



Principal Component Analysis Approach for Yield Attributing Traits in Okra (*Abelmoschus esculentus* L.) Genotypes

P V Abhilash¹, Nilesh Talekar^{1*}, I. R. Delvadiya¹, Shailesh Kumar Singh²

¹Department of Genetics and Plant Breeding, School of Agriculture, Lovely Professional University, Jalandhar – Delhi G. T. Road, Phagwara-144411, (Punjab), India

²Department of Horticulture, School of Agriculture, Lovely Professional University, Jalandhar – Delhi G. T. Road, Phagwara-144411, (Punjab), India

*Corresponding author's E-mail: nilesh.23498@lpu.co.in

Article History	Abstract
Received: 06 June 2023 Revised: 05 Sept 2023 Accepted: 26 Oct 2023	<p>The research work was investigated among 55 okra accessions in RCBD with three replications and was evaluated for seventeen phenotypic characteristics of okra principal component analysis at the Agriculture Research Farm, Lovely Professional University, Phagwara (Punjab). According to principal component analysis (PCA), six of the seventeen (PC1 to PC6) PCs had eigenvalues above 1.0 and a cumulative variance of around 75.52. PC 1 alone accounted for the highest variance of 25.38 by PC1, followed by PC 2 with 15.98%. The outcomes of this investigation might be used as a foundation for defining and implementing subsequent okra breeding initiatives. Days to the 1st flowering, days to the 50% flowering and days to the first fruit harvest appear above the average variance contribution of each PC in the screen plot. Biplot analysis of PC1 and PC2 in contribution dim.1 revealed 25% and dim.2 revealed 16%.</p>
CC License CC-BY-NC-SA 4.0	Keywords: Eigenvalue, Eigenvector, PCA, Biplot, Scree plot, Okra

1. Introduction

Okra, scientifically known as *Abelmoschus esculentus* (L.) Moench, holds paramount importance as a highly nutritious vegetable crop in India and is a member of the Malvaceae family. This plant is classified as a warm-season annual vegetable with a cultivation period spanning 90 to 100 days. It thrives in tropical and subtropical regions, where it exhibits adaptability to a wide range of soil types, ranging from sandy to clay compositions (Melaku *et al.*, 2020; Mohammed *et al.*, 2022). India is the world's largest okra producer, accounting for 73.96% of the total okra area (NHB 2021-22). In India, the area under cultivation for okra is 555 thousand ha, with a yield of 6819 thousand MT, with an average productivity of 12.07 tonnes per hectare (Agriculture Statistic 2022).

Okra is a tropical African native that originated in northeastern Africa, primarily in Ethiopia and Sudan, and has since spread to Asia, the Mediterranean, North, South, and Central America, as well as southern parts of Europe Temam *et al.*, (2020).

Principal Component Analysis (PCA) is a multivariate technique that involves the transformation of a set of potentially correlated variables into a reduced set of variables known as principal components. The main advantage of Principal Component Analysis (PCA) is to measure the significance of each dimension in explaining the variability of a dataset. It relies upon eigen vectors and eigen values to signify data (Mishra *et al.*, 2017). The eigenvalues of principal components in PCA represent the amount of variation present in traits and are indeed useful for further breeding programs as they help identify important traits that can be targeted for selective breeding in okra.

2. Materials And Methods

Study area

During the summer of 2021–2022, the experiment was conducted at the Agricultural Research Farm, Lovely Professional University Department of Genetics and Plant Breeding, in Phagwara, Punjab.

Experimental material and design

The study consisted of 55 different genotypes arranged using a randomized complete block design (RCBD). These genotypes are sown with a spacing of 45 (row) X 15 (plant) cm. All packages of

practices were followed as per standard recommendations under irrigated conditions, which indicates that in the study or experiment related to okra (*Abelmoschus esculentus* L.) genotypes.

