Genome-wide identification, evolution, regulation, and alternative shearing profile of peroxiredoxin genes in cotton

Yulong Feng [Equal first author, 1], **Renhui Wei** [Equal first author, 2], **Aiying Liu** [2], **Senmiao Fan** [2], **Jincan Che** [1], **Zhen Zhang** [2], **Baoming Tian** [1], **Youlu Yuan** [2], **Gongyao Shi** [Corresp., 1], **Haihong Shang** [Corresp. 1, 2]

[1] Zhengzhou University, Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou, China

[2] Chinese Academy of Agricultural Sciences, State Key Laboratory of Cotton Biology, Key Laboratory of Biological and Genetic Breeding of Cotton, The Ministry of Agriculture, Institute of Cotton Research, Anyang, China

Corresponding Authors: Gongyao Shi, Haihong Shang
Email address: shigy@zzu.edu.cn, shanghaihong@caas.cn

Peroxiredoxin (*PRX*) is widely known to eliminate excessive free radicals produced by stress in the body and protect cells from oxidative damage. At present, *PRX* family identification and evolution research have been carried out in some plants, however, there is a lack of systematic research about this families in cotton. In this study, a total of 47 *PRX* genes were identified in the cotton genome. Phylogenetic and gene structure analyses showed that the *PRX* gene family was divided into six subfamilies according to the conserved active site (PxxxTxxC…*S…W/F*). Segmental duplication and polyploid events were the main ways that the *PRX* family has expanded, and loss and translocation occurred during the evolution of *PRX* family. qRT-PCR analysis confirmed that the cis-acting elements could effectively regulate the expression of *PRX* genes. And differential shearing of *GhPRX14-D* occurred in roots and leaves. Subcellular localization of *PRX* proteins showed that most *PRX* members were located in chloroplasts, and a few members were located in the cell membrane and nucleus. Our results provide systematic support for the understanding of the *PRX* genes in cotton and provide a starting point for further studies of specific functions of *PRX* gene in cotton.

Genome-wide identification, evolution, regulation, and alternative shearing profile of peroxiredoxin genes in cotton

Yulong Feng[1*], Renhui Wei[2*], Aiying Liu[2], Senmiao Fan[2], Jincan Che[2], Zhen Zhang[2], Baoming Tian[1], Youlu Yuan[2], Gongyao Shi[1], Haihong Shang[1,2]

[1] Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou University, Zhengzhou, Henan 450001, China

[2] State Key Laboratory of Cotton Biology, Key Laboratory of Biological and Genetic Breeding of Cotton, The Ministry of Agriculture, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, Henan, China

Corresponding Author:

Gongyao Shi[1]

100Kexue Road, Zhengzhou, Henan P . R .China

Email address: shigy@zzu.edu.cn

Youlu Yuan[2]

38Huanghe Avenue, Wenfeng District, Anyang City, Henan P . R .China

Email address: yuanyoulu@caas.cn

Haihong Shang[1,2]

100Kexue Road, Zhengzhou, Henan P . R .China

Email address: shanghaihong@caas.cn

* These anthors contributed equally to this work.

Introduction

The peroxiredoxin (*PRX*) protein (EC 1.11.1.15) is a member of the thioredoxin-scaffold enzyme family that has cysteine-dependent peroxidase activity against hydrogen peroxide substrates. *PRX* is ubiquitous among living organisms and has been found in animals, plants, protozoa, parasites, yeast, bacteria and archaea (Knoops et al., 2007). Therefore, *PRX* in an ancient protein highly conserved throughout evolution (Hofmann et al., 2002). All the proteins in the family have conserved Cys residues at the N terminus, and some members also have conserved Cys residues at the C terminus. These conserved Cys residues can be used to divide Prx proteins into two subfamilies, namely, 1-Cys Prx and 2-Cys Prx. As an antioxidant enzyme, *PRX* is widely known to eliminate excessive free radicals in the body produced by stress and protect cells from oxidative damage (Hofmann et al., 2002). *PRX* not only reduces the damage caused by oxidative stress but also enhances the activity of natural killer cells (Nonn et al., 2003); regulates the proliferation, differentiation and apoptosis of cells (Lee et al., 2003); protects free radical-sensitive proteins (Wenders et al., 2003). At present, the research on this protein has mainly focused on mammalian disease resistance, the underlying mechanism of which is the loss of the protein, leading to an imbalance in the cell redox state (Oláhová et al., 2008). A large number of studies have shown that the *PRX* protein is involved in the pathogenesis of many diseases (Oláhová et al., 2008). *PRX* protease was found to be associated with Parkinson's disease (Andres et al., 2007; Fang et al., 2007), Jacob's disease, Pick's disease, lung malignant mesothelioma (Newick et al., 2012) and squamous cell carcinoma (Park et al., 2009).

In plants, a total of 10 *PRX* family members were identified in the *Arabidopsis* (*Arabidopsis thaliana*) genome, of which 4 targeted chloroplasts (Baier et al., 1997; Horling et al., 2001), playing an important role in the detoxification of chloroplasts (Dietz et al., 2006). Some recent studies demonstrated that 2-Cys *PRX* participates in the oxidation of chloroplast enzymes in the dark in *Arabidopsis* (Ojeda et al., 2018). The Fd-FTR-Trxs and NTRC redox systems of chloroplasts are integrated via the redox balance of 2-Cys *PRX* (Pérez-Ruiz et al., 2017) and the chloroplast 2-Cys *PRX* functions as thioredoxin oxidase in the redox regulation of chloroplast metabolism (Pérez-Ruiz et al., 2017). All of these studies show that *PRX* plays an important role in the chloroplast redox system. In addition, *PRX* plays an important role in root growth (Finkemeier et al., 2005), photosynthesis protection (Lamkemeyer et al., 2006), antioxidant activity (Pulido et al., 2010) and drought tolerance (Fichman et al., 2018) in *Arabidopsis*. However, at present, there is a lack of research on this gene in cotton, and only a few studies have shown that *PRX* gene is significantly up-regulated in response to drought stress(Zhang et al., 2016).

