assigned to a cluster by discriminant functions. The discriminatory function is particularly informative because, the misclassified accessions are identified and reassigned to the appropriate group. In first group 14 accessions were 100 per cent correctly classified, in the same way 39 accessions were 97.5 per cent correctly classified, but one accession (2.75 per cent) was misclassified in the first group. Similarly, 17 accessions were 100 per cent correctly classified in third group and in the fourth group 49 out of 51 accessions were correctly classified with 96.1 per cent and two accessions were misclassified (3.9%) in the second group. Overall classification of all the accessions in the respective groups was 97.5 per cent correctly classified (Table 9). The group centroid for each cluster is depicted in Figs. 1, 2, 3 and 4, and it is concluded that the four clusters' centroids are not close to each other, hence, all the accessions were

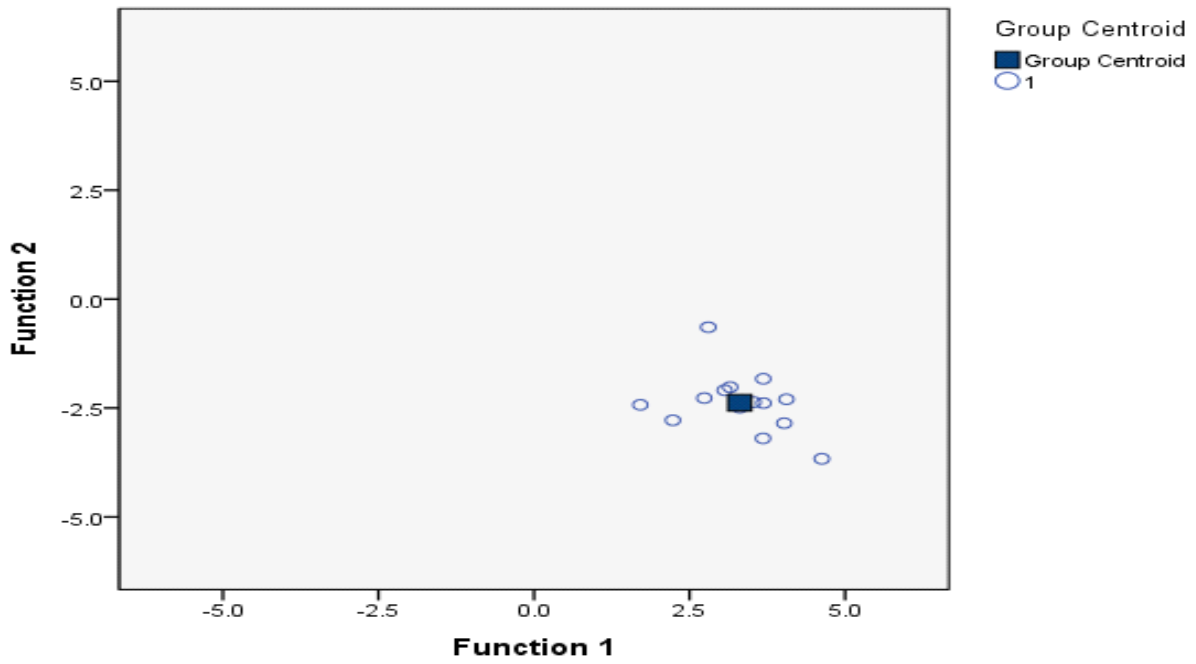


Fig. 1: Canonical discriminant function for group centroid for cluster-I.

