Correlations and Multivariate Analysis Across Non- Segregation and Segregation Generations in Two Cotton Crosses

Waleed Yehia

Agricultural Research Center

Essam Fathy El-Hashash (✉ dressamelhashash@azhar.edu.eg)

Al-Azhar University <https://orcid.org/0000-0002-0488-0610>

Research

Keywords: Generations, Correlation, Multivariate analysis, Cotton

Posted Date: June 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-587104/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Egyptian Journal of Agricultural Research on December 1st, 2021. See the published version at <https://doi.org/10.21608/ejar.2021.81571.1117>.

Abstract

Background: The cotton crop is one of the most important natural fibers crops for textile manufacture in the world. The present research uses Pearson's correlation coefficient and multivariate analysis to assess the interrelationships, similarities and dissimilarities among non-segregation (P_1 , P_2 and F_1) and segregation (F_2 , BC_1 and BC_2) generations for seed cotton yield and yield components in the two crosses Giza 92 x Pima S6 and Giza 93 x C.B. 58.

Results: The analysis of variance exhibited highly significant genetic variability among six generations for all studied traits in the two crosses. The F_1 performance was higher than the other generations for all the studied traits in the two crosses. The results had positive and high significant correlations between seed cotton yield/plant, lint cotton yield/plant and No. of bolls/plant traits across all six generations in the two crosses. A number of positive correlations were observed among the six generations for all studied traits in the two crosses. The UPGMA hierarchical clustering showed higher level of similarity coefficients among the six generations and among the studied traits. Similarity coefficients ranged from 0.96 to 0.99 and from 0.65 to 0.96, respectively. In the principal component analysis (PCA), the PCA1 extracted had Eigenvalue >1 across six generations for all studied traits in the two crosses. The PCA displayed total variation of 91.84% among the six generations contributed by PCA1 (79.47%) and PCA2 (12.38%) and had mainly distinguished the generations in different groups. The PCA1 and PCA2 were dominated by F_1 and segregation generations in the two crosses, respectively, showing high correlations with the first two PCAs. All studied traits as well as boll weight and lint percentage traits contributed positive significant component loadings for the PCA1 and PCA2, respectively. The biplot analysis of the relationship between the six generations revealed that the most appropriate generations for selecting yield traits were F_1 in the two crosses and BC_1 and BC_2 in the cross Giza 93 x C.B. 58.

Conclusion: From the obtained results, we recommend considering backcrossing may be done for 2–5 cycles (BC_2 – BC_5) at C.B. 58 parent for improving Egyptian cotton yield in the future.

Background

Cotton crop is one of the most founder crops and most important natural fibers for textile manufacture in the old and new world (Akter et al. 2019). It is a flowering plant belonging to the family Malvaceae and genus *Gossypium* (G). The G. genus is comprised of ~ 50 species. The most important cultivated species of the genus are *G. hirsutum*, *G. barbadense* (allotetraploid $2n = 52$), *G. herbaceum* and *G. arboreum* (diploid $2n = 26$) (Rehman and Farooq 2020). The annual world area harvested, yield and production of cotton were approximately 38640608 ha, 21374 hg/ha and 82589031 tonnes in 2019, respectively. As for Egypt, the total area harvested, yield and production during the 2019 growing season were 100000 ha, 30500 hg/ha and 20500 tonnes, respectively (FAOSTAT 2021). FAS/Cairo (Post) forecasts the market year 2021/22 cotton area harvested to increase by seven percent to 70,000 hectares (ha), from 65,000 ha in the market year 2020/21. Post estimates production of the market year 2021/22 at 250,000 bales compared to 215,000 bales in the market year 2020/21 (USDA 2021).

The main goal of cotton breeders is looking and selecting the genotypes with high yield traits and take advantage of the permanent untapped genetic variation of Egyptian varieties, substantial work has been carried out to develop both yield and quality traits of cotton in Egypt (Yehia and El-Hashash 2019). Through the process of breeding, progress in yield potential is produced by the gradual accumulation of genes that lead to higher yields or the elimination of unfavorable genes (Abdel-Monaem et al. 2020). The increase in yield can be possible if the existing genetic resources and information are properly utilized (Srouf and El-Hashash 2012).

A successful breeding program relies on comprehensive knowledge and understanding of relationships among yield traits and genetic diversity within and between the elite genetic materials of existing germplasm (Akter et al. 2019). The Pearson's correlation coefficient allows determining the magnitude of the relationships between essential and important traits (Abasianyanga et al. 2017). Thus, the correlation between traits can be useful in developing selection criteria for improving cotton yield, but correlation can also represent a swamp of interrelationships (Kloth 1998). Cotton breeders study the associations between the yield and its components traits, because the selection of one trait may lead to improvement or deterioration of associated traits, depending on nature and size (Nandhini et al. 2018). Several cotton researchers have reported a seed cotton yield/plant was related positively significantly with lint cotton yield/plant, No. of bolls/plant, boll weight, lint percentage and seed index traits (for example Kerby et al. 1990; El-Hashash 2013; Latif et al. 2015; Jarwar et al. 2019; Yehia and El-Hashash 2019; Eldessouky et al. 2021).

Multivariate analysis such as principal component and cluster analysis are mostly used to evaluate the magnitude of genetic diversity among the germplasm. The principal component analysis (PCA) explains the contribution of the most important traits that account for the total genetic variability. Cluster analysis allows the classification of individuals into groups in order to maximize their homogeneity within a

group and show heterogeneity among groups. It is complementary to PCA (El-Hashash 2016). Many researchers have used the PCA and cluster analysis to assess the relationship and diversity between several cotton germplasm, in addition to knowing the relationships between seed cotton yield and its components traits (Shah et al. 2018; Akter et al. 2019; Jarwar et al. 2019; Muminov et al. 2020; Abdel-Monaem et al. 2020; Sarwar et al. 2021). Scree plot is usually used for visual assessment of factor, which explains high amount of the changes in the data (Jarwar et al. 2019), also it further highlighted the partitioning of the principal components (Abasianyanga et al. 2017).

Due to the increasing demand for Egyptian cotton, this present investigation was conducted to study the interrelationships, similarities and dissimilarities among non-segregation and segregation generations for seed cotton yield and yield attributes in the two crosses by adopting Pearson's correlation coefficient and multivariate analysis.

