

## MULTIVARIATE ANALYSIS IN INDIAN MUSTARD

Neeraj Kumar\*, Ram Avtar, Nisha Kumari, Rakesh Punia and Manjeet Singh

Department of Genetics and Plant Breeding, College of Agriculture, CCS Haryana Agricultural University, Hisar 125004 (Haryana), India

Email: [neerajkummar8@gmail.com](mailto:neerajkummar8@gmail.com)

Received-03.06.2022, Revised-14.06.2022, Accepted-24.06.2022

**Abstract:** Principal component and hierarchical cluster analyze were carried out with ten quantitative traits in 50 hybrids of Indian mustard [*Brassica juncea* (L.) Czern&Coss.]. Principal factor analysis identified three principal components which cumulatively explained about 63.7% variability. PC 1 explained the most variability, accounting for 30.6%, PC 2 for 12.7%, and PC 3 for 11.4% of the overall variation. PCA correlation circle revealed that main shoot length, number of siliqua on the main shoot, secondary branches per plant, primary branches per plant and seed yield per plant were positively correlated with each other. Hierarchical cluster analysis was performed to see the grouping pattern of parents of hybrids. Fifty hybrids were grouped into three clusters. Maximum of 27 hybrids were grouped in cluster II and showed characteristic of lesser main shoot length. Cluster 1<sup>st</sup> and 3<sup>rd</sup> had nine and 13 hybrids, respectively. Pedigree of hybrids showed that Cluster I and Cluster II comprised of majority of hybrids having female parents OA-RH 8812 and OA-RH 0749 while Cluster III comprised of OA-RH 0555 and OA-RH 30. Female parent, OA-RH 0630 was distributed among all clusters. Male parents were uniformly distributed among all clusters and no clear-cut pattern was found. Hence, in the future hybrid development programme we may use one female parent from each group in addition to the OA-RH 0630.

**Keywords:** *Brassica juncea*, Hybrid, Plant, Siliqua

## INTRODUCTION

Indian mustard [*Brassica juncea* (L.) Czern & Coss.], commonly known as raya, is one of the major oilseed crops cultivated in the Indian sub-continent. It belongs to the predominant oilseeds Brassicaceae family that plays a pivotal role in vegetable oil economy. *B. juncea* (AABB; 2n=36) is a natural amphidiploid arisen from hybridization between *B. rapa* (AA; 2n=20) and *B. nigra* (BB; 2n=16) followed by chromosome doubling which led to its evolution (U, 1935). The wild form of *B. rapa* and *B. nigra* is found in the Middle East hence, this region is the most probable place of origin of *Brassica juncea*. Oil content in this crop is varied from 38-42%.

In India, Rajasthan has the largest cultivated area of 3.05 million hectares (44.05%), followed by Madhya Pradesh (11.70%), Uttar Pradesh (10.55%), Haryana (8.50%), West Bengal (7.38%) and Assam (4.63%). Among all states, Haryana has the maximum productivity (1842 kg/ha) (Jat *et al.* 2019). During Rabi, 2018-19 Indian mustard was grown in India in an area of 6.94 million hectares with an output of 7.20 million tons. The average productivity of Indian mustard in India was 1037.5 kg/ha. Yet, India meets 57% of the domestic edible oil requirements through imports and ranked 7<sup>th</sup> largest importer of edible oils in the world. As the human population continues to grow and living standards improve, per capita oil consumption will rapidly rise. Thus, there is an urgent need to increase the yield potential through genetic interventions to meet the present and future oil requirements.

A genetic study based on multivariate analysis is a powerful tool for determining the diversity and the

relative contribution of different components to this diversity (Ceolin *et al.* 2007). Hence, the present study was undertaken to study the diversity, to identify yield contributing important characters and to classify genotypes suitable for Indian mustard breeding programme.

