Genetic variability for yield and fiber related traits in genetically modified cotton

SAHAR Adeela[1†], ZAFAR Muhammad Mubashar[1,2†], RAZZAQ Abdul[2,3†], MANAN Abdul[1], HAROON Muhammad[4], SAJID Sunaina[1], REHMAN Abdul[2,5], MO Huijuan[2], ASHRAF Muhammad[1], REN Maozhi[2*], SHAKEEL Amir[1*] and YUAN Youlu[2*]

Abstract

Background: Cotton (*Gossypium hirsutum* L.) is grown for fiber and oil purposes in tropical and sub-tropical areas of the world. Pakistan is the 4th largest producer of cotton. It has a significant contribution in the GDP of Pakistan. Therefore, the present study was performed to assess the genetic variations and genetic diversity of yield and fiber quality traits in cotton and to analyze the associations present among them.

Results: Analysis of variance exhibited significant variation for all studied traits except total number of nodes and the height to node ratio. The phenotypic coefficient of variation was higher than the genotypic coefficient of variation for all studied traits. Plant height, monopodial branches, total number of bolls, lint index, seed index, and seed cotton yield displayed high heritabilities in a broad sense with maximum genetic advance. Correlation analysis revealed that seed cotton yield had a significant positive association with plant height, the number of monopodial branches, the number of sympodial branches, ginning outturn (GOT), the number of bolls, seed per boll, seed index, uniformity index, the number of sympodial branches, reflectance, and seed index at the genotypic level while a significant positive relationship was observed with plant height, the number of sympodial branches, boll number, and GOT. Plant height, monopodial branches, GOT, boll weight, seeds per boll, and short fiber index exerted direct positive effects on seed cotton yield. The first 6 principal component analysis (PCs) out of the total fourteen PCs displayed eigenvalues (> 1) and had maximum share to total variability (82.79%). The attributes that had maximum share to total divergence included plant height, uniformity index, the number of sympodial branches, seed per boll, GOT, seed cotton yield, and short fiber index.

Conclusion: The genotype AA-802, IUB-13, FH-159, FH-458, and CIM-595 were genetically diverse for most of the yield and fiber quality traits and could be utilized for the selection of better performing genotypes for further improvement.

Keywords: Upland cotton, Fiber trait, Cotton yield, Genetic variability, Yield components

* Correspondence: renmaozhi01@caas.cn; dramirpbg@gmail.com; yuanyoulu@caas.cn
†Sahar A, Zafar MM and Razzaq A contributed equally to this work.
[2]Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China
[1]Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan
Full list of author information is available at the end of the article

SAHAR et al. Journal of Cotton Research   (2021) 4:19

Page 2 of 10

Background

Cotton (*Gossypium hirsutum* L.) is mainly grown for fiber and oil purposes and is crucial for the elevation of a country's economy and also called "White gold" (Komala et al. 2018). It is recognized globally as an important fiber crop and is cultivated in tropical and subtropical regions of the world. Cotton is an important crop for an agricultural country like Pakistan. Pakistan is the 4th largest producer of cotton. Moreover, cotton seeds are also used to produce edible oil and feed in the form of seed cake (Shakeel et al. 2015). The shares of the cotton are 0.8% in the GDP and 4.5% in the agriculture sector, but the yield per acre is low in Pakistan due to pest attack and climate change. Cotton production in Pakistan has been reduced to 9.861 million bales and the area under cotton cultivation is 2 373 thousand hectares. The production has been declined up to 17.5%, causing a decrease in cotton ginning by 12.74% (Anonymous 2019–20).

Hence, there is a dire to develop disease- and pest-resistant cotton varieties with high yield and superior fiber quality. Climate change is also adversely affecting the productivity of cotton. Many high-yielding cultivars are available but do not possess better fiber quality as required by the textile industry. The decline in yield and quality has been recorded over the last decade because of the loss in genetic diversity of existing genotypes (McCarty Jr. et al. 2013).

Mainly, four types are cultivated across the world including two diploids, *G. herbaceum* and *G. arboreum*, and two tetraploid species, *G. hirsutum* and *G. barbadense*. The genetic base of the cultivated cotton species should be broadened by conducting breeding programs (Ali et al. 2017; El-Kady et al. 2015; Shabbir et al. 2016). Genetic diversity is an important tool that is helpful in breeding programs for the conservation of germplasm. Genetic diversity should be exploited to produce cultivars with desired characters. The information about the nature and extent of variation, genetic advances, and heritability is required for the selection of superior genotypes and the study of the relation among yield components and fiber characters.

The correlation coefficient assesses simple associations among the plant attributes but does not help us to conclude anything about the selection of the plants. Determination of correlation coefficients should be calculated to devise the best possible combinations among different attributes to get expected results (EL-Mohsen and Amein 2016). The correlation coefficient is divided into direct and indirect effects and is useful to examine the reason for the observed correlation (Manonmani et al. 2019). It provides a better understanding of the character's association. Principal component analysis (PCA) and cluster analysis have been used to uncover the similarities and differences among different genotypes and placed them into various clusters (Naik et al. 2016). The principal component analysis is the most appropriate statistical method to partition total variation that facilitates the selection of elite parental lines. It also revealed the importance of the main contributors to total variability. Moreover, genetic variability among existent genotypes can be assessed by principal component analysis which will help in calculating phenotypic variability (Amna et al. 2013). The knowledge of variability provided by PCA is helpful to select genetically- and agronomically-important genotypes (Isong et al. 2017). The main objective of the present study is to assess the genetic variations and diversity in yield and fiber traits in cotton genotypes and to analyze the associations among them.

Materials and methods

The present study was conducted to assess the genetic variability among 15 cotton genotypes that were obtained from Cotton Research Group, Department of Plant Breeding and Genetics, University of Agriculture Faisalabad, Pakistan (Table 1). These genotypes were planted in three replications under a randomized complete block design (RCBD) in the research area of Department of Plant Breeding and Genetics, University of Agriculture Faisalabad on 11th June, 2019.

Five plants of each genotype were selected from each replication and data was recorded for characters including plant height, the total number of nodes per plant, height to node ratio, the number of monopodial branches per plant, the number of sympodial branches per plant, the number of bolls per plant, boll weight, the total number of seeds per boll, seed index, lint index, GOT, upper half mean length (UHML), fiber strength, fiber fineness, uniformity index, short fiber index reflectance, and seed cotton yield.

