Forage Res., **40** (1) : pp. 7-13 (2014) **http://forageresearch.in**

EVALUATION OF PHENOTYPIC VARIABILITY IN *SESBANIA* USING MULTIVARIATE ANALYSES

O. P. YADAV* AND SAWINDER SINGH

Department of Genetics & Plant Breeding
CCS Haryana Agricultural University,
Hisar-125 004 (Haryana), India
**(opyadav@hau.ernet.in)*

(Received : 14 March 2014; Accepted : 19 May 2014)

SUMMARY

Principal factor analysis was done in *Sesbania* (using 57 genotypes and 17 variables) which identified five principal factors explaining 70.97 per cent variability altogether. The principal factor analysis without rotation failed to provide clear picture, hence, varimax rotation was applied which resulted in clear cut loading of almost similar type of variables on a common principal factor permitting to designate them as grain yield factor, green manuring factor, growth rate factor, etc., according to the type of variables it is loaded with. Genotypes LJ 32, Ses H 3, LJ 37, Ses H 22 and Ses H 33 were found to be better performers with regard to grain yield and its components when all the principal factors were considered simultaneously. These genotypes can further be utilized in breeding programmes for improving grain yield. Further, hierarchical cluster analysis (UPGMA with city block distances) was carried out to classify these 57 *Sesbania* genotypes on the basis of 17 grain yield and green manuring related variables which resulted in formation of seven clusters having 1 to 14 genotypes. The results of hierarchical cluster analysis and principal factor analysis confirmed the findings of each other.

Key words : Cluster analysis, *Sesbania*, germplasm, principal factor analysis

Sesbania belonging to the family Leguminosae and found throughout the tropics is chiefly used as green manure for low-land transplanted paddy and sugarcane fields. With the alarming depletion of natural resources, continuing energy crisis, growing ecological concern, problems of unbalanced or excessive/inadequate supply of fertilizers and fear of high fertilizer prices, there is a renewed interest in utilization of green manure for sustaining the soil fertility and productivity of crops.

Incorporation of *Sesbania* in the soil adds 60-80 kg of nitrogen and also has buffering action, since it contains large amount of succulent organic matter with a low carbon : nitrogen ratio. This allows quick liberation of atmospheric nitrogen in the available form. This also helps to improve the soil physical structure, prevent leaching and loss of nutrients conserve soil moisture and create access to deep soil layers and prevents growth of weeds. Utilization of genetic resources in developing sustainable solutions to basic crop constraints has been suggested from time to time but these genetic resources could not be exploited fully due to their inherent problems of large size and lack of sufficient evaluation and classification (Dahlberg, 1995). For effective utilization of these resources, it is necessary to evaluate and characterize them (Beuselinck and Steiner, 1992; Ordas *et al.*, 1994). Hamman (1972) suggested that the use of multivariate techniques could resolve several phenotypic measurements even of large collections into fewer, more interpretable and more easily visualized dimensions. As the investigator, initially unaware of the relative importance of variables, tries to include all the possible variables which are likely to have some connection with the problem and makes the resultant data matrices unmanageable and complicated. Principal component analysis helps in identifying most relevant characters that can be used as descriptors by explaining as much of the total variation in the original set of variables as possible with as few of the components as possible and reduces the complexity or dimension of the problem (Johnson and Wichern, 1988). Further the collections that have not been systematically characterized can contain duplicated accessions or too many unique or rare types (Stiener and Poklemba, 1994). Cluster analysis offers solution to this problem by defining degrees of relatedness in the gene bank samples and the best basis to define commonness, thereby eliminating redundancy

and characterizing degree of diversity (Peeters and Martinelli, 1989; Ordas *et al.*, 1994; Smith *et al.*, 1995).

Therefore, the present investigation was undertaken in *Sesbania* with the objectives of evaluation, categorization and classification of germplasm and computation of principal factors to determine the degree of similarity among the genotypes and relative importance of the principal factors and characters involved in them.