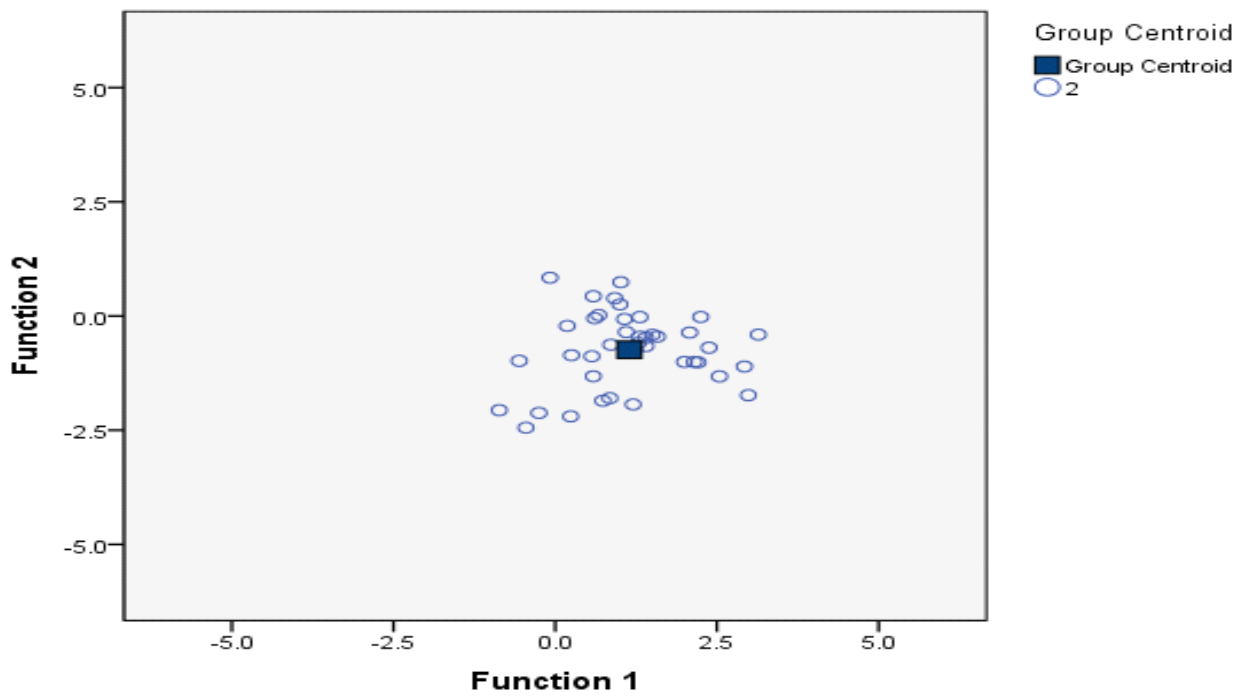


Fig. 2: Canonical Discriminant function for group centroid for cluster-II.

