

RESEARCH

Open Access



Genome-wide identification and expression profiling of photosystem II (*PsbX*) gene family in upland cotton (*Gossypium hirsutum* L)

RAZA Irum^{1†}, PARVEEN Abida^{2†}, AHMAD Adeel¹, HU Daowu¹, PAN Zhaoe¹, ALI Imran^{3*} and DU Xiongming^{1*} 

Abstract

Background Photosystem II (*PSII*) constitutes an intricate assembly of protein pigments, featuring extrinsic and intrinsic polypeptides within the photosynthetic membrane. The low-molecular-weight transmembrane protein *PsbX* has been identified in *PSII*, which is associated with the oxygen-evolving complex. The expression of *PsbX* gene protein is regulated by light. *PsbX*'s central role involves the regulation of *PSII*, facilitating the binding of quinone molecules to the Qb (*PsbA*) site, and it additionally plays a crucial role in optimizing the efficiency of photosynthesis. Despite these insights, a comprehensive understanding of the *PsbX* gene's functions has remained elusive.

Results In this study, we identified ten *PsbX* genes in *Gossypium hirsutum* L. The phylogenetic analysis results showed that 40 genes from nine species were classified into one clade. The resulting sequence logos exhibited substantial conservation across the N and C terminals at multiple sites among all *Gossypium* species. Furthermore, the orthologous/paralogous, *Ka/Ks* ratio revealed that cotton *PsbX* genes subjected to positive as well as purifying selection pressure might lead to limited divergence, which resulted in the whole genome and segmental duplication. The expression patterns of *GhPsbX* genes exhibited variations across specific tissues, as indicated by the analysis. Moreover, the expression of *GhPsbX* genes could potentially be regulated in response to salt, intense light, and drought stresses. Therefore, *GhPsbX* genes may play a significant role in the modulation of photosynthesis under adverse abiotic conditions.

Conclusion We examined the structure and function of *PsbX* gene family very first by using comparative genomics and systems biology approaches in cotton. It seems that *PsbX* gene family plays a vital role during the growth and development of cotton under stress conditions. Collectively, the results of this study provide basic information to unveil the molecular and physiological function of *PsbX* genes of cotton plants.

Keywords Photosystem, Phylogenetic, Synteny, RNA seq, Gene expression, Orthologous

[†]Raza I and Parveen A contributed equally to this work.

*Correspondence:

Ali Imran
drimranali@kust.edu.pk
Du Xiongming
dxm630723@163.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Cotton serves as a vital cash crop, primarily cultivated for its fibers, while also offering the potential for extracting edible oil from its seeds (Raza et al. 2021; Zafar et al. 2002a; Chen et al. 2007). The upland cotton (*Gossypium hirsutum* L.) dominates global cotton production, constituting approximately 95% of the total output due to its superior fiber quality, high yield, and wide adaptability traits (Zhang et al. 2008; Song et al. 2018). The prime objective of the cotton breeding programs is to enhance both fiber quality and yield simultaneously (Zahid et al. 2016). Cotton, a multi-purpose crop is highly vulnerable to biotic and abiotic stresses (Zafar et al. 2002b). A significant reduction has been witnessed in the production of cotton due to abiotic stresses such as salinity and drought (Zafar et al. 2002b, c). In the midst of the twenty-first century, the demand for food is on the rise, driven by a rapid increase in the global population (Beddington et al. 2012). However, the challenges posed by climate change, notably elevated temperatures (Zafar et al. 2022a), coupled with the severity of abiotic stressors, disrupted the growth and development of crops, led to a significant decline in cotton yields (Nouri et al. 2015; Sasi et al. 2018).

Furthermore, plants, being sessile in nature, are more exposed to the adverse effects of abiotic stresses on both growth and development as opposed to other organisms (He et al. 2018; Magwanga et al. 2018; Xu et al. 2019). Major abiotic stresses such as heat, salinity, and drought result in the over-reduction of the electron transport chain (ETC), leading to photo-oxidation (Nishiyama and Murata 2014). Globally, agricultural productivity has faced a decline due to the combined impact of drought and salinity (Zhu 2001; Dong 2012).

Moreover, salinity, intense sunlight, drought, and heat stress have the combined effect of diminishing the CO₂ assimilation rate, causing an increase in reactive oxygen species (ROS) (Pintó-Marijuan and Munné-Bosch 2014). Previous studies have shown that abiotic stresses account for a 50% loss in yield (Nath et al. 2013). Similarly, the reduced activity of photosynthesis is a major contributor to significant yield losses in crops (Nouri et al. 2015). Drought and salinity disrupt plant growth by impacting various biochemical and physiological attributes, such as chlorophyll production, photosynthetic rate, nutrient metabolism, ion uptake and translocation, carbohydrate metabolism, and respiration (Hussain et al. 2018). Plant genes affecting the photosynthesis process, including ribulose biphosphate carboxylase large chain (RBCL) (Berry et al. 2013) light-harvesting chlorophyll a/b-binding protein (LHC) (Zhao et al. 2020), and cytochrome P450, have been identified (Magwanga et al. 2019). Furthermore, genes encoding structural proteins involved

in photosystem II (*PsbA*, *PsbE*, *PsbF*, *PsbH*, *PsaN*, and *PsbX*) have been identified (Zhang et al. 2020). Plastocyanin, an enzyme integral to the photosystem II (*PSII*), oxidizes water in the presence of light. A small hydrophobic protein named 'PSII-X (*PsbX*)' having a molecular weight of 4.1 kDa is generally found in the core complex of *PSII* in both plants and cyanobacteria (Shi et al. 1999; Katoh and Ikeuchi 2001). The *PsbX* protein serves as a key regulator of the *PSII* and aids in binding the quinone molecules with the Q_b (*PsbA*) site. Moreover, *PsbX* also regulates the efficiency of photosynthesis as well as biomass accumulation. While several aspects of *PSII* structure remain unknown, the function of the *PsbX* gene family under abiotic stresses in cotton has yet to be elucidated. However, the sequencing of whole genome information of *G. raimondii* (Wang et al. 2012), *G. arboreum* (Huang et al. 2020), and *G. hirsutum* (Hu et al. 2019) has provided the framework to conduct the functional analysis of protein related to *PsbX* genes in three cotton species.

In the present study, we performed an *in-silico* characterization of *PsbX* gene family members in *G. arboreum*, *G. hirsutum*, *G. barbadense*, and *G. raimondii* to unveil the functional significance of *PsbX* gene family. Subsequently, we conducted a comprehensive array of bioinformatics analyses, encompassing multi-sequence alignment, gene loci examination, analysis of gene structures, scrutiny of promoter *cis*-elements, examination of conserved protein motifs, assessment of phylogenetic relationships, and exploration of gene expression profiles. The outcomes of our study may offer valuable insights for a more in-depth characterization of the *PsbX* gene in cotton. Additionally, our investigation provides an avenue for understanding the molecular foundation of the regulatory effects of *PsbX* across diverse developmental stages and sheds light on the response mechanisms of cotton plants under stress conditions.

Material and methods

Identification of the *PsbX* gene family in *Gossypium* Spp.

The genomes of *G. hirsutum* v1.1 and *G. barbadense* were obtained from the Group of Cotton Genetic Improvement (GCGI), Huazhong Agricultural University (HZAU) (<http://cotton.hzau.edu.cn>). Sequencing data of *Gossypium raimondii* was obtained from ftp://ftp.bioinfo.wsu.edu/species/Gossypium_raimondii/BYU-Improved_G.raimondii and *G. arboreum* was downloaded from Cottonfgd (www.cottonfgd.org). Genomes of *Arabidopsis thaliana* Araport 11, *Populus trichocarpa*, *O. sativa*, and *Vitis Vinifera* were obtained from Phytozome v13. The genome sequence data of *Theobroma cacao* v2 was downloaded from Ensemble Plants (<https://plants.ensembl.org>) for computational analysis. The Hidden

Markov Model profile for *PsbX* (PF06596) was downloaded from PFAM (<http://pfam.xfam.org>) (Finn et al. 2016). HMMER 3.2.1 was used to identify *PsbX* protein sequences against all the protein sequences of the above genomes. Protein sequences with an E-value less than $1e^{-5}$ were selected for further analysis. In addition, to approve the protein sequences obtained from cotton and other genomes *PsbX* genes, candidate protein sequences were confirmed by using the CD-search (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>).

Sequence alignment and phylogenetic tree construction

All the protein sequences of *PsbX* confirmed by SMART (<http://smart.embl-heidelberg.de/>) (Schultz et al. 2000) and The Conserved Domain Database (CDD) were aligned by MUSCLE alignment with the Neighbor-joining method. The alignment results were used in phylogenetic and evolutionary analyses and imported into MEGA X (<https://www.megasoftware.net/>). The phylogenetic tree was built using the Neighbor-joining (NJ) method with 1 000 bootstrap values and pairwise deletion. The phylogenetic tree was visualized with Figtree.