MATERIALS AND METHODS

The experiment involving 57 genotypes was grown in randomized block design during **kharif** season of 2001. Observations were recorded on 17 variables viz., green weight (45 and 60 DAS), dry weight (45 and 60 DAS), nodule count (45 and 60 DAS), nodule weight (45 and 60 DAS), grains/pod, pods/cluster, clusters/plant, pod length, grain yield/plant, growth rates at 45, 60 DAS and at maturity and hard seed percentage.

To evaluate the growth rate, plant height was recorded in centimeters at 15, 45, 60 days after sowing (DAS) and at maturity. Growth rate per day was calculated by dividing the height gained during that period by initial height of that particular growth period. For evaluating hard seed percentage 100 seeds from each genotype taken in three replications were put to standard germination test and number of hard seeds was counted.

Data Analyses

Average of the data recorded on all the five plants was computed for all the characters. Principal factor and cluster analyses were carried out on 57 genotypes and 17 variables using computer programme SPSS.

Principal factor analysis was carried out as it has many added advantages over principal component analysis. It is closely related to principal component analysis, but differs in that it assumes a definite model, where each observed variable is expressed linearly in terms of common factor and unique factor. The common factors account for the correlation among the variables, while each unique factor accounts for the remaining variance (including error) of that variable. Moreover, in principal component analysis total variation contained in a set of variables is considered, whereas in factor analysis interest centers only on that part of variance, which is shared by the common factors. Principal component method was used for factor extraction as it does not require assumption of normal distribution of population. For deciding number of principal factors to be retained, Kaiser's (1958) suggestion of dropping those principal factors with eigen roots less than one, was followed. As the initial factors loading were not clearly interpretable, the factor axes were rotated using Varimax rotation (Kaiser, 1958). Principal factor scores were calculated using Anderson-Rubin method in SPSS.

Unweighted pair-group method using arithmetic averages (UPGMA) method of hierarchical cluster analysis was utilized with city block distances to classify 57 genotypes and dendogram was prepared using the rescaled distances. Based on the method suggested by Romesburg (1984) the dendogram was cut to form the clusters.

RESULTS AND DISCUSSION

In the present investigation, principal component (PC) method was used to extract the principal factors (PF) and the principal components which had eigen values greater than one were retained (Kaiser, 1958). The first five principal components showed eigen values more than one and they altogether explained 70.97 per cent cumulative variability (Table 1). The first PC explained 27.5 per cent of the total variation. The second, third, fourth and fifth principal components explained 16.5, 11.8, 8.1 and 7.1 per cent variation, respectively. The first one absorbed and accounted for maximum proportion of total variability in the set of all variables and the remaining ones accounted for progressively lesser and lesser amount of variation. Veasey *et al.* (2001) conducted principal component analysis in *Sesbania* and obtained similar trend.

TABLE 1
Total variance explained by different principal components

Principal components	Eigen value	Variability (%)	Cumulative variability (%)
1	4.683	27.548	27.548
2	2.803	16.489	44.037
3	2.005	11.796	55.833
4	1.371	8.065	63.898
5	1.202	7.071	70.968

Initially the data were analyzed without any rotation to derive clear picture of interaction of variables among themselves and with the principal factors. But it

failed to provide much information regarding the idea of correlation between the variables and the principal factors. To select the relevant characters, those correlation values ≥0.6 were considered as relevant for that principal factor (Matus *et al.*, 1999). Factor loading of different variables without rotation indicated that the three out of the five principal factors were having high loading of nine variables in total and the last two principal factors were having high loading for none. Also, eight out of the 17 variables were left without being highly loaded on any of the principal factors.

The failure of principal factor analysis without rotation to draw sensible conclusions prompted to go for analysis with rotation. Factors loadings of different variables obtained through Varimax rotations are presented in Table 2. Fifteen variables showed high loading on different principal factors and two were left after rotation of the principal factor axes. Moreover, it clearly grouped the similar type of variables by loading them together on a common principal factor. The first principal factor (PF) showed high loading for five grain yield variables i. e. grains/pod, growth rate at maturity, pods/cluster, pod length and yield/plant and thus can easily be designed as grain yield factor. The principal factors 2, 4 and 5 ascribed for six variables in total related to green and dry weight at different stages, and growth rate at 45 and 60 DAS. The PF 3 showed high loading for nodule parameters i. e. nodule count and nodule weight at 45 and 60 DAS and can be designated as nitrogen fixing factor. The clear cut grouping of similar type of variables by getting loaded on common principal factor elaborates the successful transformation of 17 interrelated variables into five independent principal factors explaining 70.97 per cent of the variability of the original set. Application of factor analysis in *Sesbania* could not be traced in the literature to endorse the results of the present investigation.

