

Principal Component and Cluster Analysis of Fibre Yield in Roselle (*Hibiscus sabdariffa* L.)

Hariram Kumar Bandi¹ and A. Appalaswamy²

¹Department of Genetics and Plant Breeding, Agricultural College, Naira, AP- 532 185, INDIA

²Agricultural Research Station, Amadalavalasa, AP- 532 185, INDIA

Email: hrk.bandiazu@gmail.com

Paper No. 206 Received: February 16, 2014 Accepted: May 17, 2014 Published: May 28, 2014

Abstract

Genetic diversity plays an important role in crop improvement, because the segregants between lines of diverse origin generally display an improved performance than those between closely related parental genotypes. Sixty genotypes of roselle (*Hibiscus sabdariffa* L.) were evaluated to study genetic divergence of fibre yield contributing quantitative characters by using principal component and cluster analysis. Principal component analysis (PCA) identified four principal components with eigen values more than one which contributed 73.859 percent of cumulative variance. The first principal component (PC₁-) contributed maximum towards variability with significant loading of plant height, fibre length per plant, and fibre yield per plant. The plot of PC₁ and PC₂ showed characters differentiation of genotypes according to their cluster membership for each cluster. The mean scores of genotypes were used as input for clustering in order to group the genotypes into various clusters. Hierarchical clustering procedure (complete linkage dendrogram) was followed to group the genotypes into 8 clusters, the clustering pattern of genotypes was to be independent of their eco-geographical origin. The comparative study of PCA and Complete Linkage Dendrogram indicated the similarity in the clustering pattern, resulted to clustering of divergent genotypes into different clusters against the geographical origin. The genotypes REX-1, ER-1, AHS-188, AMV-5, AMV-4, AHS-161, HS-4270, AR-104, JRRM-9-2 and R-83 in the different clusters if inter crossed may generate wider variability for sustainable improvement of fibre yield.

Highlights

- 60 genetically divergent genotypes of roselle (*Hibiscus sabdariffa* L.) of different geographical origin were studied.
- 4 principal components, showing contribution of 73.859% of cumulative variance.
- The first principal component (PC₁-) contributed maximum towards variability with significant loading of plant height, fibre length per plant, and fibre yield per plant.
- 2 dimensional space scattering of genotypes observed by loading of PC₁ and PC₂.
- 8 clusters were formed by Hierarchical clustering (Complete Linkage Dendrogram).
- 10 potential genotypes were identified by the comparative study for future breeding program for improvement of fibre yield.

Keywords: Roselle (*Hibiscus sabdariffa* L.), Cluster Analysis, Genetic Divergence, Principal Component Analysis, Complete Linkage Dendrogram.



Roselle (*Hibiscus sabdariffa* L.) is an important fibre yielding crop in India next to jute and the fibre is extracted from the bast region of stem. Fibre yield is quantitatively inherited and influenced by genetic factors as well as environments. The success of any breeding programme depends upon the availability of adequate genetic diversity. The major factor responsible for limited success in increasing the roselle (*Hibiscus sabdariffa* L.) yield has been the narrow genetic base of the material available. Genetic diversity plays an important role, because segregants between lines of diverse origin generally, display an improved performance than those between closely related parents. Genetic diversity is the basic criterion for the continuous improvement of the crop whether through natural selection or direct plant breeding. The present study aims at assessment of genetic divergence in sixty genotypes of roselle through principal component analysis (PCA) and hierarchical cluster analysis (Complete Linkage Dendrogram). Crosses can be made between genetically diversified genotypes for character improvement and fibre yield.

Materials and Methods

Sixty genotypes of Roselle were sown in the kharif (rainy) season of 2011-12 in a randomized block design with three replications at Agricultural College, Naira, Srikakulam (A.P). Each genotype was planted in three rows of three meters length with a spacing of 30X10 cm. The observations were recorded on five randomly selected plants in each genotype and in each replication on nine yield component characters viz., days to 50% flowering, plant height (cm), basal stem diameter(cm), number of nodes per plant, internodal length per plant(cm), green plant weight(g), fibre length per plant(cm), fibre wood ratio, fibre yield per plant(g) and mean values were used for statistical analysis. The data were analyzed using principal component analysis (Jackson, 1991) and hierarchical cluster analysis (Complete Linkage Dendrogram) (Anderberg, 1993).

