Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

In Silico Analysis and Expression Profiling of Expansin A4, BURP Domain protein RD22- like and E6-like Genes Associated with Fiber Quality in Cotton

Farzana Ashraf
IPBB: Institute of Plant Biology and Biotechnology
Asif Ali Khan
IPBB: Institute of Plant Biology and Biotechnology
Nadia Iqbal
IPBB: Institute of Plant Biology and Biotechnology
Zahid Mahmood
Central Cotton Research Institute
Abdul Ghaffar
MNSUA: Muhammad Nawaz Shareef University of Agriculture
Zulqurnain Khan (zulqurnain.khan@mnsuam.edu.pk)
Muhammad Nawaz Shareef University of Agriculture https://orcid.org/0000-0002-6910-7389

Research Article

Keywords: DNA sequencing, Expression analysis, Fibre genes, In silico analysis, Cotton

Posted Date: February 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1327190/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Background

Wild *Gossypium* species and races are rich source of genetic polymorphism due to environmental dispersal and continuous natural selection. These genetic resources hold mass of outclass genes that can be used in cotton improvement breeding programs to exploit possible traits such as fibre quality, abiotic stress tolerance, and disease and insect resistance. Therefore, use of new molecular techniques such as genomics, transcriptomics and bioinformatics is very important to utilize the genetic potential of wild species in cotton improvement programs.

Methods

Interspecific lines and *Gossypium* species used in the study were grown at Central Cotton Research Institute (CCRI), Multan. After retrieving DNA sequence of the genes from NCBI, the primers for gene expression and full-length gene sequence were designed. Expression profiling of *Expansin A4*, *BURP Domain protein RD22-like* and *E6-like* fibre genes was performed through Real Time PCR. BLAST and DNA sequence alignment was conducted for sequence comparison of interspecific lines and *Gossypium* species. Different *in silico* analysis were used for characterization of fibre genes and identification of cis acting promoter elements in promoter region.

Results

Variable expression of genes related to fibre development was observed at different stages. BLAST and DNA sequence alignment exhibited resemblance of interspecific lines with *G. hirsutum*. *In silico* analysis on the sequence data also confirmed the role of *Expansin A4*, *BURP Domain protein RD22-like* and *E6-like* fibre genes in fibre development. Similarly, several stress tolerant and light responsive cis acting elements were identified through promotor analysis, which may contribute for fibre development in the breeding programs.

Conclusion

Expansin A4, *BURP Domain RD22-like* and *E6-like* have positive role in fibre development with variable expression at fiber length and strength associated stages.

Introduction

Globally, synthetic fibre consumption is continuously increasing and projected to reach at 130 million tons by 2030. The consumption of synthetic fibre is 62.7% compared to 24.3% cotton fibre consumption [1]. Competition of cotton fibre with polyester is creating negative influence on the demand of cotton. Genetic improvement of cotton for fibre traits is very crucial to meet the challenges of the textile industry. So, there is need to devise clear-cut policies for cotton breeding program to enhance the quality cotton production. In a breeding program, germplasm collection, its conservation and utilization, trait specific screening programs and modern genomics have key role in variety development [2]. Cotton genetic resources have been extensively studied over the last many decades to introduce valuable traits in cotton [3–5]. These genetic resources include wild *Gossypium* germplasm, innovative cytogenetic stocks with specific chromosomes additions or deletions in different species, large mapping families, recombinant inbred lines, near isogenic lines and interspecific lines. While there are some queries about narrow genetic base of these cultivars and most breeders would admit that in breeding programs maximum utilization of genetic diversity within their material should be ensured. Breeders will have to utilize wild cotton relatives, as well as advance lines or cultivars to develop cotton varieties with superior traits.

At the cellular level, cotton fibre development is supported by several genes which facilitates the elongation process, for example, *Expansins* are involved in fibre elongation at various development stages [6]. High transcript abundance of *GhEXP1* was observed in cotton fibre during the elongation phase of fibre development, which steadily decreased from 16 to 20 DPA [7, 8]. In cotton, *GhEXPA1* along with *GhRDL1* showed an increase in fiber length and an enlargement of endopleura cells of ovules [9]. The BURP Domain is a plant-specific protein characterized by repetitive units of amino acid [10]. This protein is mainly involved in promoting the fibre cells elongation when over-expressed. Because *GhRDL1* directly interacts with cotton α-*Expansin* fibre gene therefore, *Expansin*s mediate *GhRDL1's* effect on overall fibre cell enlargement [9, 11]. It was suggested that E6 protein is involved in fibre development, but no support was present to justify this hypothesis as no conclusive evidence was presented [12]. When E6 antisense suppression construct was used, there was knockdown to uncover a phenotype *E6-like*. E6 proteins play a comprehensive role in cell wall, and are deposited during fibre elongation, which give high transcripts in fibre cell during transcriptomic analysis [13].

Transcriptional profiling is a unique tool to gain knowledge about gene mechanisms, regulatory pathways, and gene expression [14, 15]. Number of techniques are used for specific gene expression studies but Real Time PCR is the most reliable technology for absolute and comparative quantification of the gene transcription [16]. This comprehensive wide-ranging gene expression study is supportive to sightsee the role of genes, which are up regulated, entirely expressed, or down regulated during different cotton fibre development stages. Through transcriptomic data, one can explain the fibre expansion process and can discover highly expressed genes for the development of transgenic cotton varieties with superior fibre traits. Profiling of fibre genes in interspecific lines will enable us to unravel variable expression pattern of selected fibre genes.

Application of *in silico* methods along with expression profiling is important for characterization of fibre genes. DNA sequence of interspecific lines and *Gossypium* species were aligned to have information about differences and similarities. Diploid and tetraploid genomes of various *Gossypium* species have repeatedly sequences making their entire genome sequences. These valuable repeatedly sequenced data revealed the evolutionary history of the cotton with polyploidization and decaploidization leading to the of the formation of genus *Gossypium* [17]. Multiple sequence alignment approaches envisage algorithmic explanation about evolutionarily sequences alignments. Fibre genes were subjected to BLAST analysis for expression validation and multiple DNA sequence alignment for similarities and differences of interspecific lines and parent species. Genomics combines recombinant DNA technology, DNA sequencing and bioinformatics sequence to analyze the structure and function of genes [18]. Bioinformatics is a systematic field that utilizes advance approaches for computational analysis of biological data [18]. Bioinformatics also aids to recognize different promoters involve in fibre yield and quality, abiotic stress tolerance and disease resistance. Strength and specificity related character of promoter sequence can be exhibited through expression profiling. Strong promoters predict high expression and vice versa. Fibre genes protein *E6*, *Expansin A4* and BURP *Domain RD22-like* also have strong promoters, which can be used in future breeding program.

Cotton breeders have extensively carried out interspecific hybridization for utilization of desirable genes from wild species to cultivated cotton and developed interspecific cotton varieties. Among them, a lot of upland cotton lines with improved traits including fibre quality and insect pest resistance have been developed [19–22]. All these upland cotton lines are designated as introgression lines of interspecific hybridization. These interspecific lines with their practical value in cotton breeding program have changed genetic basis from narrow line to a wide broad base in the present upland cotton germplasm and have broken the bottlenecks of breeding. However, the full potential of interspecific lines have not yet been obtained for beneficial traits exploitation in traditional and advanced breeding programs [23]. Therefore, this study was designed to evaluate the expression of fibre genes in diverse interspecific lines and *Gossypium* species and their role in different fibre development stages. Results of this study will be directive for development of high-quality cotton varieties.