In this study, we identified all of the *PRX* genes and their evolutionary relationships and performed physical mapping onto their chromosomes in four cotton species (*G. hirsutum*, *G. raimondii*, *G. barbadense* and *G. arboreum*). In addition, we systematically analyzed the gene structure, conserved active sites and cis-acting elements of all identified *PRX* genes in the four cotton species. According to abiotic stress and hormone-induced cis-acting elements in the *PRX* gene promoter, we carried out experiments on the expression trends of *PRX* genes under different stresses. The expression levels of *PRX* genes in cotton tissues and organs were analyzed by *G. hirsutum* TM-1 transcriptome data (Zhang et al., 2015). The alternative splicing (AS) profile of the *PRX* genes was identified and verified by RT-PCR. Finally, to determine the

localization of *PRX* proteins in cells, we carried out subcellular localization experiments with *Agrobacterium tumefaciens*-infected tobacco. These results provide a solid foundation for the study of the distribution, structure and evolution of the cotton *PRX* gene family, and the regulatory mechanism, transcript abundance and cellular location of the *PRX* gene family will provide important data information for the follow-up study of the functional differentiation and application.

Materials and methods

Sequence source

G. raimondii (*Gossypium raimondii* L., BGI) (Paterson et al., 2012) and *G. hirsutum* (*Gossypium hirsutum* L., NAU) (Zhang et al., 2015) genomic data files were obtained from the JGI database (http://www.phytozome.net/); *G. arboreum* (*Gossypium arboreum* L., CRI) (Du et al., 2018) and *G. barbadense* (*Gossypium barbadense* L., Hau) (Wang et al., 2019) genomic data files were downloaded from CottonGen database (https://www.cottongen.org/).

The following protein sequence data were obtained from the corresponding databases. *Arabidopsis* (*Arabidopsis thaliana* L.) (Berardini et al., 2004) protein sequence data were obtained from the Arabidopsis Information Resource (http://www.arabidopsis.org). Rice (*Oryza sativa* L.) (Ouyang et al., 2006) protein sequence data were downloaded from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/index.shtml). Cacao (*Theobroma cacao* L.) (Motamayor et al., 2013) protein sequence data were obtained from the JGI database (http://www.phytozome.net/), and grapevine (*Vitis vinifera* L.) (Jaillon et al., 2007) protein sequence data were obtained from the Ensembl Plants database (http://plants.ensembl.org/index.html).

Identification and conserved active site analysis of the *PRX* gene family in cotton

PRX gene Pfam domain ids (PF08534.9, PF00578.20, PF10417.8) in cotton were identified using the *PRX* protein sequence of *G. hirsutum* on the EMBL-EBI (https://www.ebi.ac.uk/Tools/hmmer/) website. The Pfam model files were downloaded from the Pfam database (http://pfam.xfam.org/) (Finn et al., 2019). The protein databases of four cotton species (*G. raimondii*, *G. arboreum*, *G. hirsutum*, and *G. barbadense*) and *Arabidopsis*, rice, *Cacao*, and grapevine were searched by HMMER (v3.2.1) (http://hmmer.org/) (E <= 0.001). All possible *PRX* genes were identified in the 8 crop species. The BLAST program was used to compare the protein database of *Arabidopsis* ($p <= 1E^{-10}$), and then the false positive *PRX* genes were deleted using the *Arabidopsis* protein annotation file (https://www.arabidopsis.org/). Finally, according to the conserved active site of the *PRX* protein given by the *PRX* protein database (http://csb.wfu.edu/prex/) (Nelson et al., 2011), to further ensure the accuracy of each candidate *PRX* gene, the MEGA7 (https://www.megasoftware.net/) software ClustalW was used to compare the candidate *PRX* protein sequences and delete the candidate *PRX* sequences in which there was no conserved active site (PxxxTxxC…*S*...*W/F*). Blastp alignment of the identified PRX proteins from the reference protein database of NCBI was performed (percent identity ≥90%). The theoretical isoelectric points (pI) and molecular weights (MW) of the *PRX* proteins were investigated within ExPASy (http://web.expasy.org/protparam/) (Finn et al., 2013).

Phylogenetic and gene structure analysis

The MEGA 7.0 software ClustalW (Kumar et al., 2016) was used to compare the protein sequences, and the maximum likelihood (ML) method with 1000 boot-strap replicates was used to construct a phylogenetic tree.

The exon/intron structure of the *PRX* genes was extracted from the corresponding cotton GFF files, and the *PRX* gene structure map was drawn by GSDS 2.0 (http://gsds.cbi.pku.edu.cn/) (Hu et al., 2014). MEME (4.11.4) (http://meme-suite.org/) (Bailey et al., 2006) was used to identify the conserved motif of cotton *PRX* with the following parameters: the maximum number of motifs was 10, and the optimal width was 6≤250. Then, TBtools (http://www.omicshare.com/forum/thread-1062-1-1.html) was used to construct a vector graph from the xml file generated by MEME.

Chromosomal mapping and gene duplication

To reveal the chromosomal mapping and duplication relationship of *PRX* genes in cotton, the genome databases of *G. raimondii*, *G. arboreum*, *G. hirsutum* and *G. barbadense* were constructed. The physical positions of *PRXs* in cotton were fetched from the corresponding GFF files. The *PRX* family duplicate data were identified by the MCScanX program ($p<=1E^{-20}$) (Ding et al., 2015), and comparative genome analysis of PRX family genes in four cotton species was carried out ($p<=1E^{-20}$) (Wang et al., 2013). The visualization was carried out with the CIRCOS (http://circos.ca/) tool (Krzywinski et al., 2009). The substitution rates of synonymous (Ks) and nonsynonymous (Ka) sites were calculated by the KaKs calculator program (Suyama et al., 2006). The divergence time was calculated by the formula $T=Ks/2^{\lambda}$ ($\lambda=1.5\times10^{-8}$) (Zhang et al.2006), where Ks was the synonymous substitution of each locus and r was the divergence rate of the plant genes. For dicotyledonous plants, r is considered to be 1.5×10^{-8} synonymous substitutions per site per year (Koch et al., 2000).

Analysis of cis-acting elements in the promoter region

By using the genome data files (GFF3) of the four cotton species, the promoter sequence of *PRX* genes (2500 bp upstream of the initiation codon "ATG") was extracted from the genome sequence of cotton (Wang et al., 2012). PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/search_CARE.html) (Lescot et al., 2002) was used to predict the cis-acting elements of the promoter sequence, and the abiotic stress response, plant growth and development, and hormone-induced cis-acting elements were analyzed (Table S2).