Methods

Genetic Material and Field Procedure:

Four genotypes were used in this study namely Giza 92, Giza 93 (Egyptian varieties), Pima S₆ (Egyptian American) and C.B. 58 (USA barbadense), which all belong to the specie *Gossypium barbadense*, L. The experiments were carried out during the three successive growing seasons from 2018 to 2020 at Sakha Agricultural Research Station, Kafr El-Shiekh Governorate, Egypt. In 2018 season, the four parental varieties were crossed to produce F₁ hybrid seeds for the two crosses Giza 92 x Pima S₆ and Giza 93 x C.B. 58. At 2019 season, each F₁ was backcrossed to both parents to obtain BC₁ and BC₂; the parents were also crossed for more hybrid seeds and the F₁ plants was selfed to obtain F₂ seeds. The six populations i.e., P₁, P₂, F₁, F₂, BC₁ and BC₂ for the two crosses were evaluated separately in a randomized complete blocks design with three replications during 2020 season. Each replicate consisted of 10 rows for F₂, 5 rows for BC₁ and BC₂ crosses (segregating generations), and 3 rows for each non-segregating generations P₁, P₂ and F₁. Each row 4 meters in length and 0.60 m in width as well as comprised 10 hills. Hills were spaced at 40 cm apart and thinned to one plant per hill. All recommended agronomic practices of cotton were applied from sowing to harvesting to get a good and healthy plant population as usual.

Traits measurement:

The studied traits on a ten individual guarded plant basis were taken of the six non-segregation and segregation generations in the two crosses studied. The data were recorded for boll weight in grams (B.W., g), seed cotton yield/plant in grams (S.C.Y./P, g), lint cotton yield/plant in grams (L.C.Y./P, g), lint percentage (L.%), number of bolls/plant (No. of B./P) and seed index (S.I., g) traits. The collected data were statistically analyzed.

Statistical Approaches:

The data of studied traits were subjected to a one-way ANOVA test following the method of Steel and Torrie (1997) to determine the significant differences and the coefficient of variation (CV%) among six generations using XLSTAT software as described by Addinsoft (2021). According to Gomes (2009), the estimates of CV% were classified as very high (CV $\geq$ 21.0%), high (15.0% $\leq$ CV $\leq$ 21.0%), moderate (10% $<$ CV $\leq$ 20%) and low (CV $<$ 10%). The significance test was done with the least significant difference test (L.S.D) at 0.05 and 0.01 levels of probability according to Steel and Torrie (1997). Quantity and plot Pearson's correlation coefficient as well as multivariate analysis (principal component and cluster analysis) were performed for a better understanding of the relationship among studied traits across six generations using the computer software program PAST version 4.03 (Hammer et al. 2020).

Results

ANOVA and Mean Performances:

The results of one-way ANOVA showed statistically significant differences ($P < 0.01$) among non-segregation (P₁, P₂ and F₁) and segregation (F₂, BC₁ and BC₂) generations for yield and yield components traits studied in the two crosses Giza 92 x Pima S₆ and Giza 93 x C.B. 58 (Fig. 1). The values of generations/error variances ratio were greater than unity (G/E ratio > 1) for all traits studied in the two crosses. According to classification by Gomes (2009), the values of CV% recorded for studied traits across the six generations in the two crosses were low (CV $< 10\%$). The lint percentage and seed index traits recorded the lowest values of CV % in the two crosses (Fig. 1).

Highly significant differences among the two parents and their F₁, F₂ and backcross generations means were observed for all studied traits in the two crosses (Fig. 1). The variety Giza 93 exhibited the best performance for all traits studied compared with the other three varieties,

except for the seed index trait (variety C.B. 58). The F_1 mean performance was determined to have a high value compared with the assessed respective parents and three segregation generations for all the studied traits in the two crosses. When comparing within the segregation generations, the BC_1 generation showed superiority in most studied traits in the two crosses.

Correlation analysis:

The Pearson's correlation coefficient of seed cotton yield and its components are separated for an evident understanding of relationships between these traits in each generation as presented in Table 1. Out of 180 Pearson correlation between studied traits, 56 (ranged from 6 in P_1 to 13 in BC_1) and 57 (ranged from 7 in P_1 to 13 in F_2) positive correlation coefficients were seen within the six generations in the two crosses Giza 92 x Pima S_6 and Giza 93 x C.B. 58, respectively. Seed cotton yield/plant, lint cotton yield/plant and No. of bolls/plant traits had strong positive and highly significant correlations ($p < 0.01$) across all six generations in the two crosses. Boll weight was significantly and positively correlated ($p < 0.01$) with seed and lint cotton yields/plant in F_2 generation during the two crosses studied, and with lint percentage trait in BC_1 generation under the cross Giza 93 x C.B. 58. The seed index shows a positive and significant correlation with seed cotton yield/plant, lint cotton yield/plant and lint percentage traits in BC_1 generation across the cross Giza 92 x Pima S_6 . Lint percentage displayed a positive association with seed index in P_2 and BC_1 generations ($p < 0.05$), as well as with lint cotton yield/plant in BC_1 , BC_2 ($p < 0.05$), P_2 , F_1 and F_2 ($p < 0.01$) generations during the cross Giza 93 x C.B.58.

Table 1
Pearson correlation coefficient among studied traits in the four cotton cultivars and their F_1 , F_2 and backcross generations.