## MATERIALS AND METHODS

This study was carried out in Oilseeds Research Area, Department of Genetics and Plant Breeding, Chaudhary Charan Singh Haryana Agricultural University, Hisar during Rabi, 2017-18. For this, 50 hybrids were evaluated in RBD with three replications (Table 1). Each genotype was grown in paired rows of 4m length with a spacing of 45 x 15 cm row to row and plant to plant, respectively. During the crop season recommended package of cultural practices was followed to raise a good crop. Field observations were recorded for plant height (cm), number of primary branches/plants, number of secondary branches/plants, main shoot length (cm), number of siliqua on main shoot, siliqua length (cm), number of seeds/siliqua, 1000-seed weight (g), seed yield per plant (g) and oil content (%). Data on five randomly selected plants from each genotype in each replication were recorded for the above traits. Clustering of genotypes into similar groups was performed using Ward's hierarchical algorithm based on squared Euclidean distances. To identify the patterns of morphological variation, PCA was conducted. The observations recorded on ten traits were analyzed by using the Fact extra package of R software.

\*Corresponding Author

## RESULTS AND DISCUSSION

Principal components and hierarchical clustering are important aspects of multivariate analysis. Principal component analysis is a very effective technique for extracting the most important information from the vast data set. PCA is carried out via a series of steps, many of which involve subjective judgments. The numbers of extracted component are equal in number of variable under study. The first component is likely to account for the highest portion of the total variance. Each successive component manifested a smaller portion of the variation. According to Jolliffe and Cadima (2016), components with an eigenvalue greater than one and contributing at least 10% of the variation, were relevant for explaining variation. Here first three PC's having more than one eigenvalue and account for 30.6 %, 12.7% and 11.4%, respectively, of the total variation (Table 2). These first three components cumulatively explain 63.7% of the total variance, which is a decent approximation of the original variation among the hybrids under study. These findings are in accordance with the earlier findings of Nachimuthu *et al.* (2014) and Pachauri *et al.* (2017). A scree plot is a graph between eigenvalues and principal component numbers. It represents the distribution of variance among principal components (Fig. 1).

The correlation circle provides insight into how different variables (characters) correlate with each other and also with the principal component. Coordinates of a variable on a PC plot represent the correlation between variable and principal component. The degree of angle between variable loadings or characters indicates the degree of correlation between characters. Positively correlated variables are grouped; while negatively correlated variables are positioned on opposite sides of the plot origin. There was a very little angle among the characters such as main shoot length, number of siliqua on the main shoot, secondary branches per plant, primary branches per plant and seed yield per plant, hence these are highly positively correlated with each other. These correlation results are supported by the previous finding of Kumar *et al.* (2016); Bineeta Devi (2018); Roy *et al.* (2018) and Rout *et al.* (2018). These characters also have contributed to the PC1. While characters such as oil content (%), plant height (cm) and seed weight (g) were correlated with each other and also contributed towards PC 2 (Fig. 2).

The value of  $\text{Cos}^2$  (square cosine) indicates the relevance of the principal component for a certain trait. PC 1 was important for the traits such as number of primary branches/plant, number of secondary branches/plant, number of siliqua on main

shoot, siliqua length, number of seed/siliqua and seed yield per plant. PC 2 was important for the oil content (%). While for main shoot length and plant height, PC 3 is important (Fig. 3).

In a PCA biplot, an individual's score is represented by a black dot while a coordinate represents the variable (characters). Similar hybrids were placed together in a biplot. Individual's scores corresponding to the PC 1 and PC 2 were plotted against the X-axis and Y-axis, respectively. In this study, first two PC's were responsible for ~53% of the overall variation. Hybrids having high seed yield and related characters are present in the lower right corner of the biplot. Those found on the upper side of the upper right exhibited high oil content (%), while those on the left side of the lower-left having the characteristic of a poor seed yield. PC 1 was positively influenced by main shoot length, number of siliqua on main shoot, secondary branches per plant, primary branches per plant, and seed output per plant. This suggested that hybrids in PC1 had a high value for the aforementioned characters. PC2 was positively influenced by the number of seeds/siliqua, siliqua length, plant height, seed weight, and oil content whereas rest of traits had unfavorable effects.