Observations recorded

Research on a PCA study involving 55 accessions and 17 quantitative traits of okra the data on these traits were recorded following the descriptor list developed for okra by the International Plant Genetic Resources Institute (IPGRI) in 1991. Which includes FFn: first flowering nodes; Ffrn: first fruiting nodes; DfF: days to the first flowering; D50f: days to the 50% flowering; DfFh: days to the first fruit harvest; PtH: plant height; InL: inter-nodal length; NbPp: number of branches per plant; NnPp: number of nodes per plant; Fl: fruit length (cm); Fd: fruit diameter (cm); NrPf: number of ridges per fruit; NmFp: number of marketable fruits per plant; NPP: number of picking plants; NfPp: number of fruits per plant; AFw: average fruit weight; and FYp: fruit yield per plant.

Statistical Analysis

Reduce dimensionality and analyse the contribution of traits, "ggbiplot" package in R is utilized to create a PCA biplot. The screen plot might help you decide how many principal components to keep for further study Lever *et al.* (2017) and Gabriel (1971).

3. Results and Discussion

Principal Component Analysis:

The entire collection of data is divided into 17 Principal Components. Table 1 and Figure 1 reveal that the first six principal components (PCs) with eigenvalues >1 account for around 75.52% of the total variability in the traits under consideration. In our analysis, the top three cumulative PCs accounted for 53.78 percent of the variation, consistent with a finding from earlier researchers (Das *et al.*, 2022). With a significant proportion of overall variation that varied from 25.38 to 6.47%, PC1 had the highest variance (25.38%), followed by 15.98% (PC2), 12.45% (PC3), 8.49% (PC4), 6.73% (PC5), and 6.47 (PC6), a similar variance range recorded in PC1 to PC6 by (Solankey and Singh 2018). The principal component analysis found six fundamental variables with eigenvalues between 4.315 and 1.101 (PC 1 to PC 6). PC1 had an eigenvalue of 4.315, PC2 have an eigenvalue of 2.717, PC3 had an eigenvalue of 2.116, PC4 has an eigenvalue of 1.445, PC5 had an eigenvalue of 1.145, and PC6 has an eigenvalue of 1.101 (Fig. 1). Due to Gutten's lowest limit principle, eigenvalues of 1 should have been discarded (Alemu *et al.*, 2022). From the first PC, a semi-curve figure was formed, with minimal variance within every PC suggested, implying that characters beneath the very first PC might well become preferable. Principal Component Analysis (PCA) is a powerful multivariate technique used to analyze data tables with inter-correlated quantitative dependent variables. Its primary objective is to extract essential information from the statistical data and represent it as a set of new orthogonal variables known as principal components. PCA helps in displaying the patterns of similarity between observations and variables as points on spot maps. This dimensionality reduction technique is commonly used in various fields to simplify data visualization and analysis (Mishra *et al.*, 2017).

Table 1. Eigen values, % variance and cumulative variance of okra genotypes

Traits	PC	eigenvalue	percentage of variance	cumulative percentage of variance
First flowering node	PC1	4.315	25.384	25.384
First fruiting node	PC2	2.717	15.983	41.367
Days to the first flowering	PC3	2.116	12.45	53.817
Days to the 50% flowering	PC4	1.445	8.499	62.316
Days to the 1 st fruit harvest	PC5	1.145	6.736	69.053
Plant height	PC6	1.101	6.476	75.529
Inter nodal length	PC7	0.867	5.103	80.632
No.of primary branches/ plant	PC8	0.691	4.067	84.698
Number of nodes/ plants	PC9	0.579	3.407	88.105
Fruit length	PC10	0.472	2.778	90.882
Fruit diameter	PC11	0.426	2.507	93.389
No.of ridges/ fruit	PC12	0.385	2.263	95.652
No. of marketable fruits/ plant	PC13	0.243	1.432	97.084
No.of pickings/ plant	PC14	0.203	1.195	98.28
No.of fruits/ plant	PC15	0.173	1.019	99.299
Average fruit weight	PC16	0.075	0.443	99.742