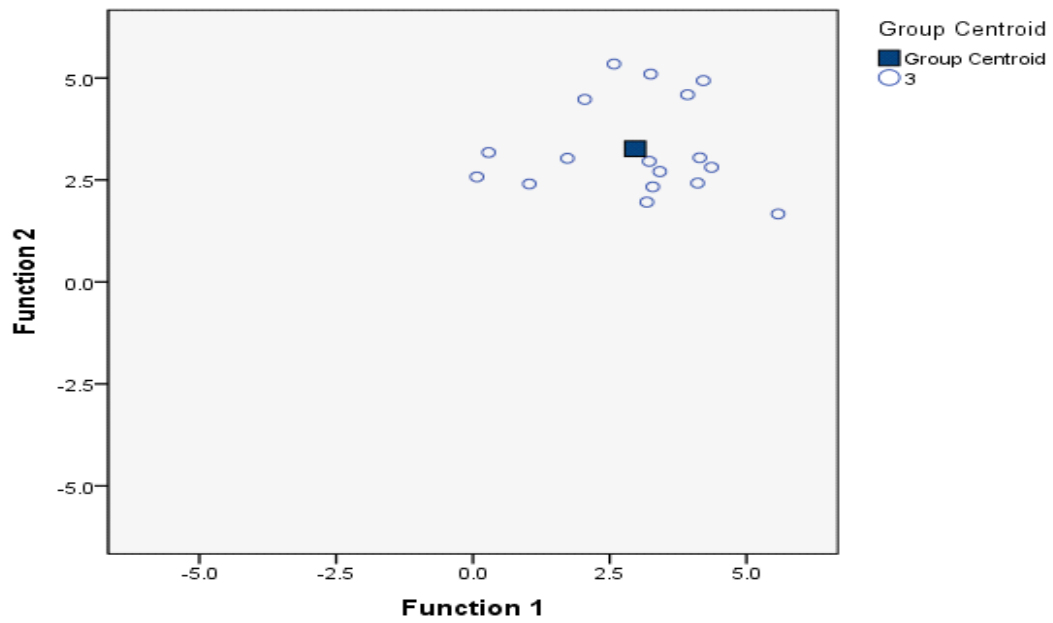


Fig. 3: Canonical Discriminant function for group centroid for cluster-III.

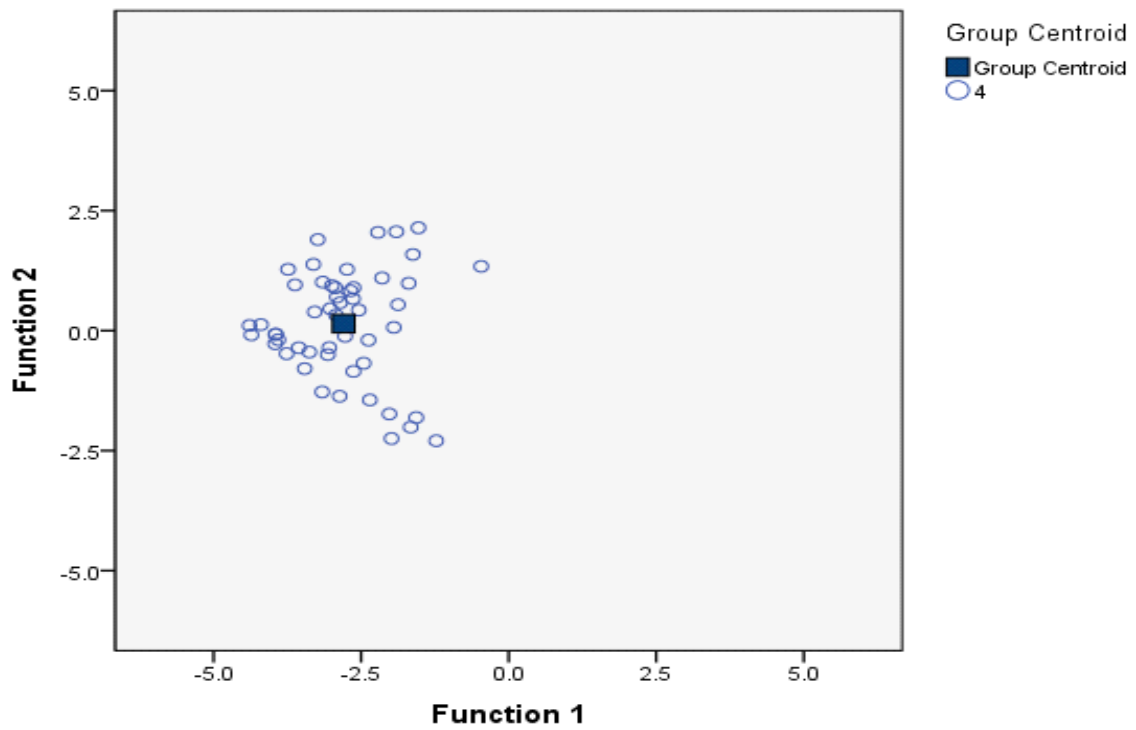


Fig. 4: Canonical Discriminant function for group centroid for cluster-IV.

Table 6: Stepwise discriminant function analysis.