Generations		Giza 92 x Pima S_6						Giza 93 x C. B. 58					
Traits		P_1	P_2	F_1	F_2	BC_1	BC_2	P_1	P_2	F_1	F_2	BC_1	BC_2
B.W.	S.C.Y./P	0.22	0.01	-0.13	0.28**	0.09	-0.02	0.11	-0.11	-0.31	0.38**	0.02	-0.04
	L.C.Y./P	0.22	0.01	-0.23	0.28**	0.12	0.02	0.07	-0.17	-0.29	0.38**	0.06	-0.11
	L.%	-0.13	-0.07	-0.12	-0.04	0.20	0.17	-0.23	-0.15	-0.10	0.11	0.35**	-0.28*
	No.B./P	-0.01	-0.11	-0.74**	0.01	-0.33**	-0.68**	-0.52**	-0.48**	-0.74**	0.04	-0.35**	-0.43**
	S.I.	0.27	-0.23	-0.06	0.07	0.11	0.09	-0.02	-0.20	0.14	-0.05	-0.07	-0.01
S.C.Y./P	L.C.Y./P	0.99**	1.00**	0.84**	0.98**	0.99**	0.97**	0.98**	0.81**	0.91**	0.99**	0.99**	0.98**
	L.%	-0.46*	0.23	-0.56**	-0.02	0.01	-0.25*	-0.10	0.04	0.28	0.11	0.11	0.06
	No.B./P	0.97**	0.99**	0.76**	0.95**	0.91**	0.73**	0.79**	0.93**	0.86**	0.93**	0.93**	0.92**
	S.I.	-0.19	0.21	0.02	0.04	0.23*	-0.46**	-0.04	-0.19	0.33	-0.02	-0.10	0.04
L.C.Y./P	L.%	-0.36	0.31	-0.03	0.14	0.18	-0.01	0.11	0.61**	0.64**	0.26**	0.27*	0.25*
	No.B./P	0.97**	0.99**	0.73**	0.93**	0.89**	0.69**	0.80**	0.78**	0.79**	0.92**	0.90**	0.93**
	S.I.	-0.20	0.22	0.14	0.02	0.27*	-0.44**	-0.12	0.09	0.32	0.01	-0.04	-0.05
L.%	No.B./P	-0.44*	0.23	-0.30	-0.04	-0.05	-0.27*	0.06	0.09	0.25	0.10	0.01	0.19
	S.I.	-0.01	0.26	0.17	-0.08	0.25*	0.17	-0.39**	0.42*	0.15	0.10	0.25*	-0.45**
No.B./P	S.I.	-0.25	0.22	0.06	0.01	0.17	-0.42**	-0.02	-0.09	0.16	0.01	-0.06	0.01
B.W.: Boll weight; S.C.Y./P: Seed cotton yield/plant; L.C.Y./P: Lint cotton yield/plant; L.%: Lint percentage; No.B./P: Number of bolls/plant; S.I.: Seed index; P_1 and P_2 : First and second parents; F_1 and F_2 : First and second generations; BC_1 and BC_2 : First and second back crosses, respectively. According to the Pearson correlation coefficient, the asterisks (*) and (**) denote a significant difference at 0.05 and 0.01 level of significance, respectively.													

The intergeneration correlations for seed cotton yield and its components traits are shown in Table 2. A number of positive correlations were observed across both of them generations comparisons for studied traits in the two crosses Giza 92 x Pima S_6 (54) and Giza 93 x C.B. 58 (36). The P_1 generation positively correlated with P_2 generation for lint percentage and seed index traits ($p < 0.05$ and 0.01) in the

cross Giza 93 x C.B. 58 and Giza 92 x Pima S₆, respectively, with F₂ generation for lint cotton yield/plant ($p < 0.05$) in the cross Giza 92 x Pima S₆ and with BC₂ generation for No. of bolls/plant ($p < 0.01$) in the cross Giza 93 x C.B. 58. The P₂ generation had a positive and significant correlation with F₁ generation for No. of bolls/plant ($p < 0.01$), with BC₁ ($p < 0.01$) and BC₂ ($p < 0.05$) generations for lint percentage and with BC₂ generation for boll weight ($p < 0.05$) in the cross Giza 92 x Pima S₆; as well as with F₂ and BC₁ generations for boll weight ($p < 0.05$) and seed index ($p < 0.01$) traits in the cross Giza 93 x C.B. 58, respectively. The F₁ generation had a significant positive association with F₂ generation for lint percentage ($p < 0.05$) and with BC₁ generation for lint cotton yield/plant ($p < 0.01$), No. of bolls/plant and seed index ($p < 0.05$) traits across the cross Giza 92 x Pima S₆; while with F₂ generation for seed index ($p < 0.01$), with BC₁ for boll weight ($p < 0.05$) and with BC₂ for seed cotton yield/plant ($p < 0.01$) under the cross Giza 93 x C.B. 58. The BC₁ generation showed a significant and positive correlation with BC₂ generation for lint percentage ($p < 0.01$) in the two crosses and for lint cotton yield/plant ($p < 0.05$) traits in the cross Giza 92 x Pima S₆.

Table 2

Pearson correlation coefficient between the non-segregation and segregation generations for studied traits in the two hybrids of cotton.

Crosses	Yield and yield components traits											
	B.W.		S.C.Y./P		L.C.Y./P		L.%		No.B./P		S.I.	
	I	II	I	II	I	II	I	II	I	II	I	II
r(P ₁ , P ₂)	0.06	0.16	0.23	-0.52**	0.25	-0.17	-0.25	0.37*	0.30	0.04	0.46**	-0.67**
r(P ₁ , F ₁)	-0.50**	-0.54**	-0.12	-0.11	-0.10	-0.03	-0.11	-0.18	0.32	-0.24	0.33	-0.02
r(P ₁ , F ₂)	0.25	0.15	0.30	0.19	0.36*	0.02	-0.45*	-0.45*	0.24	0.09	-0.62**	-0.03
r(P ₁ , BC ₁)	0.34	-0.06	-0.28	0.04	-0.10	-0.02	0.02	0.21	-0.30	-0.54**	0.14	-0.24
r(P ₁ , BC ₂)	-0.06	-0.05	0.11	0.11	0.24	0.27	0.13	0.17	0.01	0.70**	-0.04	-0.17
r(P ₂ , F ₁)	-0.49**	-0.20	0.33	-0.58**	0.08	-0.73**	-0.16	-0.65**	0.47**	-0.53**	0.08	-0.30
r(P ₂ , F ₂)	-0.12	0.44*	-0.53**	-0.36*	-0.48**	-0.36*	-0.30	-0.55**	-0.59**	-0.34	-0.25	-0.10
r(P ₂ , BC ₁)	-0.37*	0.11	-0.11	0.14	-0.06	0.01	0.59**	0.19	0.06	-0.02	-0.01	0.54**
r(P ₂ , BC ₂)	0.40*	-0.29	0.27	-0.29	0.30	-0.06	0.42*	0.28	0.32	-0.24	-0.13	-0.25
r(F ₁ , F ₂)	0.35	0.13	-0.33	-0.09	-0.28	-0.05	0.45*	0.24	-0.19	-0.21	0.15	0.53**
r(F ₁ , BC ₁)	0.06	0.38*	0.34	-0.51**	0.49**	-0.34	0.01	-0.55**	0.41*	-0.17	0.37*	0.32
r(F ₁ , BC ₂)	0.09	0.22	0.04	0.52**	0.06	0.15	0.09	-0.83**	0.28	0.22	0.26	-0.43*
r(F ₂ , BC ₁)	0.23	0.26	-0.14	-0.06	-0.18	-0.05	-0.30	-0.04	-0.08	-0.09	-0.36*	0.10
r(F ₂ , BC ₂)	0.17	-0.08	-0.29	0.08	-0.29	0.12	0.05	0.12	-0.38*	0.17	0.32	-0.55**
r(BC ₁ , BC ₂)	0.12	-0.14	0.34	-0.90	0.42*	-0.83**	0.73**	0.68**	0.15	-0.70**	0.25	-0.51**
B.W.: Boll weight; S.C.Y./P: Seed cotton yield/plant; L.C.Y./P: Lint cotton yield/plant; L.%: Lint percentage; No.B./P: Number of bolls/plant; S.I.: Seed index; I: G. 92 x Pima S ₆ ; II: Giza 93 x C.B. 58. P ₁ and P ₂ : First and second parents; F ₁ and F ₂ : First and second generations; BC ₁ and BC ₂ : First and second back crosses, respectively. According to the Pearson correlation coefficient, the asterisks (*) and **) denote a significant difference at 0.05 and 0.01 level of significance, respectively.												