Principal factor scores (PF scores) for all the 57 genotypes were estimated in all the five factors. These scores can be utilized to propose precise selection indices whose intensity can be decided by variability explained by each of the principal factors. Using these scores, all the genotypes (Names and numbers of genotypes are given in Fig. 3) were plotted for PF 1 and PF 2 and then for PF 1 and PF 3 which cumulatively explained 55.8 per cent variability and accounted for the most important characters (Figs. 1 and 2). These plots clearly indicated the separation of high yielding genotypes towards the positive side of PF 1 axis, which is factor for grain yield. The genotypes LJ 35, LJ 32, Ses H 3, LJ 37, Ses H 22 and Ses H 33 which stand out towards the positive position of factor 1 axes in both the plots were found to be having high yield/plant and genotypes Ses H 28, Ses H 36, Ses H 43, PDCSR-1 and LJ 10 had good nitrogen

TABLE 2
Factor loadings of different characters with respect to different principal factors (varimax rotation)

Characters/principal factors	PF 1	PF 2	PF 3	PF 4	PF 5
Grains/pod	0.729*	0.226	0.038	-0.165	-0.032
Growth rate at maturity	0.728*	-0.320	-0.043	-0.278	0.072
Pods/cluster	0.654*	0.254	0.165	0.049	0.224
Pod length	0.628*	0.104	0.093	0.235	-0.105
Yield	0.581*	-0.012	0.163	0.052	0.524
Clusters/plant	0.519	0.378	-0.101	0.181	-0.448
Green weight 60 DAS	0.123	0.898*	0.170	0.095	0.187
Dry weight 60 DAS	0.114	0.878*	0.110	0.130	0.222
Nodule weight 45 DAS	-0.012	-0.038	0.778*	0.269	-0.110
Nodule count 45 DAS	0.081	0.160	0.745*	0.336	0.004
Nodule weight 60 DAS	0.189	0.530	0.659*	-0.068	0.069
Nodule count 60 DAS	0.172	0.530	0.575*	-0.191	0.086
Green weight 45 DAS	-0.030	0.073	0.179	0.948*	-0.068
Dry weight 45 DAS	0.008	0.050	0.165	0.947*	-0.080
Growth rate 60 DAS	-0.173	0.269	0.052	-0.205	0.704*
Growth rate 45 DAS	-0.222	-0.256	0.320	-0.005	-0.691*
Test weight	0.415	0.173	0.432	0.027	0.440

*Significant at P=0.05 level.