Based on the data of the cis-acting elements in the promoter sequence identified by PlantCARE, we used GSDS 2.0 (http://gsds.cbi.pku.edu.cn/) (Hu et al., 2014) to display the physical sites of the cis-acting elements in the *G. hirsutum* promoter sequence.

Plant material treatments and expression analysis

To analyze the expression pattern of *PRX* genes, PEG, NaCl and salicylic acid (SA) were used to simulate drought stress, osmotic stress and SA hormone induction, respectively, in plants. The seeds of *G. hirsutum* sGK9708 were soaked in flasks overnight, germinated in fine sand at 28°C

in the dark for 2 days, and transferred to a greenhouse for hydroponic growth. Hydroponic greenhouse conditions were as follows: 28°C day / 25°C night, with a 14 hour photoperiod and 70% relative humidity. Plants were grown in Hoagland nutrient solution (Xing et al., 2019). At the three-leaf stage, the seedlings were subjected to stress treatments. The cotton seedlings were transferred to nutrient solutions supplemented with 200 mM sodium chloride (NaCl), 15% PEG-6000, 0.1 mM SA and a blank control for salt, drought, hormone induction and control treatment, respectively. A total of 20 cotton seedlings received each treatment, and the treatment was repeated three times. After 0, 1, 3, 6, and 12 hours of treatment, the leaves (the second true leaf) and the rhizomes were removed from the seedlings and immediately frozen in liquid nitrogen. The leaves were stored at -80°C before RNA extraction.

Total RNA was isolated using an EasyPure Plant RNA Kit (TransGen, Beijing, China), and RNA reverse transcription was performed using a TransScript reverse transcription system (AT341). *Gh-PRX* specific primers (Table S3) were used to find candidate specific primers on the qPrimerDB-qPCR Primer Database (https://www.ncbi.nlm.nih.gov/) website. To ensure the specificity of *Gh-PRX* gene primers, candidate specific primers were used for BLAST homologous comparison in the Primer-BLAST database of NCBI (https://www.ncbi.nlm.nih.gov/) (National Center for Biotechnology Information) to ensure that specific primers amplified only the target gene fragment (Table S3). The general fluorescent dye mixture used was SYBR Green I (TransStart), and the reactions were performed with a 7500 Rapid Real-time PCR system (Roche). UBQ7 was used as the internal standard reference to measure the expression levels of the cDNA genes. The volume of each reaction was 10 μl, and the reaction conditions were as follows: 94°C for 5 min and 40 cycles of 94°C for 5 s, 55°C for 30 s and 72°C for 30 s. During the extension step, the fluorescence signal was measured, and the acquisition time was set to 30 s. Each cDNA sample was repeated three times, and the results were analyzed by the $2^{-\triangle\triangle CT}$ method (Khan-Malek et al., 2011).

Analysis of tissue expression and alternative shearing profiles of *PRX* genes

We used published *G. hirsutum* (TM-1) (Zhang et al., 2015) transcriptome data, including data from the roots, stems, leaves, cotyledons, petals, stamens, ovules, seeds, and 5, 10, 15 and 20 DPA fibers. The expression levels of *Gh-PRX* genes in various tissues were calculated using Log2 (FPKM) values. The expression values were normalized by Genesis software and illustrated with a heatmap (Sturn et al., 2002).

Samples of *G. hirsutum* sGK9708 RNA without treatment at -80°C were obtained. Total RNA was isolated using the EasyPure Plant RNA Kit (TransGen, Beijing, China), and RNA reverse transcription was performed using the TransScript reverse transcription system (AT321). The obtained first-strand cDNA was used for subsequent RT-PCR amplification. The full-length primers were designed using Primer Premier 5.0 software (Premier Biosoft International, Palo Alto, CA). After RT-PCR, the amplification products were detected by 1.2% agarose gel electrophoresis. After recovery and purification, the PCR products were ligated with the pEASY-Blunt vector (TransGen, Beijing, China) and transformed into *E. coli* (*Escherichia coli*)-competent DH5α cells (TransGen, Beijing, China). Thirty monoclonals were selected for PCR positive identification. The positive monoclonals were sequenced (Sangon, Shanghai), and the full-length sequences were obtained. These sequences were compared by Splign

(https://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi) and GSDS 2.0 (http://gsds.cbi.pku.edu.cn/index.php), and the AS events were displayed on the gene structure.

Subcellular localization of *PRX* proteins

The TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/) server website (Emanuelsson et al., 2007) was used to predict and analyze the localization of *G. hirsutum PRX* protease in plant cells.

To verify the distribution of *PRX* in plant cells, five pairs of homologous genes of 16 *Gh-PRX* genes were selected, and the CDS of the *PRX* genes was amplified by full-length primers with no terminator codon and restriction site and subcloned into the pcambia2300-eGFP (CAMBIA) transient expression vector. The accuracy of 10 recombinant vectors was ensured by primer-based PCR and sequencing analysis. Then, we cultured *Nicotiana benthamiana* seedlings at 28°C day/25°C night with a 14 hour photoperiod and 70% relative humidity for 1 month. When the third or fourth leaf was dark green and sufficiently thick, *Agrobacterium tumefaciens* with the recombinant vector was injected into the tobacco leaves (Poulsen et al., 2016), and the recombinant vector was introduced into tobacco leaf epidermal cells, which were cultured in darkness at 25°C for 24 hours and under light for 24 hours. A square slice of 1 cm^2 was cut next to the injection area, placed on a slide with PBS buffer, and then gently covered with a cover glass. The position of the fluorescent protein in tobacco epidermal cells was observed by laser confocal microscopy FV1200 (OLYMPUS), and the pcambia2300-eGFP (CAMBIA) empty vector without the *Gh-PRX* gene was used as a control. The nuclei were stained with DAPI solution (Solarbio) as a control (Kapuscinski, 1995).

Results

Genome-wide identification and conserved motif analysis of the *PRX* gene family in cotton

To identify the *PRX* gene of cotton, we used the Redoxin domain, 1-cys-PRX_C domain and AhpC-TSA domain of the cotton *PRX* protein sequence as references. The protein databases of four cotton species were searched by Hmmsearch (Bhaduri et al., 2004). According to the annotation of *Arabidopsis* and considering the existence of conserved active sites of *PRX* proteins, 46 *PRX* genes were retrieved (Table 1), of which 16, 8, 8 and 14 *PRX* genes were identified from *G. hirsutum*, *G. raimondii*, *G. arboreum* and *G. barbadense*, respectively. The *PRX* gene family was divided into six subfamilies according to the conserved motif and the related literature (Nelson et al., 2011). The MW and pI of *PRXs* were calculated using the Compute pI/Mw tool on the ExPASy (https://web.expasy.org/compute_pi/) website. The encoded protein lengths of 46 *PRX* genes varied from 100 to 300 amino acids, the predicted MWs were 13-30 kD (kilodaltons), and the pI values were approximately 4.0-10.0. Subcellular localization analysis predicted that 17 of 46 proteins localize to chloroplasts (Table S4), these members of the *PRX* family may be involved in regulating the redox status of protein in chloroplasts.