Plant height

The height of five plants of each genotype from each replication was calculated from the zero node to the

Table 1 List of Genotypes

Sr. No.	Genotype	Sr. No.	Genotype
1	FH-458	9	FH-444
2	CIM-595	10	AGC-501
3	MNH-888	11	FH-187
4	FH-159	12	AA-802
5	AGC-999	13	VH-330
6	VH-333	14	FH-175
7	VH-259	15	IUB-13
8	BS-80		

apical bud in cm. The average height for each genotype was calculated for statistical analysis.

The number of nodes per plant
As the cotton plant grows, nodes are developed on the main stem. The total number of nodes was recorded for each plant. Average nodes of each genotype were calculated for statistical analysis.

Height to node ratio
The height to node ratio was calculated by dividing plant height by the total number of nodes for each plant. The average height to node ratio of each genotype was calculated from five replicates.

The number of monopodial branches per plant
Monopodial branches are the indirect fruiting branches that were present at the lower part of the plant. Five plants were selected from each replication and the monopodial branches were counted.

The number of sympodial branches per plant
Sympodial branches are the fruiting branches that were present all over the stem of plant. Sympodial branches were counted on five selected plants per replication and then averaged.

The number of bolls per plant
Bolls were picked from five plants of each genotype per replication. The average number of bolls was counted.

Boll weight (g)
The seed cotton yield of a plant was divided by the total number of bolls to calculate the boll weight.

The number of seeds per boll
The total number of seeds per boll was calculated for five plants of each replication by using the total seed weight, seed index, and the total number of bolls. Then the average number of seeds per boll was calculated for each genotype.

Seed cotton yield per plant (g)
The seed cotton yield was collected from bolls of five plants of each genotype per replication. The seed cotton was weighed with electronic balance and then average seed cotton yield was computed for each genotype.

Seed index (g)
Seed index was calculated by weighing 100 healthy seeds and the average was calculated.

Ginning out turn (GOT, %):

Electrical ginner was used for ginning of seed cotton yield. Then the weight of lint was calculated by using an electronic balance. The ginning out turn was calculated by using the following formula.

$$GOT\% = \frac{Sample\ lint\ weight}{Seed\ cottonweight} \times 100$$

Lint index (g)
Lint index is the weight of fiber produced by 100 seeds and calculated by using the following formula.

$$Lint\ index = \frac{Weight\ of\ 100\ seed \times GOT \times 100}{100 - GOT \times 100} \times 100$$

Fiber traits
Fiber length was measured in mm by using HVI. Then the average fiber length was calculated for all samples for statistical analysis. Fiber fineness was also measured by using HVI. After test compilation the sample was removed from the chamber. The average fiber fineness of each sample was calculated for statistical analysis. The uniformity index was calculated by dividing the mean length by upper half mean length (UHML). It is the percentage of uniformity. Fiber strength was determined by placing the fiber between two clamps. The extent of the whiteness of light reflected by cotton fibers is called reflectance. It is used to measure the color grade of cotton in conjunction with the yellowness of fiber. Short fiber index is the percentage of fibers in a sample, by weight, less than half inch in length. It was also measured by HVI.

Statistical analysis
Analysis of variance was performed by using Statistics 8.1 following Steel et al. (1997) to estimate the genotypic differences among all genotypes. The coefficient of variation was also calculated for all characters for the comparison of variability. Genotypic, phenotypic and environmental variances were divided by the means of respective traits to compute coefficients of variation. Heritability in a broad sense was calculated by diving genotypic variance with phenotypic variance. The genetic advance was also calculated at 20% selection intensity. The genotypic and phenotypic correlations were computed by using agricolae packages implemented in R software (R Development Core Team 2015). The correlation coefficients were partitioned into direct and indirect effects by using path coefficient analysis following Dewey and Lu (1959). The divergence was calculated by the principal component.

Table 2 Mean squares for yield and fiber quality related traits in *G. hirsutum* genotypes

SOV	DF	PH	TN	HNR	MB	SB	TB	BW	GOT	LI
Rep.	2	153.58	17.38	0.22	0.028	6.52	0.78	0.54	4.42	0.16
Geno.	14	444.33**	123.74	0.19	0.253*	81.84**	14.33**	0.47*	13.79*	0.89*
Err.	28	126.08	67.59	0.23	0.073	11.57	3.96	0.15	3.86	0.29

Note: PH: Plant Height (cm), TN: Total number of nodes, HNR: Height to node ratio, MB: the number of monopodial branches, SB: the number of sympodial branches, TB: Total number of bolls, BW: Boll weight (g), GOT: Ginning out turn (%), and LI: Lint index (g),
significant (*), highly significant (**)

Results

Analysis of variance revealed highly significant ($P \geq 0.01$) differences among genotypes for plant height, the number of sympodial branches, the number of bolls per plant, seed cotton yield, and fiber strength. The number of monopodial branches, boll weight, GOT, lint index, seed index, the number of seeds per boll, upper half mean length (UHML), fiber strength, short fiber index, uniformity index, and reflectance while the results were not significant for the total number of nodes and height to node ratio (Tables 2 and 3)

Phenotypic variances were slightly higher than genotypic variances. Estimates of genetic variability revealed that the highest genotypic coefficient of variation for seed cotton yield was followed by the number of monopodial branches, the number of sympodial branches, plant height, the number of bolls per plant, and boll weight. A similar trend was also observed for the phenotypic coefficient of variation. The values of environmental coefficients were low as compared with the genotypic coefficient of variation. This indicated that the influence of environment was less on these characters. Plant height, monopodial branches, the number of bolls, lint index, seed index, and seed cotton yield displayed high heritability with maximum genetic advance per percent mean (Table 4).

Correlation and path coefficient analysis

Genotypic and phenotypic correlation coefficients were estimated among all possible combinations of different characters. Correlation study displayed that seed cotton yield was positively and significantly associated with plant height, the number of monopodial branches, the number of sympodial branches, GOT (%), the number of bolls, seed per boll, seed index, uniformity index, and reflectance while negative significant relation was

observed for short fiber index at genotypic level (Table 5). At the phenotypic level, seed cotton yield was positively and significantly associated with sympodial branches, plant height, GOT (%) and the number of bolls while the significantly negative association with short fiber index (Table 6).