Cluster analysis was performed on new variables obtained from the PCA. Hierarchical method was used to construct a dendrogram. Cluster analysis showed that 50 hybrids were grouped into three clusters. Cluster I comprised nine, cluster II of 27 and cluster III of 13 genotypes (Figure 4). In comparison to the overall mean value Cluster, I had the characteristic of low seed yield and related traits. Cluster II had a characteristic of the low height of main shoot. Cluster III showed the characteristic of high mean seed yield per plant and related characters (Table 3). Pedigree of hybrids showed that Cluster I and Cluster II comprised of majority of hybrids having female parents OA-RH 8812 and OA-RH 0749 while Cluster III comprised of OA-RH 0555 and OA-RH 30 (Table 4). Female parent, OA-RH 0630 was distributed among all the clusters. There is some deviation also which may be due to the interaction between the male and female parents of the hybrid. These results showed that female parents, OA-RH 8812 and OA-RH 0749 are similar but different from OA-RH 0555 and OA-RH 30 and vice-versa. Hence, in the future breeding hybrid development programme, we may use one female parent from each group in addition to the OA-RH 0630. Male parents were distributed among all the clusters and no clear pattern was found. This may be due to the interaction of male and female parents or any unknown reason and for any conclusion further study is necessary.

**Table 1.** List of Indian mustard hybrids along with pedigree

Sr. No	Hybrid pedigree	Sr. No	Hybrid pedigree
1	OA-RH 8812 x IOR 1	26	OA-RH 0630 x IOR 11
2	OA-RH 8812 x IOR 2	27	OA-RH 0630 x IOR 12
3	OA-RH 8812 x IOR 3	28	OA-RH 0630 x IOR 18
4	OA-RH 8812 x IOR 4	29	OA-RH 0630 x IOR 19
5	OA-RH 8812 x IOR 6	30	OA-RH 0630 x IOR 22
6	OA-RH 8812 x IOR 11	31	OA-RH 30 x IOR 1
7	OA-RH 8812 x IOR 12	32	OA-RH 30 x IOR 2
8	OA-RH 8812 x IOR 18	33	OA-RH 30 x IOR 3
9	OA-RH 8812 x IOR 19	34	OA-RH 30 x IOR 4
10	OA-RH 8812 x IOR 22	35	OA-RH 30 x IOR 6
11	OA-RH 0749 x IOR 1	36	OA-RH 30 x IOR 11
12	OA-RH 0749 x IOR 2	37	OA-RH 30 x IOR 12
13	OA-RH 0749 x IOR 3	38	OA-RH 30 x IOR 18
14	OA-RH 0749 x IOR 4	39	OA-RH 30 x IOR 19
15	OA-RH 0749 x IOR 6	40	OA-RH 30 x IOR 22
16	OA-RH 0749 x IOR 11	41	OA-RH 0555 x IOR 1
17	OA-RH 0749 x IOR 12	42	OA-RH 0555 x IOR 2
18	OA-RH 0749 x IOR 18	43	OA-RH 0555 x IOR 3
19	OA-RH 0749 x IOR 19	44	OA-RH 0555 x IOR 4
20	OA-RH 0749 x IOR 22	45	OA-RH 0555 x IOR 6
21	OA-RH 0630 x IOR 1	46	OA-RH 0555 x IOR 11
22	OA-RH 0630 x IOR 2	47	OA-RH 0555 x IOR 12
23	OA-RH 0630 x IOR 3	48	OA-RH 0555 x IOR 18
24	OA-RH 0630 x IOR 4	49	OA-RH 0555 x IOR 19
25	OA-RH 0630 x IOR 6	50	OA-RH 0555 x IOR 22

**Table 2.** Eigen values and Percentage of Variance, Cumulative Percentage of Variance

Principal Component	Eigenvalue	Percentage of variance	Cumulative percentage of variance
Comp 1	3.96	39.61	39.61
Comp 2	1.27	12.66	52.27
Comp 3	1.14	11.42	63.69
Comp 4	0.94	9.38	73.07
Comp 5	0.87	8.74	81.82
Comp 6	0.54	5.43	87.24
Comp 7	0.49	4.90	92.15
Comp 8	0.39	3.94	96.09
Comp 9	0.24	2.44	98.53
Comp 10	0.15	1.47	100.00

