RESEARCH ARTICLE Indian Journal of Agricultural Research, Volume 59 Issue 4: 560-565 (April 2025)

Principle Component Analysis and Diversity Studies in Sunflower Lines (*Helianthus annuus*)

Harshavardan J. Hilli[1], Shobha Immadi[1]

10.18805/IJARe.A-5996

ABSTRACT

Background: Sunflower crop being a highly cross pollinated with high yield potential, can be grown in all seasons with its suitability in variety of inter and series cropping systems due to its photo insensitive nature and can be adapted to a variety of environmental conditions also. So keeping in mind of its potential and future use, research was conducted with available lines to study PCA and its diversity that might be used in hybridization programmes.

Methods: The 45 lines of sunflower were grown in a randomised block design with three replications at the Research Farm, Department of Genetics and Plant Breeding, College of Agriculture, Vijayapur during summer 2018-19 for study on principal component analysis and diversity analysis.

Result: At the genotypic stage, the analysis of variance showed highly significant variations among the genotypes for all of the characters. Whereas Divergence analysis using principal component analysis and hierarchical cluster analysis has been shown to be effective in determining which genetically distant parents should be used for hybridization. Divergence in genotypes were divided into nine major groups or clusters and out of nine components, four components had more than one eigen value (Table-2). The first principal component explained 33.40 per cent of total variation, while the second, third, fourth, fifth, sixth, seventh, eighth and ninth principal components explained 23.30, 14.10, 11.19, 6.30, 4.90, 3.60, 1.90 and 0.60 per cent, respectively.

Key words: Cluster, Correlation, Diversity, PCA, Sunflower.

INTRODUCTION

Sunflower (*Helianthus annuus* L.), the world's second-largest oilseed crop after soybean, is a member of the Asteraceae/ Compositae family that originated in temperate North America and has a high content of unsaturated fatty acids with a desirable consistency. It's a diploid crop with a 3000 Mb haploid genome and a diploid chromosome number of 2n=34. Sunflowers have been successfully grown in a wide variety of locations around the world. It is a highly cross pollinated crop with high yield potential that can be adapted to a variety of environmental conditions. It can be grown in all seasons and taken in a variety of inter and series cropping systems due to its photo insensitivity. Presently, it is cultivated in an area of 20.00 million hectares globally with production of 30.00 million tonnes and productivity of 1,500 kg ha^{-1}. Asia accounts for nearly 20-22 per cent of the total sunflower area in the world, contributing 18 per cent of the total production.

India is the leading country in the Asian sub-continent with an area of 0.29 million hectares and annual production of 0.21 million tonnes and productivity of 738 kg ha^{-1}. The leading sunflower growing states in India are Karnataka, Punjab, Bihar, Orissa andhra Pradesh and Maharashtra and in recent years, its cultivation is increasing to non-traditional states like Haryana, Uttar Pradesh, Gujarat and West Bengal. Karnataka is leading state in India contributing 64 per cent to the total area and 54 per cent to the total production and is popularly known as "sunflower state". In Karnataka, it is cultivated over an area of 0.44 million hectares with annual production and average productivity of 0.29 million tonnes and 530 kg ha^{-1} respectively.

[1]Department of Genetics and Plant Breeding, University of Agricultural Scieneces, Dharwad-580 025, Karnataka, India.

Corresponding Author: Harshavardan J. Hilli, Department of Genetics and Plant Breeding, University of Agricultural Scieneces, Dharwad-580 025, Karnataka, India.
Email: harshajh1995@gmail.com

How to cite this article: Hilli, H.J. and Immadi, S. (2025). Principle Component Analysis and Diversity Studies in Sunflower Lines (*Helianthus annuus*). Indian Journal of Agricultural Research. 59(4): 560-565. doi: 10.18805/IJARe.A-5996.