Materials And Methods

DNA Sequence retrieval and primer designing

DNA sequences of selected fibre genes (*Expansin A4, BURP Domain protein RD22- like and E6-like*) were retrieved from NCBI website https://www.ncbi.nlm.nih.gov/. RT-PCR Primers were designed using PRIMER 3.0 software (Table 1).

Collection of fibre tissues

Three interspecific lines (SL-19, SL-79 and SL-369) of varying fibre length categorized as long fibre (34.7mm), medium fibre (28.5 mm) and short fibre (24.0 mm) along with three parent species (*G. arboreum*, *G. anomalum* and *G. hirsutum*) were used for fibre tissue collection. Cotton bolls were collected at different stages (0, 05, 10, 15 and 20 days after anthesis). Collected bolls were rinsed with diethyl pyro carbonate (DEPC) treated water and were stored in liquid nitrogen. These frozen bolls were further used for RNA extraction.

Plant RNA extraction and cDNA synthesis

RNA was extracted following Gynidium isothiocynate method [24,25]. RNA quality was observed by electrophoresis and monitored under UV light. RNA samples were quantified through nanodrop (Thermo Scientific ND 2000) and concentrations was optimized prior to cDNA synthesis. Extracted RNA from fibre tissues was used for cDNA synthesis.

Real Time PCR analysis

To certify the sequence for specific gene, BLAST short (http://www.ncbi.nlm.nih.gov) was used. For expression analysis, Real Time PCR was performed by with SYBR Green Super Mix (Bio-Rad, USA) and 10 ng/µl of both set of primers. 18S rRNA constitutive gene primers were used as data normalizer in this assay.

Table 1 Primers used for Real Time qPCR Assay

Gene annotation	Primer pair	Primer sequence (5'-3')	Primer length	Product length (bp)	Accession No.
18S rRNA	RT18S - F	AAACGGCTACCACATCCAAG	20	153	**U42827.1**
	RT18S R	CCTCCAATGGATCCTCGTTA	20		
E6-like	RTE6-F	ATGGCTTCCTCACCAAAACTCTTCT	25	211	**DQ023519**
	RTE6-R	TTTCAGGGATGAACCTTGGCTCTT	24		
Expansin A4	RT EXPF	ATGGCAACCAAAACGATGATGT	22	220	**DQ204495**
	RT EXPR	AAGCTGCTGTGCTCGTTCCAT	21		
BURP Domain RD22-like-like	RTRD22-F	ATGAAGGTTCTCTCCCCAATTCT	23	198	**XM_016894801**
	RTRD22-R	GACGTTTACACCACCACCTCCT	22		

Full length gene specific primer designing

Full length primers (Table 2) were retrieved from phytozome https://phytozome.jgi.doe.gov/pz/portal.html.

Table 2 Detail of full-length fibre genes

Gene	Accession No	Size	5'F	5'R
E6-like	DQ023519	726	ATGGCTTCCTCACCAAAACTCTTCT	TCAGGGTTCGAACTCTTCCTCGCTT
Expansin A4	DQ204495	777	ATGGCAACCAAAACGATGATGT	TTAAAACTGGCCTCCTTCAAAAGT
RD-22	XM_016894801	1008	ATGAAGGTTCTCTCCCCAATTCT	TTACTTAGGGACCCAAACAATGT

Sequencing of PCR product

PCR products of full-length primers were sent to Macrogen Korea for Sanger sequencing. Sequencing PCR was performed using gene specific forward primers.

Sequencing comparison of interspecific lines and species

Multiple alignment of predicted DNA sequences and phylogenetic tree analysis was performed at https://www.ebi.ac.uk/Tools/msa/clustalo [26].

In silico analysis of fibre genes

Sequence of Sus gene was taken from NCBI database (https://www.ncbi.nlm.nih.gov/) by searching accession number in all data bases. Coding sequences were identified with amino acid residues. Translation of gene sequence into amino acid sequences was done through EXPASY (https://web.expasy.org/translate/) into six reading frames.

Theoretical computation of physicochemical properties

Basic physiochemical properties and hydropathy index of protein sequences were computed through Expasy's ProtParam Proteomic server (http://web.expasy.org/protparam/).

Functional annotation of protein

For Subcellular Location DeepLoc-1.0 (http://www.cbs.dtu.dk/services/DeepLoc) databases was used. Moreover, SignalP 4.0 (http://www.cbs.dtu.dk/services/SignalP/) was used to check existence of signal peptide.

Promoter sequence analysis

Promoter analysis was carried out at http://bioinformatics.psb.ugent.be/webtools/plantcare/html/.

Results

Expression profiling of *Expansin A4*, *BURP Domain protein RD22-like* and *E6-like*

Overall expression of *Expansin A4* gene was remarkably high in rapid elongating fibre during 10 DPA in all interspecific lines and *Gossypium* species. Maximum transcripts were found in SL-19 (Fig. 1). Expression of *BURP Domain protein RD22-like* was almost remained constant from 10-20 DPA fibre in all genotypes except in *Gossypium anomalum*. Transcripts of *BURP Domain protein RD22-like* gene were maximum in 10 DAP fibre as compared to 5 DPA. In all three interspecific lines, highest expression was detected at 15 and 20 DPA fibre stages in SL-19, SL-79 and SL-369 respectively (Fig. 2). Expression pattern of *E6-like* showed that high expression was detected at 10 and 15 DPA fibre stages predicting its main role in fibre elongation. In interspecific lines, transcripts of *E6-like* gene were varied from 0 DPA till 20 DPA. In SL-19, expression of fibre gene starts to increase from 0 DPA and reached at maximum level at 15 DPA and after that slightly decreases at 20 DPA (Fig. 3).

To validate expression results, the target gene transcriptomic profiles (E*6-like*, *Expansin A4* & *BURP Domain protein* RD22-like) were validated by using existing RNA-seq data on Cotton FGD. The results of available fibre specific genes were generally similar with our expression analysis results. Heat map was created on the basis of RNA-seq data of related expressed in transcript per Million (TPM) during different fibre development stages. *E6-like*, *Expansin A4* and *BURP Domain RD22-like* showed similarity with gene Gh-D05G160200, Gh_A10G149600 and Gh_D05G052400 respectively. (Fig. 4). An expression trend of gradual increasing from 5 DPA to 10 DPA were identified, while similar tendencies were also observed in our experiment.

values of log2, day post anthesis and fragments per kilobase of transcript per million mapped reads**.**

Sequence comparison of interspecific lines and species

In *E6-like* DNA sequencing, all interspecific lines exhibited sequences more similar to *G. hirsutum* as depicted at nucleotide positions 213, 217 and 221-226. In *Expansin A4*, interspecific lines were also more closely related to *Gossypium hirsutum* predicted at 390, 393, 507, 519 & 657bp which also confirm its breeding history. In B*BURP Domain RD22-like*, it was also predicted that almost all dissimilar nucleotide (241-300, 301-360, 361-420) were observed in *G. anomalum* as compared to other species of cotton (Fig. 5).

