

Unravelling the Complex Interactions among Yield Attributes in Upland Cotton (*Gossypium hirsutum* L.) through Principal Component Analysis

Bana Venkata Ravi Prakash Reddy*, Kalapati Mohan Vishnuvardhan, U. Nikhil Sagar, D. Lakshmi Kalyani, M. Siva Ramakrishna, K. Amarnath and N.C. Venkateswarlu

Acharya NG Ranga Agricultural University, Regional Agricultural Research Station, Nandyal 518 502, India

(Received 13 September, 2023; Accepted 6 October, 2023)

ABSTRACT

Present investigation was carried out with 34 upland cotton genotypes to assess the genetic divergence for seven yields and yield attributes through principal component analysis to identify the factors contributing to yield variability. The PCA showed that the first two principal components explained 59.89% of the total variation. PC1 components showed positive relationship with ginning percentage whereas PC2 showed positive relationship with plant height, bolls per plant and seed cotton yield. Hence, these characters will be given priority in future for upland cotton breeding programme. Cluster analysis using Ward's method was performed with seven quantitative traits displayed that four clusters were formed in which largest contribution was shared by cluster I with a maximum of 15 genotypes followed by cluster II consists of 10 genotypes. Based on PCA scores and cluster mean values, the genotypes NH 702, NDLH 2051-3, NH 703, AKH 2013-2 and CPD 1951 were considered to be desirable for utilization in upland cotton improvement.

Key words: Upland cotton, Yield and yield attribute, Principal component analysis, Genetic divergence and Biplotanalysis

Introduction

Cotton is an important commercial crop providing lint as a raw material to the textile industry. It is an important agricultural commodity grown in a wide range of ecosystems under varying temperatures and water regimes. Globally, cotton is cultivated over an area of 33.18 m ha with a production of 25.73 m tonnes and productivity of 775 kg ha⁻¹. In India, it is grown in an area of 12.65 m ha with a production of 5.9 m tonnes and productivity of 466 kg ha⁻¹ (International Cotton Advisory Committee (ICAC) data portal, 2021-22). There are four cultivated species of cotton viz., *Gossypium arboreum*, *G. herbaceum*, *G. hirsutum* and *G. barbadense*. Among

these four cultivated species, *G. hirsutum* L. is known for its production potential, as demonstrated by the release of number of stable varieties and hybrids. *Gossypium hirsutum*, also called upland cotton, represents 95% of global cotton fiber production (Wendel *et al.*, 1994). Yield is the most important character and ultimate objective of breeding in cotton. To study yield potential, the correlation among different characters and knowledge of their contribution to yield is important in planning a breeding program (Thiyagu *et al.*, 2010). Further, the efficiency of selection in any breeding programme mainly depends upon the knowledge of association of the characters. For the assessment of genetic diversity, breeders have widely used principal compo-

nent analysis in exploration of promising genotypes for future breeding programmes (Sun *et al.*, 2019). Principle Component Analysis (PCA), a multivariate technique used to classify the genetic relationships between the traits in multi-trait systems and for identifying the patterns of data by reducing the number of dimensions. In the present investigation, correlation, PCA and cluster analysis were used to study the complex interaction among yield attributes in upland cotton.

Materials and Methods

The present investigation was carried out using 34 upland cotton genotypes at Regional Agricultural Research Station, Nandyal in 2022 using Randomized Block Design with three replications. Each genotype was grown in 2 rows with row length of 5 m and spacing of 60 cm between the rows and 30 cm between the plants. Recommended ANGRAU package of practices were followed during crop growth period uniformly to all genotypes. The data related to days to 50% flowering, plant height, number of bolls per plant, boll weight, ginning percentage, halo length, seed cotton yield were recorded in this study. Further, the recorded data were subjected to the analysis of variance, descriptive statistics, correlation analysis and Principal component analysis using STAR (Statistical tool for Agricultural Research) 2.1.0 software (Gulles *et al.*, 2014).

Results and Discussion

A set of 34 upland cotton genotypes were evaluated for yield and yield attributing traits under field conditions. The mean values ranged from 4.03 (g plant⁻¹) for boll weight to 2176.69 (kg ha⁻¹) for seed cotton yield per plant among the evaluated genotypes (Table 1). The values of coefficient of variation ranges from 2.32 for days to 50% flowering to 15.01

for bolls per plant. The maximum value of heritability is seen for halo length (0.85) whereas the minimum value of heritability is seen for days to 50% flowering (0.24). The maximum value was observed for seed cotton yield (2176.69 kg ha⁻¹) with values ranging from 1228.02 to 2981.15 whereas the minimum value was observed for the trait, boll weight (4.03 g plant⁻¹) with values ranging from 3.3 to 4.8. Standard deviation values range from 0.41 for boll weight to 413.38 for seed cotton yield. Similar results were reported by Chaudhry *et al.* (2022) and Khattak *et al.* (2022).