The results in Fig. 2 cleared that 13 and 10 Pearson's correlation coefficients exhibited significantly ($p < 0.05$ or 0.01) among the yield and its components traits across six generations in the two crosses Giza 92 x Pima S₆ and Giza 93 x C.B. 58, respectively. Positive and significant correlations ($p < 0.05$ or 0.01) between all possible pairs for investigated traits were found during the cross Giza 92 x Pima S₆, except boll weight had positive and insignificant correlation with the number of bolls/plant and seed index traits. In the cross Giza 93 x C.B. 58, except the correlations that comprise the lint percentage trait, all correlations between yield and its components traits were positive and significant ($p < 0.05$ or 0.01). During the two crosses, there was a high, positive and significant correlation ($p < 0.01$) among all

combinations of the six generations across studied traits (Fig. 1B). Generally, the seed cotton yield/plant positively and significantly correlated with all studied traits in the two crosses, except the lint percentage in the cross Giza 93 x C.B. 58.

Cluster analysis:

Based on the six generations data in the two crosses (Fig. 3A), the cluster analysis separated the studied traits into three main clusters at 80% similarity. The first cluster contains No. of bolls/plant, seed cotton yield/plant and lint cotton yield/plant traits with 96% similarity. While the second (seed index) and third (lint percentage and boll weight traits) clusters comprised of the rest studied traits at 80% similarity. Figure (3B) comprising of six generations could be further divided into three clusters based on the data of yield and yield component traits in the two crosses. The first cluster consists of one generation (F_1) at 96% similarity. The second cluster consisted of three generations (P_1 , BC_1 and BC_2) with 99% similarity. While, the third cluster comprised of two generations (P_2 and F_2) at above 99% similarity. The greatest similarity occurred between P_1 and BC_1 generations followed by between P_2 and F_2 generations, while the lowest similarity occurred between F_1 and F_2 generations.

Principal component analysis:

Principal component analysis (PCA) has been used to estimate the similarities and dissimilarities relationships between the studied traits across six generations variables in the two crosses of cotton, which are graphically displayed in a biplot of PCA1 and PCA2 (Fig. 4). Out of five PCAs, the first main PCA1 extracted had eigenvalues larger than one (Eigenvalue > 1) with a value of 4.77 (Fig. 4). While the rest four PCAs had eigenvalues less than one (Eigenvalue < 1). The first two PCAs contributed 91.84% of the total variation existing among six generations regarding studied traits in the two crosses. The contributions of PCA1 to the total variance were higher than that of the other components, with PCA1 describing only about 79.47% of the measured data total variability.

The five PCAs for six generations based on the studied traits during the two crosses are shown in Table 3. The PCA1 was dominated by F_1 generation in the two crosses and by P_1 , BC_1 and BC_2 generations in the cross Giza 93 x C.B. 58. On the other hand, the segregation generations influenced the PCA2 in the two crosses, which explained 12.38% of the total variability. The PCA3 was controlled by P_2 and F_1 generations across the two crosses as well as by BC_1 and BC_2 generations in the cross Giza 92 x Pima S_6 . As for, the PCA4 was affected by non-segregation generations in the cross Giza 92 x Pima S_6 as well as by P_1 and F_2 generations in the cross Giza 93 x C.B. 58.

Table 3
Results of principal component analysis (PCAs) for six generations based on the studied traits during the two crosses.

PCAs	Giza 92 x Pima S_6					Giza 93 x C.B. 58				
Generations	PCA1	PCA2	PCA3	PCA4	PCA5	PCA1	PCA2	PCA3	PCA4	PCA5
P_1	-1.89	-1.82	-0.55	0.32	0.00	0.61	0.10	-1.05	0.20	0.00
P_2	-1.40	-0.76	0.49	0.35	0.01	-2.28	-0.38	0.70	-0.53	-0.02
F_1	4.18	0.05	0.49	0.28	-0.03	4.01	-0.89	0.27	-0.23	0.00
F_2	-0.96	0.15	-0.16	-0.46	-0.01	-0.64	1.46	-0.02	0.59	-0.02
BC_1	-0.45	0.66	0.74	-0.02	0.06	0.68	0.65	-0.75	-0.63	0.00
BC_2	-2.36	0.52	0.17	0.06	-0.03	0.49	0.24	-0.32	0.06	0.03
P1 and P2: First and second parents; F1 and F2: First and second generations; BC1 and BC2: First and second back crosses, respectively.										

Based on the six generations variables in the two crosses (Table 4), high positive correlations of factors loadings were recorded for seed and lint cotton yields/plant traits with PCA1 (0.45), for lint percentage with PCA2 and PCA4 (0.63), for No. of bolls/plant with PCA5 (0.68), and for seed index with PCA3 (0.63). The six generations in PCA1 presented positive loadings values for all studied traits. Also, all the variables of the yield and its components contributed to the PCA1. The PCA2 has identified boll weight and lint percentage traits possessing positive loading factors and contribution to the variables. While, the boll weight and seed index traits have positive loadings on PCA3. The PCA4 was dominated by lint percentage and seed index traits with positive factor loadings, while all studied traits in PCA5 obtained positive loading factors, except seed and lint cotton yields/plant traits.