In silico analysis of *E6-like*, *Expansin A4* and *BURP Domain RD-22*

Physicochemical properties

Expasy's Protpam analysis of predicted protein showed that Protein *E6-like* and *RD-22* was characterized as unstable as value of instability index was 47.75 and 44.72 respectively (Table 3). *Expansin A4* was characterized as a stable protein with value of instability index of 29.01.

Table 3 Physicochemical properties of fibre genes

Physicochemical properties	*E6-like*	*Expansin A4*	*BURP Domain RD-22*
Number of amino acids	241	258	335
Total negatively amino acid charged residues (Asp + Glu)	37	13	35
Total positively amino acid charged residues (Arg + Lys)	25	16	34
Molecular weight	28223.37	27936.46	36595.05
Theoretical pI	5.00	8.36	6.89
Aliphatic index	32.37	62.83	75.64
Grand average of hydropathicity (GRAVY)	-1.356	-0.090	-0.266
Instability index (II)	47.75	29.01	44.72

Subcellular Localization

DeepLoc analysis designated that protein. Proteins *E6-like*, *Expansin A4* and *BURP Domain RD22-like* were a membrane soluble protein family. Location in different organelles with the approximate values (Table 4) predicted the probability of protein location in different organelles. Highest Extracellular values of Proteins *E6-like*, *Expansin A4* and *BURP Domain RD22-like* (0.819, 0.729 and 0.843 respectively) showed that these proteins are extracellular.

Table 4 Predicted subcellular localization of *E6-like*, *Expansin A4* and *BURP Domain RD-22*

Fibre gene	Extracellular	Lysosome	Endoplasmic reticulum	Cell membrane	Golgi apparatus	Cytoplasm
E6-like	0.8195	0.1706	0.0083	0.0013	0.0002	0.0002
Expansin A4	0.7293	0.2373	0.0329	0.0005	0	0
BURP Domain RD-22	0.8435	0.1316	0.0237	0.0008	0	0.0003

Signal peptide analysis

In *E6-like*, *Expansin-A4* and BURP Domain RD22 were characterizes as extracellular membrane that's why signal peptide was present in protein coding sequence. Score values of C, S, 3Y is more than 0.45 (Table 5) that shows that peptide signal is present.

Table 5 Signal peptide Analysis of *E6-like*, *Expansin A4* and *BURP Domain RD22-like*

Fibre gene	Measure	Position	Value	Cut Off	Signal Peptide
E6-like	max.C	26	0.792		
	max.Y	26	0.840		
	max.S	15	0.941		
	Mean S	1-25	0.891		
	D	1-25	0.868	0.450	Yes
Expansin A4	max.C	30	0.427		
	max.Y	30	0.586		
	max.S	9	0.950		
	Mean S	1-29	0.821		
	D	1-29	0.713	0.45	yes
BURP Domain RD22-like	max.C	30	0.427		
	max.Y	30	0.586		
	max.S	9	0.950		
	Mean S	1-29	0.821		
	D	1-29	0.713	0.450	Yes

Promoter sequence Analysis

Sequence analysis of cotton *E6-like*, *Expansin A4* and *BURP Domain protein RD22-like*promoter using PlantCARE predicted many vital motifs in this region related to gene expression (Fig. 4). There are few transcriptions activation related motifs along with core promoter elements like TATA and CAAT boxes. These motifs are light responsive, hormone and stress regulated cis elements. These motifs are involved in the light, stress and hormones responsiveness. There were other vital core promoter elements required for promoter activity including TATA box and CAAT box (Tale-6). Cis-acting essential element for the abscisic acid reaction (*Hordeum vulgare*), light response

elements (*Arabidopsis thaliana*), gibberellin-enhancer element (*Brassica oleracea*) and element for variation of the palisade mesophyll cells (*Arabidopsis thaliana*) were present in *E6-like* promoter region. Similarly, in *Expansin A4* various cis acting premotor elements were identified. Abscisic acid responsiveness elements were identified in *Arabidopsis thaliana*, light responsiveness in *Zea mays*, element responsive for transcription start in *Brassica oleracea* and MeJA-responsiveness in *Hordeum vulgare*. In *BURP Domain RD22-like*, elements essential for light responsiveness were present in *Petroselinum crispum* while promoter and enhancer regions were identified in *Arabidopsis thaliana*. MYBHv1 binding site, MeJA and anaerobic induction responsive elements were present in *Hordeum vulgare* and *Zea mays* respectively.

Table 6 Cis acting promoter elements in promoter region

E6-like						
Site Name	**Organism**	**Position**	**Strand**	**Score.**	**Sequence**	**Function**
ABRE	*Hordeum vulgare*	425	-	9	GCAACGTGTC	cis-acting element involved in the abscisic acid responsiveness
AE-box	*Arabidopsis thaliana*	748	+	8	AGAAACAA	part of a module for light response
CAAT-box	*Arabidopsis thaliana*	638	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Pisum sativum*	852	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
GARE motif	*Brassica oleracea*	615	-	7	TCTGTTG	gibberellin-responsive element
HD-Zip 1	*Arabidopsis thaliana*	564	-	8	CAAT(A/T) ATTG	element involved in differentiation of the palisade mesophyll cells
TATA-box	*Arabidopsis thaliana*	575	-	4	TATA	core promoter element around -30 of transcription start
TC-richrepeats	*Nicotiana tabacum*	380	+	9	GTTTTCTTAC	cis-acting element involved in defense and stress responsiveness
TCT-motif	*Arabidopsis thaliana*	384	+	6	TCTTAC	part of a light responsive element
Expansin A-4						
G-Box	*Pisum sativum*	507	-	6	CACGTT	cis-acting regulatory element involved in light responsiveness
ABRE	*Arabidopsis thaliana*	508	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
ABRE	*Arabidopsis thaliana*	508	+	5	ACGTG	cis-acting element involved in the abscisic acid responsiveness
ATC-motif	*Zea mays*	384	-	9	TGCTATCCG	part of a conserved DNA module involved in light responsiveness
CAAT-box	*Pisum sativum*	361	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Arabidopsis thaliana*	581	-	8	CCCAATTT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Petunia hybrida*	694	-	7	TGCCAAC	common cis-acting element in promoter and enhancer regions
TATA-box	*Arabidopsis thaliana*	527	-	4	TATA	core promoter element around -30 of transcription start
TGACG-motif	*Hordeum vulgare*	532	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
BURP Domain RD22-like						
ABRE	*Triticum*	181	-	9	GACACGTGGC	cis-acting element involved in the

	aestivum					abscisic acid responsiveness
ARE	*Zea mays*	542	+	6	AAACCA	cis-acting regulatory element essential for the anaerobic induction
Box 4	*Petroselinum crispum*	450	-	6	ATTAAT	part of a conserved DNA module involved in light responsiveness
CAAT-box	*Arabidopsis thaliana*	55	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
CCAAT-box	*Hordeum vulgare*	440	+	6	CAACGG	MYBHv1 binding site
CGTCA-motif	*Hordeum vulgare*	515	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
TATA-box	*Arabidopsis thaliana*	291	+	4	TATA	core promoter element around -30 of transcription start
TGACG-motif	*Hordeum vulgare*	512	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
TGACG-motif	*Hordeum vulgare*	515	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness

Discussion

Realistic genetic resources are accessible for innovative cotton breeders to make more perfection in crop improvement. Transcriptomic analysis of interspecific lines and *Gossypium* species for fibre traits identified in this study will improve our understanding of fibre genes that have key role in fibre development. Transcriptomic analysis simplifies the breeding through expression profiling of highly expressed genes. Transcriptomic analysis was performed for the identification of differentially expressed genes at different fibre growth stages in interspecific lines and three *Gossypium* species. Our study predicts expression analysis of selected fibre genes during 0, 5, 10, 15 and 20 DPA fibre stages. High level variable regulation of genes encoding for fibre development was observed at different stages. Transcriptomic profiling has been effectively used for gene identification in cotton crop [27–31]. Here, we describe transcriptome profiling of genes in cotton fibre through quantitative Real Time PCR.