The Pearson's correlation coefficients between the evaluated seven traits in 34 cotton genotypes were presented in Table 2. From the correlation analysis, it is observed that the trait bolls per plant showed significant and positive correlation ($r = 0.510$) with plant height. These findings are similar to those results reported by Mawblei *et al.* (2022) and Manan *et al.* (2022). The trait halo length recorded significant and positive correlation ($r = 0.360$) with boll weight. Seed cotton yield recorded significant and positive correlation with both plant height ($r = 0.561$) and bolls per plant ($r = 0.734$). However, ginning percentage showed significant and negative correlation with bolls per plant ($r = -0.351$) and halo length ($r = -0.522$). Similar findings were earlier reported by Rizwan *et al.* (2022) and Yousaf *et al.* (2023).

In principal component analysis, the number of variables is reduced to linear functions called canonical vectors which accounts for most of the variation produced by the characters under study. A scree plot is a graph which orders the eigen values from largest to smallest and also showed the % of variability in terms of eigen values and principal components. The scree plot showing eigen values against respective principal components is represented in Figure 1 (a). PC1 showed 36.91% variability with eigen value 2.58 and then the graph gradually declined for other PCs. The eigen values, per

Table 1. Estimation of basic statistical values for seven quantitative traits in 34 cotton genotypes

Traits	Mean	SE	SD	Minimum	Maximum	CV	Heritability
Days to 50% flowering	58.55	0.19	1.09	56.00	60.00	2.32	0.24
Plant height (cm)	85.71	1.54	8.99	63.33	101.00	9.60	0.46
Bolls per plant	13.17	0.29	1.74	9.53	16.27	15.01	0.30
Boll weight (g/plant)	4.03	0.07	0.41	3.3	4.8	8.00	0.56
Ginning percentage	32.05	0.42	2.47	24.5	35.8	3.81	0.79
Halo length	27.57	0.35	2.04	23.9	31.7	3.04	0.85
Seed cotton yield (kg/ha)	2176.69	70.89	413.38	1228.02	2981.15	13.21	0.63

Table 2. Pearson phenotypic correlation coefficient among seven quantitative traits in cotton genotypes

Traits	DF	PH	BP	BW	GP	HL	SCY
DF	1.000						
PH	0.062	1.000					
BP	0.183	0.510**	1.000				
BW	0.093	0.027	0.100	1.000			
GP	-0.325	-0.158	-0.351*	-0.139	1.000		
HL	0.207	-0.087	-0.014	0.360*	-0.522**	1.000	
SCY	0.256	0.561***	0.734***	0.322	-0.235	0.073	1.000

Note: DF = days to 50% flowering; PH = plant height; BP = bolls per plant; BW = boll weight; GP = ginning percentage; HL = halo length; SCY = seed cotton yield

Table 3. Eigenvectors, eigenvalues, percentage variance and cumulative variance of seven quantitative traits in cotton genotypes

Traits	PC1	PC2
Days to 50% flowering	-0.277	-0.242
Plant height (cm)	-0.388	0.405
Bolls per plant	-0.498	0.287
Boll weight (g/plant)	-0.243	-0.311
Ginning percentage	0.376	0.383
Halo length	-0.214	-0.632
Seed cotton yield (kg/ha)	-0.528	0.230
Eigen value	2.58	1.61
Variance (%)	36.91	22.98
Cumulative Variance (%)	36.91	59.89

cent variance, per cent cumulative variance are presented in Table 3. Biplot against PC1 and PC2 for the studied characters of 34 genotypes is represented in the Figure 1 (b). The biplot analysis revealed that

genotypes are diverse for the characters under PC1 and PC2. The first two principal components with eigen value more than one accounted for 59.89% of the total variation for discriminating the lines. PC1 showed highest amount of variance (36.91%) with mostly related trait like ginning percentage (0.376). As a result, the first principal component analysis identifies the characters related to ginning percentage. PC2 showed second highest amount of variability (22.98%) with cumulative variance (59.89%) with mostly related to traits like plant height, bolls per plant, ginning percentage, seed cotton yield. Therefore, PC2 identifies the characters related to ginning percentage and seed cotton yield. The scores of 34 cotton genotypes into two principal components are represented in the Table 4. These scores can be used for the purpose of precise selection indices whose intensity is based on the variability showed by the respective components. High score of a particular

