

**MULTIVARIATE ANALYSIS AND CLUSTERING OF *CUPHEA*
PROCUMBENS INBRED LINES**

R. Ramesh KUMAR and S.P. SINGH

Department of Genetics and Plant Breeding,
National Botanical Research Institute (CSIR), Lucknow, India.

Kumar S. R. Ramesh and S. P. Singh (2006): *Multivariate analysis and clustering of Cuphea procumbens inbred lines*. - *Genetika*, Vol. 38, No. 1, 23-30.

15 *C. procumbens* inbred lines of different eco-geographical origin maintained at National Botanical Research Institute, Lucknow, India were evaluated for seed yield and its 6 component traits to assess genetic divergence among them. The standardized mean values of different traits were subjected to principal component analysis and cluster analysis was performed based on two different clustering strategies *i.e.* UPGMA and Wards. Ward's method that showed relatively high cophenetic correlation coefficient and significant Wilk's Lambda was identified as the best clustering solution. The first four principal components (PC) with eigenvalues >1 contributed 91.56% of variability among the inbreds. First PC was related with fruits/plant and branches/plant; second PC with yield/plant, seeds/fruit and test weight; third PC with plant height; and fourth PC with days to flowering. The genotypes were grouped into five clusters and cluster II was largest with 5 genotypes followed by clusters I, III and Clusters IV, V. Cluster IV exhibited highest mean for seed yield (14.77g) followed by cluster III (14.53g) and the former incorporated inbred lines, NBCP-53 and NBCP-58 that were highly divergent among themselves and from genotypes in other clusters. The inbreds in cluster IV and cluster III

with good amount of genetic divergence and superior agronomic traits were identified as promising inbreds to develop superior recombinants with desirable agronomic traits.

Key words: *Cuphea procumbens*, Multivariate analysis, Ward's Method, Cophenetic correlation coefficient, Wilk's Lambda

INTRODUCTION

C. procumbens of family Lythraceae is the most promising species under Indian conditions (ANONYMOUS, 1998) and is rich (87.7 - 93.6%) in capric acid ($C_{16:0}$) (PANDEY *et al.*, 2000) that has high nutritive and medicinal value imparting potentiality to the species to be developed as new oilseed crop. *C. procumbens* is a cross-pollinated species (HIRSINGER and KNOWLES, 1984) and is characterized by highly heterozygous and heterogenous population.

Exploitation of heterosis has been the major objective in cross-pollinated crops for which hybrid development or population improvement are commonly adapted breeding procedures. In both the cases the homozygosity attained through initial selfing is converted to heterozygosity at the end of breeding programme. Selection of genetically diverse parents with high seed yield and yield components is important to improve the desirable traits in the genetic make up of the resulting population. So the present investigation was undertaken to determine the degree of divergence among *C. procumbens* inbred lines through multivariate analysis and identification of suitable parents for hybridization programme.

MATERIAL AND METHODS

C. procumbens accessions obtained from United States Department of Agriculture (USDA), Plant Introduction Center, Ames, USA, were evaluated for agronomical traits based on which 15 superior accessions were identified. These accessions were subjected to five generations of inbreeding and the resulting inbreds were evaluated in a randomized block design with three replications during 2003-04 at National Botanical Research Institute, Lucknow, India, situated between 26° 40' N latitude and 80° 45' E longitude and at an altitude of 129 m above sea level. Spacing adopted was 60 x 30 cm with 3 m long rows. NPK @ 90 Kg, 40 Kg and 40 Kg per hectare was applied. 10 plants were randomly selected in each plot/replication and observations on 7 characters i.e. days to flowering, plant height (cm), number of branches/plant, number of fruits/plant, number of seeds/fruit, test weight (g) and seed yield/plant (g) were recorded.

The mean observations of yield and yield traits of each inbred were standardized by subtracting from each observation, the mean value of the character and then dividing by its respective standard deviation so as to eliminate the biasness due to the different magnitudes of the units of measurement. The standardized values were used to perform principal component analysis utilizing Ky plot statistical programme. The genotypes were clustered based on two hierarchical

cluster algorithms, UPGMA and minimum variance method of WARD (1963), both of them well described in the related literature (KAUFMAN and ROUSSEUW, 1990). Briefly, the UPGMA method computes the distance between two groups, i and j , as the arithmetic mean of the $n_i \times n_j$ distances among all individuals of the two groups. The ward method optimizes an objective function; that is, it minimizes sum of squares within groups and maximizes the sum of squares between groups.