Fruit yield / plant	PC17	0.044	0.258	100
----------------------------	------	-------	-------	-----

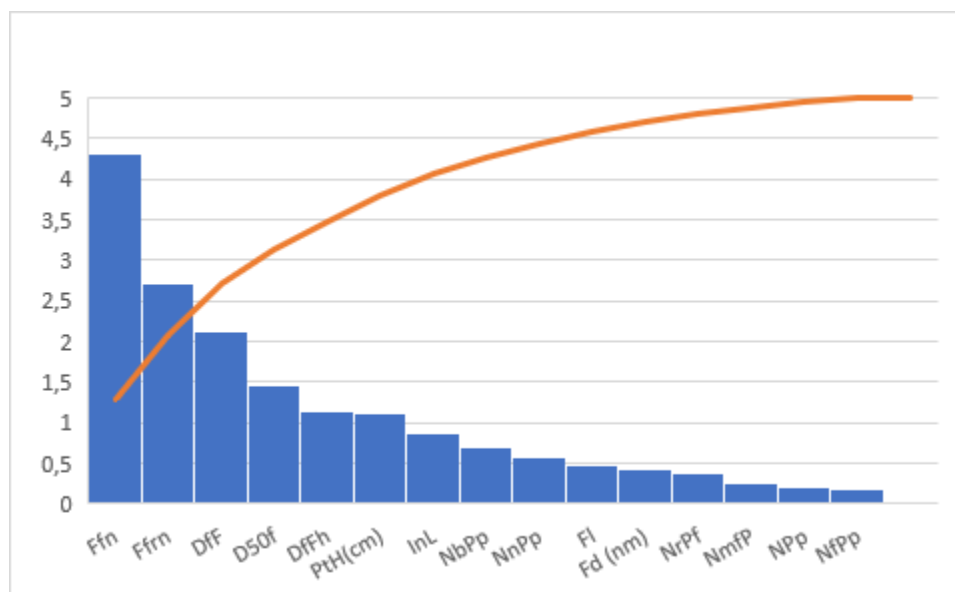


Fig. 1. Eigen values scree plot of okra

Screen plot

The scree plot, is a graphical representation of the overall variance contribution of each PC. The largest variance contribution is shown in increasing order in the line graph. According to this, PC1 is more closely related to variance, and from PC2 to PC17, this relationship begins to decline. Principal Component Analysis (PCA) findings from PC1 to PC6 are discussed above since they contain a lot of variances. While PC7 to PC17 showed less fluctuation. It's depicted in Figure 2.

Within this same PC1, qualities with the highest utter and total numbers closest to one have a greater effect on the cluster than those with the generally lowest scores closest to zero (Kenaw *et al.*, 2023). Therefore, this divergence of variations in many clusters has been due to the combined effect of those many phenotypes rather than the same enormous contribution of a limited number of traits.

PC scores were generated in six principal components within individual genotypes, which were then used to determine which varieties performed better given different variations of morphological variables. The PCA score values are presented in Tables 2 and 3. A significant PC score associated with a unique genotype in a specific PC suggests higher values for just the parameters in the same genotype which represents component. As a result, individual values may be exploited to generate exact screening indicators, the magnitude of which could be determined through the amount of variance per PC reflects.