Table 4. Scoring of 34 cotton genotypes in first two principal components

Genotypes	PC1	PC2	Genotypes	PC1	PC2
TSH 375	-1.174	0.274	BGDS 1047	-1.855	0.412
GBHV 206	0.069	1.132	CNH 2077	2.779	-0.258
NH 702	1.324	2.771	RB 603	0.073	0.251
ARBH 1951	1.150	0.399	TSH 356	-0.089	0.098
NH 615	-1.984	-0.445	GTHV 15/36	-0.400	0.937
NDLH 2051-3	-0.793	-0.213	RHC 1409	0.492	0.516
BS 8-19	0.482	-1.703	CPD 1952	-0.674	-0.194
RHC 1419	1.427	-0.378	CCH 19-2	-0.793	-1.827
TVH 007	0.377	0.495	CNH 17393	0.643	-1.968
CNDTS 283	1.209	0.795	NDLH 1938	-1.683	0.639
RB 610	-0.629	1.342	GBHV 209	0.137	1.460
CCH 19-3	-0.914	-1.911	RAH 1045	1.330	-0.214
RAH 1046	-1.168	0.401	NDLH 2056-4	-2.495	-0.293
NH 703	0.613	2.199	NH 701	1.582	1.444
AKH 2013-2	-4.419	-0.584	BS 7-19	2.338	-2.892
CPD 1951	3.911	-0.080	CNH 1134	-0.843	-0.491
CNH 09-119	-1.006	0.468	Suraj	0.982	-2.578

Table 5. Mean values of seven quantitative traits in four clusters of 34 cotton genotypes

Clusters	Genotypes	DF	PH	BP	BW	GP	HL	SCY
I	TSH 375, NDLH 2051-3, TVH 007, RAH 1046, NH 703, CNH 09-119, BGDS 1047, RB 603, TSH 356, GTHV 15/36, RHC 1409, CPD 1952, GBHV 209, NDLH 2056-4, CNH 1134	58.78	88.31	13.83	4.18	32.9	27.25	2411.96
II	GBHV 206, NH 702, ARBH 1951, RHC 1419, CNDTS 283, RB 610, CPD 1951, CNH 2077, RAH 1045, NH 701	57.67	82.73	12.48	3.71	33.03	25.94	1899.38
III	NH 615, AKH 2013-2, NDLH 1938	59.78	98.44	15.27	4.05	27.5	28.4	2529.21
IV	BS 8-19, CCH 19-3, CCH 19-2, CNH 17393, BS 7-19, Suraj	58.83	77.78	11.63	4.18	30.55	30.67	1874.45

genotype in particular principal component indicates high variability for a particular character of that principal component. The similar types of results were reported by Isong *et al.* (2017) and Ullah *et al.* (2022).

The 34 cotton genotypes were grouped into four clusters based on D² values in which the genotypes belonging to same cluster had an average minimum D² value, than that of genotypes present in different clusters. The Dendrogram of cotton genotypes resulting from cluster analysis using Ward's method based on seven quantitative traits is represented in Figure 2. Out of 4 clusters formed, cluster I was the largest comprising of maximum of 15 genotypes. The cluster II consists of 10 genotypes, cluster III consists of 3 genotypes and cluster IV comprises of 6 genotypes. Cluster mean values for 7 traits of all the 4 clusters were presented in Table 5. Cluster mean values for plant height were ranged from 77.78 (Cluster IV) to 98.44 (Cluster III). For days to 50% flowering, it ranged from 57.67 (Cluster II) to 59.78 (Cluster IV). Number of bolls per plant was ranged from 11.63 (Cluster IV) to 15.27 (Cluster III). Boll weight ranged from 3.71 g (Cluster II) to 4.18 g (Clusters I and IV). Ginning percentage ranged from 27.5 % (Cluster III) to 33.03 % (Cluster II). Halo length ranged from 25.94 mm (Cluster II) to 30.67 mm (Cluster IV). Seed cotton yield ranged from 1874.45 kg/ha (Cluster IV) to 2529.21 kg/ha (Cluster III). The similar types of results were reported by Shakeel *et al.*, (2015) and Jawahar *et al.* (2019).

Conclusion

From the study it may concluded that, the scrutiny

of per cent contribution of different characters towards the expression of total genetic divergence showed that seed cotton yield contributed maximum to total variation followed by plant height, days to 50% flowering, ginning percentage, halo length, number of bolls per plant and boll weight. Based on correlation, PCA and cluster analysis, the genotypes NH 702, NDLH 2051-3, NH 703, AKH 2013-2 and CPD 1951 were found to be desirable for yield and fibre quality improvement in upland cotton improvement programmes.