Mapping genes *PsbX* on the chromosome, gene structure, and motif analysis

All members of *PsbX* were mapped onto *G. hirsutum* chromosomes based on their physical positions. The gene mapping and gene structures of *G. hirsutum* were constructed with TBtools using a GTF file. Conserved protein motifs were discovered by inputting protein sequences of *G. raimondii*, *G. arboreum*, *G. hirsutum*, and *G. barbadense* *PsbX* to the Meme-Suit program (<http://meme-suite.org/>) (Bailey et al. 2009). The following parameters were used: motif width was set to 6–100 and motif site distribution was set to zero or one site per sequence.

Promoter region *cis*-acting element analysis

The promoter sequences (2 kb upstream of the translation start site) of all *PsbX* genes were analyzed to predict and locate their *cis*-elements from the online tool PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Lescot 2002).

Physicochemical features and subcellular location prediction

The physicochemical properties were calculated using the ExPASy ProtParam tool (<http://us.expasy.org/tools/protparam.html>) for *PsbX* genes including the number of amino acids (protein length), pI (isoelectric point) and MW (molecular weight). Furthermore, an online tool WolfPSORT (<https://wolfsort.hgc.jp>) was used to predict the subcellular localization of *PsbX* genes in cotton.

Ka/Ks analysis

The divergence of genes was estimated by analysis of *Ka/Ks* values. *Ka/Ks* values were calculated using a Galaxy wrapper for *Ka/Ks* calculations (<http://usegalaxy.org>).

Transcriptional profile analysis

Online available transcriptome data of *G. hirsutum* L. for abiotic stress and tissue specific expression was downloaded from NCBI SRA database. The project names are PRJNA490626 and PRJNA532694. Methods from Perteau et al (2016) were adopted to analyze the RNA-seq data.

Plant stress experiment, RNA extraction and qRT-PCR analysis

The seeds of *G. hirsutum* L. (cultivar TM-1) were collected from Institute of Cotton Research, Chinese Academy of Agricultural Sciences (CAAS, Anyang, China). The collected seeds were grown in the growth room. When the plants grew to three true leaf stages, stress was applied. For drought treatment, 20% PEG solution and for NaCl treatment, 250 mmol·L⁻¹ NaCl solution was applied to the pots. Samples were taken from the leaves. Time for sampling was 0 h, 2 h, 6 h, 12 h, 24 h, 48 h, and 72 h after treatment. The collected leaf samples were stored in liquid nitrogen and subsequently stored at -80 °C. TIANGEN, RNAprep Pure Plant plus Kit (www.tiangen.com) was used for RNA extraction following manufacturer instructions.

NanoDrop 2000 was used to check the quality and concentration of RNA extracted with a standard wavelength of 260/280 nm, which must be between 1.80~2.10 and cDNA was synthesized from 500 ng total RNA using the Prime-Script[®] RT reagent kit (Takara, Dalian, China). We selected ten *PsbX* genes from the total identified *PsbX* genes for RT-qPCR. The primers were designed (Additional file 1: Table S1) by using NCBI PRIMER BLAST website (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi>).

Applied Biosystems 7500 Real-Time PCR system (Applied Biosystems, Foster City, CA, USA) was employed for qRT-PCR. The *G. hirsutum* constitutive *β-actin* gene was kept as a reference gene and *PsbX* gene primers were applied for RT-qPCR. The relative expression levels of *PsbX* gene after infiltration were calculated using the 2^{-ΔΔCT} method (Livak and Schmittgen 2001).

Results

Genome wide identification of *PsbX* gene family

Various bioinformatics analyses were conducted to explore the *PsbX* gene family. Comparative analysis of all cotton genomes was performed using the Hidden Markov Model (HMM) and the Local Blast Research tool, considering E-values ≤ 10⁻¹⁰ for the *PsbX* gene family.

Elimination of redundant *PsbX* gene sequences was carried out through the application of multiple sequence alignment tools. Furthermore, the identification of protein domains in cotton *PsbX* was achieved by utilizing various databases, including PROSITE and Pfam (<http://pfam.xfam.org/>), InterProscan 63.0 (<http://www.ebi.ac.uk/interpro/>) and CD-search (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>). The elimination of redundant protein sequences resulted in the identification of 40 genes belonging to the *PsbX* family across nine distinct plant species. The number of *PsbX* genes in different plant species is as follows: 1 in *A. thaliana*, 2 in *V. vinifera*, 2 in *O. sativa*, 2 in *T. cacao*, 3 in *P. trichocarpa*, 5 in *G. arboreum*, 10 in *G. barbadense*, 10 in *G. hirsutum*, and 5 in *G. raimondii*. Additional details are provided

in Additional file 2: Table S2. To avoid ambiguity and duplication, *PsbX* genes identified in cotton species were renamed as *GhPsbX*, *GaPsbX*, *GrPsbX*, and *GbPsbX*. The results indicate that the size of the *PsbX* gene family is comparatively smaller than other gene families.

Phylogenetic analysis of *PsbX* gene family

The *PsbX* gene family from nine species, including *A. thaliana*, *V. vinifera*, *O. sativa*, *T. cacao*, *P. trichocarpa*, *G. barbadense*, *G. arboreum*, *G. hirsutum*, and *G. raimondii*, was utilized to construct a phylogenetic tree using the Neighbor-joining (NJ) method. The resulting tree was employed to identify the phylogeny and comprehend the functional relationships among all 40 putative *PsbX* genes (Fig. 1). Due to the close relationship of the *PsbX* gene

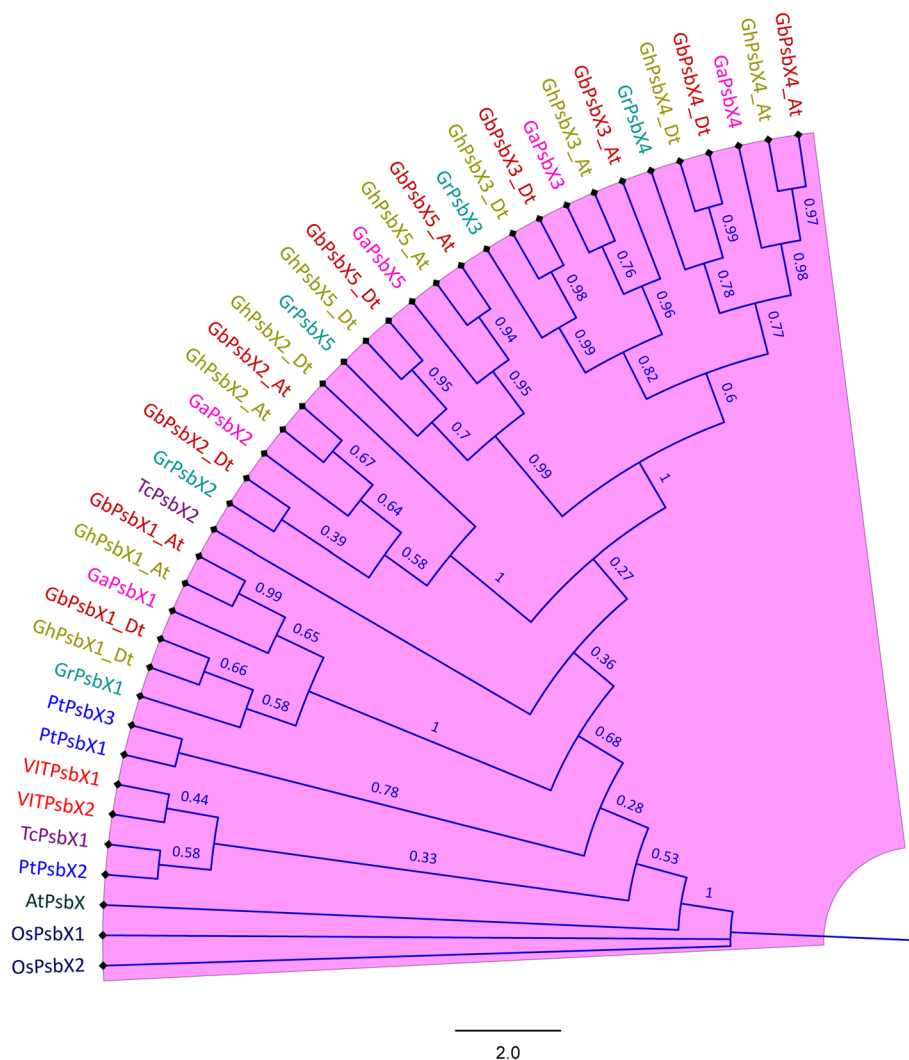


Fig. 1 Phylogenetic tree relationships between 10 *G. hirsutum*, 10 *G. barbadense*, 5 *G. arboreum*, 5 *G. raimondii*, 1 *Arabidopsis*, 2 *O. stiva*, 3 *P. trichocarpa*, 2 *T. cacao*, 2 genes from *Vitis Vinefera*. The MEGA X program using the NJ (Neighbor-joining) method created the phylogenetic tree. The bootstrap test was done with 1 000 iterations. The different species genes are shown in different colors

family, all 40 genes were categorized within the same clade.