Results and Discussion

The analysis of variance revealed highly significant differences among the sixty genotypes of roselle indicating that the existence of substantial genetic variability for all the characters under study.

Principal component analysis (PCA) identified four principal components with eigen values more than one which contributed 73.859 % of cumulative variance (Table 1). The first principal component (PC¹) contributed maximum

towards variability (42.601 percent) with significant loading of plant height (-0.472), fibre length per plant (-0.434), and fibre yield per plant (-0.419) which were negatively correlated and basal stem diameter (0.387) which was positively correlated. The second principal component (PC₂) accounted for 18.775 % of total variance and it reflected significant loading of days to 50 % flowering (0.648), which was positively correlated and green plant weight (-0.529) which was negatively correlated. The third principal component (PC₃) accounted for 12.423 percent of cumulative variance and it was characterized by conspicuously high loading for number of nodes per plant(0.517), which was positively correlated (Table 2). Based on these first three principal components mean genotype scores were computed. Principal factors scores for all the sixty genotypes were estimated in all 3 PC's and utilized to construct precise 2D editor plot by considering PCA score 1 and 2 (Fig 1). All the genotypes were plotted for PC₁ and PC₂ which cumulatively explained 61.376 percent of variability accounted for all the characters.

Table 1: Eigene values, percent and cumulative variance for four principal components in roselle.

Characters	PCA1	PCA2	PCA3	PCA4
Eigene value(root)	3.840	1.690	1.118	0.761
Per cent variance explained	42.601	18.775	12.423	8.452
Cummulative variance explained	42.661	61.436	73.859	82.311

Table 2: Character loading of four principal components for sixty genotypes of roselle.

Characters	PCA1	PCA2	PCA3	PCA4
Days to 50% flowering	0.111	0.648	0.071	0.238
Plant height	-0.472	0.129	-0.115	0.085
Basal stem diameter	0.387	-0.136	0.287	-0.406
Number of nodes per plant	0.216	0.423	0.517	-0.107
Internodal length per plant	-0.258	0.215	-0.481	-0.611
Green plant weight	-0.224	-0.529	0.378	-0.032
Fibre length per plant	-0.434	0.142	0.264	0.098
Fibre wood ratio	-0.295	0.141	0.352	-0.569
Fibre yield per plant	-0.419	0.023	0.253	0.229

The plot of PC₁ and PC₂ showed characters differentiation of genotypes according to their cluster membership for each cluster. The mean scores of genotypes were used as input for clustering in order to group the genotypes into various clusters. Hierarchical clustering procedure (complete linkage dendrogram) was followed to group the



genotypes into 8 clusters (Table 3 Fig 2). This reflects that there was no relation between geographical origin and genetic diversity.

The biggest cluster was cluster III consisting of 16 genotypes followed by cluster I consisting of 15 genotypes followed by cluster II, cluster VII, cluster VI, cluster VIII, cluster IV and cluster V consisting of 12, 6, 5, 4, 2 and 1 genotypes respectively. Based on cluster analysis the intra

Table 3: Cluster compositions of sixty genotypes of roselle – Complete Linkage Dendrogram

Cluster number	Number of genotypes	Genotypes	Origin
I	15	CRIJAFR-2, CRIJAFR-8, ER-1, ER-10, REX-1 AS-80-09, R-134, , AR-77, R-271, AS-80-31, AS-80-29, AR-72, AS-81-22 JRR-9, JRRM-9-1	CRIJAF, Barrackpore(W.B) (2) Exotic (3) Indigenous local collection (8) Orissa (2)
II	11	AHS-160, AHS-163, AHS-172, AHS-188, AMV-5 ER-63 AR-12, AR-71, R-28, R-92, R-129	ARS, Amadalavalasa(A.P) (5) Exotic (1) Indigenous local collection (5)
III	16	AHS-169, AMV-4 HS-4288, HS-4239 ER-38, ER-52 AR-20, AR-30, AR-35, AR-48, R-16, R-68, R-143, R-190, R-200, R-301	ARS, Amadalavalasa(A.P) (2) CRIJAF, Barrackpore(W.B) (2) Exotic (2) Indigenous local collection (10)
IV	2	AHS-161 AR-13	ARS, Amadalavalasa(A.P) (1) Indigenous local collection (1)
V	1	HS-4270	CRIJAF, Barrackpore(W.B) (1)
VI	5	REX-50 AR-19, AR-66, AR-104, R-318	Exotic (1) Indigenous local collection (4)
VII	6	ER-118 AR-4, AR-16, AS-80-26, R-117 JRRM-9-2	Exotic (1) Indigenous local collection (4) Orissa (1)
VIII	4	AR-43, R-83, AR-85, R-246	Indigenous local collection (4)

ARS – Agricultural Research Station.