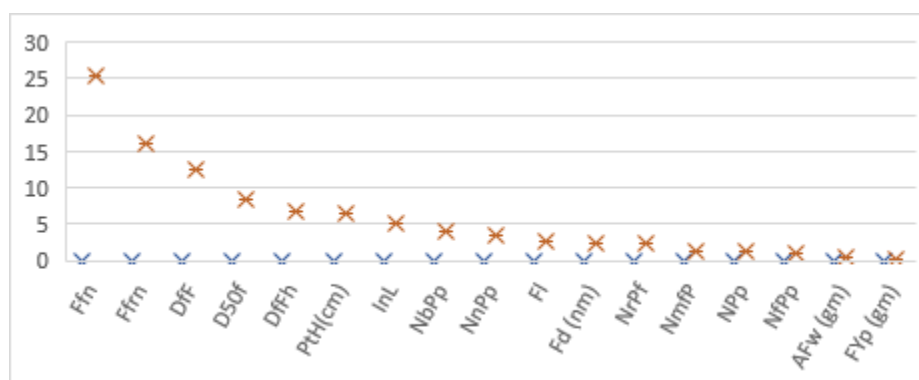


Fig. 2. Scree plot of variance in 17 PCs

Rotated component matrix

In this study focusing on okra germplasm, a PCA analysis was conducted on 55 yield traits. The characters were estimated based on the PC scores (principal component scores) and are presented in Tables 2 and 3. These scores provide valuable information that can be used to develop specific selection indices, with their intensity determined by the variability explained by each principal component. Figure 3 presents a graphical depiction illustrating the rotated component matrix in six Principal Components

(PCs) associated with distinct traits. Implying that characters beneath the very first PC might well become preferable with regard to PC1 in (table 2), NfPp (17.31) has the highest positive correlation, followed by FYp (13.54), NPp (13.50), Fl (12.54), NmFp (12.33), and NnPp (12.27). The second PC had a high positive correlated with DfF (24.70), followed by D50f (22.37) and DfFH (17.81). Here, characters such as fd (28.99) followed by NrPf (28.74) in the 3rd PC show a positive correlation. InL (44.63) and PtH (22.21) both had positive correlations with the fourth PC. The fifth PC was positively correlated with FfrN (19.31), followed by FFn (16.99). The sixth PC was positively correlated with InL (35.66), followed by NbPp (30.94) and AFw (18.08). Seventeen yield-contributing traits of okra were divided into 6 principal components, and similar high positive trait values were found in PCs by Ranga *et al.*, 2021, for fruit length and number of fruits per plant in PC1. For PC2, similar findings were observed by Kumar *et al.*, (2016) for days to the 50% flowering, days to the first flowering, days to the first fruit harvest, and PC4 for plant height. Genotypes selected on the basis of RC score in each component having the high positive values in RC1 (Table 3) were accessions such as 27 were followed by 19, 35, and 20. The second RC has high positive values for accessions like 5, 12, 4, 35, 34, 33, and 22. In the 3rd RC, genotypes such as 2, 1, 10, 25, and 27 show high positive values. In the 4th RC, genotypes such as 39,19, and 21 show high positive values. Genotypes with high positive values in the fifth RC include 21, 36, 8, 33, 23, 54, 53, and 52. The sixth RC positive value is germplasm like 48, 3, 8, 1, 12, 50, and 52.

Table 2. Principal Components for 17 yield contributing traits of Okra

Traits/ PC	PC1	PC2	PC3	PC4	PC5	PC6
First flowering node	2.08	17.00	6.82	4.57	16.99	0.31
First fruiting node	4.12	13.82	6.80	2.53	19.31	0.33
Days to the first flowering	1.40	24.70	2.26	3.26	6.18	0.52
Days to the 50% flowering	0.41	22.37	3.03	1.24	4.29	0.02
Days to the 1st fruit harvest	2.78	17.81	2.50	3.18	1.08	6.80
Plant height	0.17	0.04	0.60	22.21	5.44	35.66
Inter nodal length	0.01	0.60	1.10	44.63	1.84	0.01
Number of primary branches/ plants	2.33	1.59	2.46	4.07	2.10	30.94
Number of nodes/ plants	12.27	0.14	1.82	0.91	0.00	2.40
Fruit length	12.54	0.15	0.02	1.10	5.92	1.10
Fruit diameter	0.08	0.15	28.99	0.67	13.68	0.02
No.of ridges/ fruit	0.64	0.45	28.74	6.20	1.09	1.00
No. of marketable fruits/ plant	12.33	0.00	2.23	0.33	0.77	0.28
No.of pickings/ plant	13.50	0.02	2.16	2.10	0.28	0.39
No.of fruits/ plant	17.31	0.01	2.53	1.10	0.12	1.54
Average fruit weight	4.48	1.16	6.51	1.89	15.33	18.08
Fruit yield / plant	13.54	0.00	1.44	0.00	5.58	0.60