A notable close relationship was observed among the *PsbX* genes of the four cotton species compared with other five plant species. Moreover, the *TcPsbX2* gene occupied the same subclade but on a distinct branch within the close relationship of *GrPsbX2* and *GbPsbX2-Dt* genes. Additionally, *G. raimondii* exhibited a greater number of *PsbX* genes compared with *T. cacao*. Similarly, *PsbX* genes of monocot species like *O. sativa* were positioned at the end of the clade on separate branches in the phylogenetic tree. *AtPsbX1* was identified on a separate branch closely related to two *PsbX* genes of *O. sativa*. Nevertheless, *PsbX* genes of dicots (Arabidopsis, cotton, and cacao) were identified in the same clade. These findings suggest that the primary function of the *PsbX* gene family evolved prior to the divergence between monocots and dicots. The current results demonstrate that the nine plant species are situated in one clade, consisting of various subclades of *PsbX* gene members. This strongly indicates that the variations in these plants emerged after the expansion of the *PsbX* gene family. Furthermore, *PsbX* genes were notably more abundant in *Gossypium* genomes compared with other species.

Furthermore, a distinctive observation was noted that several subclades exclusively included members of the *PsbX* gene family derived from a specific plant. Despite the greater similarities among orthologs of the four cotton species, the orthologs of A-genome and At sub-genome of cotton clustered together. A similar pattern was observed for D genome and Dt-sub genome of cotton. Consequently, the aforementioned results indicate that *PsbX* orthologs from either genome of At-A or Dt-D may share a common ancestor.

Gene structure, conserved motif, and sequence logo analysis of *PsbX*

Gene structure analysis was conducted on the sequence file of *PsbX* genes, as depicted in Fig. 2A. The total number of exons was 1 or 2, with an average of 1.5. All genes exhibited only one exon (CDS), except for a gene that possessed two exons. Introns were identified in only two *PsbX* genes, while the remaining 28 had no introns. Typically, genes present in the same evolutionary branch and clade shared similar structures, possibly originating from a conserved gene pattern in terms of exon/intron length and numbers. Generally, untranslated regions (UTR) were found

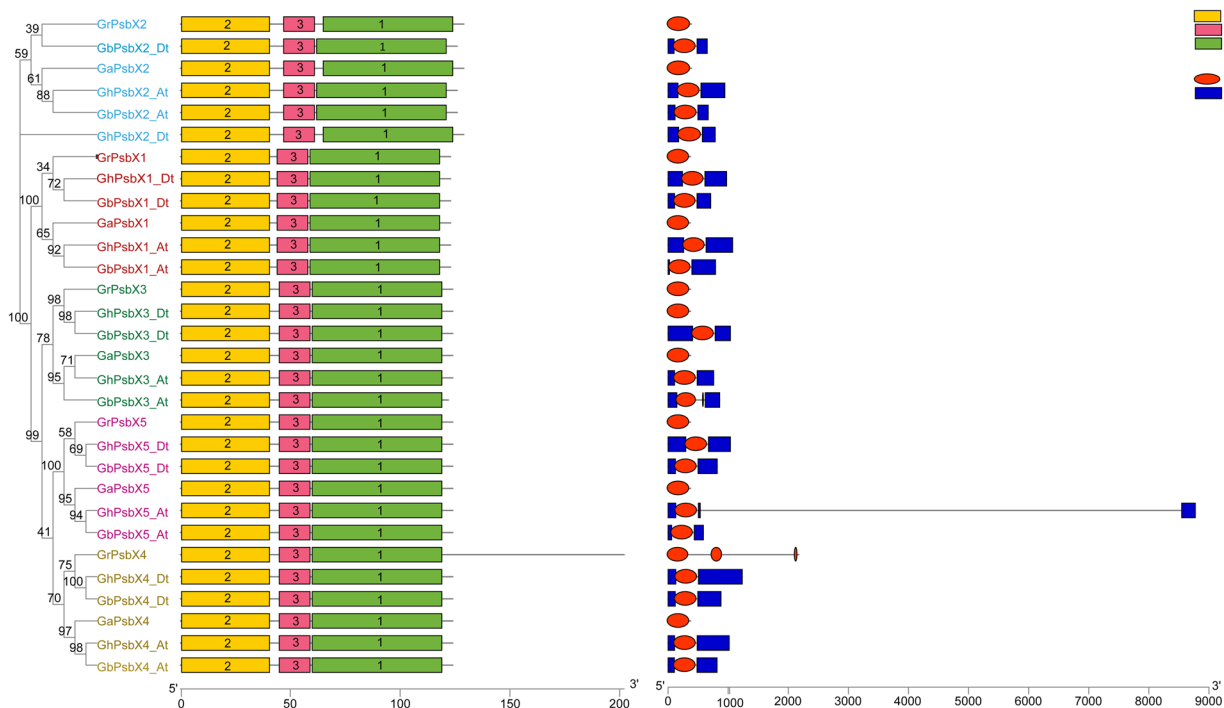


Fig. 2 Structural features of *PsbX* gene family in cotton. **A** Three motifs identified by the MEME tool are represented by colored boxes, and their consensus sequences are shown in supplementary Fig. S1, **B** The exon–intron arrangement of *PsbX* gene. The blue shapes represent UTR, red shapes indicate CDS and the black line represents introns

in *G. hirsutum* and *G. barbadence*. However, no UTR was identified in both *G. arboreum* and *G. raimondii* species.

The gene structure of *GrPsbX4* differed from that of other genes, suggesting a potential impact on evolutionary speed and function. MEME software (Additional file 3: Fig. S1) was employed to identify three conserved motifs of *PsbX* genes with a width range of 15~60 amino acids. However, both the number and type of conserved motifs in different genes were consistent within the same branch of the phylogenetic tree. Sequence logo analysis of conserved amino acid residues for the three motifs was conducted across all cotton species to confirm the conservation of *PsbX* gene family proteins during evolution. The resulting sequence logos exhibited substantial conservation throughout N and C terminals at various sites among all cotton species (Fig. 2A).

Chromosomal locations, physicochemical features and subcellular location prediction of *PsbX* gene family

The identified *PsbX* genes were assigned to their respective chromosomes. Figure 3 illustrates that five genes were assigned to chromosomes of At sub-genome, while other five genes were allocated to their corresponding chromosomes of Dt sub-genome. Chromosome A13 and its homolog D13 harbored three genes each, but most

chromosomes only had a single *GhPsbX* gene. One gene (*PsbX*) was assigned to chromosomes A05, D05, A07, and D07, respectively. Moreover, the results indicated that gene loss did not occur, as genes were evenly distributed in the cotton genome (At and Dt sub-genome) during evolution. This distribution may be a consequence of the relatively small size of the *PsbX* gene family.

Subsequently, the physicochemical properties of the *PsbX* gene family were predicted, including locus ID, the corresponding chromosome, start and end points, strand polarity, protein length (amino acid, aa), MW, pI, and cellular localization.

The details of physicochemical properties are shown in Table 1. The results revealed that five *GhPsbX* genes diverged from At sub-genome, and the other five genes originated from Dt sub-genome of cotton species. The protein length was also determined in the range of 120~200 amino acids. The MW was 12 386.5 kDa for GhPsbx1-At and 13 126.32 kDa for GhPsbx2-Dt. Similarly, the determination of cellular localization for GhPsbX proteins indicated their presence in the chloroplast. Other predicted physicochemical properties are presented in Table 1.

Identification of *cis*-regulatory element

Cis-regulatory elements were identified in the upstream (promoter) region of 1 500 bp for each gene and further

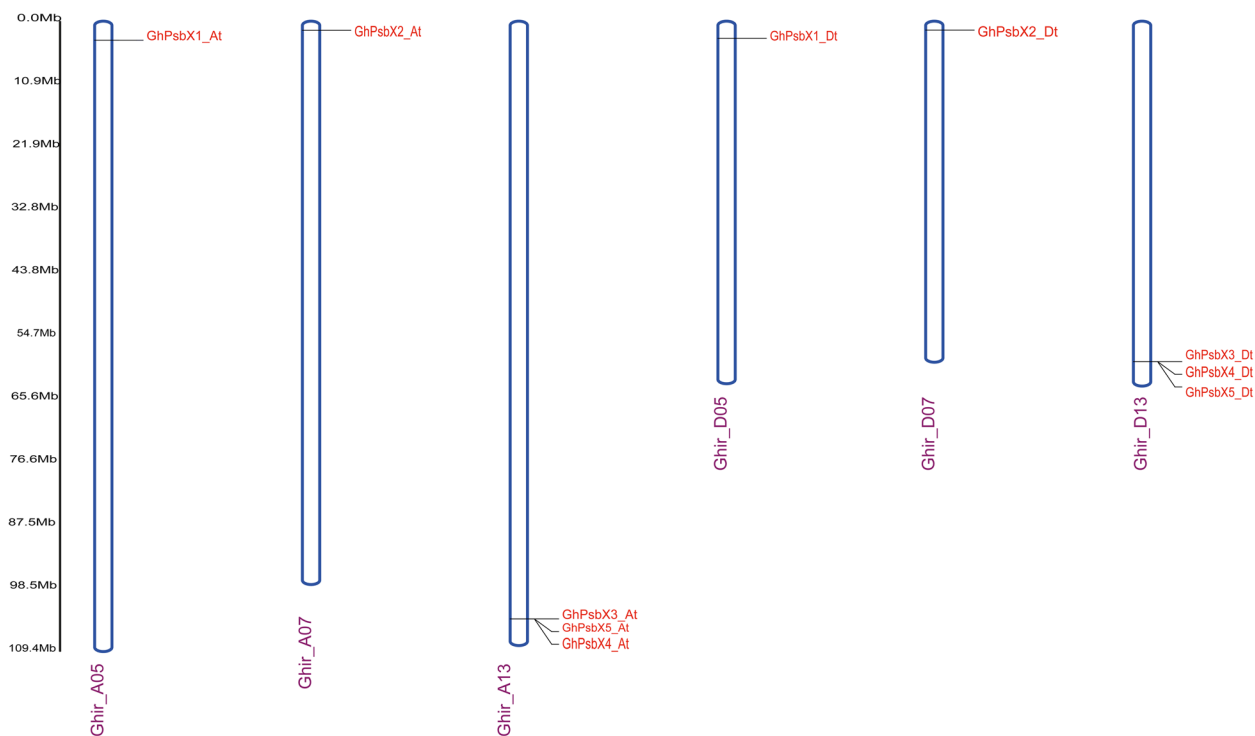


Fig. 3 Gene mapping on chromosome of *G. hirsutum*. Chromosome number is mentioned below of each bar. Gene name is indicated by red color