Best clustering strategy among these two, was identified based on cophenetic correlation coefficient and Wilk's Lambda criteria. Cophenetic correlation coefficient was obtained by estimating the correlation between the original distance matrix (Euclidean square distance among genotypes) and the cophenetic distance matrix (the distance among inbreds based on dendrogram tree obtained through different clustering methods). The clustering solution providing cophenetic correlation coefficient value closer to 1 was identified as the best strategy.

The analysis of the multivariate data variability for 7 traits was done expressing matrices of sum of squares and cross products corresponding to the total variability, T , the within groups variability, W , and the between groups variability, B . The different matrix notations have been described by FRANCO *et al.*, (1997). The determinants IWI and ITI were then used to compute Wilks's Lambda $[\Lambda]$, which is given by following:

$$[\Lambda] = IWI / ITI$$

$[\Lambda]$ was then converted to F value by means of set of procedures developed by RAO (1973). The calculated F value was compared against tabulated F value for significance.

RESULTS AND DISCUSSION

Analysis of variance showed significant differences among the inbred lines for all the characters (data not shown) indicating that further analysis of principal components and clustering of inbreds are worthwhile.

Principal Component Analysis:

The first four principal components (PC) with eigenvalues >1 contributed 91.56% of the variability amongst the inbreds evaluated for 7 quantitative traits (Table 1). Other PCs (5-7) had eigenvalues <1 . The first PC was more related to component traits like fruits/ plant and branches/ plant with PC loadings of 0.961 and 0.919 respectively. The second PC was more related to yield (0.870) and seed specific component traits like test weight (0.777) and seeds/ fruit (0.430). Strong positive correlation and high direct path of these seed related components towards yield (RAMESH KUMAR and SINGH, 2003) might have resulted in high scores of these traits in PC2 relevant to yield. The third and fourth PCs were found to be related to phenological trait (plant height) and developmental trait (days to flowering) respectively.

Table 1. - Principal Component Loadings of 7 *C. procumbens* agronomic traits on the first four principal components (PC)

Traits	PC 1	PC 2	PC 3	PC4
Days to Flowering	0.525	0.021	-0.394	0.630
Plant Height (cm)	-0.024	-0.431	0.826	-0.031
Branches/ Plant	0.919	0.073	0.179	0.092
Fruits/ Plant	0.961	-0.038	0.150	-0.093
Seeds/ Fruit	-0.641	0.430	0.271	0.558
Test Weight (g)	0.014	0.777	-0.116	-0.530
Yield/ Plant (g)	0.245	0.870	0.380	0.156
Eigenvalues	2.515	1.739	1.124	1.031
% of Variance	35.930	24.840	16.060	14.730
Cum. % of Variance	35.930	60.770	76.830	91.560

Table 2. - 15 *C. procumbens* inbred lines along with their origin, principal component (PC) scores and cluster grouping

S. No.	Inbred Lines	Origin	Cluster Number	PC Scores	
				PC1	PC2
1	NBC51	Mexico	I	-0.103	-0.018
2	NBC52	Mexico	III	0.267	0.724
3	NBC53	Mexico	IV	2.190	-0.553
4	NBC54	Mexico	V	1.356	-1.140
5	NBC55	Mexico	V	0.698	-0.733
6	NBC56	United States	I	-0.574	-0.601
7	NBC57	Germany	II	-0.998	-0.858
8	NBC58	Germany	IV	1.170	0.813
9	NBC59	Germany	II	-0.764	0.676
10	NBC60	Germany	III	-0.440	2.070
11	NBC61	Germany	III	0.390	1.577
12	NBC62	Unknown	II	-0.778	-0.952
13	NBC63	United States	I	-0.160	-0.219
14	NBC64	Hungary	II	-1.175	-1.130
15	NBC65	Mexico	II	-1.078	0.343

The origin of the 15 inbred lines along with their grouping and scores for first two PCs are presented in Table. 2. With respect to first and second PC that is related to yield and its component traits, NBCP - 58, which was identified to be the high yielding inbred line (RAMESH KUMAR *et al.*, 2002) showed moderate PC scores ranking III in both. NBCP-60 and NBCP-61, the two high yielding genotypes following NBCP-58 (RAMESH KUMAR *et al.*, 2002) ranked I and II respectively in PC2. This shows the significant contribution of component traits particularly seed specific components in imparting high yielding ability to the genotypes.