We performed multiple sequence alignment of 46 *PRX* proteins using ClustalW in MEGA7, and the specific, conserved and fully folded active sites of the *PRX* protein are presented in detail (Adak et al., 2017). The results showed that the active sites of the *PRX* protein sequence

identified by us were consistent with those of the protein active sites (Figure 1) by which the *PRX* family was classified in the *PRX* database. All *PRX* family proteins have conserved P, T, C, S and W/F amino acid sites, and the differences in the six subfamilies depend on the position of each conserved amino acid in the protein sequence. In particular, the fifth conserved amino acid in the PRX1 subfamily protein sequence is F, which is different from the fifth conserved amino acid in other subfamilies.

Phylogenetic and gene structure analysis of *PRXs* in cotton

We constructed a phylogenetic tree using 76 *PRX* genes from eight dicot species (*G. raimondii*, *G. arboreum*, *G. hirsutum*, *G. barbadense*, *Arabidopsis,* rice, *Cacao,* grapevine). The results show that cotton *PRX* genes are divided into six subfamilies: PRX1, PRX2, PRX5, PRX6, PRXQ, and TPX (Figure 2). Moreover, cotton *PRX* genes were concentrated on each subbranch, which indicated that these genes showed specific expansion in all six subfamilies.

The analysis of conserved domains helped to clarify the functional evolution and evolutionary relationships of *PRX* genes in cotton. A total of 10 conserved motifs were identified in the *PRX* proteins of the 4 cotton species. The motif logo of these conserved motifs is shown in Figure S1, and the flags for the 10 conserved motifs are shown in Figure 3c. The number of conserved motifs in each *PRX* protein ranged from 4 to 9. The conserved active structural site of the *PRX* protein exists in the motif in two forms, and the details can be seen in Figure S2. One of the conserved active structural sites of the *PRX* protein is located in motif 1 (Figure 3c and Figure S2a), which is shared by all members of the four subfamilies (TPX, PRX2, PRX5, and PRXQ). Another conserved active structural site of the *PRX* protein is motif 6 and motif 4 (Figure 3c and Figure S2b), which contain all members of the two subfamilies of PRX6 and PRX1. Therefore, according to the two forms in which the conserved active structural site of the *PRX* protein exists in the motif, we divided the *PRX* genes into two categories (PTC4 Prx and PTC2 Prx), which provides a new idea for the classification of the *PRX* family. This is different from the traditional 1-Cys Prx and 2-Cys Prx subfamilies according to the different number of conserved Cys residues (Choi et al., 2003). Compared to the motif containing a conserved active site, the other seven different motifs (2, 3, 5, 7, 8, 9, 10) play an important role in the functional differentiation of *PRX* family genes.

Exon-intron structural differences in gene families play a key role in the evolution of a polygene family. We found that the number of 46 *PRX* introns ranged from 0 to 6. *PRX* genes of the same subfamily showed a similar exon-intron structure (Figure 3); the TPX subfamily contained two intron sequences of different lengths, but none of the PRX2 subfamily genes had introns. In the PRX5 subfamily, there was an intron before the *GbPRX5-A* translation initiation site (TIS), and the remaining subfamilies had four introns. PRXQ subfamily genes all contained three introns, and most of the PRX1 and PRX6 subfamily genes also contained the same number of introns. However, there was an intron sequence before the TIS of *GbPRX12-D* and *GbPRX14-D*. The intron before the TIS is an integral part of the gene and plays an important role in regulating its expression (LI et al., 2013).

Chromosome mapping and duplication analysis

PRX genes were physically mapped on chromosomes in *G. hirsutum* and *G. barbadense.* We found that the mapping of *PRX* genes on the chromosomes of two cotton species was mirrored

(Figure 4a and 4b), which shows that the genes of this family were conserved during evolution. However, the homologs of five of these *PRX* genes (*GhPRX3-A*, *GbPRX6-A*, *GhPRX11-D*, *GbPRX8-D* and *GbPRX12-D*), marked in black, may have been lost during the evolution of the species. There are 6 pairs of homologous genes in the *G. hirsutum PRX* gene family, which are distributed on chromosomes Gh_A01, Gh_A04, Gh_A08, Gh_A10, Gh_D01, Gh_D04, Gh_D08 and Gh_D10. Among the 16 *Gh-PRX* genes, *GhPRX3-A* and *GhPRX16-D* were localized to scaffolds, and their exact positions have not yet been determined. There are five pairs of *PRX* homologous genes in *G. barbadense*, which are mapped on chromosomes Gb_A01, Gb_A08, Gb_A10, Gb_D01, Gb_D08 and Gb_D10. Compared with those of *G. hirsutum*, the *PRX* genes on Gb_A04 of *G. barbadense* were lost, so there is a lack of a pair of homologous genes. The comparative genomic analysis of *PRX* genes between *G. hirsutum* and *G. barbadense* shows that GbPRX6-A on Gb_A05 was caused by GhPRX11-D on Gh_D05 translocation (Figure 4b and Figure S3). Based on the chromosome physical mapping analysis and comparison of *PRX* genes between the *G. arboreum* and *G. raimondii* genomes, we found that the location of only *GaPRX5* and *GrPRX6* on Chr02 was relatively conserved in the two cotton species (Figure 4c, 4d and Figure S3).