Table 1 List of *PsbX* genes identified in *Gossypium* and their sequence physiochemical properties

Gene ID	Name	Chr	Start /bp	End /bp	Strand	Number of amino acid	Mw /KDa	pI	#Most-Likely-Location
D5.v1.pred_00000384-RA	GrPsbX1	D05	2 614 382	2 614 753	+	120	12 373.49	10.4	CHLOROPLAST
D5.v1.pred_00033500-RA	GrPsbX2	D07	1 394 653	1 395 042	+	120	13 124.31	10.0	CHLOROPLAST
D5.v1.pred_00021191-RA	GrPsbX3	D13	54 975 179	54 975 553	-	120	12 625.86	10.7	CHLOROPLAST
D5.v1.pred_00021192-RA	GrPsbX4	D13	54 976 963	54 979 138	-	200	21 616.23	10.1	CHLOROPLAST
D5.v1.pred_00021193-RA	GrPsbX5	D13	54 983 135	54 983 509	-	120	12 731.94	10.7	CHLOROPLAST
Ga05G0346.1	GaPsbX1	A05	3 159 541	3 159 512	+	120	12 343.46	10.4	CHLOROPLAST
Ga07G0152.1	GaPsbX2	A07	1 699 142	1 699 531	+	120	13 122.32	10.0	CHLOROPLAST
Cotton_A_03158_BGI-A2_v1.0	GaPsbX3	A13	83 186 770	83 187 144	+	120	12 656.94	11.1	CHLOROPLAST
Ga13G2390.1	GaPsbX4	A13	119 166 072	119 166 446	-	120	12 639.87	10.5	CHLOROPLAST
Ga13G2391.1	GaPsbX5	A13	119 166 072	119 170 965	-	120	12 699.94	10.7	CHLOROPLAST
Ghir_A05G003380.1	GhPsbX1_At	A05	3 298 268	3 299 342	+	120	12 386.49	10.7	CHLOROPLAST
Ghir_A07G001470.1	GhPsbX2_At	A07	1 590 068	1 591 012	+	120	12 843.05	10.0	CHLOROPLAST
Ghir_A13G020450.1	GhPsbX3_At	A13	103 747 523	103 748 282	-	120	12 656.94	11.1	CHLOROPLAST
Ghir_A13G020460.1	GhPsbX4_At	A13	103 747 544	103 756 316	-	120	12 625.84	10.5	CHLOROPLAST
Ghir_A13G020470.1	GhPsbX5_At	A13	103 750 766	103 751 783	-	120	12 699.94	10.7	CHLOROPLAST
Ghir_D05G003520.1	GhPsbX1_Dt	D05	2 964 671	2 965 644	+	120	12 403.51	10.4	CHLOROPLAST
Ghir_D07G001490.1	GhPsbX2_Dt	D07	1 556 339	1 557 123	+	120	13 126.32	10.0	CHLOROPLAST
Ghir_D13G021280.1	GhPsbX3_Dt	D13	59 090 890	59 091 264	-	120	12 639.89	10.7	CHLOROPLAST
Ghir_D13G021290.1	GhPsbX4_Dt	D13	59 093 852	59 095 086	-	120	12 535.78	10.5	CHLOROPLAST
Ghir_D13G021300.1	GhPsbX5_Dt	D13	59 098 593	59 099 629	-	120	12 701.91	10.7	CHLOROPLAST
Gbar_A05G002930.1	GbPsbX1_At	A05	2 942 757	2 943 546	+	120	12 386.49	10.7	CHLOROPLAST
Gbar_A07G001500.1	GbPsbX2_At	A07	1 673 556	1 674 223	+	120	12 843.05	10.0	CHLOROPLAST
Gbar_A13G020900.1	GbPsbX3_At	A13	104 625 505	104 626 364	-	120	12 615.94	10.8	CHLOROPLAST
Gbar_A13G020910.1	GbPsbX4_At	A13	104 629 009	104 629 823	-	120	12 625.84	10.5	CHLOROPLAST
Gbar_A13G020920.1	GbPsbX5_At	A13	104 633 708	104 634 295	-	120	12 699.94	10.7	CHLOROPLAST
Gbar_D05G003370.1	GbPsbX1_Dt	D05	2 851 525	2 852 234	+	120	12 358.47	10.3	CHLOROPLAST
Gbar_D07G001490.1	GbPsbX2_Dt	D07	1 514 984	1 515 635	+	120	12 879.12	10.0	CHLOROPLAST
Gbar_D13G021280.1	GbPsbX3_Dt	D13	55 675 317	55 676 353	-	120	12 639.89	10.7	CHLOROPLAST
Gbar_D13G021290.1	GbPsbX4_Dt	D13	55 678 883	55 679 763	-	120	12 535.78	10.5	CHLOROPLAST
Gbar_D13G021300.1	GbPsbX5_Dt	D13	55 683 341	55 684 159	-	120	12 701.91	10.7	CHLOROPLAST

searched in the Plant CARE database. Various *cis*-regulatory elements present in the promoter region of *PsbX* genes may be the reason for diverse function of putative genes. The resulting *cis*-regulatory elements within the promoter region of *PsbX* genes have been presented in Additional file 4: Table S3. Typically, regulatory sequences such as TGACG-motifs, STREs (rapid stress response elements), TATA boxes, and CAATs were identified among all genes of the *PsbX* family in upland cotton. The primary *cis*-acting elements CAATs and TATA boxes were located within the promoter region of transcriptional genes in eukaryotes. Additionally, the CAAT box plays a role in the regulation of gene expression by providing a binding site for RNA transcription factors

(TF) (Laloum et al. 2013; Shore and Sharrocks 1995; Ramji and Fok 2002).

The CAAT box also plays a significant role in modulating the nopaline synthase promoter (Dai et al. 1999). Presently, a TATA box has been reported for its supportive role in transcription, harboring a binding site for transcription factors (TF) or histones (Isogai et al. 2007; Bae et al. 2015). The primary transcriptional stress response is activated via STREs against abiotic and/or biotic stresses (Walley et al. 2007). The above-mentioned results demonstrate that *cis*-regulatory elements play a pivotal role, enabling the plant to cope with stresses such as intense light and drought.

Syntenic blocks

The hybridization of two diploid cotton species, such as *G. arboreum* and *G. raimondii*, resulted in the allotetraploid species *G. hirsutum*. However, the doubled chromosomes of *G. hirsutum* exhibited a resemblance to those of *G. raimondii* and *G. arboreum*. The expansion of gene families is attributed to various duplication events, including tandem, segmental, and whole-genome duplication. Syntenic blocks are characterized by the similar chromosomal arrangement of conserved genes among different species. In the current study, associated gene pairs among the allotetraploid *G. hirsutum* and the two diploid cotton ancestors, *G. raimondii* and *G. arboreum* were identified (Fig. 4). The genes of *G. hirsutum* on chromosomes A05, A07, A13, D05, D07, D13 were originated from the following chromosomes Ga05, Ga07, and Ga13 of conserved syntenic blocks of the ancestral species (*G. arboreum*). Additionally, orthologs of *G. hirsutum* and *PsbX* were supposed to be present on *G. raimondii* chromosomes like Gr05, Gr07, and Gr13.

Ka/Ks ratio

During the process of evolution, duplicated genes might be classified into following non-functionalization, sub-functionalization, and neo-functionalization. The non-synonymous (*Ka*) and synonymous substitution (*Ks*), rates of substitution (*Ka/Ks* value) might be used to deduce the magnitude of both positive selection as well as selective constraints. Typically, the criterion of *Ka/Ks* ratio for positive selection is greater than one, for neutral evolution equal to one, and for purifying selection less

than one. In the current study, the estimation of *Ka*, *Ks*, and *Ka/Ks* values of *PsbX* homologous gene pairs for *G. hirsutum* was presented in Table 2. The results indicated that the *Ka/Ks* < 1 for 4 *PsbX* genes of homologous pairs and one gene which had ratios of *Ka/Ks* > 1, implying that these genes might have been subjected to positive and purifying selection, subsequent segmental as well whole genome duplications.