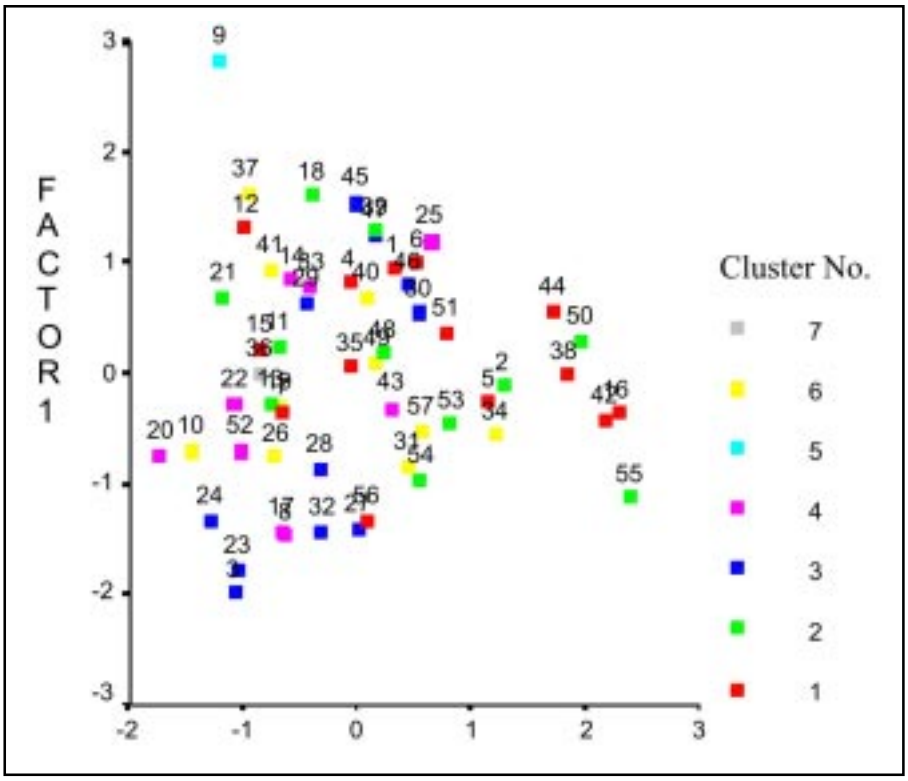

Fig. 1. Location of all entries based on PF scores w. r. t. factors 1 & 2.

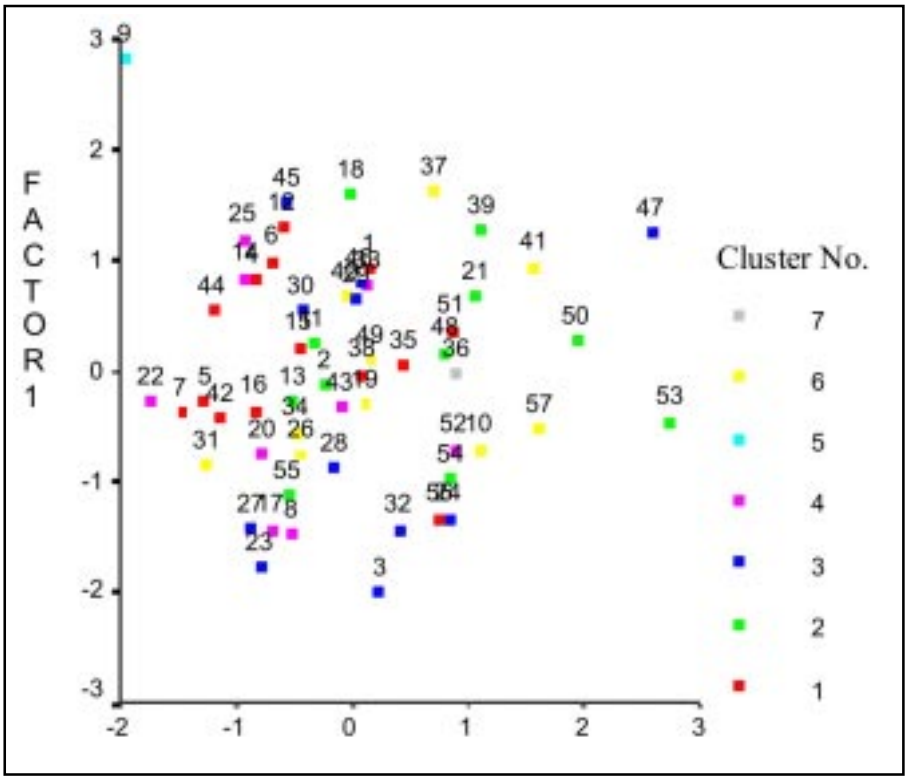

Fig. 2. Location of all entries based on PF scores w. r. t. factors 1 & 3.