Table 4
Correlation coefficient between studied traits and five main PCAs generated using six generations in the two crosses.

PCAs Traits	PCA1	PCA2	PCA3	PCA4	PCA5
B.W.	0.37	0.52	0.53	-0.50	0.23
S.C.Y./P	0.45	-0.16	-0.17	-0.23	-0.27
L.C.Y./P	0.45	-0.09	-0.17	-0.15	-0.63
L. %	0.36	0.63	-0.28	0.63	0.05
No.B./P	0.42	-0.33	-0.43	-0.14	0.68
S.I.	0.38	-0.45	0.63	0.51	0.05
B.W.: Boll weight; S.C.Y./P: Seed cotton yield/plant; L.C.Y./P: Lint cotton yied/plant; L. %: Lint percentage; No.B./P: Number of bolls/plant; S.I.: Seed index.					

The relationships among yield and its components traits across six generations in the two crosses studied are graphically displayed by the biplot of the first two PCAs (Fig. 4). The two crosses performances during the six generations displayed a positive correlation among yield and its components variables, but, they differed in their degree and consistency in quantity. Positive and significant correlations ($p < 0.05$ or 0.01) between all possible pairs for investigated traits were found across six generations in the two crosses, except between lint percentage and seed index traits had positive and insignificant correlation.

When comparing the six generations, the PCA1 and PCA2 showed that the yield and yield components variables were distributed in different regions and formed different groups (Fig. 4). The F_1 generation in the two crosses produced the highest yield and yield components. The F_1 generations in the two crosses as well as P_1 , BC_1 and BC_2 generations in the cross Giza 93 x C.B. 58 occupied Quadrants first (I) and fourth (IV) of the diagram, and which is strongly correlated with PCA1. On the other hand, the rest generations in the two crosses were positively associated with PCA2 and occurred in the second (II) and third (III) quarters. Lint percentage and boll weight traits were located in quarter I with P_1 , BC_1 and BC_2 in the cross Giza 93 x Pima S6C.B. 58 and with F_1 generation Giza 92 x Pima S₆. As for the other traits were located near the F_1 generation in quarter IV across the cross Giza 93 x C.B. 58. The F_2 generation in the two crosses as well as BC_1 and BC_2 generations in the cross Giza 92 x Pima S₆ were located near the most studied traits. The scree plot of the PCA for six generations on yield and yield components during the two crosses displayed that the first two eigenvalues correspond to the whole percentage of the variance in the dataset (Fig. 5).

Discussion

ANOVA and Mean Performances:

The mean squares due to six generations exhibited high significance for studied traits in the two studied crosses. The genetic variances were higher than the environmental variances among the six generations for those traits across the two crosses. Similar type of results in the six populations was reported by Hussain et al. (2008) in cotton and Omoigui et al. (2019) in cowpea. Significant differences observed among the F_2 , F_3 and F_4 generations for most studied traits by Srour and El-Hashash (2012), demonstrating the existence of inherent variability among the two crosses of Egyptian cotton. The analysis of variance depicted highly significant genetic variability among cotton germplasm for yield and its components (Shah et al. 2018; Yehia and El-Hashash 2019). Low values of CV% ($< 10\%$) were registered for studied traits across the six generations in the two crosses, due to the environmental influence was low for those traits, thus this trial are would be considered as of high precision. The magnitude of CV% indicated that the non-segregation and segregation generations had exploitable genetic variability during the selection of traits examined in cotton. These results are different with those identified by Raza et al. (2016) and Li et al. (2020) in cotton, where the CV% values were more than 10%.

The results of mean performances indicating the presence of genetic variability for these traits in those studied materials. The F_1 generation showed superiority in studied traits at the two crosses. These results indicated that the relation between non-segregation and segregation generations revealed that there was different behavior for studied traits in the study materials. Thus, it is possible to benefit from the selection in the segregation generations in future breeding programs of improving these traits in Egyptian cotton. El-Hashash (2017) has been reported that the mean values of non-segregation generations were better than the mean values of segregation generations in some cases and conversely in another case in the two crosses of Egyptian cotton.

Correlation analysis:

Positive and significant correlation coefficients were observed among and within the six generations for studied traits in the two crosses. The correlation coefficients among different pairs of plant traits for the two crosses across generations indicated that cotton yield can be improved and increased by increasing most yield components traits. The results could be a possibility of plants with desirable attributes of cotton yield in the next segregating generations. In this connection, Srour and El-Hashash (2012) reported similar results. Positive intergeneration correlations for studied traits in the two crosses indicated that selection for increased value of one trait will result increase in value of other. The finding of Barman and Borah (2012) in mutant rice strain and Srour and El-Hashash (2012) in cotton revealed that the correlation coefficient among F_2 , F_3 and F_4 generations were significant or highly significant for studied traits. A significant correlation ($p < 0.05$ or 0.01) between cotton yield and yield components traits across the single and double-cross hybrids were observed by El-Hashash (2013). The results showed a positive and significant correlation among all combinations of the six generations across studied traits. Generally, the seed cotton yield/plant positively and significantly correlated with all studied traits in the two crosses, except the lint percentage in the cross Giza 93 x C.B. 58. This implies that cotton yield is effective for selection in a later generation and the two crosses may be used in improving cotton yield in Egyptian cotton. Like present findings, the significant positive correlation for yield and its components traits were reported by Khokhar et al. (2017); Jarwar et al. (2019); Kumar (2020); Rehman et al. (2020) and Sarwar et al. (2021). Srour and El-Hashash (2012) and Khokhar et al. (2017) stated that more than one trait can be used as a selection standard in the next segregating generations of cotton.