Gene tandem duplication, segmental duplication and polyploid events are the main means of gene family expansion (Cannon et al., 2004). It is necessary to analyze gene duplication events at the genome level. We identified 8 pairs of duplications out of 16 *PRX* genes in *G. hirsutum* (*GhPRX8-D*:*GhPRX1-A*, *GhPRX9-D*:*GhPRX2-A*, *GhPRX10-D*:*GhPRX12-D*, *GhPRX10-D*:*GhPRX4-A*, *GhPRX12-D*:*GhPRX4-A*, *GhPRX5-A*:*GhPRX13-D*, *GhPRX15-D*:*GhPRX7-A*, and *GhPRX14-D*:*GhPRX6-A*) (Figure 4a), located on six pairs of chromosomes (Gh_A01: Gh_D01, Gh_A04: Gh_D04, Gh_D05: Gh_D08, Gh_A08: Gh_D05, Gh_A08: Gh_D08, and Gh_A10: Gh_D10). With the exception of *GhPRX10-D*:*GhPRX12-A* and *GhPRX10-D*:*GhPRX4-A*, which underwent segmental duplication, the others were produced during the process of genomic polyploidization. Compared with those in *G. hirsutum*, the duplicated genes on the *G. barbadense* Gb_A04 and Gb_D04 chromosomes changed (Figure 4b). *GbPRX8-D* on the Gb_D04 chromosome was preserved, but the corresponding homologous gene on Gb_A04 was lost during evolution. The other duplication genes were the same as those in *G. hirsutum* and remained conserved. The diploid *G. arboreum* genome is very similar to the tetraploid *G. hirsutum* genome and *G. raimondii* D subgenome, only one gene (GaPRX5) was lost in evolution, and there were also paired segmental duplication genes on the Ga_chr05 and Ga_chr08 chromosomes (Figure 4d). However, in *G. raimondii*, the physical position of the segmental duplication genes changed to the Gr_chr04 and Gr_chr09 chromosomes (Figure 4c). According to the KaKs calculation of the duplicated gene pair in the two cotton species (Table S1), the divergence time of the duplicated genes (*GrPRX3*:*GrPRX7*) on the *G. raimondii* chromosomes Gr_chr04 and Gr_chr09 was 24.81 MYA; the divergence time of the repetitive gene pairs (*GaPRX3*:*GaPRX7*) on *G. arboreum* Ga_chr05 and Ga_chr08 was 23.98 MYA. We conclude that *G. arboreum* underwent chromosomal translocation during species evolution (Time: 23.98 MYA), transferring the duplicated gene pair originally located on Gr_chr04 and Gr_chr09 to Ga_chr05 and Ga_chr08. After that, the diploid *G. arboreum* genome became the donor of the allotetraploid *G. hirsutum* genome and *G. raimondii* D subgenome, which provides evidence for *G. hirsutum* and *G. raimondii* D originating from *G. arboreum*.

Analysis of cis-acting elements in the promoter region

Cis-acting elements in plants play an important role in gene expression regulation. We found many cis-acting elements in the promoters of *PRX* genes in four cotton species (Table S2) and identified three main types of cis-acting elements.

The first type of cis-acting element is responsive to abiotic stress (Figure 5a). There were nine types of elements (ARE, DRE1, LTR, MBS, MYC, STRE, TC-rich repeats, WRE3, and WUN-motif), among which the most abundant cis-acting elements were related to drought induction. The elements related to drought induction were MBS and MYC, among which MYC was the most widespread. *GhPRX5-A*, *GhPRX6-A*, *GhPRX7-A*, *GhPRX9-D*, *GhPRX13-D*, *GbPRX1-A*, *GbPRX2-A*, *GbPRX5-A*, *GbPRX8-D*, *GbPRX9-D*, *GrPRX2*, *GrPRX8* and *GrPRX5* were the genes that contained more than five drought-induced elements among the four cotton species. Second, the cis-acting elements involved in the response to osmotic stress and defense were STRE- and TC-rich. In addition, there were also elements that responded to anaerobic induction, hypothermia and wound induction. It is speculated that *PRX* genes are sensitive to drought stress.

The second type of cis-acting element is responsive to plant hormones (Figure 5b). There are ten types of elements (ABRE, AuxRR-core, CGTCA-motif, ERE, GaRE-motif, P-box, TATC-box, TCA-element, TGaCG-motif, and TGa-element) that respond to abscisic acid, auxin, ethylene, gibberellin and SA. Among them, ABRE (ABA) and ERE were the most abundant in cotton *PRX* gene promoters and existed in 32 and 37 *PRX* genes, respectively, which means that the expression levels of *PRX* genes are regulated by abscisic acid and ethylene. We also found that there are three types of elements (GaRE-motif, P-box and TATC-box) involved in the gibberellin response in the *PRX* promoter. Among them, there were 2 and 3 GaRE-motif and TATC-box elements in 46 promoters, while there were 15 P-box elements, which indicated that the *PRX* gene had high specificity in response to gibberellin. In addition, auxin response elements (AuxRR-core, TGA-element), methyl jasmonate response elements (MeJA, TGACG-motif), and SA response elements (TCA-element) were also found in many *PRX* promoters. These results suggest that the expression of *PRX* genes is regulated by a variety of hormones.

The third type of cis-acting element is related to plant growth and development (Figure 5c). There are nine such types of elements (AACA_motif, AT-rich element, CAT-box, circadian, GCN4_motif, HD-Zip 1, MBSI, MYB31, and O2-site). Among them, only one type of cis-acting element is involved in the circadian motif, and this element exists only in the promoters of two *PRX* genes in *G. hirsutum*. The results showed that the *PRX* family genes were not particularly regulated by circadian rhythm. The largest number of cis-acting elements were responsive to Myb protein binding sequence sites, and there were two types of MYB protein sequence binding site elements (MBSI and MYB), indicating that the expression of the *PRX* genes is regulated by the MYB protein during plant growth and development. In addition, some cis-acting elements (GCN4_motif and AACA_motif) involved in endosperm expression, cis-acting regulatory elements (CAT-box) involved in meristem expression, and response elements (HD-Zip 1) involved in palisade mesophyll cell differentiation were also found. A cis-acting regulatory element (O2-site) involved in the regulation of zein metabolism was also detected.

In addition, there are some light-related cis-acting elements (Figure S4), which indicates the important role of *PRX* in biological processes and its response to abiotic stress, plant hormones and growth and development in cotton.

Expression profiles of *PRX* genes in *G. hirsutum* under stress

To investigate the regulation of *PRX* genes expression by cis-acting elements in the promoter region, we analyzed cis-acting elements in the promoter region of *Gh-PRX* genes that respond to drought (MBS, MYC), osmotic stress (STRE, TC-rich) and SA (TCA-element). The physical positions of these five cis-acting elements in the promoter region of *Gh-PRX* genes were plotted (Figure 6). An element (MYC) involved in drought induction exists in all *Gh-PRX* gene promoter regions, and the effects of drought on the expression level of this gene were predicted to be strong.