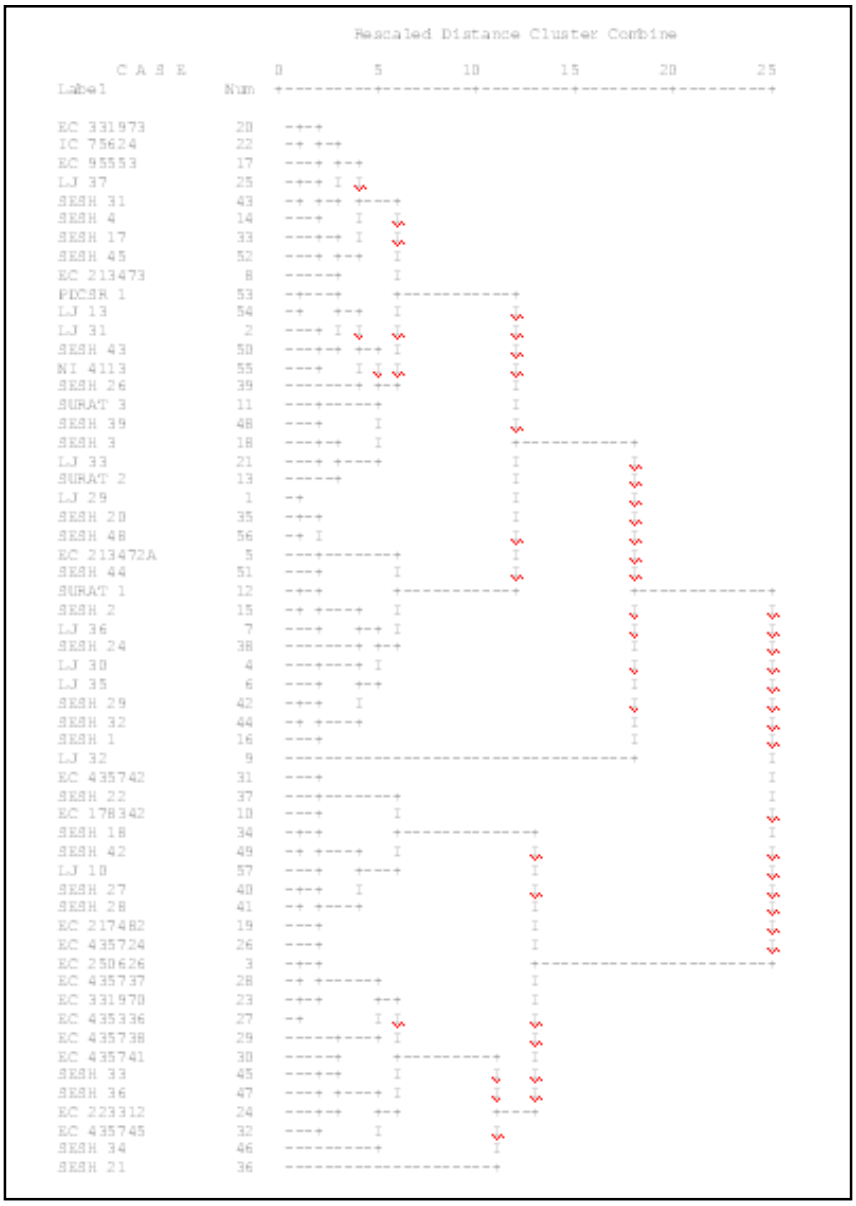

Fig. 3. Dendorgram portraying clustering pattern of different genotypes.

fixing ability. So, it can be concluded that these genotypes may prove better parents in hybridization programme, when all the factors are considered because these genotypes are repeated invariably in all the combinations.

The hierarchical cluster analysis identified seven clusters containing between 1 and 14 genotypes (Table 3). The cluster I was having the maximum number of genotypes i. e. 14 and the clusters V and VII were having the minimum i. e. one each. The clusters II, III, IV and VI comprised 11, 11, 9 and 10 genotypes, respectively. The association among the different genotypes is presented in the form of dendogram (Fig. 3) prepared using rescaled distances. The resemblance coefficient between the two genotypes is the value at which their branches join. The dendogram also showed the relative magnitude of resemblance among the different clusters. In the present study, the mean performance of different clusters calculated for different characters revealed wide range of differences among clusters with respect to these traits (Table 4). The cluster I comprised accessions mainly with medium green weight 45 DAS, dry weight 45 DAS, nodule count 45 and 60 DAS, nodule weight

TABLE 3
Cluster membership profile of different genotypes (UPGMA–City block distance)