Table 3. Correlation matrix and its contribution of traits to Principal Components

S.NO	Genotypes	RC1	RC2	RC3	RC4	RC5
1	IC052299	0.125	-0.97	3.017	0.458	0.895
2	IC052301	-2.749	0.244	4.504	0.645	0.382
3	IC052302	-4.412	-0.393	-2.706	0.343	-1.892
4	IC052303	-0.58	1.068	0.241	0.422	-0.241
5	IC052312	0.805	1.567	-0.142	-0.027	-0.698
6	IC052321	-0.266	0.565	-0.293	-0.522	-0.584
7	IC052322	0.512	-1.089	0.307	-0.554	0.834
8	IC057733	0.001	-1.207	-0.084	-0.613	1.945
9	IC058235	-0.02	-0.424	-0.716	-0.69	-0.605
10	IC058704	-0.412	-0.085	1.213	-0.209	-0.807
11	IC058710	0.208	0.569	0.56	-0.507	0.822
12	IC058712	-0.932	1.361	0.627	0.571	-1.333
13	IC058768	0.531	-1.61	0.036	-0.507	-0.741
14	IC086008	0.032	0.846	0.322	0.121	0.274
15	IC089712	0.023	-1.747	-0.195	0.809	-0.601
16	EC169509	-0.347	-0.532	-0.064	-0.128	0.445
17	EC169450	0.523	0.849	-0.494	0.566	0.122
18	EC169452	-0.243	0.205	-0.307	-0.542	-0.963

19	EC169453	1.987	0.272	-0.773	1.18	-0.19
20	EC169455	1.2122	0.291	-0.106	1.077	0.051
21	EC169456	0.56	0.029	-1.251	0.911	2.12
22	EC169459	-0.441	1.108	0.066	0.224	0.561
23	EC169461	-1.303	-0.099	-0.898	0.772	1.304
24	IC045995	-0.64	1.534	0.272	-0.621	0.968
25	IC050418	0.158	-2.046	-0.433	-0.565	0.869
26	IC046018	-0.84	-0.168	1.349	-0.569	-0.325
27	IC048281	2.042	-0.766	1.132	-0.195	-0.737
28	IC048948	0.662	-1.231	0.044	-0.932	-0.239
29	IC045993	0.47	0.158	0.467	0.072	-0.142
30	IC045994	-1.038	-0.17	-0.091	0.365	-0.454
31	IC049972	0.669	2.134	-0.524	-1.391	-1.027
32	IC049734	0.485	0.433	0.744	-1.581	-0.522
33	IC049749	-1.227	1.093	-0.259	-0.75	1.585
34	IC052298	-0.051	1.185	-0.016	-1.047	0.707
35	IC052313	1.428	1.34	0.22	-1.888	-0.311
36	IC052308	0.386	0.817	-0.154	-1.161	1.504
37	IC052320	-0.848	0.678	-0.234	-2.045	0.611
38	IC052310	-0.313	-0.877	-0.889	-1.47	-0.088
39	P7	0.559	-1.222	0.168	1.634	-0.104
40	Hari Kranti	0.424	-1.615	0.48	-0.098	-1.739
41	Plamkomal	0.659	-0.425	-0.46	-1.652	0.085
42	Anima	-0.114	-2.143	-0.031	-1.09	-0.621
43	Pusa Makmali	-0.465	-1.576	-0.235	0.346	0.428
44	Salekerthi	0.746	-0.307	0.298	0.363	-0.845
45	Kasi Kranti	0.499	0.445	0.066	1.226	-2.205
46	Pusa sawani	0.014	0.55	-0.807	2.135	-0.063
47	Hisar Naveen	0.582	0.777	-0.777	1.426	0.557
48	VRO-4	0.721	0.551	-0.408	1.33	0.899
49	ArkaAnamika	-0.244	-0.257	-0.895	0.528	-0.141
50	Kashi Pragati	0.444	0.413	-0.957	-0.162	0.311
51	Dhanvi66	-0.268	-0.626	-1.057	-1.333	-0.78
52	Hinarch	0.188	0.347	0.082	1.254	1.139
53	GFS Gold	0.101	-0.909	-0.476	1.6	1.214
54	Punjab8	-0.623	0.107	-0.07	1.112	1.07
55	Somya	0.62	0.958	0.587	1.359	-2.704