This is the initial comprehensive expression profiling that identified the differentially expressed genes with different stages contributing to fibre development in contrasting interspecific lines of cotton. Real Time PCR results predicted high expression levels specifically in the interspecific lines SL-19 (long staple line) as compared to parent species (Fig. 1-3) envisaging that when genome of two different species merge with each other, its progenitors possess more DNA content, which can be associated with fibre elongation and amplified size of single-celled fibres. It was also concluded that transgressive segregates are possible with hybrid vigor because of different genome groups of *Gossypium*, which make it possible to get interspecific lines with good fibre length, fibre strength and fibre fineness [32–35].

Expression profiling was compared with RNA sequence data submitted in different bio projects on FGD (Fig. 5). In *Expansin A 4*, our results were according to PRJNA490626 project in which transcripts were detected in 5 experiments including fibre development at various stages (0-25 DPA). Maximum expression was at 10 DPA which was similar to our results. GhEXPA4a and GhEXPA4b are specific fibre related genes that exhibited high expression during the fibre

initiation and elongation stages (0 to 15 DPA). Over-expression of *GhEXPA8* predicted that these genes have ability to improve the fibre length and fineness in cotton crop [6]. Expansin proteins indorse the spillage between different microfibrils by Hemicellulose and cellulose cleavage [36]. Moreover, our data also suggested that Expansin protein has essential role in cotton fibre development by enlargement of fibre cells through sliding apart cellulose micro fibrils. Expression levels for *E6-like* genes was also compared. *E6-like* gene has similarity with genes Gh-D05G160200 for fibre related gene. It also plays its role fibre development. *E6* gene was firstly recognized as fibre gene with high expression during cotton fibre development and similar *E6*-like was predicted in Angiosperms [13].

BURP Domain proteins are known as important proteins that has significant roles in plant growth and stress responses [37, 38]. Number of BURP proteins have been recognized and characterized on the basis of sequences features. However, different members from different subfamilies predicted variable expression patterns. In our findings, *BURP Domain RD22-like* genes actually execute main function in fibre elongation and maturation. Although low copy number of TPM of *BURP Domain RD22-like*gene were observed but this has a role in fibre development. The cotton fibre related gene (AtRD-22-Like) with over expression in elongating fibre cells, translates a BURP Domain-containing protein [9]. Cotton plants with high expression of GhRDL1 and *GhEXPA1* give more number of bolls, resulting up to 40% more lint yield plant^{-1} without disturbing fibre quality and non-reproductive growth. [9].

It is further concluded from the study that there is a direct association between *Expansin A4*, *E6-like*, *BURP Domain protein RD22-like* and fibre quality traits. Thus, these are key target for improving the fibre characteristics. Transformation of these highly expressed genes in local cotton varieties can fulfill the mechanized textile industry requirements. Moreover, genetically modified cotton produced by over expression of these genes will be the best source for use as a long staple variety or use as a parent in breeding program.

Biological sequences comparison in molecular biology and bioinformatics has been an imperative approach to supports analysis, such as prediction of protein sub-cellular localization [39], Physio chemical properties [40] and the field of taxonomy [41]. *E6-like* was characterized as unstable as value of instability index was 47.75. A protein whose instability index is less than 40 is expected as stable while a value greater than 40 indicates that the protein may be unstable. Similarly, *Expansin A4* was characterized as a stable protein with value of instability index of 29.01. An imperative step on this mode is prediction of subcellular localization of each protein. *E6-like*, *Expansin A4* and *BURP Domain RD22-like* were characterized as a membrane soluble protein family. *In silico* analysis also confirm the role of genes in fibre elongation, *Expansin*-A4, *BURP Domain protein RD22-like-like*and *E6-like* play its main role in rapid elongation and also with predominantly effect in transition stage of elongation supporting to secondary cell wall synthesis.

DNA sequence alignment is a criterion for almost all comparative genomic analyses, including documentation of well-preserved sequence motifs and investigation of genes and species historical relationships [42]. *E6-like*, *Expansin A4* and *BRUP Domain RD22-like* PCR amplified full length gene was sequenced and subjected to BLAST analysis followed by multiple sequence alignment of DNA sequence and protein sequence for similarities and differences of interspecific lines and parent species (Fig. 5). It was concluded from the sequence comparison of interspecific lines and species of cotton that tri-species introgression lines are more closely related to *Gossypium hirsutum* as compared to *Gossypium arboreum* and *Gossypium anomalum* depicted. This confirms its back crossing with *Gossypium hirsutum* for yield improvement. These interspecific lines were also originate from BC_4S_5 population {*G. hirsutum* × 2(*G. arboreum* × *G. anomalum*) developed at Cytogenetics Section, CCRI, Multan [22]. In interspecific hybrids of *Gossypium*, a greater proportion of female gametes than male gametes is generally useful with few exceptions [43], hence backcross breeding should be subjugated. Review of backcrossing with distinct reference to cotton traits improvement exhibited

that during repeated backcrossing one set of chromosomes retained with genes balanced. This technique has been used successfully in crosses of different *Gossypium* species [44–46].

In silico analysis tries to find proteins with consistent annotations about their interaction and functions in the cellular machinery. An imperative step on this mode is prediction of subcellular localization of each protein. *E6*-like, *Expansin A4* and *BURP Domain RD22-like* were characterized as a membrane soluble protein family. In *E6-like*, *Expansin-A4* and *BURP Domain RD22-like* were characterizes as extracellular membrane that's why signal peptide was present in protein coding. As validation of specific genes for crop improvement programs is also becoming popular engendering novel properties [47–49].Promoter regions *In silico* analysis of fibre related gene could be used to predict gene expression profiles in cotton plant. Many stresses resistant, light responsive which can contribute for fibre development were present in *E6-like*, *Expansin A4* and *Burp Domain RD22-like* (Fig. 6 and Table 6). To explore the molecular mechanisms regulating cotton fibre development, promoters of several cotton fibre genes have been identified. *E6* was the first of such genes to be reported, and the *E6* promoter has been used for engineering cotton fibre quality [50]. *GhRDL1*, a gene highly expressed in cotton fibre cells at the elongation stage, encodes a BURP domain-containing protein [51], and the *GaRDL1* promoter exhibited a trichome-specific activity in transgenic *Arabidopsis* plants [52]. The aim of our analysis was to predict promoter and regulatory elements of genes encoding useful stress responsive leading to fibre production. In cotton, basic information related to different cis acting elements was generated to support the effort of improving cotton plant for a stress resistant with more fibre production.