Conflict of Interest: None

References

- Chaudhry, U.F., Khalid, M.N., Aziz, S., Amjad, I., Khalid, A., Noor, H. and Sajid, H.B. 2022. Genetic studies in different F2 segregating population for yield and fiber quality traits in cotton (*Gossypium hirsutum* L.). *International Journal of Agriculture and Bioscience*. 11(11).
- Gulles, A.A., Bartolome, V.I., Morante, R.I.Z.A., Nora, L.A., Relente, C.E.N., Talay, D.T., Caneda, A.A. and Ye, G. 2014. Randomization and analysis of data using STAR (Statistical Tool for Agricultural Research). *Philippine Journal of Crop Science*. 39(1): 137.
- International Cotton Advisory Committee (ICAC) data portal. 2021-22. *Cotton Statistics*. <https://www.icac.org>.
- Isong, A., Balu, P.A. and Ramakrishnan, P. 2017. Association and principal component analysis of yield and its components in cultivated cotton. *Electronic Journal of Plant Breeding*. 8(3) : 857-864.
- Jawahar, A.H., Wang, X., Iqbal, M.S., Sarfraz, Z., Wang, L., Qifeng, M.A. and Shuli, F. 2019. Genetic divergence on the basis of principal component, correlation and cluster analysis of yield and quality traits in cotton cultivars. *Pakistan Journal of Botany*. 51(3):

- 1143-1148.
- Khattak, W.A., He, J., Abdalmegeed, D., Hu, W., Wang, Y. and Zhou, Z. 2022. Foliar melatonin stimulates cotton boll distribution characteristics by modifying leaf sugar metabolism and antioxidant activities during drought conditions. *Physiologia Plantarum*. 174: e13526.
- Manan, A., Zafar, M.M., Ren, M., Khurshid, M., Sahar, A., Rehman, A., Firdous, H., Youlu, Y., Razzaq, A. and Shakeel, A. 2022. Genetic analysis of biochemical, fiber yield and quality traits of upland cotton under high temperature. *Plant Production Science*. 25:105-119.
- Mawblei, C., Premalatha, N., Rajeswari, S. and Manivannan, A. 2022. Genetic variability, correlation and path analysis of upland cotton (*Gossypium hirsutum* L.) germplasm for seed cotton yield. *Electronic Journal of Plant Breeding*. 13(3): 820-825.
- Rizwan, M., Farooq, J., Farooq, A., Farooq, M., Sarwar, G., Nadeem, M., Riaz, M., Shahid, M.R., Abbas, H.G. and Sarwar, M.K.S. 2022. Yield and related components of cotton (*Gossypium hirsutum* L.) effected by chlorophyll contents. *Pakistan Journal of Agriculture Research*. 35(1): 29-35.
- Shakeel, A., Talib, I., Rashid, M., Saeed, A., Ziaf, K. and Saleem, M.F. 2015. Genetic diversity among upland cotton genotypes for quality and yield related traits. *Pakistan Journal of Agricultural Science*. 52(1): 73-77.
- Sun, Z., Xingfen, W., Zhengwen, L.I.U., Qishen, G.U., Yan, Z., Zhikun, L.I., Huifeng, K.E., Jun, Y., Jinhua, W.U., Liqiang, W.U., Guiyin, Z. and Zhiying, M.A. 2019. Evaluation of the genetic diversity of fibre quality traits in upland cotton (*Gossypium hirsutum* L.) inferred from phenotypic variations. *Journal of Cotton Research*. 2(22): 1-8.
- Thiyagu, K., Nadarajan, N., Rajarathinam, S., Sudhakar, D. and Rajendran, K. 2010. Association and path analysis for seed cotton yield improvement in inter-specific crosses of cotton (*Gossypium* spp.). *Electronic Journal of Plant Breeding*. 1(4): 1001-1005.
- Ullah, A., Shakeel, A., Ahmed, H.G.M., Naeem, M., Ali, M., Shah, A.N., Wang, L., Jaemko, M., Abdelsalam, N.R., Ghareeb, R.Y. and Hasan, M.H. 2022. Genetic basis and principal component analysis in cotton (*Gossypium hirsutum* L.) grown under water deficit condition. *Frontiers in Plant Science*. 13: 981369.
- Wendel, J., Rowley, R. and Stewart, J. 1994. Genetic diversity in and phylogenetic relationships of the brazilian endemic cotton, *Gossypium mustelinum* (malvaceae). *Plant Systematics and Evolution*. 192: 49-59.
- Yousaf, M.I., Hussain, Q., Alwahibi, M.S., Aslam, M.Z., Khalid, M.Z., Hussain, S., Zafar, A., Shah, S.A.S., Abbasi, A.M., Mehboob, A. and Riaz, M.W. 2023. Impact of heat stress on agro-morphological, physio-chemical and fiber-related parameters in upland cotton (*Gossypium hirsutum* L.) genotypes. *Journal of King Saud University-Science*. 35(1): 102379.