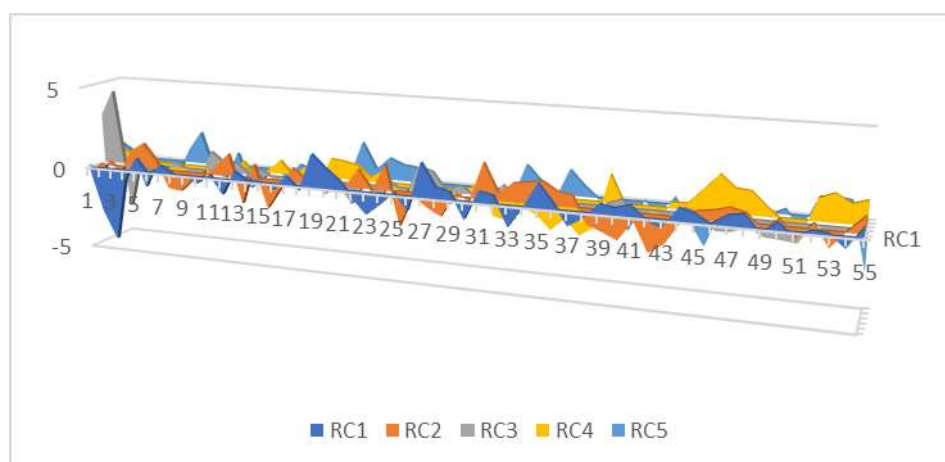


Fig. 3. Graphical demonstration of the rotated component matrix for the six RC's different traits

Biplot analysis:

The biplot analysis helps visualize the relationships among the accessions based on two principal components, allowing researchers to identify potential patterns and groupings among the accessions. Regarding the dimensions that account for the highest variance (Dim 1 on the X-axis and Dim 2 on the Y-axis), it is possible to partition the PCA biplot shown in Figure 4 into four quadrants. As biplot spots

variance between traits and genotypes, blue colour is used as an intensity of variance. Biplot analysis of PC1 and PC2 in contribution dim.1 revealed 25% and dim.2 revealed 16%. Quadrant 1 (Q-1) lies left above the center has a positive effect on Dim. 2 and a negative effect on Dim. 1. Genotypes with the highest variance, including 33, 24, 34, 22, 31, 37, 54, 14, 4.23.6, 46, and 2 (Table 3), are grouped in Q-1, along with those with the greatest variance traits such as DfF, D50f, DfFH, and NrPf. Genotype 33 is observed as the highest contributing accession, and DfF has strong and high variance as it lies away from the graph line. Comparable biplot results were documented by Solankey and Singh (2018) for days to 50% flowering. Quadrant 2 (Q-2) is the one that is directly above the centroid and has a positive effect on both Dim. 1 and Dim. 2. Accessions such as 36, 35, 5, 17, 11, 50, 47, 48, 19, 20, and 52 (Table 3) influence the variability observed in traits such as FFn, FfrN, AFw, FI, NnPP, and PtH. The third quadrant (Q-3) lies to the left of the center and has a negative influence on both Dim.1 and Dim.2, respectively. Genotypes like 26, 10, 30, 55, 51, 45, and 19 (table 3) are highly associated with Dim.1 and Dim.2 for traits Fd. The 4th quadrant (Q-4) lies to the left of the center and has a negative influence on both Dim.1 and Dim.2. Right below the centroid, genotypes 33, followed by 53, 29, 8, 44, 41, 38, 43, 25, 13, 40, 28, and 1 (Table 1), are grouped in Q-4 with traits InL, NbPp, and NPp. The trait DfF contributed greater variation than other contributors, as mentioned in Fig. 3, the line graph displays the average variance of all features. The overall variance for D50f and DfFH stands above average after DfF. Variation in okra has been observed to be heavily influenced by both fruit length and yield per plant (Amoateyet *al.*, 2015). To find out how genotypes react to overall variability, a study was conducted. When all genotypes were compared, it was found that genotype IC049749 had the highest level of variation in Dims. 1 and 2, followed by genotypes IC045995 and IC052298, according to Table 3.