Plant height was significantly positively correlated with the number of sympodial branches, uniformity index, and seed cotton yield while a significantly negatively significant correlation was observed with GOT, the number of bolls, UHML, boll weight, seeds per boll, and short fiber index at the genotypic level. The number of monopodial branches displayed a positive and significant correlation with seeds per boll and seed cotton yield. It was positively and non-significantly associated with GOT. The number of sympodial branches was highly significant and positive association with plant height, fiber fineness, seed cotton yield, and seed per boll while boll number was significantly and positively correlated with boll weight, UHML, GOT, and seed cotton yield. The number of seeds per boll was significantly and positively correlated with GOT, uniformity index, seed cotton yield, and the number of monopodial branches. The correlation coefficient revealed seed index was significantly and positively related to boll weight, uniformity index, and seed cotton yield. GOT had a significant and positive association with the number of bolls, seed cotton yield, and seeds per boll. Uniformity index depicted a significant and positive relationship with seed cotton yield, reflectance, and fiber fineness while reflectance had a significant and positive association with fiber strength and seed cotton yield.

At the phenotypic level, plant height was positively and significantly correlated with sympodial branches and seed cotton yield whereas a significantly negative correlation was observed for GOT. Sympodial branches had

Table 3 Mean squares for yield and fiber quality-related traits in *G. hirsutum* genotypes

SOV	DF	SI	SPB	SCY	UHML	FF	FS	RD	SF	UI
Rep.	2	0.77	2.09	0.513	0.006	0.001	0.03	0.33	0.017	0.32
Geno.	14	0.68*	7.18*	388.3**	0.85*	0.086*	3.02**	4.19*	0.17*	1.43*
Err.	28	0.19	2.33	84.7	0.25	0.023	0.82	1.17	0.04	0.42

Note: SI: Seed index (g), SPB: Seed per boll, RD: Reflectance, FF: Fiber fineness (ug·inch^{-1}), FS: Fiber strength (cN·tex^{-1}), UI: Uniformity index, SF: Short fiber index,
SCY: Seed cotton yield (g), and UHML: Upper half mean length (mm)
significant (*), highly significant (**)

Table 4 Estimation of parameters of genetic variability and heritability

Variables	GV	PV	EV	GCV	PCV	ECV	H$_{2b}$ /%	GA	GAM
PH	402.30	444.33	42.03	28.16	29.59	9.10	91.42	28.08	31.51
MB	0.23	0.28	0.02	40.05	42.36	27.09	82.14	0.60	63.3
SB	77.98	99.84	3.86	39.26	41.41	10.51	78.10	10.92	64.58
TB	13.01	15.33	1.32	26.44	27.75	8.42	85.45	4.68	35.27
BW	0.42	0.47	0.05	23.15	24.48	7.99	89.63	0.86	30.63
GOT	12.50	13.79	1.29	8.84	9.28	2.84	91.67	4.76	11.79
LI	0.79	0.89	0.10	19.06	20.19	6.65	89.14	1.17	25.20
SI	0.62	0.68	0.06	11.47	12.05	3.68	90.69	1.04	15.29
SPB	6.40	7.18	0.78	12.16	12.88	4.24	89	3.34	16.08
SCY	360.07	388.30	28.23	42.71	43.27	11.40	93.13	25.69	54.88
UHML	1.03	1.41	0.18	3.79	4.10	1.58	73.04	1.21	4.89
FF	0.08	0.09	0.01	6.60	6.91	2.06	91.09	0.38	8.81
FS	2.75	3.02	0.27	5.66	5.93	1.79	88.90	2.16	7.56
RD	3.80	5.19	0.39	2.54	2.67	0.81	73.21	2.33	3.39
SF	0.16	0.37	0.01	4.85	5.05	1.41	43.26	0.36	6.522
UI	1.29	1.93	0.14	1.33	1.40	0.44	66.83	1.29	1.77

Note: GV: Genotypic variance, PV: Phenotypic variance, EV: Environmental variance, GCV: Genotypic coefficient of variance, PCV: Phenotypic coefficient of variance, ECV: Environmental coefficient of variation, H$_{2b}$: Broad sense heritability (%), GA: Genetic advance, GAM: Genetic advance as % mean

a significantly positive association with the uniformity index and seed cotton yield whereas a non-significant and positive association was observed for fiber strength and fiber fineness while GOT was positively and significantly associated with the number of bolls and seed cotton yield. GOT, seed cotton yield, and fiber strength displayed a positively significant association with boll numbers. The short fiber index displayed a negative and significant relationship with seed cotton yield at both phenotypic and genotypic levels.

Path coefficient analysis determined direct and indirect effects of all the attributes on dependent variable. It revealed that traits like the number of monopodial branches, plant height, seed per boll, short fiber index,

Table 5 Genotypic correlation coefficient for yield and fiber traits in upland cotton

	PH	MB	SB	GOT	SI	TB	BW	LI	SPB	UI	RD	SF	UHML	FF	FS	SCY
PH	1	0.11	0.97**	−0.56**	−0.02	−0.68**	−0.57*	0.09	−0.41*	0.61*	0.19	−0.57*	−0.89**	−0.14	0.03	0.65**
MB		1	−0.12	0.08	−0.06	0.21	0.09	0.36	0.52**	−1.88	0.26	−0.30	−0.03	−0.59	−0.60*	0.49**
SB			1	−0.34	−0.12	−0.41*	−0.35	0.00	−0.01	0.44	−.01*	−0.69	−0.82**	0.03**	0.14	0.62**
GOT				1	0.01	0.76**	0.27	−0.04	0.53**	−0.12	0.21	−0.35	−0.10	−0.77*	−0.14	0.59*
SI					1	−0.61**	0.53**	0.08	−0.31	0.66**	−0.36	0.38	0.24	−0.28	0.04	0.53*
TB						1	0.26**	−0.18	0.01	−0.98*	0.38	−0.11	0.48**	0.26	0.72	0.31*
BW							1	0.03	−0.43*	−0.50*	0.25	0.11	−0.43*	−0.06	0.27	0.09
LI								1	0.38	−0.01	−0.36	0.35**	−0.74**	−0.05	−0.07	−0.06
SPB									1	0.35*	−0.44*	0.06	−0.20	−0.08	−0.40*	0.41*
UI										1	0.12	0.65**	0.27	−0.01	−0.89*	0.25*
RD											1	−0.64*	−0.12	−0.21	0.54*	0.58**
SF												1	−0.16	−0.19	−0.55*	−0.88*
UHML													1	0.25	0.09	−0.38
FF														1	0.60*	0.04
FS															1	−0.13
SCY																1

significant (*), highly significant (**)