Cluster Analysis:

Both the clustering methods grouped the 15 genotypes in 5 clusters. However, UPGMA strategy formed two single genotype clusters of inbred lines NBCP-53 and NBCP-58, which were grouped in to a common cluster (cluster IV)

in Wards method. Comparison of the two clustering methods utilizing cophenetic correlation coefficient and Wilk's Lambda criteria showed that Wards method is the ideal strategy. Wards method showed a higher cophenetic correlation value (0.71) than UPGMA strategy (0.69), indicating that the cophenetic distance of Wards has a strong correlation with the original Euclidean square distance.

The F value of the converted Wilk's Lambda in Ward's method was 83.64, which was highly significant indicating that the different cluster groups are significantly different across the yield and yield attributes. The F value for UPGMA could not be estimated due to negative within group sum of squares W, resulting in negative value for determinant IWI, which may be due to single genotype clusters.

Ward's method with relatively high cophenetic correlation coefficient and significant Wilk's Lambda was identified as the best clustering solution for grouping of the inbred lines (Figure 1). The 15 inbreds were grouped in to 5 clusters showing wide diversity among them. The cluster II was largest containing five inbred lines followed by clusters I, III (3 inbreds each) and clusters IV, V (2 inbreds each). 6 Mexican genotypes were distributed in all the clusters and five genotypes from Germany diversified in 3 clusters (Table 2). Two genotypes from United States were grouped in cluster I and the genotype from Hungary was included in cluster II. The tendency of genotypes belonging to different origin, grouping in same cluster shows that geographical isolation is not directly related to genetic diversity and similar observations in *C. procumbens* were made by SINGH and SINGH, 2004. *Cuphea* being a rich annual source of medium-chain fatty acids, incorporating wild traits (SINGH and RAMESH KUMAR, 2003) has been distributed to different countries for domesticating in to a potential oilseed crop. During the process of domestication that involved selection, induced mutation and hybridization along with natural selection, out crossing and spontaneous mutation, might have caused genetic diversity among the genotypes of common origin.

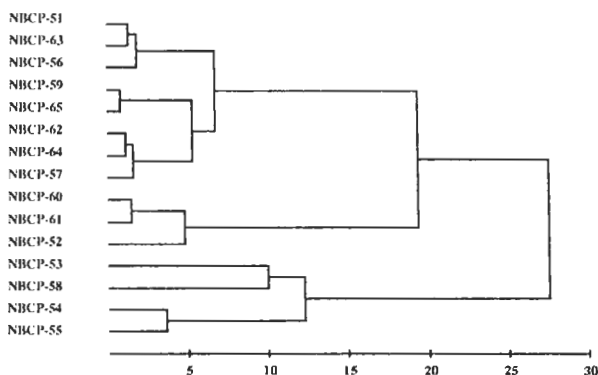


Fig. 1. Dendrogram of 15 *C. procumbens* inbred lines, based on 7 mean variables, according to minimum variance method of Ward

The intra and inter-cluster Euclidean square values of the five clusters are presented in Table 3. The intra-cluster value ranged from 3.23 to 19.09, the minimum being in cluster I (3.23), indicating the three inbreds in this cluster were

similar of which two belonged to same origin (United States). Cluster IV showed maximum intra-cluster distance (19.09), followed by cluster V (7.31) and cluster III (6.31) indicating existence of diverse genotypes in these clusters. Cluster IV with maximum intra-cluster distance, showed wide divergence from other clusters by virtue of its high inter-cluster Euclidean square distance of 26.20 from cluster II, 22.71 from cluster III, 19.94 from cluster V and 19.30 from cluster I.