Submitted: 20-04-2022 **Accepted:** 09-07-2022 **Online:** 25-07-2022

Adequate information about heritability, variability and the degree of association among different traits is needed for further yield improvement. It has been proposed that genetic tools be used to establish sustainable solutions to basic crop constraints, but this is difficult due to the large number of variation effects and a lack of adequate evaluation and classification techniques. There are a variety of statistics available for grouping lines in order to select for different types, although there are some drawbacks since data can become unmanageable at times. As a result, PCA and cluster analysis (Peeters and Martinelli, 1989) are useful tools for identifying important characters. The parents for the hybridization program should be chosen based on the magnitude of genetic distance, the contribution of different characters to total divergence and the magnitude of cluster means for distinct characters with the most heterosis. The effective hybrid breeding and selection needs presence of

large amount of variability among the traits to be studied. Hence it is necessary to study variability in respect of quantitative characters with reference to genetic parameters. PCA is a multivariate technique that analyses data and converts it into a series of new orthogonal variables called principal components by linearly combining the variables that account for the majority of the variance in the original variables (Abdi and Williams, 2010). In order to treat qualitative variables, PCA can be generalised as correspondence analysis (CA) and as multiple facto analysis (MFA). Divergence analysis using principal component analysis and hierarchical cluster analysis has been shown to be effective in determining which genetically distant parents should be used for hybridization (Choudhary *et al*., 2015).

Materials and Methods

The current investigation was conducted with 45 lines of sunflower (Table 7) at the College of Agriculture, Vijayapura, during the summer 2018 at H block which is situated at 16°N latitude and 75°E longitude and altitude of 593 MSL with a mean annual rainfall of 750 mm. The agroclimatic region and soil conditions are considered to be good for crop establishment. Each line was grown in a 3 m long row with 30cm between plants and 60 cm between rows. To grow a good harvest, recommended agronomic practises were used. Days to 50% flowering, days to maturity, relative chlorophyll content and days to 50% flowering, days to maturity, relative chlorophyll content using SPAD chlorophyll meter at 45 DAS, relative chlorophyll content using SPAD chlorophyll meter at 60 DAS, head diameter, test weight, seed yield per hectare and the oil content were obtained from five randomly tagged plants in each lines from both replications. To assess genetic diversity among lines, principal component analysis (PCA) (Hotelling, 1936) and cluster analysis were used. The data was analysed using the R programming language.

Results and Discussion

Principal component analysis

The goal of principal component analysis is to find a limited number of linear combinations that account for the majority of the variation in the data being used. ANOVA for seed yield and its components in sunflower lines is depicted in Table 1 indicating significant difference obtained in all traits studied. Principal components analysis was used in this study to classify the most important characters and view them in more visual dimensions using linear combinations of variables that account for the majority of the variance in the original set of variables. Out of nine components, four components had more than one eigen value (Table 2).

Table 1: Analysis of variance (mean sum of square) for seed yield and its components in sunflower lines.

Source	d.f	Plant height/ PH(cm)	SPAD at 45 DAS	SPAD at 60 DAS	Days to 50 per cent flowering	Days to maturity /DM	Head diameter /HD(cm)	Test weight/ TW(g)	Seed yield/SY (kg/ha)	Oil content/ OC(%)
Replications	1	2.83	5.96	0.12	0.16	36.92	1.39	0.20	31238.22	0.02
Treatments	43	855.74*	16.88*	43.45*	49.63*	62.64*	3.83*	0.43*	17259.87*	22.53*
Error	43	36.44	6.96	7.44	6.07	19.24	0.28	0.12	2799.71	4.41

Table 2: Total variance explained by different principal components in sunflower lines.

Component	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Eigen values	3.005	2.097	1.27	1.067	0.567	0.439	0.321	0.175	0.058
Proportion	0.334	0.233	0.141	0.119	0.063	0.049	0.036	0.019	0.006
	33.40%	23.30	14.10	11.19	6.30	4.90	3.60	1.90	0.60
Cumulative	0.334	0.567	0.708	0.827	0.89	0.938	0.974	0.994	0.987
proportion	33.4%	56.70	70.80	82.70	89.00	93.80	97.40	99.40	98.70

Table 3: Factor loading of different characters with respect to different principal factor in sunflower lines.