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
LJ 29	LJ 31	EC 250626	EC 213473	LJ 32	EC 178342	SES H 21
LJ 30	Surat 3	EC 331970	SES H 4		EC217482	
EC 213472	Surat 2	EC 223312	EC 95553		EC 435724	
LJ 35	SES H 3	EC 435336	EC 331973		EC 435742	
LJ 36	LJ 33	EC 435737	IC 75624		SES H 18	
Surat 1	SES H 26	EC 435738	LJ 37		SES H 22	
SES H 2	SES H 39	EC 435741	SES H 17		SES H 27	
SES H 1	SES H 43	EC 435745	SES H 31		SES H 28	
SES H 20	PDCSR 1	SES H 33	SES H 45		SES H 42	
SES H 24	LJ 13	SES H 34			LJ 10	
SES H 29	NI 4113	SES H 36				
SES H 32						
SES H 44						
SES H 48						
14	11	11	9	1	10	1

TABLE 4
Cluster means for different characters in *Sesbania* (UPGMA – City Block)

Character	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Green weight 45 DAS	53.34	56.94	59.15	47.46	26.67	57.33	34.30
Green weight 60 DAS	319.06	295.95	231.59	199.67	195.37	230.47	200.87
Dry weight 45 DAS	13.38	13.63	14.34	11.37	6.27	14.29	8.90
Dry weight 60 DAS	72.13	68.54	50.08	42.91	37.47	45.23	38.50
Nodule count 45 DAS	20.55	26.60	20.49	16.82	4.43	22.35	24.53
Nodule count 60 DAS	45.13	50.92	32.78	27.68	29.50	37.45	27.67
Nodule weight 45 DAS	0.51	0.89	0.73	0.53	0.02	0.88	0.91
Nodule weight 60 DAS	1.16	1.54	0.82	0.85	0.43	1.11	1.13
Pod length	18.83	18.57	17.92	17.53	21.10	19.06	19.30
Clusters/plant	52.80	55.46	51.43	48.91	58.83	62.30	53.30
Pods/cluster	3.45	3.58	3.09	3.32	3.70	3.11	3.33
Grains/pod	25.52	25.09	24.52	22.68	29.73	23.79	22.07
Test weight	1.63	1.53	1.41	1.41	1.54	1.53	1.39
Growth rate 60 DAS	242.81	210.62	174.02	214.77	232.40	203.93	188.17
Growth rate at maturity	149.20	144.40	138.90	146.69	293.17	146.10	157.73
Yield/plant	59.10	61.91	42.21	60.33	85.33	41.83	52.00

45 and 60 DAS, pod length, clusters/plant, pods/cluster, growth rate at maturity and yield/plant, whereas highest green weight 60 DAS, dry weight 60 DAS, nodule weight 60 DAS, and test weight and growth rate 60 DAS.

The genotypes of cluster II were characterized by highest nodule count 45 and 60 DAS, with medium green weight (45 and 60 DAS), dry weight (45 and 60 DAS), pod length, clusters/plant, pods/cluster, grains/pod, test weight, growth rates (60 DAS and maturity) and yield/plant. Clusters III and VI were characterized by genotypes with higher green weight 45 DAS, dry weight 45 DAS and clusters/plant with moderate green weight and dry weight 60 DAS, nodule count 60 DAS, nodule weight (45 and 60 DAS), clusters/plant and test weight. But genotypes had lower growth rate (60 DAS and maturity), yield/plant and pods/cluster. Cluster IV had accessions with mostly moderate to poor morphological, green manuring and seed yield characters. Thus, these genotypes were average to below average performers. Cluster V comprising only one genotype was having the highest grain yield, with maximum yield/plant pod length, pods/cluster, grains/pod and growth rate at maturity. But this genotype exhibited poor vegetative growth and green manuring characters. It

had lowest green and dry weight (45 and 60 DAS) and nodule count and weight (45 and 60 DAS). Thus, this genotype was good only for grain purpose but not for manuring purpose. Cluster VII also comprised one genotype with lowest test weight and nodule count 60 DAS with highest nodule weight 45 DAS, with most of other morphological characters being moderate.