Expression profiles of *PsbX* gene family in various tissues and organs

The growth and development of plants are regulated by gene expression. Analyzing the expression level of genes can provide information on gene ontology. Thus, the expression pattern of *PsbX* genes in various cotton tissues and organs was evaluated. Firstly, the gene expression analysis of *PsbX* was performed from online accessible transcriptomic data of *G. hirsutum* L. against abiotic stresses. The tissue-specified expression data was also downloaded from an online database named NCBI SRA. The various organs used for

Table 2 Gene evolution forms of *Ka*, *Ks* and *Ka/Ks* values for homologous *PsbX* gene pairs of *G. hirsutum*

SEQ1	SEQ2	<i>Ka</i>	<i>Ks</i>	<i>Ka/Ks</i>	Substitution
GhPsbX2_At	GhPsbX2_Dt	0.011 6	0.018 0	0.644 3	synonymous
GhPsbX4_Dt	GhPsbX4_At	0.020 6	0.017 6	1.166 7	non-synonymous
GhPsbX5_At	GhPsbX5_Dt	0.012 3	0.026 4	0.463 6	synonymous
GhPsbX1_Dt	GhPsbX1_At	0.016 3	0.026 4	0.618 3	synonymous
GhPsbX3_At	GhPsbX3_Dt	0.024 5	0.045 0	0.545 1	synonymous

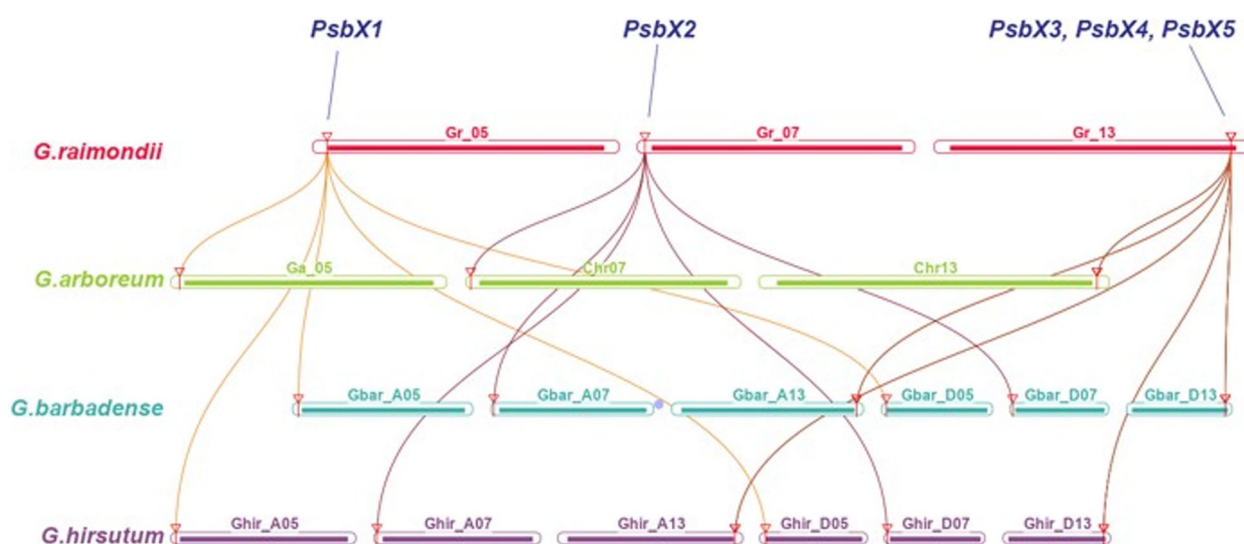


Fig. 4 Synteny analysis of *PsbX* gene family in all cotton species. Cotton species are represent by different colors. Red color represents *G. raimondii*, green color indicates *G. arboreum*, sky blue color indicates *G. barbadense*, and purple represents *G. hirsutum*. Chromosome number of *PsbX* is indicated by dark blue color

this study include leaves, roots, stems, and reproductive tissues (torus, petals, stamens, and pistils), and exposure to abiotic stresses, including salt and PEG treatments at different time periods.

A heat map was generated for gene expression, revealing that genes with similar expression levels clustered closely together. However, the expression levels of many genes varied significantly across different tissues (Fig. 5). For instance, a lower expression level of many *PsbX* genes was observed in the following organs/tissues: roots, torus, filament, anther, and sepal. Interestingly, a higher expression level was observed for the following genes, *GhPsbX2_At*, *GhPsbX2_Dt*, *GhPsbX1_At*, and *GhPsbX1_Dt*, in the bracket, stem, and leaf. Moreover, a higher expression level of *GhPsbX2_At* was observed, albeit irregularly, in various floral organs. Comparatively, low expression levels of two genes named *GhPsbX5_At* and *GhPsbX3_Dt* was observed across all tissues of cotton, while one gene, *GhPsbX3_Dt*, exhibited a high expression level in the leaves of cotton.

Plants are unable to move and become vulnerable to various abiotic stresses during growth and developmental stages. Subsequently, we examined the gene expression pattern of *GhPsbX* against abiotic stresses including salinity and drought at various time durations. The results of the heat map showed up- and down-regulation of the expression pattern of all *GhPsbX* genes against different stress treatments at various periods (Fig. 5).

The genes with similar responses formed similar clusters together. The following genes like *GhPsbX3_At*, *GhPsbX1_At*, *GhPsbX1_Dt*, *GhPsbX4_Dt*, *GhPsbX2_At*, and *GhPsbX2_Dt* were highly expressed as compared with other genes under drought and/or salt stress at various time intervals. However, two genes, *GhPsbX5_At* and *GhPsbX3*, exhibited a relatively lower expression level in response to salinity and drought treatments. Likewise, ten genes of *GhPsbX* expressed in salt and drought conditions were selected. The response of these selected genes was evaluated by qRT-PCR analysis at different periods (Fig. 6). All genes were highly expressed compared with the control at various times under each stress treatment, with some exhibiting expression or lack thereof for different treatments at specific periods. Furthermore, PEG treatment induced the up-regulation of genes, namely *GhPsbX1-At*, *GhPsbX3-At*, *GhPsbX4-At*, *GhPsbX5-At*, *GhPsbX1-Dt*, *GhPsbX2-Dt*, and *GhPsbX3-Dt* at 2 h, 12 h, and 24 h, while NaCl treatment suppressed the expression of the aforementioned genes. However, NaCl treatment up-regulated the *GhPsbX2-At* gene at 2 h and 48 h. Interestingly, *GhPsbX4-Dt* showed the highest expression level at 2 h in both PEG and NaCl treatments compared with other time points. Moreover, *GhPsbX5-Dt* was upregulated by both NaCl and PEG treatments for 2 h,

48 h, and 2 h, 24 h, and 72 h, respectively. All *PsbX* genes were downregulated at 6 h and 12 h under PEG and NaCl treatments.

Discussion

Numerous studies have aimed to understand the physiological and molecular functions of genes involved in photosystems across various plant species. Notably, the *PsbX* gene family's role in cotton plants has not been explored in the existing scientific literature. This study aims to fill this void by investigating *PsbX* genes in both diploid and allotetraploid cotton species. Our analysis encompassed the evolutionary relationships, gene duplications, selection pressures, and expression levels in response to different abiotic stresses. The availability of whole-genome sequencing data for cotton has provided an unprecedented opportunity for a thorough exploration of the *PsbX* gene family's potential functions in the photosynthetic process.

In this study, we identified and scrutinized *PsbX* genes in four cotton species: *G. hirsutum*, *G. raimondii*, *G. arboreum*, and *G. barbadense*. While data were collected for all species, our primary focus was on *G. hirsutum*. The findings from our study are poised to provide foundational knowledge and serve as a valuable resource for further investigations into the function of *PsbX* genes in cotton species. In our study, 40 *PsbX* genes were identified in nine different species of plants like *T. cacao*, *G. arboreum*, *G. hirsutum*, *G. raimondii*, *G. barbadense*, *P. trichocarpa*, *V. vinifera*, *A. thaliana*, and monocotyledons (*O. sativa*). While orthologs of monocot species formed distinct clusters, orthologs of dicots exhibited similar clustering, suggesting that the fundamental function of the *PsbX* gene family originated before the divergence of dicots and monocots (Chen et al. 2018).