Cluster analysis:

For comparison and determination of the similarities and differences between traits and generations studied on the basis of the data in the two crosses, UPGMA hierarchical clustering with correlation similarity index was used. By the cluster analysis, both studied traits and six generations were divided into three clusters. The maximum similarity occurred between seed and lint cotton yield/plant traits as well as between P_1 and BC_1 generations, while the minimum similarity occurred between seed index and lint percentage traits as well as between F_1 and F_2 generations. Based on 100 cotton genotypes of *Gossypium hirsutum* L., the mean performances of yield and yield components traits were grouped into six different clusters (Akter et al. 2019). The tree diagram had exhibited the highest correlation between the traits or generations inside each cluster. While the lowest correlation of traits or generations were found among the clusters. In amphidiploid cotton hybrid plants, Muminov et al. (2020) mentioned that the cluster analysis grouped the parents and F_1 - F_6 generations into four different clusters. According to the report of Abdel-Monaem et al. (2020), the parents and F_1 generation plants had divided into four and seven major clusters, respectively, and the clusters of F_1 generation were wide divergent from parental populations in Egyptian cotton.

Principal component analysis:

Principal component analysis (PCA) is a multivariate statistical technique. The PCA simplifies the complex data by transforming the number of correlated variables into a smaller number of variables in the data collection in a breeding and selection process called principal components (PCAs). The PCA1 had eigenvalues higher than one. Therefore, the PCA1 was kept for the final analysis, in which, the PCA1 explains variance (79.47%) more than an individual attribute (Sharma, 1996) and it expresses more variability and support to select the trait with a positive loading factor. The results indicating that the first two PCAs may be used to summarize the original variables in any further analysis of the data, as well as to explain 91.84% of the total variation and the grouping of the PCAs. Based on all measured data in the two crosses, the first two PCAs had mainly distinguished the generations in different groups. Therefore, the first two PCAs were employed to draw a biplot (Fig. 4). According to Abasianyanga et al. (2017); Shah et al. (2018); Akter et al. (2019); Abdel-Monaem et al. (2020) and Sarwar et al. (2021), the eigenvalues had higher than one for the first six, five, four, eight and four PCAs, and which contributed for 70.39%, 78.66%, 53.23%, 85.28% and 79.90% of the accumulative variation of the original variables under examination in cotton, respectively. The PCA1 and PCA2 explained 99.77% and 69.30% of the total variation between six generations in sugar beet (Bayomi et al. 2019) and between 230 F_2 populations in cotton (Nandhini et al. 2018) based on all studied traits, respectively.

From the results of PCAs for six generations based on the studied traits during the two crosses, Selection during generations that have high PCA1 and PCA2 for studied traits are suitable and effective in the two crosses. In practice, the choice of F_2 vs. backcross based populations in "second cycle" breeding is complicated by the fact that the breeder regards not only a single trait but several characters simultaneously (Meichinger 1987). The six generations in PCA1 presented positive loadings values for all studied traits. Also, all the variables of the yield and its components contributed to the PCA1. It is evident that these components can be interpreted as a response related to the yield and its components traits and which possessing positive and negative contributions to the variables. Similarly, Abasianyanga et al. (2017); Shah et al. (2018); Rathinavel (2019) and Jarwar et al. (2019) reported that all or some yield and yield components traits showed a positive correlation of factor loadings on the five PCAs, especially PCA1 that divulged great contribution towards total variability. El-Hashash and EL-Agoury (2019) stated that the acute angle (below 90 degrees) between the two variables

indicates the significance of these variables. The two crosses across six generations displayed a positive correlation among seed cotton yield and its components traits, but, they differed in their degree and consistency in quantity. These results indicate that selection based on these traits would result in an increasing cotton yield in both crosses. Hence, emphasis must be placed on these materials in a breeding program to improve the Egypt cotton. The biplot showed the degree of correlation amongst cotton traits (Sarwar et al. 2021). When comparing the six generations, the PCA1 and PCA2 showed that the yield and yield components variables were distributed in different regions and formed different groups, and therefore these results indicate that there are differences between these variables across the six generations in the two crosses. The respective variable distances from the first two PCAs demonstrate the contribution of different variables at the total variability (Sarwar et al. 2021).

The biplot diagram depicted the contribution of yield and its components traits in creating a variability of six generations. The biplot analysis of the relationship between the six generations revealed that the most appropriate generations for selecting yield traits were F_1 in the two crosses and BC_1 and BC_2 in the cross Giza 93 x C.B. 58. Meichinger (1987) declared that F_2 , BC_1 and BC_2 offer equal alternatives with respect to time, work, inbreeding level, and the amount of genetic variation released within lines in subsequent selfing generations if linkage and epistasis are of small importance. Therefore, the choice to separate a population can be based on the characteristics of the first segregating generations. The Fig. 1 shows that there is a divergence among the six populations, thus these diversity can be used to improve the yield and its components in cotton. During the biplot diagram of the first two PCAs, the plants close to the ideal type would be selected (Abdel-Monaem et al. 2020). The extent of variation in each trait between genotypes showed greater divergence (Rathinavel 2019) and most traits contributed to more variance (Nandhini et al. 2018), thus these traits could be considered essential for selection in large populations as well as which may be useful for an effective program to cotton breeding. In the scree plot, the first two eigenvalues correspond to the whole percentage of the variance in the dataset. These results were consistent with Abdel-Monaem et al. (2020) among parental cotton genotypes and their F_1 's cross combinations in Egyptian cotton. Our results were harmonic with El-Hashash (2016) who reported that there is a break in the plot that separates the meaningful components from the trivial components. Most researchers would agree that the first two PCAs are probably meaningful. This finding is in agreement with those reported by Abasianyanga et al. (2017); Jarwar et al. (2019) and Sarwar et al. (2021).

Conclusions

Significant divergences among the six generations for all studied traits in the two crosses were observed by ANOVA. The F_1 performance was higher than the other generations for all the studied traits in the two crosses. The results of Pearson's correlation coefficient and multivariate analysis from our study could be useful and use in breeding programs for cotton yield improvement. Therefore, we recommend considering backcrossing may be done for 2–5 cycles ($BC_2 - BC_5$) at C.B. 58 parent for improving Egyptian cotton yield in future.

Abbreviations

P_1 and P_2 : First and second parents; F_1 and F_2 : First and second generations; BC_1 and BC_2 : First and second back crosses, respectively; G/E ratio: mean squares of generations/ mean squares of error; C.V%: Coefficient of variation; B.W.: Boll weight; S.C.Y./P: Seed cotton yield/plant; L.C.Y./P: Lint cotton yied/plant; L.%: Lint percentage; No.B./P: Number of bolls/plant; S.I.: Seed index; I: Giza. 92 x Pima S6; II: Giza 93 x C. B. 58.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

All data and materials are available.