We used PEG-6000, NaCl and SA to treat *G. hirsutum* hydroponic seedlings. Gene samples were extracted from *G. hirsutum* for qRT-PCR verification (Figure 7 and 8). The results showed that the three treatment factors could significantly induce or inhibit the expression of *Gh-PRX* genes with response elements. Strangely, in the leaves of plants under salt stress (Figure 8), the three genes with TC+rich cis-acting elements were induced (*GhPRX11-D*, *GhPRX14-D*, and *GhPRX15-D*), and the gene expression levels increased. The seven genes with STRE cis-acting elements were repressed under salt stress (*GhPRX4-A*, *GhPRX5-A*, *GhPRX6-A*, *GhPRX9-D*, *GhPRX10-D*, *GhPRX12-D*, and *GhPRX13-D*). This may be due to the different response mechanisms of TC+rich and STRE cis-acting elements to stress. However, in the roots of plants under salt stress (Figure 7), the expression levels of three genes with TC+rich cis-acting elements increased rapidly at 1 hour after stress induction (*GhPRX11-D*, *GhPRX14-D*, and *GhPRX15-D*) and then decreased slowly. The expression levels of the seven genes with STRE cis-acting elements decreased at first and then increased (*GhPRX4-A*, *GhPRX5-A*, *GhPRX6-A*, *GhPRX9-D*, *GhPRX10-D*, *GhPRX12-D*, and *GhPRX13-D*). Under salt stress, the difference in *PRX* gene expression between roots and leaves may be due to the direct contact of plant roots with the salt stress environment, while the leaves are indirectly regulated by salt stress.

The expression levels of four genes with TCA cis-acting elements in roots and leaves were repressed under SA stress (*GhPRX2-A*, *GhPRX4-A*, *GhPRX6-A*, and *GhPRX8-D*) (Figure 7 and 8). This indicates that the response of the TCA element in the promoter region can inhibit the expression of *PRX* genes.

MYC cis-acting elements exist in all *Gh-PRX* genes, while MBS cis-acting elements exist in the promoter regions of six genes (*GhPRX4-A*, *GhPRX7-A*, *GhPRX10-D*, *GhPRX12-D*, *GhPRX13-D*, and *GhPRX15-D*). Under drought stress simulated by PEG-6000, the expression of all *Gh-PRX* genes in roots first increased and then decreased (Figure 7). Especially under treatment with PEG-6000 for 1 hour, the expression level of 12 of the 16 *PRX* genes increased significantly and then decreased rapidly, and the expression levels of the other four genes (*GhPRX3-A*, *GhPRX7-A*, *GhPRX9-D*, and *GhPRX14-D*) also showed a trend of first increasing and then decreasing at other time points. However, the expression level of only 3 of 16 *PRX* genes in leaves increased continuously (*GhPRX7-A*, *GhPRX14-D*, and *GhPRX3-A*) (Figure 8). The expression levels of eight genes decreased continuously (*GhPRX2-A*, *GhPRX5-A*, *GhPRX6-A*, *GhPRX8-D*, *GhPRX12-D*, *GhPRX13-D*, *GhPRX10-D* , and *GhPRX16-D*). After 1 hour of drought stress treatment, the expression levels of four genes decreased sharply (*GhPRX1-A*, *GhPRX4-A*, *GhPRX11-D*, and *GhPRX15-D*) and then returned to normal expression levels. The expression of only one gene (*GhPRX9-D*) increased rapidly and decreased rapidly under drought stress

treatment. This indicates that the MYC and MBS cis-acting elements responding to stress differentially regulate genes in different tissues of plants.

Expression and alternative shearing profiles of *PRX* genes in different tissues of *G. hirsutum*

It is important to study the function of genes to analyze the expression of genes in plant tissues and organs. We analyzed the available transcriptome data of various tissues and organs of *G. hirsutum* TM-1 (Zhang et al., 2015), including the following: roots, stems, leaves, cotyledons, petals, stamens, ovules, seeds, and 5, 10, 15 and 20 DPA fibers. A heatmap was used to show the expression of the *PRX* genes (Figure 9a). The homologous genes *GhPRX4-A* and *GhPRX12-D* were highly expressed in all tissues, but there were differential expression patterns between petals and stamens. The homologous genes *GhPRX2-A* and *GhPRX9-D* were highly expressed in roots, leaves, cotyledons, ovules, and 5 DPA fibers. The homologous genes *GhPRX7-A* and *GhPRX15-D* were highly expressed in 20 DPA fibers. The homologous genes *GhPRX1-A* and *GhPRX8-D* were highly expressed in cotyledons. The homologous genes *GhPRX6-A* and *GhPRX14-D* were differentially expressed in roots, leaves, cotyledons, petals, stamens, ovules, and 5 DPA fibers. *GhPRX3-A*, *GhPRX5-A*, *GhPRX10-D*, *GhPRX11-D*, *GhPRX13-D* and *GhPRX16-D* were expressed at low levels or were not expressed in all tissues.

Understanding and verifying the AS events of the *G. hirsutum PRX* family is an important step in the study of gene functional differentiation. Five main types of AS events were used for further analysis: exon skipping (ES), intron retention (IR), 5' or 3' alternative splice sites (A5SS or A3SS), alternative first exon (AFE), and alternative last exon (ALE). Because only a change in the coding sequence (CDS) can cause a protein change and influence the function of the gene, the AS events located on the CDS of the genes were used for further analysis. Using the full-length primers designed for the amplification of *PRX* gene cDNA, RT-PCR amplification was performed using the leaf and root cDNA of *G. hirsutum* sGK9708 as a template, and it was found that the amplified bands of *GhPRX14-D* were not single [Figure 9b and 9c], there were a few bands with different levels of brightness and unequal size. We speculate that *GhPRX14-D* may have multiple transcripts. The PCR products of *GhPRX14-D* were recovered, purified and ligated into the clone vector for cloning and sequencing. Through the analysis of the obtained sequence, it was found that there were two different transcripts of *GhPRX14-D* in the leaves of *G. hirsutum* (Table 2) and three in the roots. *GhPRX14-D* uses three AS methods, IR, ES and A3SS. *GhPRX14-Leaf-AS2* and *GhPRX14-Root-AS2* use one nonstandard AS site: 5'-AG. AA-3' (3 clones and 2 clones), which retains the first intron 3' 15-bp base [Figure 9c]. *GhPRX14-Root-AS3* use two nonstandard AS sites: 5'-AG. AT-3' (2 clones) and 5'-GG. TT-3' (2 clones), which retain the fourth intron 3' 14-bp base. The second, third, and fourth exons were skipped [Figure 9c].