CRIJAF – Central Research Institute on Jute and Allied Fibres.

A.P – Andhra Pradesh.

W.B – West Bengal.

Table 4: Inter and Intra (diagonal) cluster average Euclidean² and Euclidean values (parenthesis) of sixty genotypes of roselle – Complete linkage dendrogram.

Clusters	I	II	III	IV	V	VI	VII	VIII
I	29.008 (5.386)	48.309 (6.951)	39.409 (6.278)	77.940 (8.828)	180.350 (13.430)	146.066 (12.086)	51.540 (7.179)	83.346 (9.129)
II		19.909 (4.462)	47.087 (6.862)	69.638 (8.345)	142.433 (11.935)	186.921 (13.672)	104.903 (10.242)	111.451 (10.557)
III			23.566 (4.855)	55.733 (7.466)	104.240 (10.210)	108.894 (10.435)	53.474 (7.313)	50.026 (7.073)
IV				71.198 (8.438)	101.672 (10.083)	151.632 (12.314)	93.194 (9.654)	85.753 (9.260)
V					0.000 (0.000)	179.975 (13.416)	197.800 (14.064)	91.216 (9.551)
VI						38.853 (6.233)	78.941 (8.885)	64.096 (8.006)
VII							18.047 (4.248)	62.146 (7.883)
VIII								37.032 (6.085)

cluster values range from 0.00 (cluster V) to 71.198 (cluster IV), the maximum inter cluster distance was observed between cluster V and cluster VII (197.800) followed by cluster II and cluster VI (180.350) and cluster IV and cluster VI (151.632) as shown in (Table 4). Cluster II is characterized by high mean value for plant height (407.611 m), basal stem diameter (2.27 cm), green plant weight (713.51 g), fibre length per plant (378.77 m) and fibre yield per plant (35.24 g). Based on these studies crosses may be effective between the genotypes of these clusters to obtain better and desirable segregants.

Utilization of principal component analysis combined with hierarchical cluster analysis in genetic diversity studies was reported by earlier workers. The comparative study of PCA and Complete Linkage Dendrogram (Fig. 1) indicated the similarity in the clustering pattern, resulted to clustering of divergent genotypes into different clusters against the geographical origin.

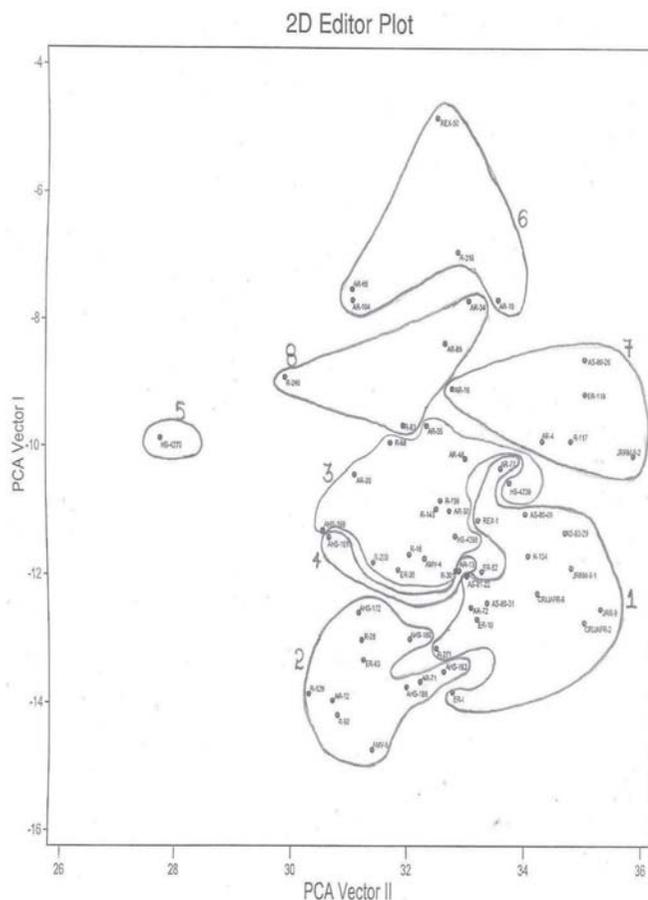


Fig. 1: Comparative study of PCA and Complete linkage dendrogram in roselle.