Table 6 Phenotypic correlation coefficient for yield and fiber traits in upland cotton

	PH	MB	SB	GOT	SI	TB	BW	LI	SPB	UI	RD	SF	UHML	FF	FS	SCY
PH	1	− 0.17	0.84**	− 0.29*	0.01	− 0.35*	− 0.01	0.02	0.14	0.22	0.12	−0.14	− 0.16	0.08	− 0.09	0.54**
MB		1	−0.22	0.12	0.13	0.20	− 0.04	0.7	0.52	−0.18	0.26	−0.30	−0.03	− 0.59*	− 0.60*	0.10
SB			1	−0.27	− 0.10	− 0.40	−0.07	− 0.06	− 0.01	0.44*	−0.01	− 0.69	−0.26	0.03	0.14	0.46**
GOT				1	− 0.17	0.30*	0.05	0.20	0.53	−0.12	0.21	−0.35	−0.09	− 0.77*	− 0.14	0.12*
SI					1	0.08	0.25	0.13	−0.31	0.66	−0.36	0.38	0.09	− 0.28	0.03	0.16
TB						1	0.14	0.02	0.01	−0.98**	0.38	−0.10	0.10	0.26	0.72*	0.10*
BW							1	0.29*	− 0.43	− 0.50	0.25	0.11	−0.01	− 0.06	0.27	0.03
LI								1	0.38	−0.00	− 0.35	0.35	−0.31	− 0.05	− 0.07	− 0.02
SPB									1	0.27	− 0.30*	− 0.03	0.11	0.09	−0.39*	0.15
UI										1	− 0.00	0.08	0.23	−0.14	− 0.39*	− 0.21
RD											1	− 0.28*	0.11	−0.05	− 0.20	0.20
SF												1	−0.06	−0.11	−0.04	− 0.28*
UHML													1	−0.14	−0.15	− 0.38
FF														1	0.23	−0.12
FS															1	−0.04
SCY																1

significant (*), highly significant (**)

GOT, and boll weight impacted positively and directly on yield. The remaining traits exerted negative direct effects on yield (Table 7).

Principal component analysis
Mean data of all the studied traits was analyzed for PCA to study the genetic divergence among 15 cotton genotypes by using XLSTAT. It provided information about factor loading and the share of each attribute to total variability. The total variance was divided into 14 components. Two-dimensional representation was illustrated by using the first two PCs. The first 6 PCs out of the total of the fourteen PCs displayed eigen-values (> 1) and had maximum share to total variability. PC-I, II, III, IV, V, and VI had a share of 22.63, 16.43, 14.37, 11.12, 9.61, and 8.60% to total variability, respectively (Table 8). PC-

Table 7 Direct and Indirect effects for various yield and quality-related traits in upland cotton

	PH	MB	SB	GOT	SI	TB	BW	LI	SPB	UI	RD	SF	UHML	FF	FS	rg
PH	0.82	0.08	−0.25	−0.40	0.03	0.24	−0.81	−0.06	− 0.12	0.22	0.10	0.33	−0.47	0.07	0.01	0.65
MB	0.09	0.72	0.03	0.06	0.08	−0.38	0.13	−0.25	0.16	0.18	0.01	−0.16	−0.52	0.30	0.46	0.49
SB	0.81	−0.09	−0.26	− 0.24	0.15	0.76	−0.50	− 0.00	−0.00	− 0.43	−0.00	−0.37	−0.50	− 0.01	−0.11	0.62
GOT	−0.47	0.06	0.09	0.71	−0.01	−0.39	0.38	0.02	0.16	0.12	0.01	−0.19	0.05	0.38	0.11	−0.60
SI	−0.01	−0.04	0.03	0.01	−0.26	0.08	0.25	−0.05	−0.09	0.23	−0.18	− 0.22	0.13	0.14	0.01	0.53
TB	−0.56	0.15	0.10	0.54	0.77	−0.82	0.14	0.12	−0.01	−0.53	0.25	−0.19	−0.38	0.04	0.73	−0.30
BW	−0.47	0.07	0.09	0.19	−0.68	−0.48	0.40	−0.02	0.18	−0.78	0.17	0.02	0.34	−0.01	0.28	0.09
LI	0.07	0.26	−0.00	− 0.02	−0.10	0.34	0.05	−0.68	0.17	−0.00	−0.44	−0.13	0.44	− 0.62	−0.03	− 0.06
SPB	−0.21	− 0.47	−0.01	− 0.16	−0.16	0.01	0.14	0.20	0.03	−0.34	0.14	0.34	−0.03	−0.37	0.14	−0.39
UI	0.32	0.17	0.26	0.03	0.35	−0.85	0.07	−0.00	0.12	− 0.35	−0.02	- 0.01	0.04	0.17	0.01	0.25
RD	0.10	−0.24	−0.01	− 0.06	−0.19	0.32	−0.08	− 0.19	−0.08	− 0.02	−0.41	0.14	−0.02	− 0.02	0.02	0.58
SF	−0.29	0.28	−0.41	0.11	0.20	−0.95	−0.04	0.18	0.03	0.01	−0.17	0.34	−0.02	− 0.01	−0.01	− 0.88
UHML	−0.50	− 0.01	−0.61	0.07	0.13	0.41	0.14	−0.39	0.01	0.09	−0.05	0.05	−0.17	0.06	0.01	−0.43
FF	−0.07	0.54	0.02	0.23	−0.15	0.22	0.01	−0.56	0.04	0.23	−0.03	0.02	0.04	−0.26	− 0.00	0.04
FS	0.01	0.55	0.08	0.04	0.02	0.65	−0.09	−0.03	− 0.03	0.02	0.09	0.05	0.01	−0.01	- 0.12	−0.13

Where, PH: Plant Height (cm), TN: Total number of nodes, HNR: Height to node ratio, MB: Monopodial branches, SB: Sympodial branches, TB: Total number of bolls, BW: Boll weight (g), GOT: Ginning out turn (%) and LI: Lint index, (g) SI: Seed index (g), SPB: Seed per boll, RD: Reflectance, FF: Fiber fineness (ug·inch^{-1}), FS: Fiber strength (g·tex^{-1}), UI: Uniformity index, SF: Short fiber index, SCY: Seed cotton yield (g), and UHML: Upper half mean length (mm)