Competing interests

No competing of interest.

Funding

The study was funded by the two authors.

Authors' contributions

WY and EE suggested the research idea, designed the experiments, collected field data and equally contributed to interpreting results, writing and revising the manuscript, as well as approved the final manuscript to be published.

Acknowledgements

The authors are thankful for workers helped in field labor.

Authors details

1. Cotton Breeding Department Cotton Research Institute, Agriculture Research Center, Egypt
2. Department of Agronomy, Faculty of Agriculture, Al-Azhar University, Cairo, Egypt.

References

1. Abasianyanga I, Balu PA, Ramakrishnan P. Association and principal component analysis of yield and its components in cultivated cotton. *Elec J of Plant Breed*. 2017;8(3):857–64. DOI:10.5958/0975-928X.2017.00140.5.
2. Abdel-Monaem MA, Abido WAE, Hadházy A, Ghoneima MH, EL-Mansy YM, EL-Shazly MW. Genetic divergence among Egyptian cotton genotypes under water deficit conditions. *Acta Ecol Sin*. 2020. <https://doi.org/10.1016/j.chnaes.2020.11.007>.
3. Addinsoft. (2021) XLSTAT statistical and data analysis solution. New York, USA. <https://www.xlstat.com>.
4. Akter T, Islam AKMA, Rasul MG, Kundu S, Khalequzzaman, Ahmed JU. Evaluation of genetic diversity in short duration cotton (*Gossypium hirsutum* L.). *J Cotton Res*. 2019;2:1. <https://doi.org/10.1186/s42397-018-0018-6>.
5. Barman D, Borah SP. Effect of Selection Response on F3 and F4 Generation for Yield and Yield Component Characters in Mutant Rice Strain (*Oryza sativa* L.). *APCBEE Procedia*. 2012;4:183–7. <https://doi.org/10.1016/j.apcbee.2012.11.031>.
6. Bayomi KEM, El-Hashash EF, Moustafa ESA. Comparison of Genetic Parameters in Non-segregating and Segregating Populations of Sugar Beet in Egypt. *Asian Jo of Res in Crop Sci*. 2019;3(4):1–12. <https://doi.org/10.9734/ajrcs/2019/v3i430055>.
7. Behere GT, Tay WT, Russell DA, Kranthi KR, Batterham P. Population Genetic Structure of the Cotton Bollworm *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) in India as Inferred from EPIC-PCR DNA Markers. *PLoS ONE*. 2013;8(1):e53448. <https://doi.org/10.1371/journal.pone.0053448>.
8. Eldessouky SEI, El-Fesheikawy ABA, Baker KMA. Genetic variability and association between oil and economic traits for some new Egyptian cotton genotypes. *Bull Natl Res Cent*. 2021;45:43. <https://doi.org/10.1186/s42269-020-00438-1>.
9. El-Hashash EF. American-Eurasian. Heterosis and Gene Action among Single and Double-Cross Hybrids Performances in Cotton. *J Agric Environ Sci*. 2013;13(4):505–16. DOI:10.5829/idosi.ajeaes.2013.13.04.1955.
10. El-Hashash EF. Genetic Diversity of Soybean Yield Based on Cluster and Principal Component Analyses. *J of Adv in Biol Biot*. 2016;10(3):1–9. <https://doi.org/10.9734/JABB/2016/29127>.
11. El-Hashash EF. Comparison of various methods for estimating of heritability and number of genes affecting the quantitative traits in cotton. *Egypt J Plant Breed*. 2017;21(5):612–28.
12. El-Hashash EF, EL-Agoury RYA. Comparison of grain yield-based drought tolerance indices under normal and stress conditions of rice in Egypt. *Scholars J of Agric Vet Sci*. 2019;6(1):41–54. DOI:10.21276/sjavs.2019.6.1.6.
13. FAOSTAT. (2021) Food and agriculture organization of the United Nations. <http://faostat.fao.org>. FAO Statistics Division. Crops Primary data last updated April 20, 2021.
14. Gomes FP. (2009) Curso de estatística experimental. 15.ed. Piracicaba: Esalq, 2009. 477p.
15. Hammer Ø, Harper DAT, Ryan PD. Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*. 2020;4(1):9 pp.
16. Hussain M, Azhar FM, Khan AA. Genetic basis of variation in leaf area, petiole length and seed cotton yield in some cotton (*Gossypium hirsutum*) genotypes. *Int J Agric Biol*. 2008;10:705–8.