The *PsbX* protein, encoded by the nucleus in Photosystem II (*PSII*), is known for its relatively smaller size. This gene family exhibits a modest number of genes. Our study revealed distinct gene compositions in diploid species, with *T. cacao* displaying fewer gene members, approximately half the number found in its close relative *G. raimondii*. These observations suggest an expansion of the *PsbX* gene family during the evolutionary trajectory of cotton. Previous research showed that both cacao as well as cotton were subjected to paleohexaploidisation events shared by the eudicots. However, a recent duplication event was observed in cotton, but no such event was observed in *T. cacao*. It can be concluded that diploid cotton possesses more *PsbX* genes (Wang et al. 2012). It is interesting to note that *PsbX* gene numbers in *G. hirsutum*, revealing a twofold increase compared with its diploid counterparts, *G. raimondii* and *G. arboreum*. *G.*

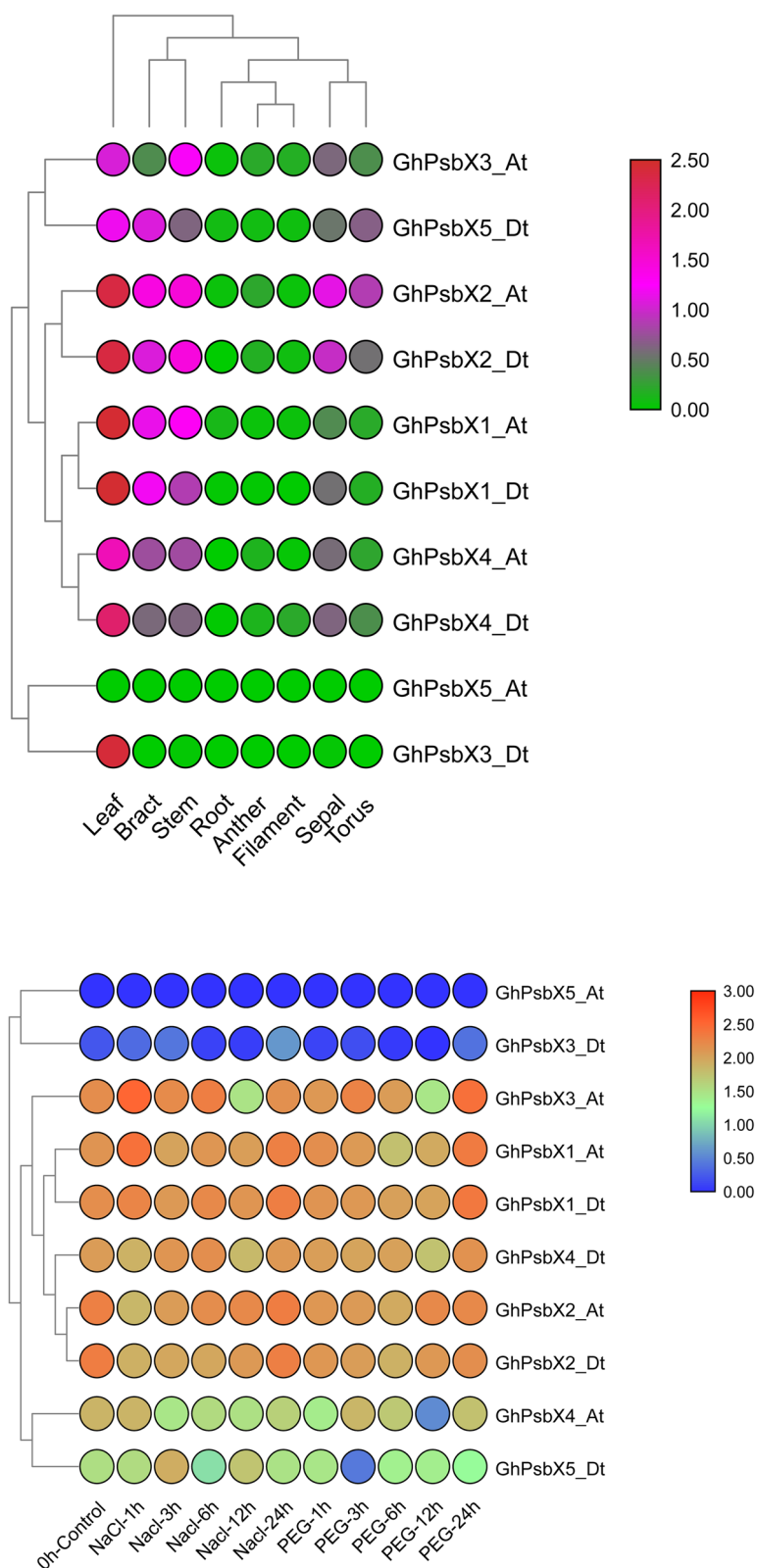


Fig. 5 Expression profiles of *PsbX* genes under drought and salinity treatments. (A) Heat map show expression profile in different tissues (B) Genes are shown on the right, and the phylogenetic relationships are shown on the left

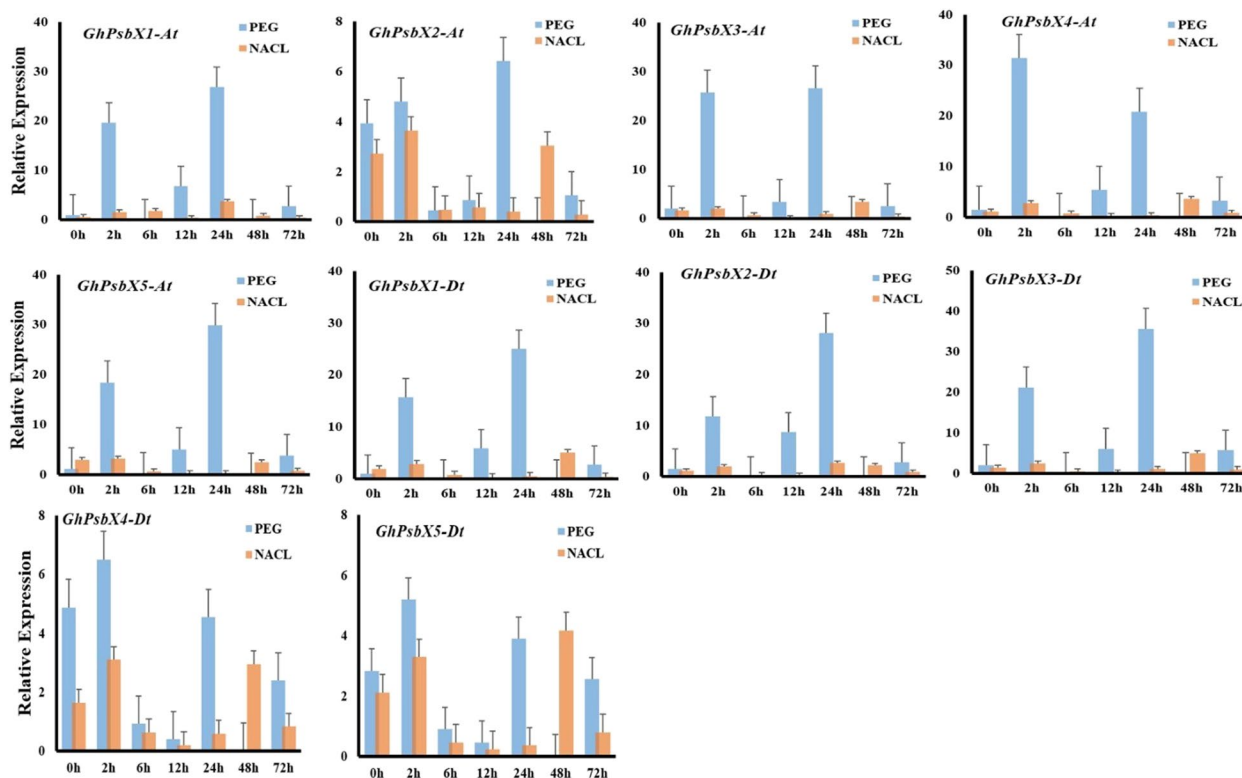


Fig. 6 Relative expression level by qRT-PCR analysis for selected *GhPsbX* genes. Error bars indicate the standard deviations (SD) among three independent biological repeats

hirsutum, being a typical allotetraploid species resulting from the hybridization of A and D genomes, experienced chromosome doubling around 1–2 million years ago (Renny-Byfield et al. 2016). Similarly, upland cotton underwent conservation following the polyploidization event, with orthologs of At and Dt genomes maintaining significant collinearity.

The evolution of multigene families was driven by structural diversity (Sánchez-Gracia et al. 2011; Davies et al. 2014; Suarez et al. 1998). The exon/intron arrangements in cotton were studied to elucidate the structural diversity of *PsbX* genes. The diverse number of exons/introns among genes might lead to the functional diversity of *PsbX* genes. It could possibly occur due to exon gain or loss in *PsbX* gene family during evolution. Various reports, based on gene structure analyses, indicate that introns have wielded substantial influence during the evolutionary phases in various plant species (Suarez and Gilbert 2006). A greater number of introns were present during the early phase of expansion but later decreased with the passage of time (Roy and Penny 2007). These reports revealed that higher species had a smaller number of introns in their genomes (Roy and Gilbert 2005). The greater number of introns in the genome might pose a significant burden on the organism, and introns

could possibly change the activities of genes and lead to new functions of genes. Likewise, the smaller number of introns were found in mostly genes of cotton species such as *G. hirsutum*, *G. barbaence*, *G. arboreum*, and *G. raimondii*. The absence of introns might suggest that the gene expansion possibly occurred independently. The genome wide analysis also revealed that extensive loss/gain of introns occurred via eukaryotic diversification event (Roy and Penny 2007; Rogozin et al. 2003). Moreover, the results indicated that the presence of three conserved motifs with long sequences of *PsbX* gene family also implied the conservation of motifs of *PsbX* gene family in cotton plants.