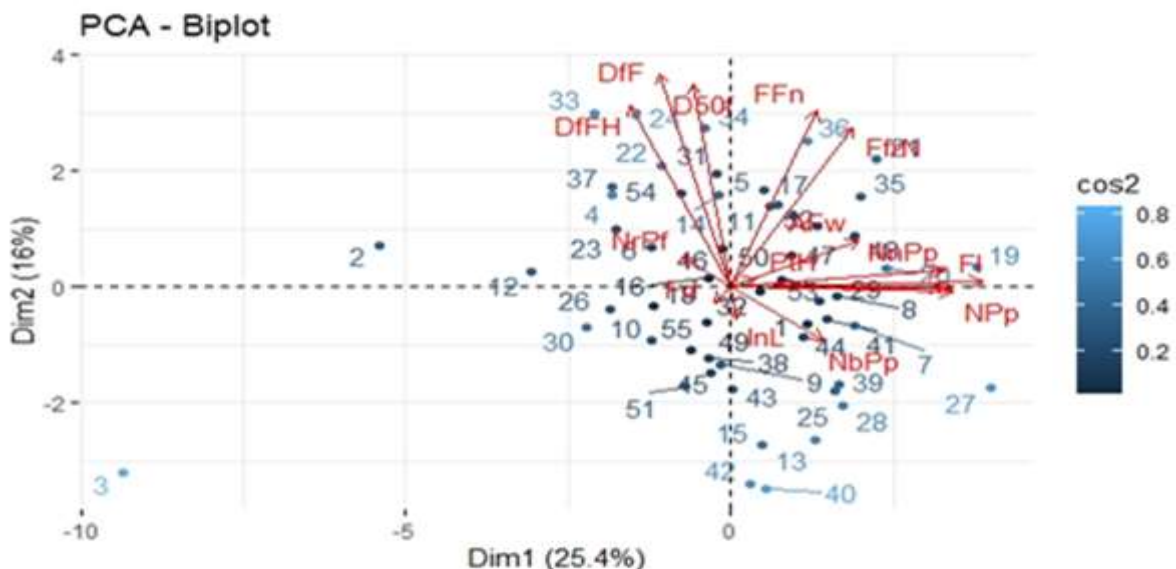


Fig. 4 Biplot analysis of PC1 and PC2 in contribution of variance to traits and genotypes of study

4. Conclusion

Through the PCA, researchers learn how much variation each genotype and study characteristic contribute. Every breeder should use highly variable genotypes for their breeding programs. However, biplot connects genotypes to traits to ensure that there is a connection between them in variability. Biplot analysis detected variation in traits like fruit length, fruit weight, and the number of fruits per plant. These are crucial in contributing to total yield. When all genotypes were compared, it was found that genotype IC049749 had the highest level of variation. PCA should look into this and provide information regarding the selection of variable traits that increase crop productivity.

Conflict Of Interest

We have no conflicts of interest to disclose.

References:

Agriculture Statistics at a Glance 2022. Government of India, Ministry of Agriculture & Farmers Welfare, Department of Agriculture & Farmers Welfare, Directorate of Economics & Statistics. www.agricoop.nic.in.