Variables Entered/Removed		Wilks' Lambda										
Step	Entered	Exact F					Approximate F					
		Statistic	df1	df2	df3	Statistic	Sig.	Statistic	df1	df2	Sig.	
1	Pods per plant	0.243	1	3	118	122.6	< 0.01					
2	Days to physiological maturity	0.059	2	3	118	121.6	< 0.01					
3	Days to 50% flowering	0.045	3	3	118			80.619	9	282.464	< 0.01	
4	Pod bearing length	0.037	4	3	118			62.497	12	304.553	< 0.01	

Table 7: Standardized canonical discriminant function coefficients.

Character	Function		
	1	2	3
Days to 50 % flowering	0.2	0.603	0.451
Days to physiological maturity	0.687	0.332	-0.202
Pods per plant	0.788	-0.771	0.055
Pod bearing length	-0.053	0.536	0.922

Table 8: Classification result of predicted group.

Cluster number of case	Predicted group membership				Total
	1	2	3	4	
Original Count	1	14	0	0	14
	2	1	39	0	40
	3	0	0	17	17
	4	0	2	0	49
%	1	100	0	0	100
	2	2.5	97.5	0	100
	3	0	0	100	100
	4	0	3.9	0	96.1

classified correctly with no error in classification.

Number of pods per plant, number of primary branches per plant, plant height, days to flowering and days to maturity and their positive association with seed yield indicated that these are major yield contributing traits in pigeonpea (Rao et al. 2013).

In conclusion, the results from the study demonstrated that, the linear discriminant function is a useful tool for screening and evaluating the variation among the pigeonpea accessions and also reported that 122 pigeon pea accessions differed in days to 50 per cent flowering, pods per plant,

days to physiological maturity and pod bearing length. However, the level of variation existing within the accessions can make them suitable as source materials for future breeding programme (Cobbinah et al. 2012).

## REFERENCES

- Ae, N., Arihara, J., Okada, K., Yoshihara, T. and Johansen, V. 1990. Phosphorus uptake by pigeonpea and its role in cropping systems of the Indian subcontinent. *Science*, 284: 477-480.
- Asante, I. K., Addy, R. and Carson, A. G. 2006. Use of linear discriminant function analysis in seed morphotype relationship study in 31 Lima beans (*Phaseolus lunatus* L.) accessions in Ghana. *Ghana J. Agric. Sci.*, 39: 87-91.
- Barone, E., Di Marco, L., Marra, F. P. and Sidari, M. 1996. Isozymes and canonical discriminant analysis to identify pistachio (*Pistacia vera* L.) germplasm. *Hort. Sci.*, 31(1): 134-138.
- Bennet-Lartey, S.O. and Ofori, I. 1999. Variability studies in some qualitative characters of cowpea (*Vigna unguiculata* (L.) Walp) accessions from four cowpea growing regions of Ghana. *Ghana J. Agric. Sci.*, 32: 3-9.
- Bolanga, C. J. and Timko, M. P. 2005. An improved genetic linkage map for cowpea (*Vigna unguiculata* (L.) Walp). Department of Biology, University of Florida, Florida, USA.
- Cobbinah, F. A., Asanteand, I. K. and Addo-Quaye, A. A. 2012. Use of linear discriminant function analysis in five yield sub-characters relationship study in 134 Cowpea (*Vigna unguiculata* (L.) Walp) accessions. *West African Journal of Applied Ecology*, 20(2): 19-22.
- FAO 2007. <http://faostat.fao.org>
- Fisher, R.A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics.*, 7(2): 179-188.
- Nene, Y. L. 1987. Overview of pulses research at ICRISAT. Adaptation of chickpea and pigeonpea to abiotic stresses. *Patancheru*, 502(324): 7-12.
- Rao, P. J. M., Malathi, S., Reddy, D. V. V. and Upender, M. 2013. Genetic studies of association and path coefficient analysis of yield and its component traits in Pigeon pea (*Cajanus Cajan* L. Millsp.) *International Journal of Scientific and Research Publications*, 3(8): 1-5.
- Singh, B.B., Mohan Raj, D.R., Dashiell, K.E. and Jackai, L.E.N. (eds.) 1997. *Advances in Cowpea Research*. IITA-JIRCAS, Ibadan, Nigeria.