The present study depicted the relative divergence in morphological and yield traits. The clustering pattern could be utilized in identifying the best cross combinations for generating variability with respect to various traits under study. The genotypes REX-1, ER-1, AHS-188, AMV-5, AMV-4, AHS-161, HS-4270, AR-104, JRRM-9-2 and R-83 in the different clusters if inter crossed may generate wider variability. Some combinations of crossing of the genotypes may also exhibit high heterosis for fibre yield and transgressive segregants for yield and yield components may also be expected.

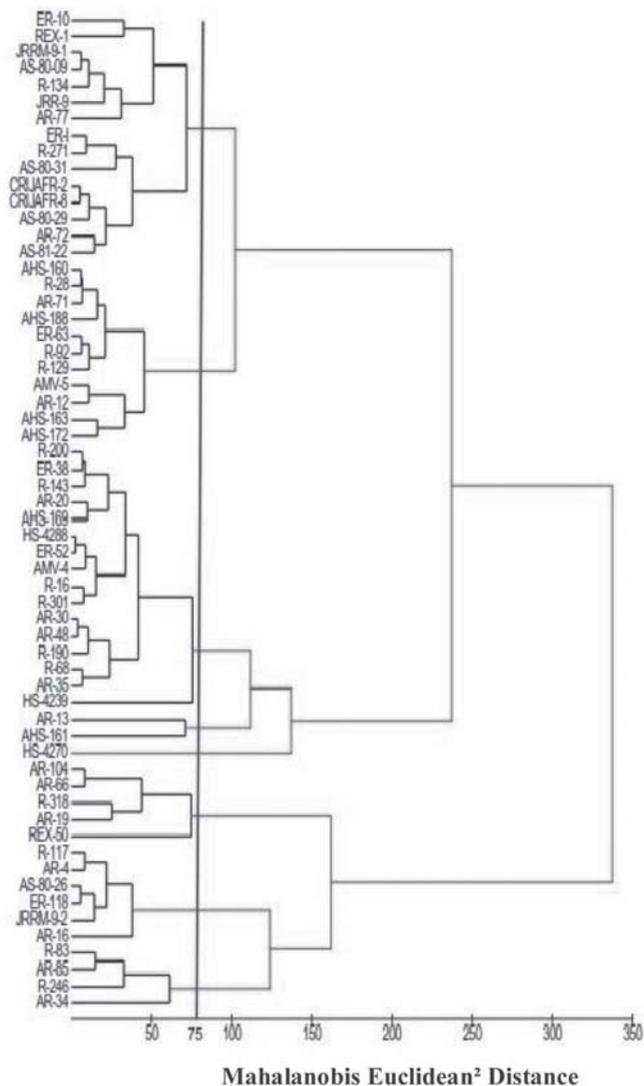


Fig. 2: Complete linkage dendrogram in roselle.



Acknowledgements

The first author is grateful to Acharya N G Ranga Agricultural University, Hyderabad for providing financial assistance in the form of fellowship during my course of study.

References

- Anderberg, M.R. 1993. Cluster analysis for applications. Newyork Academic Press.
- Jackson, J.E. 1991. A user's guides to principal components. New York, John Wiley and Sons.
- Lasalita-Zapico, F.C., Aguilar, C.H.M and Aujero, J.M. 2010. Phenotypic variability of manila hemp (*Musa textilis* L. Nee) genotypes in southern mindanao island, Philippines. *Journal of Tropical Agriculture* **48** (1-2): 68-70.
- Pulli Bai, P, Razia Sultana, Panduranga Rao, C and Srinivasa Rao, V. 2005. Genetic divergence studies in roselle hemp (*Hibiscus sabdariffa* L.). *M.Sc. (Ag) Thesis*. Acharya N G Ranga Agricultural University, Rajendranagar, Hyderabad, India.
- Steven Brown, J. 1991. Principal component and cluster analyses of cotton cultivar variability across the U.S. cotton belt. *Crop Science* **31**: 915-922.
- Usha Rani, P. 1997. Application of principal component analysis in chilli (*Capsicum annum* L.). *Mysore Journal of Agricultural Science* **31**: 137-142.