Conclusion

The SL-19 appeared to be a promising source for cotton quality improvement with maximum expression for all fibre genes. To address the negative correlation between yield and fibre quality, use of genetic engineering is recommended to break this linkage by transferring *E6-like*, *Expansin A4* and *BURP Domain RD22-like* genes in local cotton cultivars.

Declarations

Acknowledgements

Authors are the whole cotton group working at MNS University of Agriculture, Multan and Central Cotton Research Institute, Multan, for providing technical support and germplasm for this study.

Author contributions

AAK, NI and ZK designed the research plan. FA carried out the experiments and drafted the manuscript. ZM and AG supported in experimentation and manuscript review and improvement. ZK and NI helped in data analysis. AAK, NI and ZK reviewed the final manuscript.

Funding

Funding for this study was provided by Pakistan Science Foundation (PSF), Islamabad under PSF-NSFC-IV/Agr/P-MNSUAM (30). The authors are thankful to PSF for supporting this research work.

Declarations

Conflict of Interests

The authors declare that they have no conflict of interests.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Sandin G, Peters GM (2018) Environmental impact of textile reuse and recycling–A review. J Clean Prod 184:353–365
2. Stelly DM, Lacape J-M, Dessauw D, Kohel RJ, Mergeai G et al (2008) International genetic, cytogenetic and germplasm resources for cotton genomics and genetic improvement; Omnipress
3. Campbell B, Saha S, Percy R, Frelichowski J, Jenkins JN et al (2010) Status of the global cotton germplasm resources. Embrapa Recursos Genéticos e Biotecnologia-Artigo em periódico indexado (ALICE)
4. Lubbers EL, Chee PW (2009) The worldwide gene pool of G. hirsutum and its improvement. Genetics and genomics of cotton: Springer. pp. 23-52
5. Percy RG (2009) The worldwide gene pool of Gossypium barbadense L. and its improvement. Genetics and genomics of cotton: Springer. pp. 53-68
6. Lv L-M, Zuo D-Y, Wang X-F, Cheng H-L, Zhang Y-P et al (2020) Genome-wide identification of the expansin gene family reveals that expansin genes are involved in fibre cell growth in cotton. BMC Plant Biol 20:1–13
7. Orford SJ, Timmis JN (2000) Expression of a lipid transfer protein gene family during cotton fibre development. Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids. 1483:275–284
8. Ruan Y-L, Llewellyn DJ, Furbank RT (2001) The control of single-celled cotton fiber elongation by developmentally reversible gating of plasmodesmata and coordinated expression of sucrose and K+ transporters and expansin. Plant Cell 13:47–60
9. Xu B, Gou J-Y, Li F-G, Shangguan X-X, Zhao B et al (2013) A cotton BURP domain protein interacts with α-expansin and their co-expression promotes plant growth and fruit production. Mol Plant 6:945–958
10. Xu H, Li Y, Yan Y, Wang K, Gao Y et al (2010) Genome-scale identification of soybean BURP domain-containing genes and their expression under stress treatments. BMC Plant Biol 10:1–16
11. Park J, Cui Y, Kang B-H (2015) AtPGL3 is an Arabidopsis BURP domain protein that is localized to the cell wall and promotes cell enlargement. Front Plant Sci 6:412
12. John ME, Crow LJ (1992) Gene expression in cotton (Gossypium hirsutum L.) fiber: cloning of the mRNAs. Proceedings of the National Academy of Sciences 89: 5769-5773
13. Doucet J, Truong C, Frank-Webb E, Lee HK, Daneva A et al (2019) Identification of a role for an E6-like 1 gene in early pollen–stigma interactions in Arabidopsis thaliana. Plant Reprod 32:307–322
14. Ruan Y-L, Llewellyn DJ, Furbank RT (2003) Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. Plant Cell 15:952–964
15. Qin Y-M, Zhu Y-X (2011) How cotton fibers elongate: a tale of linear cell-growth mode. Curr Opin Plant Biol 14:106–111
16. Yuan JS, Reed A, Chen F, Stewart CN (2006) Statistical analysis of real-time PCR data. BMC Bioinformatics 7:1–12
17. Pan Y, Meng F, Wang X (2020) Sequencing multiple cotton genomes reveals complex structures and lays foundation for breeding. Front Plant Sci 11:1377
18. Ogbe RJ, Ochalefu DO, Olaniru OB (2016) Bioinformatics advances in genomics-A review. Int J Curr Res Rev 8:05–11

19. Zhou B, Chen S, Shen X, Zhang X, Zhang Z (2003) Construction of gene pools with superior fiber properties in Upland cotton through interspecific hybridization between Gossypium hirsutum and Gossypium wild species. Acta Agron Sinica 29:514–519
20. Zhenglan L, Ruqin J, Wennan Z, Jianxing H, Chuanwei S et al (2002) Creation of the technique of interspecific hybridization for breeding in cotton. Sci China Ser C: Life Sci 45:331–336
21. Anjum ZI, Azhar MT, Hayat K, Ashraf F, Shahzad U et al (2014) Development of high yielding and CLCuV resistant ulpand cotton variety “CIM-608”. Pakistan J Phytopathol 26:25–34
22. Anjum Z, Hayat K, Celik S, Azhar M, Shehzad U et al (2015) Development of cotton leaf curls virus tolerance varieties through interspecific hybridization. Afr J Agric Res 10:1612–1618
23. Pang C, Du X, Ma Z (2005) The progress of enhancement and utilization of upland cotton elite germplasm with wild cotton genes. Cotton Sci (in Chinese) 17:171–177
24. Logemann J, Schell J, Willmitzer L (1987) Improved method for the isolation of RNA from plant tissues. Anal Biochem 163:16–20
25. Dolferus R, Jacobs M, Peacock WJ, Dennis ES (1994) Differential interactions of promoter elements in stress responses of the Arabidopsis Adh gene. Plant Physiol 105:1075–1087
26. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA et al (2007) Clustal W and Clustal X version 2.0. bioinformatics 23: 2947-2948
27. Arpat A, Waugh M, Sullivan JP, Gonzales M, Frisch D et al (2004) Functional genomics of cell elongation in developing cotton fibers. Plant Mol Biol 54:911–929
28. Wilkins TA, Arpat AB (2005) The cotton fiber transcriptome. Physiol Plant 124:295–300
29. Shi Y-H, Zhu S-W, Mao X-Z, Feng J-X, Qin Y-M et al (2006) Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. Plant Cell 18:651–664
30. Gou J-Y, Wang L-J, Chen S-P, Hu W-L, Chen X-Y (2007) Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. Cell Res 17:422–434
31. Wang H, Guo Y, Lv F, Zhu H, Wu S et al (2010) The essential role of GhPEL gene, encoding a pectate lyase, in cell wall loosening by depolymerization of the de-esterified pectin during fiber elongation in cotton. Plant Mol Biol 72:397–406
32. Van't Hof J (1999) Increased nuclear DNA content in developing cotton fiber cells. Am J Bot 86:776–779
33. Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78:139
34. Senchina DS, Alvarez I, Cronn RC, Liu B, Rong J et al (2003) Rate variation among nuclear genes and the age of polyploidy in Gossypium. Mol Biol Evol 20:633–643
35. Soltis PS, Marchant DB, Van de Peer Y, Soltis DE (2015) Polyploidy and genome evolution in plants. Curr Opin Genet Dev 35:119–125
36. McCann MC, Knox JP (2018) Plant cell wall biology: polysaccharides in architectural and developmental contexts.Annual Plant Reviews online:343–366
37. Yamaguchi-Shinozaki K, Shinozaki K (1993) The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of rd22, a gene responsive to dehydration stress in Arabidopsis thaliana. Mol Gen Genet MGG 238:17–25
38. Phillips K, Ludidi N (2017) Drought and exogenous abscisic acid alter hydrogen peroxide accumulation and differentially regulate the expression of two maize RD22-like genes. Sci Rep 7:1–12
39. Zhao Y, Li X, Qi Z (2014) Novel 2D graphic representation of protein sequence and its application. J Fiber Bioeng Inf 7:23–33

40. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD et al (2005) Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook: 571-607
41. Huang D-S, Yu H-J (2013) Normalized feature vectors: a novel alignment-free sequence comparison method based on the numbers of adjacent amino acids. IEEE/ACM Trans Comput Biol Bioinf 10:457–467
42. Kumar S, Filipski A (2007) Multiple sequence alignment: in pursuit of homologous DNA positions. Genome Res 17:127–135
43. Harland SC, Atteck OM (1941) The genetics of cotton. XVIII. Transference of genes from diploid North American wild cottons (Gossypium thurberi Tod., G. armourianum Kearney, G. aridum comb. nov. Skovsted) to tetraploid New World cottons (G. barbadense L. and G. hirsutum L.). J Genet 42:1–19
44. Deodikar G (1949) Cytogenetic studies on crosses of G. anomalum with cultivated cottons. I (G. hirsutum× G. anomalum) doubled× G. hirsutum. Indian J Agric Sci 19:389–399
45. Marappan P, Santhanam V (1962) Breeding behaviour of some arboreum-anomalum backcrosses. Indian Cot Gr Rev 16:24–30
46. Mehetre SS (2010) Wild Gossypium anomalum: a unique source of fibre fineness and strength.Current Science:58–71
47. Lata C, Prasad M (2011) Role of DREBs in regulation of abiotic stress responses in plants. J Exp Bot 62:4731–4748
48. Puranik S, Sahu PP, Srivastava PS, Prasad M (2012) NAC proteins: regulation and role in stress tolerance. Trends Plant Sci 17:369–381
49. Singh RK, Deshmukh R, Muthamilarasan M, Rani R, Prasad M (2020) Versatile roles of aquaporin in physiological processes and stress tolerance in plants. Plant Physiol Biochem 149:178–189
50. John M (1996) Metabolic pathway engineering in cotton: Biosynthesis of polyester in fiber;
51. Li C-H, Zhu Y-Q, Meng Y-L, Wang J-W, Xu K-X et al (2002) Isolation of genes preferentially expressed in cotton fibers by cDNA filter arrays and RT-PCR. Plant Sci 163:1113–1120
52. Wang E, Hall JT, Wagner GJ (2004) Transgenic Nicotiana tabacum L. with enhanced trichome exudate cembratrieneols has reduced aphid infestation in the field. Mol Breeding 13:49–57

Figures

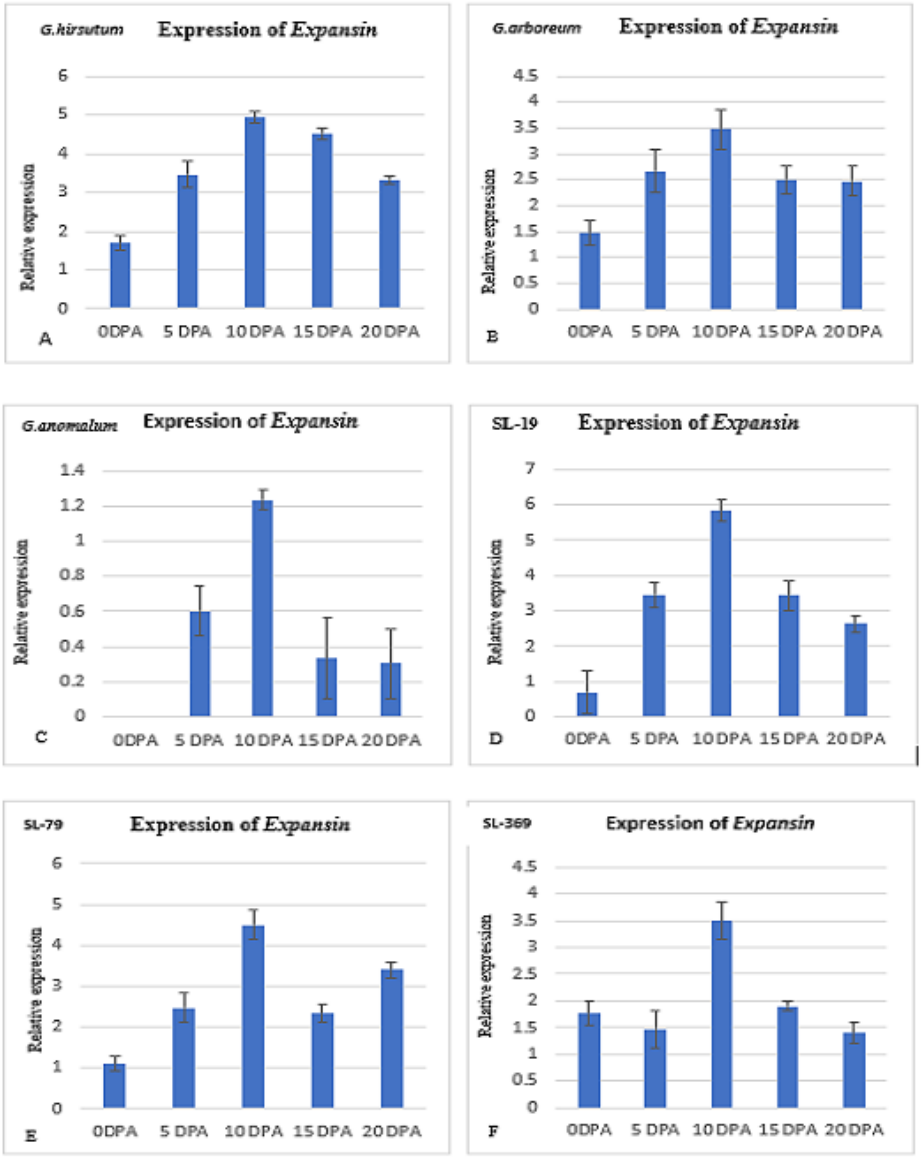

Figure 1

Expression profiling of *Expansin A4* in *Gosssypium* species and interspecific lines: **A** (Expression of *Expansin A4* in *G. arboreium*), **B** (Expression of *Expansin A4* in *G. hirsutum*), **C** (Expression of *Expansin A4* in *G. anomalum*), **D** (Expression of *Expansin A4* in SL-19) **E** (Expression of *Expansin A4* in SL-79), **F** (Expression of *Expansin A4* in SL-369).