The even allocation of *GhPsbX* genes to At- and Dt-sub genomes confirmed the equal distribution of chromosomes as both genomes had five genes. The physicochemical properties such as predicted subcellular localization, pI, and MW of genes were almost similar because of the equal allocation of genes in both sub-genomes. All *GhPsbX* genes had a similar distribution of *cis*-elements of promoter associated with growth, development, and stress responses. It is reported that light intensity plays a pivotal part in the growth and differentiation of plants (Qanmber et al. 2019; Fankhauser and Chory 1997). *Cis*-elements like TCA and CGTCA are engaged in the

regulation of gene expression when exposed to MeJA and SA stress, correspondingly (Wen et al. 2014; Maestrini et al. 2009). In addition, the existence of W-box mediated ABA responses under drought stress (Singh et al. 2002). The above-mentioned *cis*-elements present in the majority of *PsbX* genes were used for identification of predicted function during growth, development, and various stresses. Abiotic stresses adversely affect the photosynthesis process in result of reducing the photosynthetic activity (Nouri et al. 2015). The response of the photosynthetic system is diverse under drought stress in seedlings as well as mature crop. In mature crops, drought stress disturbs photosynthetic activity in turn ROS species are formed due to excessive light absorption eventually impairing the photosynthetic apparatus. However, chlorophyll biosynthesis genes tended to downregulate, *PSI* and *PSII* ceased to stop the sunlight capturing to minimize the damage when young seedlings were exposed to drought stress (Dalal et al. 2018). Chloroplast is a powerhouse of plants; therefore, it has prime importance in plant sciences. However, it is also vulnerable to both abiotic as well as biotic stresses as it shows the actual level of plant response toward stress (Li et al. 2020; Liu et al. 2013). *PsbX* protein is part of *PSII* and plays a critical role in photosynthesis. The function of *PsbX* gene family has not been elucidated yet in cotton. Identification of the molecular function of *PsbX* gene family will aid in studying the function of *PsbX* gene family in various important crops.

The tissue expression pattern data of RNA seq were downloaded from NCBI SRA data to figure out the possible function of *GhPsbX* genes in specific tissues under abiotic stresses. The selected ten genes expressed in specific tissues revealed that these genes (*PsbX*) contributed positively to the development of leaf, bracket, and stem. The current finding indicates that *PsbX* gene family possibly play a vital role in photosynthesis in cotton plant. Additionally, all genes were highly upregulated positively as compared with the control under PEG and NaCl stress treatments at different time intervals confirmed by *GhPsbX* genes analysis. PEG induced the positive expression of many genes at 2 h, 12 h, and 24 h, however, *GhPsbX-At2* was upregulated under NaCl treatment at 2 h and 48 h.

Overall, our results indicate that *GhPsbX* expression levels varied among different organs/tissues like stem, leaf, and bracket. Similarly, the gene expression of *GhPsbX* might possibly be regulated under drought and salinity stresses. So, *GhPsbX* genes could be the potential candidates for the cotton breeding program.

Conclusions

In the current study, *PsbX* genes were identified and characterized in four cotton species. We performed genome-wide identification of *PsbX* gene family in cotton, for instance, identification of gene structure, evolutionary analyses, and expression patterns of genes. The result indicates that *PsbX* gene family was highly conserved in cotton and in other plant species. The existence of *cis*-elements within promoter regions having typical characteristics exhibited that they may have a potential role in plant growth as well as stress responses. The expression level of *PsbX* genes differed in various tissues during various abiotic stresses. The results showed that *PsbX* gene might be regulated by abiotic stresses and this information could be useful in future cotton breeding programs.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42397-023-00160-z>.

Additional file 1: Table S1. Specific primers used in relative quantitative real-time RT-PCR.

Additional file 2: Table S2. Gene IDs of cotton species.

Additional file 3: Figure S1. Consensus sequences of motif identified by MEME tool.

Additional file 4: Table S3. *Cis*-element analysis of *PsbX* gene promoters.

Acknowledgements

Not applicable.

Authors' contributions

Du XM, Ali I, and Raza I designed the experiment; Du XM, Ali I supervised the research. Raza I and Hu DW collected the samples; Raza I and Ahmad A worked on the analysis. Raza I, Parveen A, and Pan ZE interpreted the results and wrote the manuscript. All authors read, edited, and approved the final manuscript.

Funding

This work was supported by National Natural Science Foundation of China (32060466) and Chinese Academy of Agricultural Sciences.

Availability of data and materials

The datasets used and analyzed during current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute of Cotton Research, Chinese Academy of Agricultural Sciences, National Key Laboratory of Cotton Bio-Breeding and Integrated Utilization, Anyang, Henan 455000, China. ²Department of Biotechnology and Genetic

Engineering, Kohat University of Science and Technology, Kohat, Pakistan.
³Department of Botany, Kohat University of Science and Technology, Kohat, Pakistan.

Received: 1 August 2023 Accepted: 20 November 2023
 Published online: 10 January 2024