Table 8 Eigenvalues and principal components for yield and fiber attributes in cotton genotypes

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Eigenvalue	3.622	2.631	2.300	1.780	1.538	1.376	0.934	0.668	0.424	0.350	0.202	0.134	0.038	0.002
Variability %	22.638	16.443	14.377	11.126	9.611	8.602	5.839	4.172	2.651	2.187	1.265	0.838	0.241	0.011

I and PC-II contributed 22.63 and 16.43% to total variation. PC-I was mainly related to plant height, uniformity index, the number of sympodial branches and seed cotton yield while PC-II revealed higher and positive values for the number of bolls, seed index, UHML, fiber fineness and fiber strength. PC-III and PC-IV had 14.37 and 11.12% share to total variability, respectively. PC-III had high and positive values for attributes including the number of monopodial branches, reflectance, seed cotton yield, boll weight, and lint index while negative values were observed for seed per boll, UHML, the number of sympodial branches, and short fiber index. PC-IV displayed positive and higher values for boll weight, lint index, seed index, and fiber strength. PC-V and PC-VI had contribution of 9.61 and 8.60% to total variation, respectively. The number of bolls, seed index, the number of monopodial branches and the number of seeds per boll displayed high and positive eigen-vectors for PC-V while for PC-VI positive and high values were demonstrated on UHML, uniformity index, boll weight,

reflectance, the number of monopodial branches, and seed index (Table 9).

Scree plot
Scree plot demonstrated the variance percentage in accordance with all principal components illustrated by a graph between the eigenvalues and principal components. PC1 displayed the highest variability of 22.63% with the eigenvalue of 3.622. Minimum variability was observed in PC13 and PC14 with the eigenvalues of 0.038 and 0.002, respectively. PC1 had maximum variability so the genotypes in PC1 should opt for selection (Fig. 1).

Biplot
Biplot displayed that variables were represented on the plot in the form of vectors. The relative distance of the variables from the origin regarding PC-1 and PC-2 depicted the share of each variable to the total variation of the germplasm. Boll weight, the number of seeds per boll, the number of monopodes and reflectance showed minimum differences as they were close to the origin whereas GOT, plant height, the number of sympodial branches, lint index, UHML, fiber fineness, short fiber index, seed index, fiber strength, boll number, seed cotton yield, and fiber attributes displayed maximum differences as they were at a greater distance from the origin (Fig. 2).

Table 9 Contribution of variables to total variability

	PC1	PC2	PC3	PC4	PC5	PC6
PH	18.687	2.721	0.056	0.022	1.365	1.931
MB	0.084	4.753	8.986	12.873	19.528	1.330
SB	22.265	3.116	0.078	0.002	0.124	5.971
GOT	14.432	2.129	6.352	0.260	3.030	6.391
SI	0.900	8.303	0.042	4.926	19.798	12.855
TB	8.603	5.640	7.488	1.465	9.095	3.810
BW	0.300	0.873	6.758	17.687	1.815	4.671
LI	1.024	15.665	3.982	5.432	2.819	2.064
SPB	0.030	0.619	7.445	8.318	16.340	17.307
UHML	7.293	4.826	2.304	3.837	6.998	24.773
RD	0.101	0.421	25.167	5.919	9.251	6.933
UI	13.454	4.480	1.828	0.141	1.235	2.209
SF	0.612	12.999	16.980	1.653	0.188	2.363
FF	0.112	24.054	4.121	0.328	4.001	1.424
FS	0.027	7.162	0.046	35.349	0.581	2.447
SCY	12.076	2.239	8.366	1.786	3.831	3.523

Where, PH: Plant Height (cm), TN: Total number of nodes, HNR: Height to node ratio, MB: Monopodial branches, SB: Sympodial branches, TB: Total number of bolls, BW: Boll weight (g), GOT: Ginning out turn (%) and LI: Lint index (g), SI: Seed index (g), SPB: Seed per boll, RD: Reflectance, FF: Fiber fineness (ug·inch^{-1}), , FS: Fiber strength (g·tex^{-1}), UI: Uniformity index, SF: Short fiber index, SCY: Seed cotton yield (g), and UHML: Upper half mean length (mm)
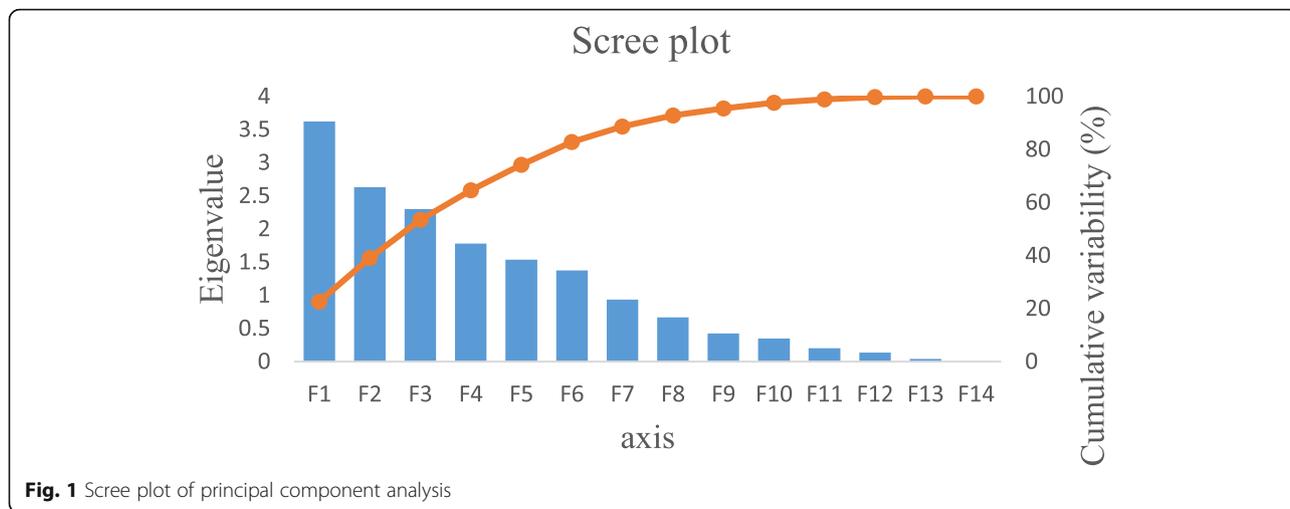
Discussion
The exploitation of genetic diversity is important to select diverse genotypes for breeding programs. Heritability and genetic variability help us to point out the extent of improvement that can be made through selection on the basis of phenotype (Magadum et al. 2012). Genetic variability provides the basis to exploit the genes for the improvement of the desired attributes (Baloch et al. 2014). The divergence of the available germplasm should be utilized to produce genetically diverse and better-quality genotypes (Ullah et al. 2017). Selection is done based on the association present between yield characters and fiber properties. Variation from genetically diverse plant resources can be utilized to develop favorable and pest-resistant varieties. Therefore, the variation present among different morphological and fiber characters should be studied for crop betterment (Memon et al. 2017).