Subcellular localization of *PRX* proteins

Determining the distribution of proteins in cells is an important step to verify the function of proteins. Using the TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/) website to predict and analyze proteins, we found that *PRX* proteins are mainly located in chloroplasts and plasma membranes. However, the specificity of a few gene prediction results was too low to determine the protein locations (Table S4).

To verify the predicted subcellular localizations of *PRX* proteins, we performed tobacco subcellular localization on five pairs of homologous genes of the *PRX* gene family, constructed a transient expression 35 s-PRX-eGFP vector, and carried out subcellular localization analysis. The results showed that the fluorescence signal of eGFP in the control group was dispersed in the whole tobacco leaf cells, while (Figure 10a) the fluorescence signals of homologous gene pair 1 (*GhPRX4-A* and *GhPRX12-D*) of the *PRX* family were mainly distributed in the cell membrane (Figure 10a). The fluorescence signals of homologous gene pair 2 (*GhPRX7-A* and *GhPRX15-D*) proteins were mainly distributed on the cell membrane and nucleus (Figure 10b). The fluorescence signals of homologous gene pair 3 (*GhPRX1-A* and *GhPRX8-D*) and homologous gene pair 4 (*GhPRX2-A* and *GhPRX9-D*) proteins were mainly distributed in the chloroplasts of cells (Figure 10c). For the homologous pair 5 (*GhPRX6-A* and *GhPRX14-D*), whose expression position was uncertain in the predicted results, the localization results showed that these proteins were mainly located on the cell chloroplasts (Figure 10c). Most of the localization results were consistent with those predicted by TargetP 1.1. The results of comprehensive prediction and verification showed that *PRX* proteins were mainly located in the chloroplast, and a few members were located in the cell membrane and nucleus.

Discussion

PRX is a ubiquitous and abundant protein and is very important for resisting oxidation and regulating the cell signal transduction pathway. Recent Advances: Peroxiredoxin involvement in the initiation and progression of human cancer (Hampton et al., 2018). In plants, PRX1 in chloroplasts regulates the redox status of the protein oscillates diurnally between hyperoxidation and reduction, thus protecting plants from a myriad of harsh environmental stresses (Lee et al., 2018). Most plastids 2-Cys PRX in plants are involved in key signaling processes, such as photosynthesis deactivation at night (Cerveau et al., 2019). These recent advances in *PRX* function indicate that this family genes are involved in photosynthesis and antioxidation pathways in chloroplasts, and our subcellular localization of *PRX* genes also shows that *PRX* genes is mainly distributed in chloroplast [Figure 10C]. In addition to their antioxidant function, these proteins also protect DNA from damage *in vivo* or *in vitro* (Banmeyer et al. 2005). These important functions of PRX indicate that they are an ancient and important enzyme family. However, no genome-wide identification and evolutionary process analysis of *PRX* gene family has been reported in cotton although four cotton species have been sequenced. According to our results, we identified 16, 9, 8 and 14 *PRX* genes from *G. hirsutum*, *G. raimondii*, *G. arboreum* and *G. barbadense*, respectively. All *PRX* protein sequences in cotton were aligned in the NCBI reference protein (refseq_protein) database. We found that the study of this family of proteins in cotton was still in the prediction stage (Table S5), which is most similar to the protein sequences of *Durio zibethinus* and *Citrus sinensis*.

We used DASP method to classify the subfamily of *PRX* in cotton. The correct subfamily classification of the *PRX* gene family has always been a problem because the classification of the *PRX* subfamily is usually independent of phylogenetic distribution, which is different from the phylogenetic classification based on the PSI-BLAST method. Using the DASP method to classify proteins according to their conserved active functional sites can provide a more accurate subfamily classification (Nelson et al., 2011). We found that all cotton *PRX* proteins contain an absolutely conserved active site cysteine, called cysteine peroxide (CP), and there are some highly conserved active sites, PXXXTXXCp, which is consistent with the conserved site of the

PRX gene family previously found by the DASP method (Nelson et al., 2011). The difference between the subfamilies is mainly the fourth and fifth active conserved sites [Figure 1]. By constructing a phylogenetic tree among eight dicotyledonous species, we found that *PRX* genes from cotton within the same subfamily always cluster together [Figure 2], which indicated that *PRX* showed specific amplification in each subfamily. Gene structure of *PRX* in the same subfamily is similar [Figure 3a], different *PRX* subfamily genes have different gene structures, and the number of exons is between 1 and 7 [Figure 3b]. As well as different cotton *PRX* subfamily genes shared similar motifs. Motif 1 contains conserved active structural sites (PxxxTxxC…*S*...*W/F*) and exists in 33 *PRX* proteins. Motif 4 and motif 6 also constitute conserved active structural sites (PxxxTxxC…*S*...*W/F*), which exist in 13 PRX proteins. As shown in figure [Figure 3c], All PRX genes share motif containing conserved structural sites.

Gene replication, transfer and deletion play an important role in the evolution of gene families. In order to better identify the evolution of *PRX* in cotton, we performed chromosomal mapping and comparative genome analysis of PRX family genes in cotton. The results showed that there were six pairs of homologous genes in *G. hirsutum PRX* family, but only five pairs of *PRX* homologous genes in *G. barbadense*. In the process of evolution, the *PRX* gene on *G. barbadense* Gb_A04 was lost, and GbPRX6-A on Gb_A05 was caused by GhPRX11-D on Gh_D05 translocation [Figure 4a and 4b]. These deletions and translocations in the evolutionary process led to slight differences between the *PRX* family of *G. hirsutum* and *G. barbadense*. There are segmental duplication events in all four cotton species, in which the segmental duplication events of *G. hirsutum* and *G. barbadense* occur on Chr05 and Chr08 of D subgenome, and the segmental duplication events of *G. arboreum* also occur on Chr05 and Chr08, but in *G. raimondii*, the segmental duplication occurs on the Chr04 and Chr09 [Figure 4c], according to the divergence time of duplication genes calculated by KaKs (Table S1), the divergence time of duplication gene pair of *G. raimondii* was earlier than *G. hirsutum*, *G. barbadense* and *G. arboretum*, and comparative genomic analysis among the four cotton species showed that the duplication genes was orthology relationship [Figure S3]. Based on the above analysis, we concluded that two pairs of duplication *PRX* genes in Chr05 and Chr08 of *G. hirsutum*, *G. barbadense* and *G. arboretum* were translocated and evolved from two pairs of duplication *PRX* genes on *G. raimondii* Chr04 and Chr09. Identifying the evolution of *PRX* genes in cotton is an essential step for understanding *PRX* family in cotton.