References

- Bae SH, Han HW, Moon J. Functional analysis of the molecular interactions of TATA box- containing genes and essential genes. *PLoS ONE*. 2015;10(3):e0120848. <https://doi.org/10.1371/journal.pone.0120848>.
- Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37:W202–8. <https://doi.org/10.1093/nar/gkp335>.
- Beddington JR, Asaduzzaman M, Clark ME, et al. Agriculture: What next for agriculture after Durban? *Science*. 2012;335(6066):289–90. <https://doi.org/10.1126/science.1217941>.
- Berry JO, Yerramsetty P, Zielinski AM, Mure CM. Photosynthetic gene expression in higher plants. *Photosynth Res*. 2013;117(1):91–120. <https://doi.org/10.1007/s11120-013-9880-8>.
- Chen ZJ, Scheffler BE, Dennis E, et al. Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol*. 2007;145:1303–10.
- Chen Y, Yang Z, Xiao Y, et al. Genome-wide analysis of the *NF-YB* gene family in *Gossypium hirsutum* L and characterization of the role of *GhDNF-YB22* in embryogenesis. *Int J Mol Sci*. 2018;19:483. <https://doi.org/10.3390/ijms19020483>.
- Dai Z, An K, Edward GE, An G. Functional role of CAAT box element of the nopaline synthase (nos) promoter. *J Plant Biol*. 1999;42:181–5.
- Dalal VK, Tripathy BC. Water-stress induced downsizing of light-harvesting antenna complex protects developing rice seedlings from photo-oxidative damage. *Sci Rep*. 2018;8(1):10–6. <https://doi.org/10.1038/s41598-017-14419-4>.
- Davies KTJ, Tsagkogeorga G, Bennett NC, et al. Molecular evolution of growth hormone and insulin-like growth factor 1 receptors in long-lived, small-bodied mammals. *Gene*. 2014;549:228–36. <https://doi.org/10.1016/j.gene.2014.07.061>.
- Dong H. Underlying mechanisms and related techniques of stand establishment of cotton on coastal saline-alkali soil. *Chin J Appl Ecol*. 2012;23(2):566–72.
- Fankhauser C, Chory J. Light control of plant development. *Annu Rev Cell Dev Biol*. 1997;13:203–29. <https://doi.org/10.1146/annurev.cellbio.13.1.203>.
- Finn RD, Coghill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44:D279–85. <https://doi.org/10.1093/nar/gkv1344>.
- He M, He CQ, Ding NZ. Abiotic stresses general defenses of land plants and chances for engineering multistress tolerance. *Front Plant Sci*. 2018;9:1771. <https://doi.org/10.3389/fpls.2018.01771>.
- Hu Y, Chen J, Fang L. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet*. 2019;51(4):739–48. <https://doi.org/10.1038/s41588-019-0371-5>.
- Huang G, Wu Z, Percy RG. Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nat Genet*. 2020;52(5):516–24. <https://doi.org/10.1038/s41588-020-0607-4>.
- Hussain HA, Hussain S, Khaliq A. Chilling and drought stresses in crop plants: implications, cross talk, and potential management opportunities. *Front Plant Sci*. 2018;9:393. <https://doi.org/10.3389/fpls.2018.00393>.
- Isogai Y, Keles S, Prestel M, et al. Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev*. 2007;21:2936–49. <https://doi.org/10.1101/gad.1608807>.
- Katoh H, Ikeuchi M. Targeted disruption of *psbX* and biochemical characterization of photosystem II complex in the thermophilic cyanobacterium *Synechococcus elongatus*. *Plant Cell Physiol*. 2001;42(2):179–88. <https://doi.org/10.1093/pcp/pce024>.
- Laloum T, De Mita S, Gamas P, et al. CCAAT-box binding transcription factors in plants. *Trends Plant Sci*. 2013;18:157–66. <https://doi.org/10.1016/j.tplants.2012.07.004>.
- Lescot M. Plantcare, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res*. 2002;30:325–7. <https://doi.org/10.1093/nar/30.1.325>.
- Li XW, Zhu YL, Chen CY, et al. Cloning and characterization of two chlorophyll A/B binding protein genes and analysis of their gene family in *Camellia sinensis*. *Sci Rep*. 2020;10(1):1–9. <https://doi.org/10.1038/s41598-020-61317-3>.
- Liu R, Xu YH, Jiang SC, et al. Light-harvesting chlorophyll a/b-binding proteins positively involved in abscisic acid signalling require a transcription repressor *WRKY40* to balance their function. *J Exp Bot*. 2013;64(18):5443–56. <https://doi.org/10.1093/jxb/ert307>.
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using realtime quantitative PCR and the 2^{-ΔΔCT} method. *Methods*. 2001;25(4):402–8.
- Maestrini P, Cavallini A, Rizzo M, et al. Isolation and expression analysis of low temperature-induced genes in white poplar (*Populus alba*). *J Plant Physiol*. 2009;166:1544–56. <https://doi.org/10.1016/j.jplph.2009.03.014>.
- Magwanga RO, Lu P, Kirungu JN, et al. Whole genome analysis of cyclin dependent kinase (CDK) gene family in cotton and functional evaluation of the role of *CDKF4* gene in drought and salt stress tolerance in plants. *Int J Mol Sci*. 2018;19(9):2625. <https://doi.org/10.3390/ijms19092625>.
- Magwanga RO, Kirungu JN, Lu P, et al. Genome wide identification of the trihelix transcription factors and overexpression of *Gh_A05G2067 (GT-2)*, a novel gene contributing to increased drought and salt stresses tolerance in cotton. *Physiol Plant*. 2019;167(3):447–64. <https://doi.org/10.1111/ppl.12920>.
- Nath K, Kajoo A, Poudyal RS. Towards a critical understanding of the photosystem II repair mechanism and its regulation during stress conditions. *FEBS Lett*. 2013;587(21):3372–81. <https://doi.org/10.1016/j.febslet.2013.09.015>.
- Nishiyama Y, Murata N. Scheme for the mechanism of photoinhibition and its application to enhance the abiotic stress tolerance of the photosynthetic machinery. *Appl Microbiol Biotechnol*. 2014;98(21):8777–96. <https://doi.org/10.1007/s00253-014-6020-0>.
- Nouri MZ, Moumeni A, Komatsu S. Abiotic stresses: insight into gene regulation and protein expression in photosynthetic pathways of plants. *Int J Mol Sci*. 2015;16(9):20392–416. <https://doi.org/10.3390/ijms160920392>.
- Pertea M, Kim D, Geo MP, Jeffrey T. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11:1650–67. <https://doi.org/10.1038/nprot.2016.095>.
- Pintó-Marjuan M, Munné-Bosch S. Photo-oxidative stress markers as a measure of abiotic stress-induced leaf senescence: advantages and limitations. *J Exp Bot*. 2014;65(14):3845–57. <https://doi.org/10.1093/jxb/eru086>.
- Qanmber G, Ali F, Lu L, et al. Identification of histone *H3 (H3)* genes in *Gossypium hirsutum* revealed diverse expression during ovule development and stress responses. *Genes*. 2019;10(5):355. <https://doi.org/10.3390/genes10050355>.
- Ramji DP, Fok P. CCAAT/ENHANCER-binding proteins: structure, function and regulation. *Biochem J*. 2002;365(3):561–75. <https://doi.org/10.1042/BJ20020508>.
- Raza I, Hu DW, Ahmad A, et al. Correlation analysis of stem hardness traits with fiber and yield-related traits in core collections of *Gossypium hirsutum*. *J Cotton Res*. 2021;4:8. <https://doi.org/10.1186/s42397-021-00082-8>.
- Renny-Byfield S, Page JT, Udall JA, et al. Independent domestication of two old world cotton species. *Genome Biol Evol*. 2016;8:1940–7. <https://doi.org/10.1093/gbe/evw129>.
- Rogozin IB, Wolf YI, Sorokin AV, et al. Remarkable interkingdom conservation of intron positions and massive lineage-specific intron loss and gain in eukaryotic evolution. *Curr Biol*. 2003;13:1512–7. [https://doi.org/10.1016/S0960-9822\(03\)00558-x](https://doi.org/10.1016/S0960-9822(03)00558-x).
- Roy SW, Gilbert W. Complex early genes. *Proc Natl Acad Sci USA*. 2005;102:1986–91. <https://doi.org/10.1073/pnas.0408355101>.
- Roy SW, Penny D. A very high fraction of unique intron positions in the intron-rich diatom *Thalassiosira pseudonana* indicates widespread intron gain. *Mol Biol Evol*. 2007;24:1447–57.
- Sánchez-Gracia A, Vieira FG, Almeida FC, Rozas J. Comparative genomics of the major chemosensory gene families in arthropods. In: *Encyclopedia of Life Sciences (ELS)*. Chichester: John Wiley & Sons, Ltd; 2011. <https://doi.org/10.1002/9780470015902.a0022848>.
- Sasi S, Venkatesh J, Daneshi RF, Gururani M. Photosystem II extrinsic proteins and their putative role in abiotic stress tolerance in higher plants. *Plants*. 2018;7(4):1–15. <https://doi.org/10.3390/plants7040100>.

- Schultz J, Copley RR, Doerks T, et al. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* 2000;28:231–4. <https://doi.org/10.1093/nar/28.1.231>.
- Shi LX, Kim SJ, Marchant A, et al. Characterization of the PsbX protein from photosystem II and light regulation of its gene expression in higher plants. *Plant Mol Biol.* 1999;40:737–44. <https://doi.org/10.1023/a:1006286706708>.
- Shore P, Sharrocks AD. The MADS-box family of transcription factors. *Eur J Biochem.* 1995;229:1–13. <https://doi.org/10.1111/j.1432-1033.1995.tb20430.x>.
- Singh KB, Foley RC, Onate-Sanchez L. Transcription factors in plant defense and stress responses. *Plant Biol.* 2002;5:430–6. [https://doi.org/10.1016/s1369-5266\(02\)00289-3](https://doi.org/10.1016/s1369-5266(02)00289-3).
- Song C, Li W, Wang Z, et al. Genome resequencing reveals genetic variation between the parents of an elite hybrid upland cotton. *Agronomy.* 2018;8:305. <https://doi.org/10.3390/agronomy8120305>.
- Suarez CE, Gilbert W. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat Rev Genet.* 2006;7:211–21. <https://doi.org/10.1038/nrg1807>.
- Suarez CE, Palmer GH, Hötzel I, McElwain TF. Structure sequence and transcriptional analysis of the *Babesia bovis* rap-1 multigene locus. *Mol Biochem Parasitol.* 1998;93:215–24.
- Walley JW, Coughlan S, Hudson ME, et al. Mechanical stress induces biotic and abiotic stress responses via a novel *cis*-element. *PLoS Genet.* 2007;10:1800–12. <https://doi.org/10.1371/journal.pgen.0030172>.
- Wang K, Wang Z, Li F, et al. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet.* 2012;44(10):1098–103. <https://doi.org/10.1038/ng.2371>.
- Wen F, Zhu H, Li P, et al. Genome-wide evolutionary characterization and expression analyses of WRKY family genes in *Brachypodium distachyon*. *DNA Res.* 2014;21:327–39.
- Xu Y, Magwanga RO, Cai X, et al. Deep transcriptome analysis reveals reactive oxygen species (ROS) network evolution, response to abiotic stress, and regulation of fiber development in cotton. *Int J Mol Sci.* 2019;20(8):1863. <https://doi.org/10.3390/ijms20081863>.
- Zafar MM, Mustafa G, Shoukat F, et al. Heterologous expression of *cry3Bb1* and *cry3* genes for enhanced resistance against insect pests in cotton. *Sci Rep.* 2022;12(1):1–11.
- Zafar MM, Rehman A, Razzaq A, et al. Genome-wide characterization and expression analysis of erf gene family in cotton. *BMC Plant Biol.* 2022;22(1):1–18.
- Zafar MM, Shakeel A, Haroon M, et al. Effects of salinity stress on some growth, physiological, and biochemical parameters in cotton (*Gossypium hirsutum* L.) germplasm. *J Nat Fibers.* 2022;19(14):8854–86.
- Zahid KR, Ali F, Shah F. Response and tolerance mechanism of cotton *Gossypium hirsutum* L. to elevated temperature stress: a review. *Front Plant Sci.* 2016;7:1–13. <https://doi.org/10.3389/fpls.2016.00937>.
- Zhang JY, Cun Z, Chen JW. Photosynthetic performance and photosynthesis-related gene expression coordinated in a shade-tolerant species *Panax notoginseng* under nitrogen regimes. *BMC Plant Biol.* 2020;20:273. <https://doi.org/10.1186/s12870-020-02434-z>.
- Zhang HB, Li Y, Wang B, Chee PW. Recent advances in cotton genomics. *Int J Plant Genom.* 2008;742–304. <https://doi.org/10.1155/2008/742304>.
- Zhao L, Lü Y, Chen W, et al. Genome-wide identification and analyses of the *AHL* gene family in cotton (*Gossypium*). *BMC Genomics.* 2020;21(1):1–14. <https://doi.org/10.1186/s12864-019-6406-6>.
- Zhu JK. Plant salt tolerance. *Trends Plant Sci.* 2001;6(2):66–71. [https://doi.org/10.1016/S1360-1385\(00\)01838-0](https://doi.org/10.1016/S1360-1385(00)01838-0).

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