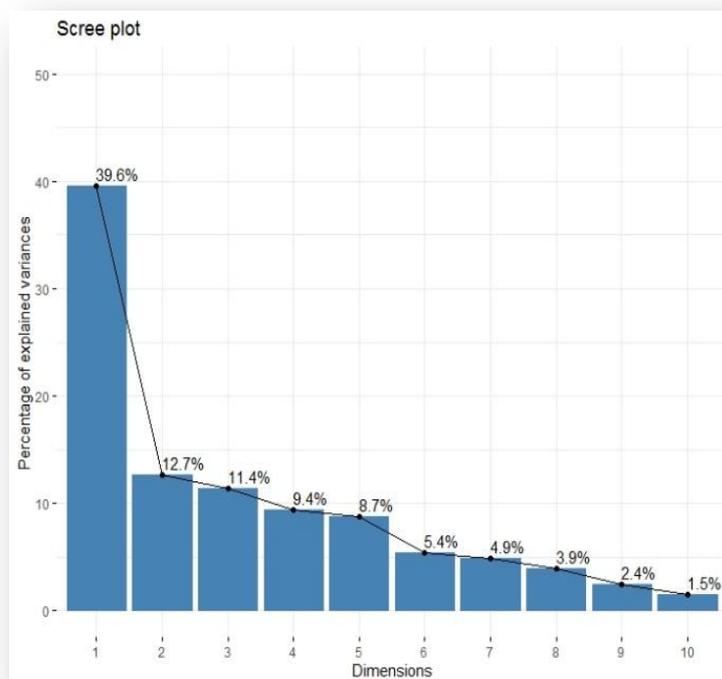
**Table 3.** Comparison of different clusters on the basis of mean value

Traits	Clusters I	Clusters II	Clusters III	Overall mean
Number of primary branches/plant	211.22	211.33	212.93	211.76
Number of secondary branches/plant	<b>5.76</b>	7.09	<b>7.83</b>	7.056
Main shoot length (cm)	<b>13.11</b>	15.54	<b>19.06</b>	16.09

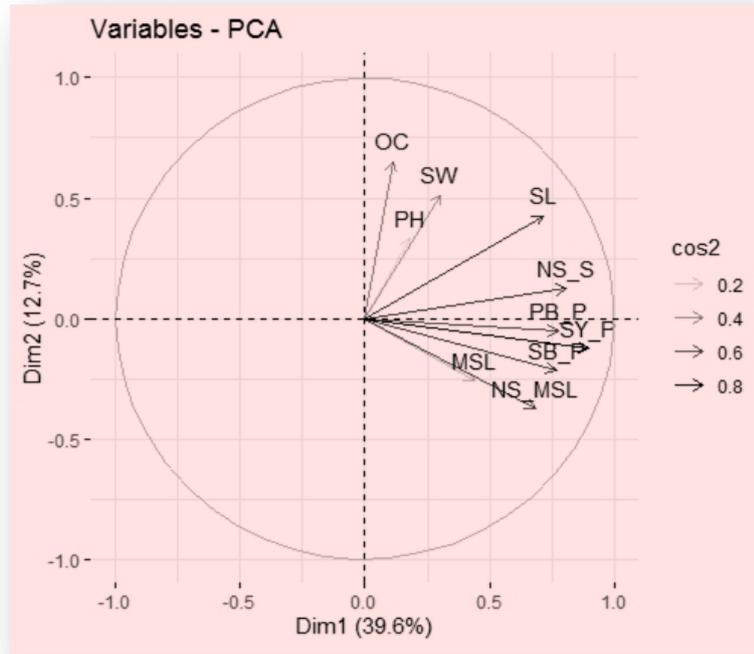
Number of siliqua on main shoot	77.46	<b>78.43</b>	<b>85.64</b>	80.28
Siliqua length (cm)	<b>48.51</b>	53.61	<b>58.75</b>	54.13
Number of seeds/siliqua	<b>3.82</b>	4.16	<b>4.29</b>	4.13
1000-seed weight (g)	<b>11.60</b>	12.47	<b>13.56</b>	12.62
Seed yield /plant (g)	<b>4.44</b>	4.83	4.84	4.76
Oil content (%)	<b>18.13</b>	23.57	<b>29.77</b>	24.33
Number of primary branches/plant	38.55	38.79	38.64	38.70