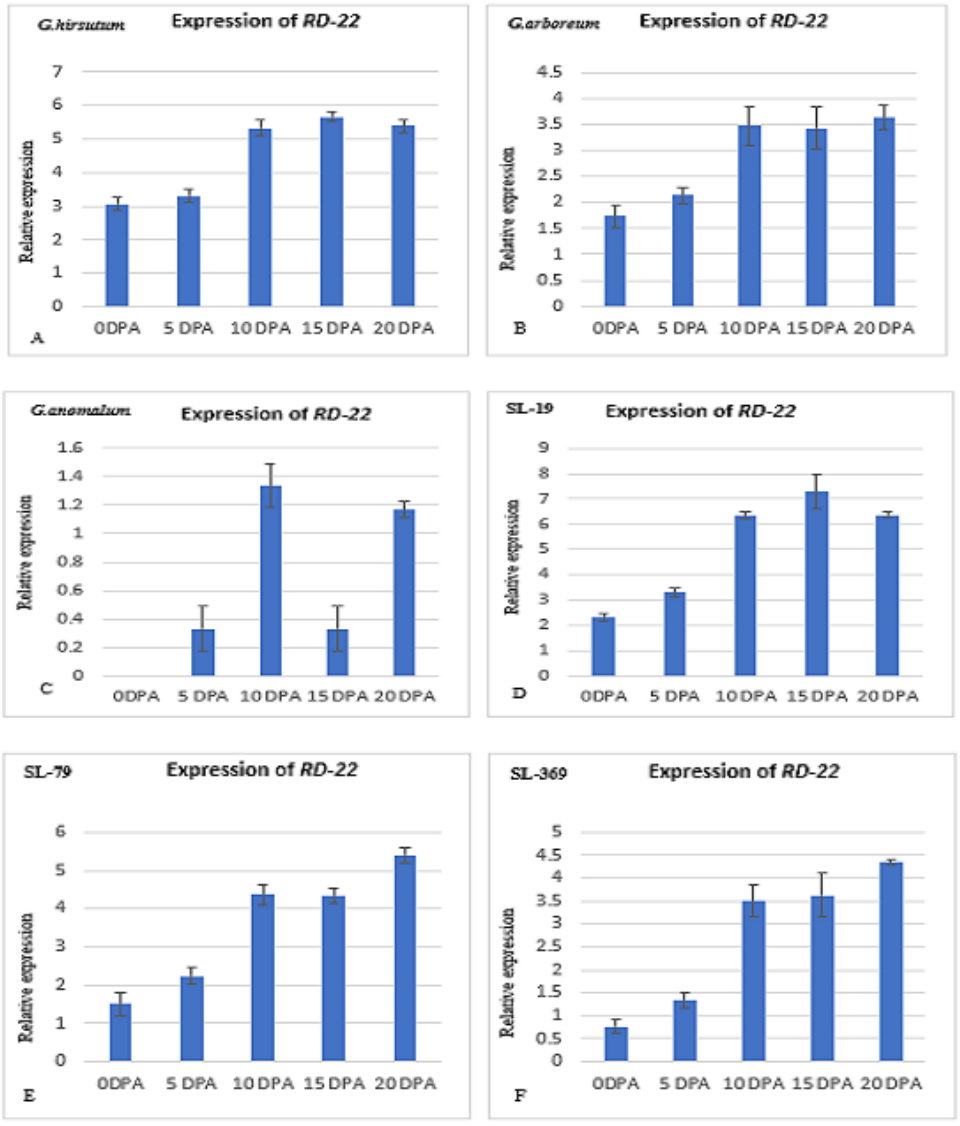

Figure 2

Expression profiling of *BURP Domain RD-22* in *Gosssypium* species and interspecific lines: **A** (Expression of *RD-22* in *G. arboreium*), **B** (Expression of *RD-22* in *G. hirsutum*), **C** (Expression of *RD-22* in *G. anomalum*), **D** (Expression of *RD-22* in SL-19) **E** (Expression of *RD-22* in SL-79), **F** (Expression of *RD-22* in SL-369).