- Alemu, F., & Mohammed, W. 2022. Genetic Diversity Assessment of Okra (*Abelmoschus esculentus* (L.) Moench) Collections in Ethiopia Using Multivariate Analysis. *Middle East Journal of Agriculture Research*, 11(03):757-769.
- Amoatey, H.M., Klu, G.Y.P., Quartey, E.K., Doku, H.A., Sossah, F.L., Segbefia, M.M. & Ahiakpa, J.K. 2015. Genetic Diversity Studies in 29 Accessions of Okra (*Abelmoschus* spp L.) Using 13 Quantitative Traits. *American Journal of Experimental Agriculture*, 5(03): 217-225.
- Das, A., Yadav, R. K., Choudhary, H., Lata, S., Singh, S., Kumar, C., & Talukdar, A. 2022. Population structure, gene flow and genetic diversity analyses based on agro-morphological traits and microsatellite markers within cultivated and wild germplasms of okra [*Abelmoschus esculentus* (L.) Moench.]. *Genetic Resources and Crop Evolution*, 69(2): 771-791.
- Gabriel KR. 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3):453-467. <https://doi.org/10.1093/biomet/58.3.453>.
- International Plant Genetic Resource Institute (IPGRI). 1991. Okra Descriptor, Diversity for Development, Inter-national Plant Genetic Resource Institute, Rome, Italy.
- Kenaw, W., Mohammed, W., & Woldetsadik, K. 2023. Morpho-agronomic variability of okra [*Abelmoschus esculentus* (L.) Moench] genotypes in Dire Dawa, eastern Ethiopia. *Plos one*, 18(7). <https://doi.org/10.1371/journal.pone.0288534>.
- Kumar, A., Solankey, S. S., Nand, N., Adarsh, A., & Verma, R. B. 2016. Assessment of genetic diversity in Okra (*Abelmoschus esculentus* L. Moench) for yield and yellow vein mosaic virus incidence. *International Journal of Agriculture, Environment and Biotechnology*, 9(4): 485-491.
- Lever, J., Krzywinski, M., & Altman, N. 2017. Points of significance: Principal component analysis. *Nature methods*, 14(7), 641-643.
- Melaku, A.B., W. Mohamed., & V. Kumar. 2020. Variability, heritability and genetic advance in indigenous and exotic okra [*Abelmoschus esculentus* (L.) Moench] genotypes for yield and yield related traits at Dire Dawa. Eastern Ethiopia. *MOJ Eco Environ Sci*, 5(4): 164-169.
- Mishra, S. P., Sarkar, U., Taraphder, S., Datta, S., Swain, D., Saikhom, R., & Laishram, M. 2017. Multivariate statistical data analysis-principal component analysis (PCA). *International Journal of Livestock Research*, 7(5): 60-78.
- Mohammed, J., Mohammed, W., & Shiferaw, E. 2022. Performance and genetic variability of okra (*Abelmoschus esculentus* (L.) Moench) genotypes in Ethiopia for agromorphology and biochemical traits. *Advances in Agriculture*, 1-8.
- NHB, 2022. National Horticulture Board, Government of India, Ministry of Agriculture & Farmers Welfare.
- Ranga, A. D., Kumar, S., & Darvhankar, M. S. 2021. Variability among different yield and yield contributing traits of Okra (*Abelmoschus esculentus* L. Moench) genotypes. *Electronic Journal of Plant Breeding*, 12(1):74-81.
- Solankey, S. S., & SINGH, A. K. 2018. Genetic Evaluation and Bi-plot Analysis in Okra Hybrids for Yield, Quality Components and YVMV Resistance: Evaluation of Okra Hybrids Yield and YVMV Resistance. *Journal of Agri Search*, 5(4):240-246.
- Temam, N., Mohamed, W., & Aklilu, S. 2020. Agro morphological characterization and evaluation of okra [*Abelmoschus esculentus* (L.) Moench] genotypes for yield and other variability components at Melkassa, Central Ethiopia. *MOJ Ecology & Environmental Sciences*, 5(2), 80-87.