Based on the results of the present study it was recommended to use LJ 32 as one of the parents for improving yield and component traits. However, for improvement of specific traits like green manuring and nitrogen fixation, genotypes from cluster II should be involved. The hybridization among diverse parents is likely to produce heterotic hybrids and desirable transgressive segregants in further generations. Romesburg (1990) opined that findings of similar alternatives reduced the decision problem to two stages i. e. first, to select the cluster that can best achieve the planning objective, and second select the best alternative within the best cluster.

The results of hierarchical cluster analysis and principal factor analysis confirmed the findings of each other. The plots of PF 1-PF 2 and PF 1-PF 3 accounting for about 55 per cent variation (Figs. 1 and 2) showed clear differentiation of genotypes according to their cluster membership denoted by different colours. Genotypes belonging to a common cluster have fallen nearer to each other and *vice-versa* thereby confirming the results of clustering. In other plots, the genotypes have intermingled as they accounted for lesser variability and the clusters were formed on the basis of total variability. Similarly, the genotypes like LJ-32, etc. found to be superior using principal factor analysis were also found to be members of the best performing clusters i. e. cluster V. Such confirmatory results were also obtained by Bisht *et al.* (1998) in greengram.

Hence, the present study has been proved to be successful in classifying different genotypes based on various morphological characters, reducing large number of variables into only five principal factors and identifying different genotypes better for different combinations of characters. The results of the present study can be used as a stepping stone for evolving well defined approach based on evaluation and characterization of genetic variation in *Sesbania* and can be utilized in various breeding programmes depending on their specific objectives.

REFERENCES

Beuselinck, P. R., and J. J. Steiner, 1992 : A proposed framework for identifying, quantifying and utilizing plant germplasm resources. *Field Crops Res.,* **29** : 261-272.

Bisht, I. S., R. K. Mahajan, and T. G. Kawalkar, 1998 : Diversity in greengram [*Vigna radiata* (L.) Wilczek] germplasm collection and its potential use in crop improvement. *Ann. Appl. Biol.,* **132** : 301-312.

Dahlberg, J. A. 1995 : *Operation Handbook for Sorghum Germplasm Maintenance*. ARS Special Publication, Beltsville, MD, USA.

Hamman, H. K., 1972 : Utilization of multivariate statistics for discriminating among flue-cured tobacco varieties. Technical Bulletin Number 212, North Carolina Agricultural Experimental Station (Raleigh, NC).

Johnson R. A., and D. W. Wichern, 1988 : *Applied Multivariate Statistical Analysis*. Prentice Hall Inc., Englewood Cliffs, New Jersey.

Kaiser, H. F. 1958 : The varimax criterion for analytic rotation in factor analysis. *Psychometrika,* **23** : 187.

Matus, I., M. I. Gonzalez, and A. Del Pozo, 1999 : Evaluation of phenotypic variation in Chilean collection of garlic (*Allium sativum* L.) clones using multivariate analyses. *Plant Genet. Resour. Newsl.* **117** : 31-36.

Ordas, A., R. A. Malvar, and A. M. de Ron, 1994 : Relationship among American and Spanish populations of maize. *Euphytica,* **79** : 149-161.

Peeters, J. P., and J. A. Martinelli, 1989 : Hierarchical cluster analysis as a tool to manage variation in germplasm collections. *Theor. Appl. Genet.* **78** : 42-48.

Romesburg, H. C. 1984 : *Cluster Analysis for Researchers.* Krieger Publishing Co., Malabar, Florida.

Ramesburg, H. C. 1990 : *Cluster Analysis for Researchers.* Krieger Publishing Co., Malabar, Florida.

Smith, S. E., L. Guarino, A. Al-Doss, and D. M. Conta, 1995 : Morphological and agronomic affinities among Middle Eastern alfalfa accessions from Oman and Yemen. *Crop Sci.,* **35** : 1188-1194.

Stiener, J. J., and C. J. Poklemba, 1994 : *Lotus corniculatus* classification by seed globulin polypeptides and relationship to accession pedigree and geographic origin. *Crop Sci.,* **34** : 255-264.

Veasey, E. A., E. A. Schammass, R. Vencovsky, P. S. Martins, and G. Bandel, 2001 : Germplasm characterization of *Sesbania* accessions based on multivariate analyses. *Genet. Resour. and Crop Evol.* **48** : 79-90.