**Table 4.**Characteristic of different clusters

Cluster	No. of Hybrids	Hybrids	Representative Hybrids
I	9	OA-RH 8812 x IOR 2,OA-RH 8812 x IOR 4, OA-RH 8812 x IOR 18,OA-RH 0749 x IOR 1, OA-RH 0749 x IOR 2, OA-RH 0749 x IOR 3, OA-RH 0630 x IOR 2, OA-RH 0630 x IOR 6 and OA-RH 0555 x IOR 11	OA-RH 8812 x IOR 2, OA-RH 0555 x IOR 11, OA-RH 0630 x IOR 2, OA-RH 8812 x IOR 18 and OA-RH 0749 x IOR 3
II	27	OA-RH 8812 x IOR 1, OA-RH 8812 x IOR 3, OA-RH 8812 x IOR 6, OA-RH 8812 x IOR 11, OA-RH 8812 x IOR 12, OA-RH 0749 x IOR 11, OA-RH 8812 x IOR 22, OA-RH 0749 x IOR 4, OA-RH 0749 x IOR 6, OA-RH 8812 x IOR 19, OA-RH 0749 x IOR 12, OA-RH 0749 x IOR 18, OA-RH 0749 x IOR 19, OA-RH 0749 x IOR 22, OA-RH 0630 x IOR 1, OA-RH 0630 x IOR 3, OA-RH 0630 x IOR 4, OA-RH 0630 x IOR 11, OA-RH 30 x IOR 4, OA-RH 0630 x IOR 19, OA-RH 30 x IOR 3, OA-RH 0630 x IOR 18, OA-RH 30 x IOR 18, OA-RH 30 x IOR 19, OA-RH 30 x IOR 22, OA-RH 0555 x IOR 3 and OA-RH 0555 x IOR 6	OA-RH 0749 x IOR 19, OA-RH 8812 x IOR 3, OA-RH 0749 x IOR 12, OA-RH 8812 x IOR 22, OA-RH 30 x IOR 4
III	14	OA-RH 0630 x IOR 12, OA-RH 0630 x IOR 22, OA-RH 30 x IOR 1, OA-RH 30 x IOR 2, OA-RH 30 x IOR 6, OA-RH 30 x IOR 11, OA-RH 30 x IOR 12, OA-RH 0555 x IOR 1, OA-RH 0555 x IOR 2, OA-RH 0555 x IOR 4, OA-RH 0555 x IOR 12, OA-RH 0555 x IOR 18, OA-RH 0555 x IOR 19 and OA-RH 0555 x IOR 22	OA-RH 0555 x IOR 4, OA-RH 30 x IOR 1, OA-RH 0555 x IOR 22, OA-RH 0630 x IOR 22, OA-RH 30 x IOR 2,

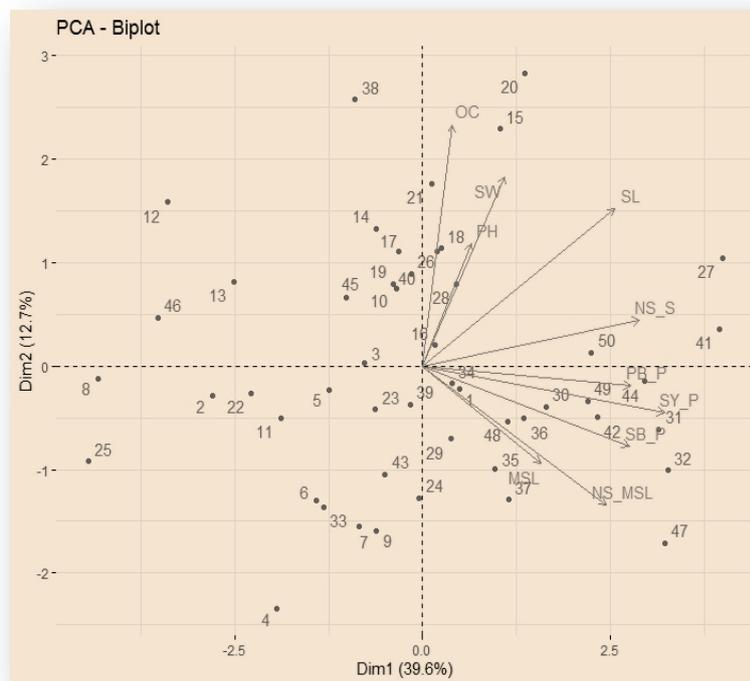


**Figure 1:** Screen plot of top 10 principal components



**Figure 2.**The plot above is also known as variable correlation plots. It shows the relationships between all variables (traits).