Fig. 1 Scree plot of principal component analysis

Analysis of variance displayed significant variation for all studied traits except total nodes and height to node ratio. Similar results for the mean square of genotypes were also reported by Baloch et al. (2015), Dhamayanathi et al. (2010) and Djaboutou et al. (2017).

Estimates of genetic variability revealed the highest genotypic coefficient of variation for seed cotton yield followed by the number of monopodial branches, the number of sympodial branches, plant height, the number of bolls, and boll weight. A similar trend

was observed for the phenotypic coefficient of variation (Shakeel et al. 2015a). Plant height, the number of monopodial branches, the number of bolls, lint index, seed index, and seed cotton yield displayed high heritability with maximum genetic advance per percent mean. Similar results were reported by many researchers including Dhivya et al. (2014), Khan et al. (2017), Shar et al. (2017), and Hayat and Bardak (2020). Fiber attributes displayed high phenotypic variance values in comparison with genotypic variance as reported by Shakeel

Fig. 2 Principal component biplot for the contribution of traits
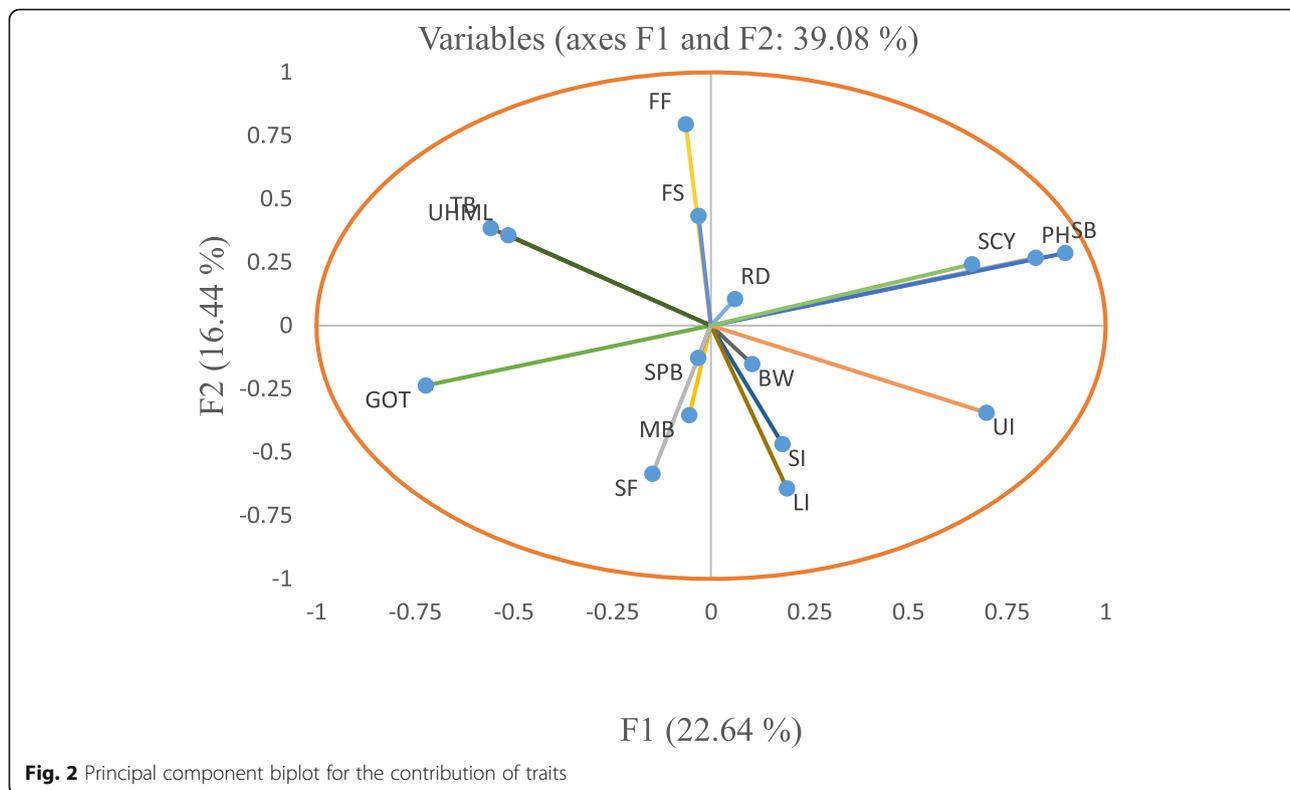
et al. (2015). Heritability was maximum for micronaire and fiber strength with low genetic advance as revealed by Nawaz et al. (2019).