Components	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
PH	0.448	-0.133	-0.134	-0.344	-0.077	-0.213	-0.758	0.091	-0.108
SM@45	0.235	0.588	0.207	0.044	0.154	0.133	0.041	0.12	-0.705
SM@60	0.219	0.595	0.204	0.081	0.159	0.043	-0.158	0.142	0.691
HD	0.458	0.086	-0.05	-0.074	-0.555	-0.456	0.491	0.125	0.047
DFF	-0.449	0.072	0.038	-0.476	0.178	-0.316	0.081	0.653	-0.013
DM	-0.25	0.153	0.537	-0.517	-0.35	0.027	-0.066	-0.48	0.012
TW	0.365	-0.184	-0.08	-0.527	0.136	0.638	0.31	0.127	0.101
Yl/pl	0.291	-0.299	0.442	-0.037	0.617	-0.411	0.189	-0.21	0.008
OC	0.066	-0.352	0.635	0.307	-0.289	0.225	-0.126	0.473	-0.015

The first principal component explained 33.40 per cent of total variation, while the second, third, fourth, fifth, sixth, seventh, eighth and ninth principal components explained 3.30, 14.10, 11.19, 6.30, 4.90, 3.60, 1.90 and 0.60 per cent, respectively (Table 2 and Fig 1).

The principal factor study (Kaiser, 1958) did not yield a simple image of the characters' interactions. As some factors had very high variable loading and others had low (Table 3). The data in the table says that PF-1 was loaded on plant height and SPAD readings @ 45 days, similarly PF-8 loaded on SPAD @ 60 days and test weight respectively. The highest loading recorded was PF-9 and PF-8.

Principal component analysis was also used in oat by Vaisi *et al.* (2013) and Krishna *et al.* (2014), who proposed transferring several associated variables into a few separate principal components, which explained much of the heterogeneity in the original collection. Hemavathy (2020) performed a sweet corn principal component research study and found similar results. In the Fig 2 wider angle indicates maximum diverse which indicates negative correlation and lesser or acute angle indicating more correlation. Results in figure view also depicted at the end.

PCA will allow for the depiction of individual differences as well as the identification of possible groups. The original variables are linearly transformed into a new set of uncorrelated variables known as principle components to achieve the reduction. Maruthi Sankar *et al.* (1999) have assessed the variability of eight plant traits for growth of sunflower and reduced the dimensionality to two principal components, which extracted about 80% of variance in the original data.

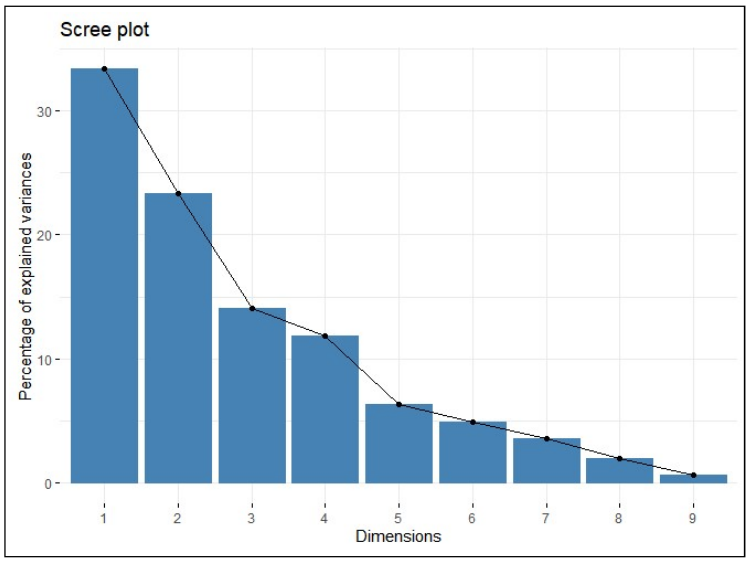


Fig 1: Variances shown by different components.