PH= Plant height (cm), PB\_P=Number of primary branches/plant, SB\_P=Number of secondary branches/plant, MSL=Main shoot length (cm), NS\_MSL= Number of siliqua on main shoot, SL=Siliqua length (cm),NS\_S=Number of seeds/siliqua, SW=1000-seed weight (g), SY\_P=Seed yield /plant (g), OC =Oil content (%)



**Figure 3.** Principal component analysis (PCA) biplot of individuals (hybrids) and explanatory variables (traits).

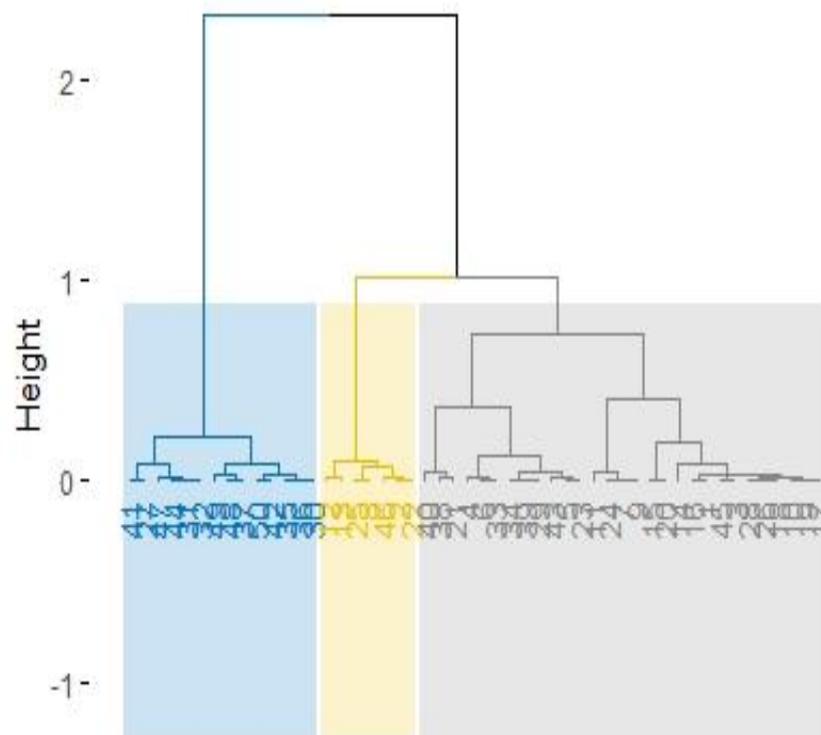


Figure 4. Hierarchical clustering on principal components

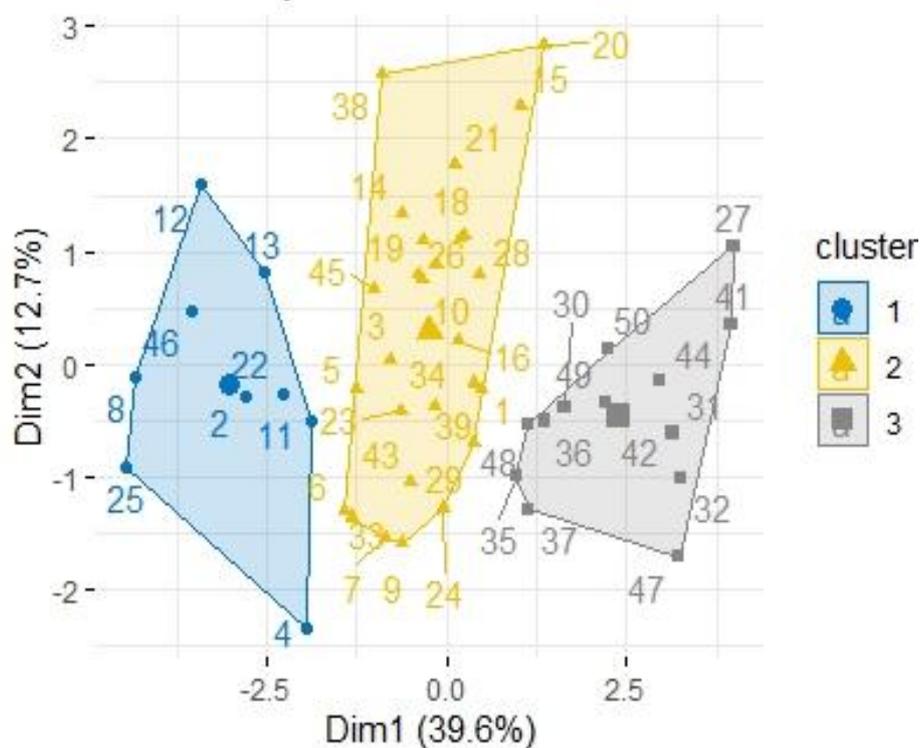


Figure 5. Visualization of individuals on the cluster map

## REFERENCES

**Bineeta Devi** (2018). Correlation and path analysis in Indian mustard (*Brassica juncea* L.) in agro-climatic conditions of Jhansi (U.P.). *Int. J. Pharmacogn. Phytochem.*, **7**(1): 1678-1681.

[Google Scholar](#)  
**Ceolin, A. C. G., Goncalves-Vidigal, M. C., Vidigal Filho, P. S., Kvitschal, M. V., Gonela, A. and Scapim, C. A.** (2007). Genetic divergence of *Phaseolus vulgaris* L. using morpho-agronomic traits by multivariate analysis. *Hereditas*, **144**: 1-9.

- [Google Scholar](#)  
**Jat, R.S., Singh, V.V., Sharma, P. and Rai, P.K.** (2019). Oilseed *brassica* in India: Demand, supply, policy perspective and future potential. *OCL*, **26**: 1-8.
- [Google Scholar](#)  
**Jolliffe, I. T. and Cadima, J.** (2016). Principal component analysis: a review and recent developments. *Philos. Trans. A. Math. Phys. Eng. Sci.*, **374**(2065): 20150202.
- [Google Scholar](#)  