Table 3. - Intra (Bolded) and inter-cluster Euclidean square values in *C. procumbens*

Clusters	I	II	III	IV	V
I	3.23				
II	6.45	4.46			
III	11.57	14.40	6.31		
IV	19.30	26.20	22.71	19.09	
V	12.58	17.83	18.06	19.94	7.31

This observation reveals that the two inbreds, NBCP-53 and NBCP-58 belonging to this cluster were not only highly diverse among themselves but also widely divergent from genotypes in other clusters. Plotting of the first two principal components to observe relationship between clusters (Figure 2) showed prominent separation of the inbreds into different clusters with wide displacement of cluster IV from other clusters.

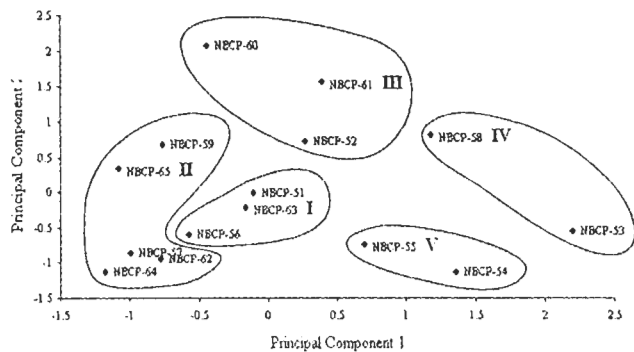


Fig. 2. Scattered diagram of *C. procumbens* inbred lines for first two principal components

While selecting genotypes to be included in the hybridization programme on the basis of genetic diversity, cluster means for seed yield and its components should also be given due importance. The mean values of different clusters for yield and yield components are given in Table 4. Cluster IV that showed the maximum intra- and inter-cluster distances had highest mean for seed yield (14.77g). The high mean yield in cluster IV may be due to higher mean values for branches/plant (122.50) and fruits/plant (203.00), which is also evident from high PC1 scores of the inbred lines NBCP-53 (2.19) and NBCP-58 (1.17) belonging to this cluster (Table 2).

Table 4. Cluster means of seed yield and its 6 component traits in *C. procumbens* inbreds

Traits	Clusters				
	I	II	III	IV	V
Days to Flowering	80.17	83.25	86.58	96.00	81.50
Plant Height (cm)	52.58	48.20	43.33	50.75	45.13
Branches/ Plant	74.08	63.30	78.25	122.50	103.25
Fruits/ Plant	149.81	122.83	151.18	203.00	190.00
Seeds/ Fruit	25.58	30.35	26.17	25.38	17.13
Test Weight (g)	3.13	2.84	3.78	2.87	2.91
Yield/ Plant (g)	11.95	10.64	14.53	14.77	9.43

Cluster III showed high mean values for seed yield (14.53g) and fruits/plant (151.18) following cluster IV. They had the highest mean value for test weight, which was reflected by high PC2 scores of inbreds (NBCP-52, NBCP-60 and NBCP-61) belonging to this cluster.

The results of this study indicated that there is significant variation for agronomic traits among the clusters of *C. procumbens* inbreds derived by Ward's method. Intra and inter- cluster Euclidean square distances showed NBCP-53 and NBCP-58 in cluster IV with high mean seed yield to be highly divergent among themselves and from inbreds in other clusters. This is also evident from partitioning of these inbreds in two separate (single genotype) clusters by UPGMA method. Also the inbred lines, NBCP-52, NBCP-60 and NBCP-61 in cluster III that ranked second for seed yield showed wide divergence from inbreds in cluster IV. So these inbred lines with good amount of genetic divergence and superior agronomic traits will be of practical utility and can be used to develop superior derivatives. Crossing of these divergent genotypes may also result in breaking of the wild characters in addition to yield improvement.

Acknowledgements. - Authors thank G.Robellen, Institute of Plant Breeding and Genetics, Goettingen, Germany who suggested this crop for research. Also we thank United States Department of Agriculture (USDA), Ames, USA, for providing *C. procumbens* germplasm lines. Thanks are also due to the Director, NBRI for providing facilities 2002 (2.73 %). In both study years, path coefficient analysis revealed a significant direct effect of stem length on green mass yield per plant (9.09* and 6.23*, respectively).