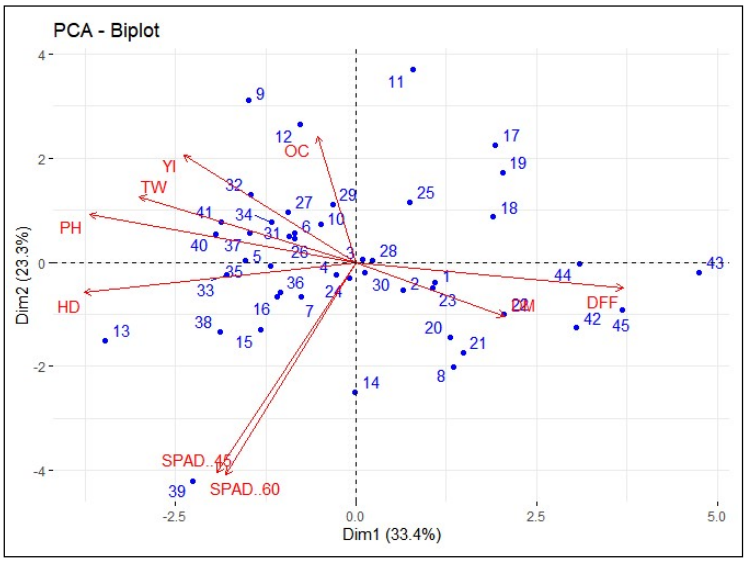


Fig 2: Figure depicting PCA and cluster distance between different components.

Genotypic and phenotypic correlation

Correlations are useful for identifying the main factors that determine final grain production, they only give a partial picture of the relative importance of direct and indirect impact on the individual elements. Yield is a complicated trait that is influenced by the number of other characteristics that can have both positive and negative effects on it. As a result, understanding the mechanism of interaction, consequences and cause of relationship will aid in the selection of breeding methods for increasing yield in sunflower. The magnitude of phenotypic correlation is lower than genotypic correlation, but they have a similar trend in direction, according to correlation analysis. Low phenotypic association indicates that the environment has an effect on the expression of these traits. Certain traits had significant positive correlation and certain had significant positive correlation. Yield per plant and plant height had significant positive correlation and rest had significant negative correlation. Similar results were obtained by Arunkumar *et al.* (2014). This indicates that selection of traits like plant height and yield per plant can be used as major criteria for improving the yield (Table 4 and 5).

The lines which are having high heritability and greater amount genotypic coefficient of variation need to be selected and also the lines having high plant height and high seed yield per plant leads to increase in total yield per hectare basis. Thus there might be certain traits depicting negative correlation and certain positive correlation, but breeders

Table 4: Cluster groups and average intra cluster distance.

Cluster groups	Genotypes	Number of genotypes	Average intra cluster distance
1	34, 37, 35, 31, 33, 29, 40, 6, 5, 16, 10, 36, 41, 38, 26, 28, 7, 27, 4, 15	20	40.6014
2	3, 30, 2, 23, 25, 21, 20, 14, 1, 42, 22, 18, 19, 44	14	49.8334
3	45, 43	2	57.0137
4	17, 11, 12	3	62.7988
5	32, 9	2	35.3073
6	8	1	0
7	24	1	0
8	39	1	0
9	13	1	0

Table 5: Genotypic correlation coefficient among various traits in sunflower lines.

Traits	PH	SM@45	SM@60	HD	DFF	DM	TW	YL/PL	OC
SM@45	0.076								
SM@60	0.081	0.902**							
HD	0.573**	0.320**	0.289**						
DFF	-0.420**	-0.219*	-0.194	-0.510**					
DM	-0.210*	0.108	0.097	-0.214*	0.480**				
TW	0.492**	0.042	-0.049	0.298**	-0.208*	-0.067			
Yl/pl	0.339**	-0.033	-0.049	0.200	-0.270*	-0.057	0.276**		
OC	0.339**	-0.033	-0.049	0.200	-0.270*	-0.057	0.276**	1.000**	
Yl/ha	0.221*	-0.173	-0.170	0.006	-0.259*	0.091	0.014	0.380**	0.380**

Table 6: Phenotypic correlation coefficient among various traits in sunflower lines.

Traits	PH	SM@45	SM@60	HD	DFF	DM	TW	YL/PL	OC
SM@45	0.081								
SM@60	0.101	0.972**							
HD	0.583**	0.371**	0.334**						
DFF	-0.474**	-0.238*	-0.239*	-0.566**					
DM	-0.339**	0.060	0.059	-0.249*	0.677**				
TW	0.982**	0.028	-0.025	0.684**	-0.590**	-0.336**			
Yl/pl	0.422**	-0.005	-0.019	0.278**	-0.333**	-0.186	0.603**		
OC	0.422**	-0.005	-0.019	0.278**	-0.333**	-0.186	0.603**	1.000**	
Yl/ha	0.314*	-0.259*	-0.261*	-0.010	-0.301**	0.181	-0.002	0.563**	0.563**

need to concentrate on the traits which are having significant positive correlation in order to improve crop yield.