**Kumar, R., Gaurav, S. S., Jayasudha, S. and Kumar, H.** (2016). Study of correlation and path coefficient analysis in germplasm lines of Indian mustard (*Brassica juncea* L.). *Agri. Sci. Dig.*, **36**(2): 92-96.
- [Google Scholar](#)  
**Nachimuthu, V. V., Robin, S., Sudhakar, D., Raveendran, M., Rajeswari, S. and Manonmani, S.** (2014). Evaluation of rice genetic diversity and variability in a population panel by principal component analysis. *Indian J. Sci. Technol.*, **10**: 1555-1562.
- [Google Scholar](#)
- Pachauri, A. K., Sarawgi, A. K., Bhandarkar, S. and Ojha, G. C.** (2017). Agromorphological characterization and morphological based genetic diversity analysis of Rice (*Oryza sativa* L.) germplasm. *Int. J. Pharmacogn. Phytochem.*, **6**(6): 75-80.
- [Google Scholar](#)  
**Rout, S., Kerkhi, S. A., Chauhan, C.** (2018). Character association and path analysis among yield components in Indian mustard [*Brassica juncea* (L.) Czern and Coss]. *Int. J. Curr. Microbiol.*, **7**(1): 50-55.
- [Google Scholar](#)  
**Roy, R. K., Kumar, A., Kumar, S., Kumar, A. and Kumar, R. R.** (2018). Correlation and path analysis in Indian mustard [*Brassica juncea* (L.) Czern and Coss] under late sown condition. *Environ. Ecol.*, **36**(1): 247-254.
- [Google Scholar](#)  
**U, N.** (1935). Genome Analysis in *Brassica* with Special Reference to the Experimental Formation of *B. napus* and Peculiar Mode of Fertilization. *Jpn. J. Bot.*, **7**: 389-452.
- [Google Scholar](#)