REFERENCES

- ANONYMOUS (1998): *Cuphea* species. In: *Annual Report 1997-98*. National Botanical Research Institute, Lucknow. p: 34-36.
- FRANCO J., J. CROSSA, J. DIAZ, S. TABA, J. VILLASENOR, and S.A. EBERHART (1997): A sequential clustering strategy for classifying gene bank accessions. *Crop Sci.*, 37, 1656-1662.
- HIRSINGER F. and P.F. KNOWLES (1984): Morphological and agronomic description of selected *Cuphea* germplasm. *Econ. Bot.*, 38, 439-451.
- KAUFMAN L. and P.J. ROUSSEEUW (1990): *Finding groups in data*. John Wiley and Sons Inc. New York.
- PANDEY V., R. BANERJI, B.S. DIXIT, M. SINGH, S. SHUKLA, and S.P. SINGH (2000): *Cuphea* a rich source of medium chain triglycerides: fatty acid composition and oil diversity in *Cuphea procumbens*. *Eur. J. Lipid Sci. Technol.*, 102, 463-466.

- RAMESH KUMAR R. and S.P. SINGH (2003): Genetic Association in *Cuphea procumbens*. *Farm Sci. J.*, 12, 113-115.
- RAMESH KUMAR R., M. SINGH, and S.P. SINGH (2002): Genetic variability and heritability studies in *Cuphea procumbens*. *Ind. J. Plt. Gen. Res.*, 15, 36-39.
- RAO C.R. (1973): *Linear Statistical Inference*. John Wiley and Sons Inc. New York. p: 556.
- SINGH S.P. and M. SINGH, M (2004): Multivariate analysis in relation to genetic improvement in *Cuphea procumbens*. *J. Genet. Breed.*, 58, 105-112.
- SINGH S.P. and R. RAMESH KUMAR (2003): *Cuphea* - importance and bibliography. *App. Bot. Abs.*, 23, 60 -75.
- WARD J. (1963): Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.*, 58, 236-244.

Received November 21, 2005

Accepted February 15, 2006

ANALIZA MULTIVARIANSE I GRUPISANJA SAMOOPLODNIH LINIJA *CUPHEA PROCUMBENS*

R. Ramesh KUMAR i S.P. SINGH

Odeljenje za Genetiku i oplemenjivanje biljaka,
Nacionalni institut za botanička istraživanja Institut (CSIR), Lucknow, Indija

I z v o d

Istraživanja su obuhvatila 15 samooplodnih linija *C. procumbens* različitog eko-geografskog porekla koje se održavaju u Nacionalnom institutu za botanička istraživanja, Lucknow, Indija. Vršeno je ocenjivanje prinosa semena i šest komponenata prinosa u cilju utvrđivanja genetičke divergentnosti linija. Standardne srednje vrednosti različitih osobina su bile uključene u analizu glavnih komponenata prinosa (PC) glavne komponente (Principal components) i grupisanje (clustering) korišćenjem UPGMA i metoda Warda. Wardova metoda, koja je pokazala relativno visok kofenetički koeficijent korelacije i značajna Wilksova Lambda su identifikovani kao najbolje rešenje za grupisanje. Prve četiri glavne (PC) komponente sa vrednostima > 1 su doprinele 91.56 % varijabilnosti među samooplodnim linijama. Prva PC je vezana za broj plodova po biljci, druga PC za prinos po biljci, broj semena po plodu i težini; treća PC je vezana za visinu biljaka a četvrta PC sa brojem dana pre cvetanja. Genotipovi su se grupisali u pet grupa (cluster-a). Grupa (cluster) II je bila najveća (5 genotipova) a slede grupe I, III, i IV i V. Grupa IV pokazuje veću srednju vrednost za prinos semena (14,77 gr) a slede grupa III (14,53 gr) i uključene samooplodne linije NBCEP-53 i NBCEP-58, značajno divergentne kako međusobno tako i od genotipova koje pripadaju drugim grupama. Linije u grupi IV i grupi III velike genetičke divergentnosti sa superiornim agronomskim osobinama su identifikovane kao perspektivne linije i osnova za razvoj superiornih rekombinacija, koje će imati poželjne agronomске osobine.

Primljeno 21. novembra 2005.

Odobreno 15. februara 2006.