17. Jarwar AH, Wang X, Iqbal M, Sarfraz Z, Wang L, Qifeng MA, Shuli F. Genetic divergence on the basis of principal component, correlation and cluster analysis of yield and quality traits in cotton cultivars. *Pakistan J of Bot.* 2019;51(3):1143–8. DOI:10.30848/PJB2019-3(38).
18. Kerby TA, Cassman KG, Keeley M. Genotypes and plant densities for narrow-row cotton systems. I. Height, nodes, earliness, and location of yield. *Crop Sci.* 1990;30(3):644–9. <https://doi.org/10.2135/cropsci1990.0011183X003000030034x>.
19. Khokhar ES, Shakeel A, Maqbool MA, Anwar MW, Tanveer Z, Irfan MF. Genetic study of cotton (*Gossypium hirsutum* L.) genotypes for different agronomic, Yield and quality traits. *Pakistan J of Agric Res.* 2017;30(4):363–72. <http://dx.doi.org/10.17582/journal.pjar/2017/30.4.363.372>.
20. Kloth RH. Analysis of commonality for traits of cotton fiber. *J Cotton Sci.* 1998;2:17–22.
21. Kumar KB. Inter Relationship between Seed Cotton Yield and Yield Contributing Characters in American Cotton (*Gossypium hirsutum* L.). *Int J Curr Microbiol App Sci.* 2020;9(6):2276–9. <https://doi.org/10.20546/ijcmas.2020.906.278>.
22. Latif A, Bilal M, Hussain SB, Ahmad F. Estimation of genetic divergence, association, direct and indirect effects of yield with other attributes in cotton (*Gossypium hitsutum* L.) using biplot correlation in and path coefficient analysis. *Tropical Plant Res.* 2105;2(2):120–6.
23. Li B, Tian Q, Wang X, Han B, Liu L, Kong X, Si A, Wang J, Lin Z, Zhang X, Yu Y, Yang X. Phenotypic plasticity and genetic variation of cotton yield and its related traits under water-limited conditions. *The Crop J.* 2020;8(6):966–76. <https://doi.org/10.1016/j.cj.2020.02.003>.
24. Meichinger AE. Expectation of means and variances of testcrosses produced from F2 and backcross individuals and their selfed progenies. *Heredity.* 1987;59:105–15. DOI:10.1038/hdy.1987.101.
25. Muminov KH, Ernazarova Z, Amanov B. Cluster analysis of valuable economic traits in amphidiploid cotton hybrid plants. *EurAsian J of BioSci.* 2020;14:4973–81.
26. Nandhini K, Amala Balu P, Isong A. Genetic Analysis and Inheritance Studies in F2 Population of Upland Cotton (*G. hirsutum* L.). *Int J Pure App Biosci.* 2018;6(2):1499–505. doi:<http://dx.doi.org/10.18782/2320-7051.6472>.
27. Omoigui OL, Arrey MO, Kamara AY, Danmaigona CC, Ekeruo G, Timko MP. Inheritance of resistance to Cercospora leaf spot disease of cowpea [*Vigna unguiculata* (L.) Walp]. *Euphytica.* 2019;215:101. <https://doi.org/10.1007/s10681-019-2420-7>.
28. Pathak RS, Singh RB. Genetic analysis of the duplicate loci, cluster and short branch in *Gossypium hirsutum* L. *Theoret Appl Genet.* 1975;46:281–7. <https://doi.org/10.1007/BF00281150>.
29. Rathinavel K. Agro-morphological Characterization and Genetic Diversity Analysis of Cotton Germplasm (*Gossypium hirsutum* L.). *Int J Curr Microbiol App Sci.* 2019;8(2):2039–57. <https://doi.org/10.20546/ijcmas.2019.802.237>.
30. Rehman A, Farooq M. (2020) Morphology, Physiology and Ecology of Cotton Chap. 2. Cotton Production, First Edition. Edited by Khawar Jabran and Bhagirath Singh Chauhan. John Wiley & Sons Ltd.
31. Rehman A, Mustafa N, Du X, Azhar MT. Heritability and correlation analysis of morphological and yield traits in genetically modified cotton. *J Cotton Res.* 2020;3(23):1–9. <https://doi.org/10.1186/s42397-020-00067-z>.
32. Raza H, Khan NU, Khan SA, Gul S, Latif A, Hussain I, Khan J, Raza S, Baloch M. Genetic variability and correlation studies in F₄ populations of upland cotton. *The J of Anim Plant Sci.* 2016;26(4):1048–55.
33. Sarwar G, Nazir A, Rizwan M, Shahzadi E, Mahmood A. Genetic diversity among cotton genotypes for earliness, yield and fiber quality traits using correlation, principal component and cluster analyses. *Sarhad J of Agric.* 2021;37(1):307–14. <http://dx.doi.org/10.17582/journal.sja/2021/37.1.307.314>.
34. Shah SAI, Khan SJ, Ullah K, Sayal OU. Genetic diversity in cotton germplasm using multivariate analysis. *Sarhad J of Agric.* 2018;34(1):130–5. <http://dx.doi.org/10.17582/journal.sja/2018/34.1.130.135>.
35. Sharma S. Applied Multivariate Techniques. New York: Wiley; 1996.
36. Srour MSM, El-Hashash EF. Phenotypic and Genotypic Correlation Coefficients for Some Yield and Fiber Quality Traits of Segregating Populations (F2, F3 and F4) in Egyptian Cotton Crosses. *JKAU: Met Env Arid Land Agric Sci.* 2012;23(1):19–35. DOI: 10.4197/Met. 23 – 1.2.
37. Steel RGD, Torrie JH. Dickey DA principles and procedures of statistics: a biometrical approach. 3rd ed. New York: McGraw Hill; 1997.
38. USDA. (2021) United States Department of Agriculture Foreign Agricultural Service, Office of Global Analysis. Cotton and Products Annual, Egypt Cotton Production Rebounds. March 31, 2021, Report Number: EG2021-0005. 1–17.
39. Yehia WMB, El-Hashash EF. Combining ability effects and heterosis estimates through line x tester analysis for yield, yield components and fiber traits in Egyptian cotton. *Elixir Agric.* 2019;131:53238–46.

Figures

Figure 1

The values of means and LSD for studied traits during the non-segregation and segregation generations in the two hybrids G. 92 x Pima S6 (blue columns) and Giza 93 x C.B. 58 (green columns). P1 and P2: First and second parents; F1 and F2: First and second generations; BC1 and BC2: First and second back crosses, respectively; G/E ratio: mean squares of generations/ mean squares of error; C.V%: Coefficient of variation; LSD values denote highly significant differences between the six generations were analyzed by ANOVA test.

Figure 2

Plot describing Pearson's correlation between yield and yield components traits (A) as well as between six generations (B) in the two hybrids of cotton. BW: Boll weight; SCY: Seed cotton yield/plant; LCY: Lint cotton yied/plant; L%: Lint percentage; No.B: Number of bolls/plant; SI: Seed index; P1 and P2: First and second parents; F1 and F2: First and second generations; BC1 and BC2: First and second back crosses, respectively. The large and medium blue circles indicate a positive and significant (* p < 0.05) or highly significant (** p < 0.01), while the small blue circles indicate a positive and non-significant correlation.

Figure 3

Tree diagram constructed using UPGMA hierarchical clustering with correlation similarity index. BW: Boll weight; SCY: Seed cotton yield/plant; LCY: Lint cotton yied/plant; L%: Lint percentage; No.B: Number of bolls/plant; SI: Seed index; P1 and P2: First and second parents; F1 and F2: First and second generations; BC1 and BC2: First and second back crosses, respectively.

Figure 4

Biplot diagram among first two PCAs display contribution of studied traits in variability of six generations across the two crosses. EV: Eigen value; EV: Explained variance; BW: Boll weight; SCY: Seed cotton yield/plant; LCY: Lint cotton yied/plant; L%: Lint percentage; No.B: Number of bolls/plant; SI: Seed index; P1 and P2: First and second parents; F1 and F2: First and second generations; BC1 and BC2: First and second back crosses, respectively.

Figure 5

Scree plot of PCA between respective eigenvalues and components number.