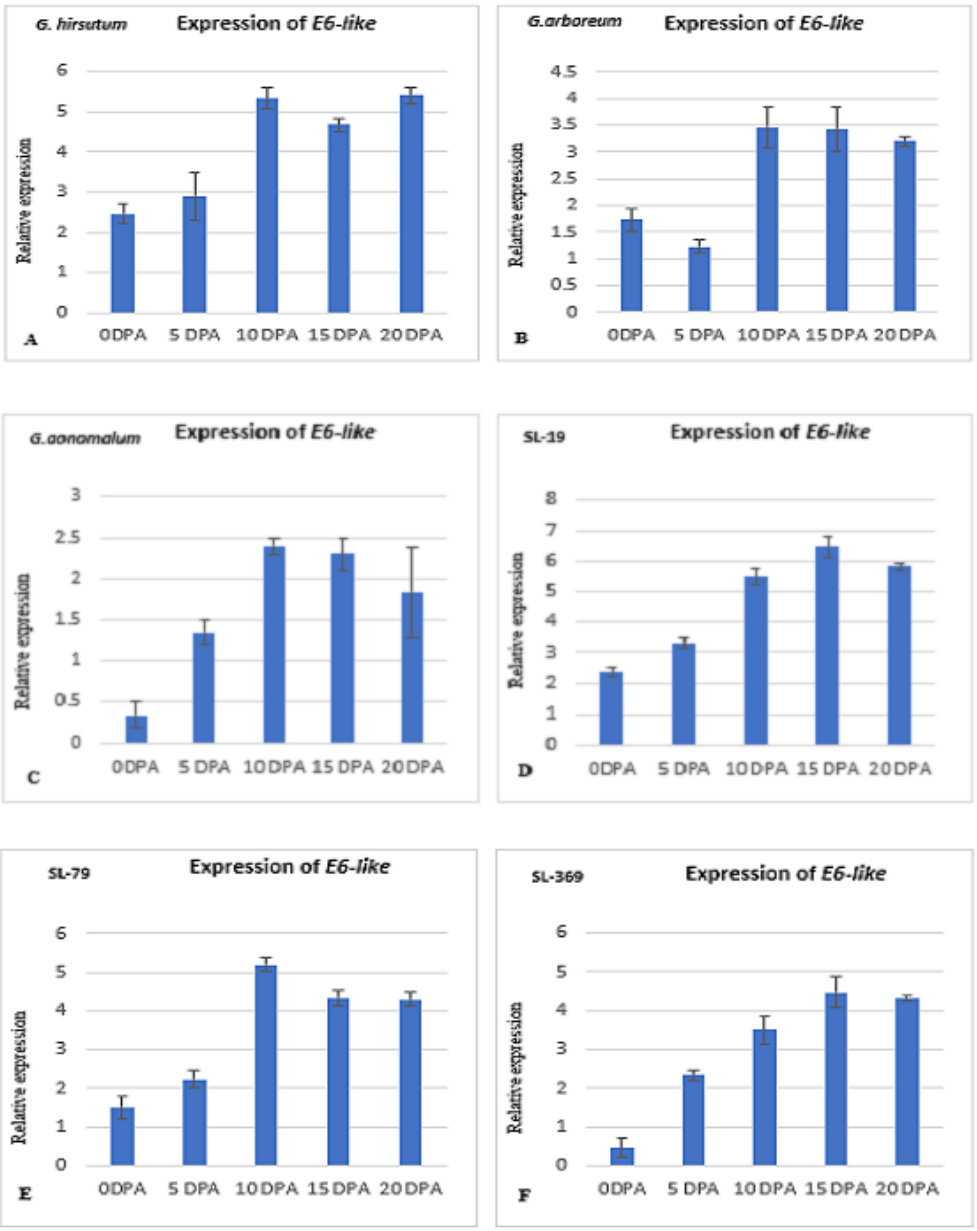

Figure 3

Expression profiling of *E6-like* in *Gosssypium* species and interspecific lines: **A** (Expression of *E6-like* in *G. arboreium*), **B** (Expression of *E6-like* in *G. hirsutum*), **C** (Expression of *E6-like* in *G. anomalum*), **D** (Expression of *E6-like* in SL-19) **E** (Expression of *E6-like* in SL-79), **F** (Expression of *E6-like* in SL-369).

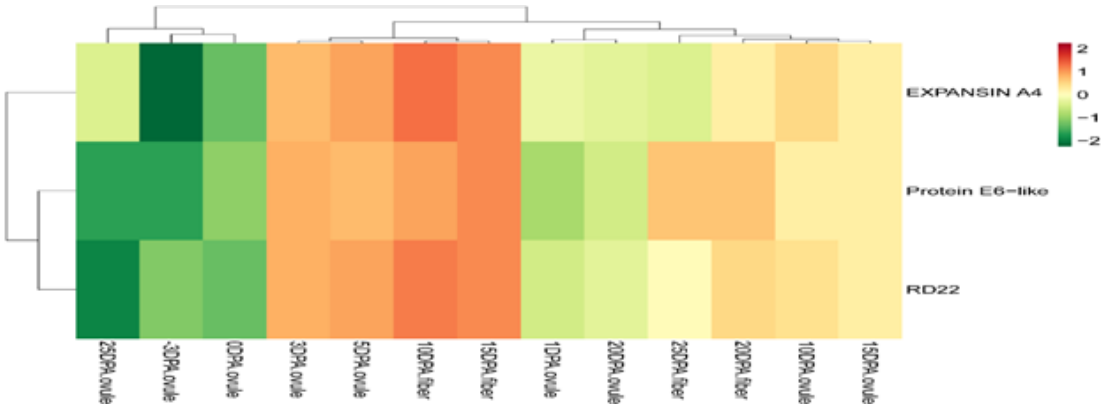

Figure 4

Heat map of expression levels (log-transformed transcript per kilobase million (TPM) values). Figure was generated based on available RNA-seq data of *BURP Domain RD2-like-2*, *Expansin* and *E6-like* submitted bio projects from cotton FGD data base. Red indicates high expression, yellow indicates intermediate expression and green indicates no expression. It is straightforward to identify highly expressed genes in specific tissues from this figure. Tissues are labeled with Days After post anthesis (DPA). Rows indicates the fibre gens and column show the fibre stages (250DPA ovule -25 DPA fibre). The data denotes the logarithm-transformed values of log2, day post anthesis and fragments per kilobase of transcript per million mapped reads.

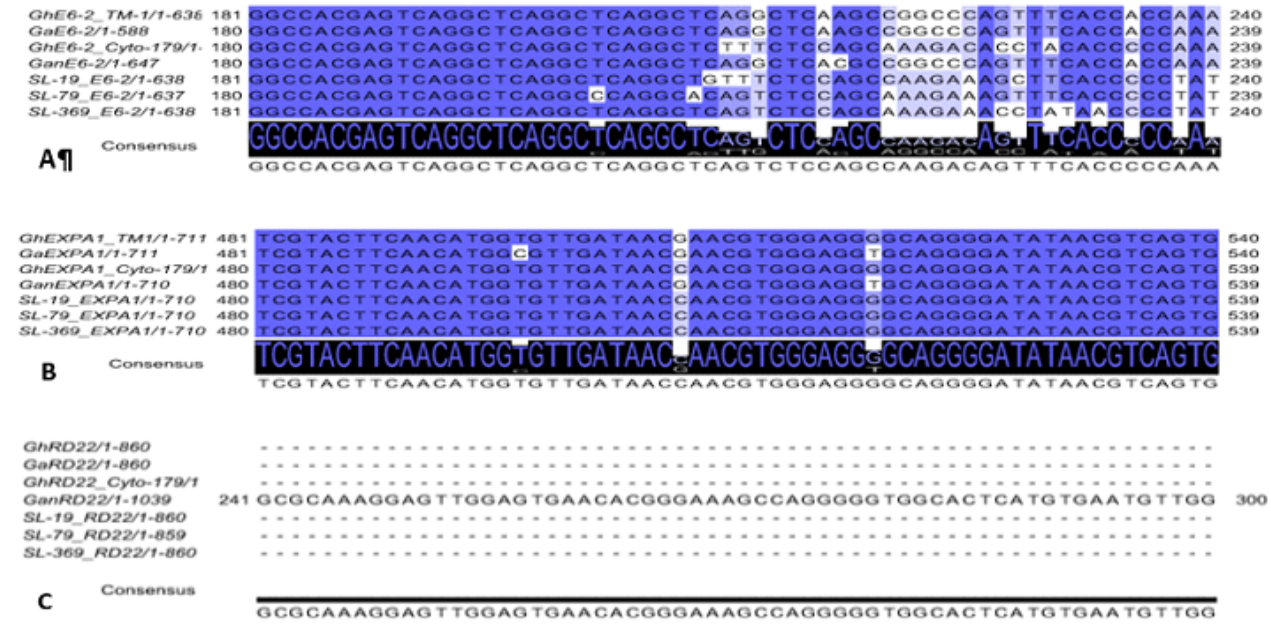

Figure 5

DNA sequence alignments of fibre genes. A (*E6-like*), B (*Expansin A4*), C (*BURP Domain RD22-like*). White shadings indicate the polymorphic nucleotides. Interspecific lines and *Gossypium* species names are indicated in the left and number of bases depicted in each line is marked by the number shown at the top right of each section.

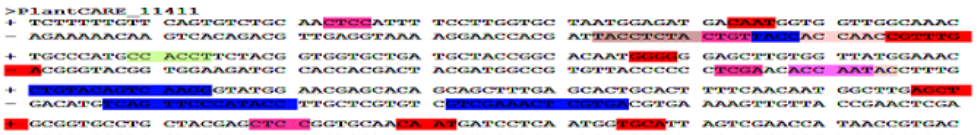

Figure 6

In silico analysis of promoter sequences of fibre genes. **A** (cis-acting regulatory elements in *E6-like*)-like, **B** (cis-acting regulatory elements in *Expansin A4*), **C** (cis-acting regulatory elements in *BURP Domain RD-22*) Highlighted regions show cis regulatory motifs present in the promoter regions with specific function.