Correlation analysis revealed that seed cotton yield was significantly and positively associated with plant height, the number of monopodial branches, the number of sympodial branches, GOT, the number of bolls, seed per boll, seed index, uniformity index, and reflectance while a negative significant relation was observed for short fibre index at the genotypic level. At the phenotypic level, seed cotton yield was positively and significantly associated with the number of sympodial branches, plant height, GOT, and the number of bolls while a significantly negative association with short fibre index. Similar results have been reported by Reddy et al. (2019) while Kumbhar et al. (2020) reported a significantly positive association of plant height with sympodial branches. Erande et al. (2014) and Nandhini et al. (2019) reported a non-significant and positive correlation of the number of monopodial branches with GOT. Salahuddin et al. (2010) found that at the phenotypic level, yield was positively associated with sympodial and bolls. Shakeel et al. (2015) reported that plant height and seed per boll was significantly and positively correlated with the number of sympodial branches. Baloch et al. (2015) reported a significant and positive relation of bolls with seed cotton yield at both genotypic and phenotypic levels. A similar association of seeds per boll with other yield and fiber traits was observed by Ali et al. (2020), Rai and Sangwan (2020), and Bhatti et al. (2020). Seed index was significantly and positively associated with yield and uniformity ratio as displayed by Ahmed et al. (2019), and Rai and Sangwan (2020). Erande et al. (2014) and Monisha et al. (2018) reported a significantly positive association of GOT with yield, the number of bolls and seeds per boll. The results of path analysis were in accordance with Nandhini et al. (2019) and Kumbhar et al. (2020), who suggested direct positive effects of boll weight on yield. Manonmani et al. (2019) and Ali et al. (2020) reported a direct positive effect of seed per boll on yield. Ahsan et al. (2015) observed that GOT had positive and direct effects on yield.

PCA divided total variance into 14 components. Two-dimensional representation was domenstrated by using the first two PCs. The first 6 PCs out of the total of fourteen PCs displayed eigen-values (> 1) and had maximum share to total variability. Isong et al. (2017) reported similar results. According to scree plot, PC1 displayed the highest variability 22.63% with an eigenvalue of 3.622. Minimum variability was observed for PC13 and PC14 with eigenvalues of 0.038 and 0.002, respectively. PC1 had maximum variability so the genotypes in PC1 should opt for selection. The results were in accordance with the findings of Shakeel et al. (2015) and Riaz et al.

(2019). The analysis revealed maximum variation in the germplasm that can be utilized in future breeding programs to get desired genotypes.

Conclusion

Heritability estimation are of great importance for plant breeders. The traits with higher heritability (> 80%) make the selection procedure easier. Plant height, the number of monopodial branches, the number of bolls, lint index, seed index, and seed cotton yield displayed a high heritability with the maximum genetic advance. Correlation study displayed significant association of seed cotton yield with sympodial branches, plant height, GOT, and the number of bolls at both genotypic and phenotypic levels. Path coefficient analysis unveiled that plant height, the number of monopodial branches, GOT, boll weight, seeds per boll, and short fiber index exerted direct positive effects on seed cotton yield while the rest of the characters had direct negative effects on yield. PCA provided information about the divergence present among genotypes. AA-802, IUB-13, FH-159, FH-458, and CIM-595 were genetically diverse and could be utilized for the selection of better-performing genotypes for further improvement.

Authors' contributions
Sahar A, Zafar MM, and Razzaq A wrote the initial draft of the manuscript. Manan A, Haroon M, and Sajid S made all necessary corrections and carried out final editing of the manuscript. Rehman A, Mo HJ, Shakeel A, and Ashraf M proofread the manuscript. Final approval for publication was given by the group leader Shakeel A, Yuan YL and Ren MZ. The authors read and approved the final manuscript.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no conflict of interest for the publication of the manuscript.

Author details
[1]Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan. [2]Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China. [3]Institute of Molecular Biology and Biotechnology, The University of Lahore-Pakistan, Lahore, Pakistan. [4]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China. [5]Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou University, Zhengzhou, China.

References

Ahmed H, Ibrahim SE, Ibrahim AE, et al. Assessment of phenotypic performance, association and path analysis of upland cotton (*Gossypium hirsutum* L.) yield and fiber quality traits in central Sudan. Gezira J Agric Sci. 2019;17:1.

Ahsan MZ, Laghari MA, Bhutto H, et al. Genetic variability in different biometric traits of upland cotton genotypes. Int J Biol Biotechnol. 2015;12:607–12.

Ali I, Khan NU, Mohammad F, et al. Genotype by environment and GGE-biplot analyses for seed cotton yield in upland cotton. Pak J Bot. 2017;49(6):2273–83.

Ali M, Kundu S, Alam M, et al. Selection of genotypes and contributing characters to improve seed cotton yield of upland cotton (*Gossypium hirsutum* L.). Asian Res J Agric. 2020;13(1):31–41. https://doi.org/10.9734/arja/2020/v13i130095.

Amna N, Jehanzeb F, Abid M, et al. Estimation of genetic diversity for CLCuV, earliness and fiber quality traits using various statistical procedures in different crosses of *Gossypium hirsutum* L. J Вестник аграрной науки. 2013;4:43.

Economic Adviser's Wing, Finance Division, Government of Pakistan. Economic survey of Pakistan, 2019–20. http://www.finance.gov.pk/survey/chapter_20/PES_2019_20.pdf. Accessed 20 Dec 2020.

Baloch A, Baloch M, Jatoi S, et al. Genetic diversity analysis in genetically modified cotton (*Gossypium hirsutum* L.) genotypes. Sindh Uni Res J. 2015; 47(3):527–30.

Baloch A, Sahito J, Ali M, et al. Association analysis of yield and fiber traits in advance Pakistani cotton cultivars (*Gossypium hirsutum* L.). Adv Appl Agri Sci. 2014;2(12):73–80.

Bhatti MH, Yousaf A, Ghani MA, et al. Assessment of genetic variability and traits association in upland cotton (*Gossypium hirsutum* L.). Int J Bot Studies. 2020; 5(2):148–51.

Dewey DR, Lu K. A correlation and path-coefficient analysis of components of crested wheat grass seed production. Agron J. 1959;51(9):515–8. https://doi.org/10.2134/agronj1959.00021962005100090002x.

Dhamayanathi K, Manickam S, Rathinavel K, et al. Genetic variability studies in *Gossypium barbadense* L. genotypes for seed cotton yield and its yield components. Electron J Plant Breed. 2010;1(4):961–5.

Dhivya R, Amalabalu P, Pushpa R, et al. Variability, heritability and genetic advance in upland cotton (*Gossypium hirsutum* L.). African J Plant Sci. 2014; 8(1):1–5. https://doi.org/10.5897/AJPS2013.1099.

Djaboutou M, Sinha M, Houedjissin S, et al. Variability and heritability of morphological traits in collection of cotton genotypes (*Gossypium hirsutum* L.) and their potential use for the selection. Eur Sci J. 2017;13:3. https://doi.org/10.19044/esj.2017.v13n3p385.