Diversity analysis

Genetic variety is required for any crop improvement program since it aids in the generation of superior recombinants by allowing for the selection of parents with greater variability for various traits. Many crops has lost its diversity in recent decades as local types have been replaced with high-yielding variants. Genetic divergence study assesses the genetic diversity of a population.

D2 statistics is a method of calculating genetic divergence in germplasm collections using a numerical technique. The parents for the hybridization program should be chosen based on the magnitude of genetic distance, the contribution of different characters to total divergence and the magnitude of cluster means for distinct characters with the most heterosis. The D^2 study of 45 sunflower lines resulted in nine clusters, with cluster I and II having the most (20) and 14 lines and cluster VI to IX had one each. Cluster III and IV had the greatest intra-cluster distance. Clusters II, III and IV had stronger cluster means for most of the characters, hence lines 17, 11, 12, 45 and 43 can be used in crossing for crop improvement (Table 6,7). Similar results were obtained by Poonia and Phogat, 2017 while working in oat.

Table 7: Sunflower lines used for experiment.

Sl. no	Genotype
1	DSR-1
2	DSR-2
3	DSR-3
4	DSR-4
5	DSR-5
6	DSR-6
7	DSR-7
8	DSR-8
9	DSR-9
10	DSR-10
11	DSR-11
12	DSR-12
13	DSR-13
14	DSR-14
15	DSR-15
16	DSR-16
17	DSR-17
18	DSR-18
19	DSR-19
20	DSR-23
21	DSR-35
22	DSR-37
23	DSR-66
24	DSR-107
25	DSR-125
26	DSR-132
27	RHA95 C-1
28	RHA6D-1
29	NS-15
30	NS-3
31	B4L-21
32	B2L-13
33	B5L-16
34	B4L-18
35	B4L-6
36	IB-03
37	IB-60
38	IB-51
39	IB-104
40	ID-2089
41	GP-9
42	GP-5
43	5 RI
44	NS-8
45	NR P1

Conclusion

Knowledge of germplasm variability and genetic relations among breeding materials could prove to be a beneficial asset in crop development efforts. A variety of approaches for analyzing genetic diversity in germplasm accessions, breeding lines and populations are currently available. Principal components analysis is a multivariate powerful data reduction methodology that removes interrelationships between components and is helpful in detecting data set structures, genotyping grouping and estimate. Also breeders need to concentrate on the traits which are having significant positive correlation in order to improve crop yield.

Conflict of interest

All authors declared that there is no conflict of interest.

References

Abdi, H. and Williams, L.J. (2010). Principal component analysis. Wiley interdisciplinary reviews: Computational Statistics. 2(4): 433-459.

Arunkumar, B., Biradar, B.D. and Salimath, P.M. (2004). Genetic variability and character association studies in *rabi* Sorghum. Karnataka J. Agric. Sci. 17(3): 471-475.

Chaudhary, S.P., Sagar, B.K., Hooda. and Arya, R.K. (2015). Multivariate analysis of pearl millet data to delineate genetic variation. Forage Res. 40: 201-208.

Hemavathy, A.T. (2020). Principal component analysis in sweet corn [*Zea mays* (L.*) saccharata.*]. Forage Res. 45(4): 264-268.

Hotelling, H. (1936). Relations between two sets of varieties. Biometrica. 28(3-4): 321-377.

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

Krishna, A., Ahmed, S., Pandey, H.C. and Kumar, V. (2014). Correlation, path and diversity analysis of oat (*Avena sativa* L.) genotypes for grain and fodder yield. Journal of Plant Science and Research. 1(2): 9. ISSN: 2349-2805.

MaruthiSankar, G., Narasimha Murthy, D., Vanaja, M., Raghuram Reddy, P. (1999). A multiple selection index for selecting sunflower genotypes using principal component analysis. Indian Journal of Dryland Agricultural Research and Development. 14(2): 93-10.

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

Poonia, A. and Phogat, D.S. (2017). Genetic divergence in fodder oat (*Avena sativa* L.) for yield and quality traits. Forage Research. 43: 101-105.

Vaisi, H., Golparvar, A.R., Resaie, A. and Bahraminejad, S. (2013). Factor analysis of some quantitative attributes in oat (*Avena sativa* L.) genotypes. Scientia Agriculturae. 3(3): 62-65.