El-Kady DA, El-Mohsen AA, Latif HMA, et al. Evaluating bivariate and multivariate statistical analysis of yield and agronomic characters in Egyptian cotton. Pak J Sci Agri. 2015;9(3):150–64. https://doi.org/10.15192/PSCP.SA.2015.9.3.150164.

Erande CS, Kalpande H, Deosarkar D, et al. Genetic variability, correlation and path analysis among different traits in desi cotton (*Gossypium arboreum* L.). Afr J Agric Res. 2014;9(29):2278–86. https://doi.org/10.5897/AJAR2013.7580.

Hayat K, Bardak A. Genetic variability for ginning outturn and association among fiber quality traits in an upland cotton global germplasm collection. Sains Malaysiana. 2020;49(1):11–8. https://doi.org/10.17576/jsm-2020-4901-02.

Isong A, Balu PA, Ramakrishnan P. Association and principal component analysis of yield and its components in cultivated cotton. Electron J Plant Breed. 2017;8(3):857–64. https://doi.org/10.5958/0975-928X.2017.00140.5.

Khan MI, Haq HA, Ullah K, et al. Genetic diversity and correlation studies for cotton leaf curl disease (CLCuD), fiber & yield related attributes in exotic lines of *Gossypium arboreum* L. Am J Plant Sci. 2017;8(3):615–24. https://doi.org/10.4236/ajps.2017.83042.

Komala M, Ganesan NM, Kumar M, et al. Genetic variability, heritability and correlation analysis in F_2 populations of ratoon upland cotton hybrids. Int J Agric Environ Biotechnol. 2018;11(6):815–27. https://doi.org/10.30954/0974-1712.12.2018.2.

Kumbhar ZM, Jatoi WA, Sootaher JK, et al. Studies on correlation and heritability estimates in upland cotton (*Gossypium hirsutum* L.) genotypes under the agro-climatic conditions of Tandojam, Sindh, Pakistan. Pure Appl Bio. 2020; 9(4):2272–8. https://doi.org/10.19045/bspab.2020.90241.

Magadum S, Banerjee U, Ravikesavan R, et al. Variability and heritability analysis of yield and quality traits in interspecific population of cotton (Gossypium spp.). Bioinfolet. 2012;9(4A):484–7.

Manonmani K, Mahalingam L, Malarvizhi D, et al. Genetic variability, correlation and path analysis for seed cotton yield improvement in upland cotton (*Gossypium hirsutum* L.). J Pharmaco Phytochem. 2019;8(4):1358–61.

McCarty Jr J, Jenkins J, Wubben M, et al. Registration of three germplasm lines of cotton derived from *Gossypium barbadense* L. accession GB713 with resistance to the *reniform nematode.* J Plant Regist. 2013;7(2):220–3. https://doi.org/10.3198/jpr2012.08.0024crg.

Memon S, Baloch AW, Gandahi N, et al. Evaluation of genetic divergence, character associations and path analysis in upland cotton genotypes. Pure Appl Bio. 2017;6(4):1516–21. https://doi.org/10.19045/bspab.2017.600163.

El-Mohsen A, Amein MM. Study the relationships between seed cotton yield and yield component traits by different statistical techniques. Int J Agron Agric Res. 2016;8(5):88–104.

Monisha K, Premalatha N, Sakthivel N, et al. Genetic variability and correlation studies in upland cotton (*Gossypium hirsutum.* L). Electron J Plant Breed. 2018;9(3):1053–9. https://doi.org/10.5958/0975-928X.2018.00131.X.

Naik BM, Satish Y, Babu DR, et al. Genetic diversity analysis in American cotton (*Gossypium hirsutum* L.). Electron J Plant Breed. 2016;7(4):1002–6. https://doi.org/10.5958/0975-928X.2016.00137.X.

Nandhini K, Premalatha N, Saraswathi R, et al. Genetic studies for yield and fibre quality related traits in upland cotton (*Gossypium hirsutum* L.) and identification of superior transgressive segregants. Int J Curr Microbiol App Sci. 2019;8(11):1609–19. https://doi.org/10.20546/ijcmas.2019.811.186.

Nawaz B, Naeem M, Malik TA, et al. Estimation of gene action, heritability and pattern of association among different yield related traits in upland cotton. Int Multidiscip Res J. 2019;3(1):25–52. https://doi.org/10.29329/ijiaar.2019.188.4.

Rai P, Sangwan O. Correlation studies in elite lines of upland cotton (*Gossypium hirsutum* L.). Electron J Plant Breed. 2020;11(1):298–300. https://doi.org/10.37992/2020.1101.052.

Reddy B, Somveer N, Sangwan RS, et al. Genetic parameters, correlation and path analysis of agro-morphological traits in elite genotypes of upland cotton (*Gossypium hirsutum* L.). Int J Crop Sci. 2019;7(5):1885–9.

Riaz B, Saeed A, Fiaz S, et al. Genetic diversity among bt cotton (*Gossypium hirsutum* L.) germplasm assessed through morphological and within-boll yield attributes. J Anim Plant Sci. 2019;29(1):226–31.

Salahuddin S, Abro M, Kandhro M, et al. Correlation and path co-efficient analysis of yield components of upland cotton (*Gossypium hirsutum* L.). World Appl Sci J. 2010;8:71–5.

Shabbir RH, Bashir QA, Shakeel A, et al. Genetic divergence assessment in upland cotton (*Gossypium hirsutum* L.) using various statistical tools. J Glob Innov Agric Soc Sci. 2016;4(2):62–9. https://doi.org/10.22194/JGIASS/4.2.744.

Shakeel A, Talib I, Rashid M, et al. Genetic diversity among upland cotton genotypes for quality and yield related traits. Pak J Agric Sci. 2015;52:73–7.

Shar T, Baloch M, Arain M, et al. Phenotypic associations, regression coefficients and heritability estimates for quantitative and fiber quality traits in upland cotton genotypes. Pak J Agric Agric Eng Vet Sci. 2017;33(2):142–52.

Steel RGD, Torrie JH, Dickey DA. Principles and procedures of statistics: a biometrical approach. 3rd ed. New York, USA: McGraw Hill; 1997.

Ullah A, Sun H, Yang X, Zhang X. Drought coping strategies in cotton: increased crop per drop. Plant Biotechnology J. 2017;15(3):271–84. https://doi.org/10.1111/pbi.12688.