In order to understand the expression pattern of *PRX* during cotton growth and development, it is necessary to study the regulation of gene expression by cis-acting elements. We found that there are abundant cis-acting elements related to abiotic stress, hormone induction and plant growth and development regulation in the cotton promoter region [Figure 5], suggesting that *PRX* is widely involved in the growth and development of cotton plants. According to the four most types of cis-acting elements (MBS\ MYC, ABRE, ERE, and MBSI\ MYB) in the promoter region of the *PRX* gene, the expression level of the *PRX* gene may be sensitive to drought (Cho et al., 2012; Haddad et al., 2015; Xu et al., 2019), abscisic acid (Baier et al., 2004; Haslekås et al., 2003), ethylene (Tovar et al., 2011) and MYB protein, and a few *PRX* genes have been studied for their roles in abiotic stress and hormone-induced regulation (Cho et al., 2012; Xu et al., 2015). In the results of our study, two cis acting elements (TC+rich and STRE) that were responsive to salt stress showed opposite trends in the regulation of prx gene expression. Three genes (*GhPRX11-D*, *GhPRX14-D*, and *GhPRX15-D*) with TC+rich cis-acting elements were induced, and the expression level increased. While the seven genes (*GhPRX4-A*, *GhPRX5-A*,

GhPRX6-A, *GhPRX9-D*, *GhPRX10-D*, *GhPRX12-D*, and *GhPRX13-D*) with STRE cis-acting elements were repressed [Figure 7 and 8]. Therefore, this may be the reason for the differential expression of *PRX* under salt stress in some existing studies, the levels of *PRX* RNA and protein in mungbean decreased due to high salinity (Cho et al., 2012), while the expression level of PRX protein in rice (*Oryza sativa* L.) was upregulated due to high salinity (Xu et al., 2015). The TCA cis-acting element in the promoter region of the *PRX* gene can lead to a decrease in *PRX* gene expression in cotton in response to SA hormone induction [Figure 7 and 8]. At present, in plants, studies on TCA cis-acting elements in the promoter region have shown that these elements can be activated by SA (Salazar et al., 2007), whether the expression of genes related to this promoter is suppressed or up-regulated in other plants is still lacking relevant data. Our research on the TCA cis-acting element in cotton provides a reference for the role of this element in other plants. Under drought stress, MYC and MBS cis-acting elements exist in all Gh-PRX genes and differentially regulate *PRX* genes in different tissues of plants. In cotton roots, all *PRX* genes increased within 12 hours after drought stress. In cotton leaves, only the expression level of *GhPRX3-A* increased continuously, and the other *PRX* genes showed a downward trend within 12 hours [Fig. 7 and 8]. In previous studies, In drought-stressed alfalfa, *PRX* genes were repressed in the shoot but induced in the root (Kang and Udvardi, 2012), which was consistent with our verification of the expression trend of the *PRX* gene under drought stress. These results show that cis-acting elements play an important role in the regulation of plant expression.

AS events are a key posttranscriptional regulatory mechanism that can produce multiple transcripts and protein isomers, enrich the diversity of proteins and increase the complexity of protein functions. Studies have shown that AS is an important way to regulate plant gene expression (Syed et al., 2012) and is involved in many physiological metabolic processes, signal transduction and responses to external biological and abiotic stresses in plants (Barbazuk et al., 2008; Mastrangelo et al., 2012). A large amount of data have shown that IR is the most common AS event (Campbell et al., 2006; Rauch et al., 2013; Zhiguo et al., 2013; Reddy et al., 2013) and ES is the lowest common event (Barbazuk et al., 2008). Our analysis of the AS events of *PRX* genes in *G. hirsutum* showed that the main AS modes were IR, ES, and A5SS or A3SS (Table 2). According to the type and location of AS, we found that *GhPRX14-D* has the same type of transcripts in plant roots and leaves as do *GhPRX14-Leaf-AS2* and *GhPRX14-Root-AS2* [Figure 9c]. There is also a specific transcript, *GhPRX14-Root-AS3*, which has two alternative shearing modes, ES and A3SS. A transcript uses two or more alternative shearing methods, which are very common in plants (Zhang et al., 2019; Chen et al., 2018), as is differential splicing of homologous genes. We found that there were no AS events in *GhPRX6-A,* the homologous gene of *GhPRX14-D*, which indicated that the homologous gene was differentially expressed in the AS events. AS transcripts use two or more AS methods, and the differential splicing of homologous genes significantly increases the complexity of transcripts and enriches the types of proteins. Therefore, verifying the alternative shearing of *PRX* genes can make us understand the distribution of transcripts of this family genes in cotton, and provide valuable reference for the follow-up study of the diversity of this family of proteins in cotton.

Conclusions

In this study, we identified 16, 9, 8 and 14 *PRX* genes from *Gossypium hirsutum*, *G. raimondii*, *G. arboreum* and *G. barbadense*, respectively. The distribution, evolution, regulation mechanism, alternative splicing and location of *PRX* gene in cotton were analyzed and identified

in detail. This is the first systematic report on *PRX* in cotton, which will provide comprehensive data support for further study of the specific function of cotton *PRX*.

Acknowledgments

We would like to thank Peng Huo (The Institute of Cotton Research, Chinese Academy of Agricultural Sciences) for assistance with laser confocal microscopy (FV1200, OLYMPUS).

Abbreviations list

PRX, Peroxiredoxin; SA, salicylic acid; IR, intron retention; ES, exon skipping; A5SS or A3SS, 5' or 3' alternative splice sites; AS, alternative shearing; pI, isoelectric point; MW, molecular weights.

References

Adak S, Begley T P. RutA-catalyzed oxidative cleavage of the uracil amide involves formation of a flavin-N5-oxide[J]. Biochemistry, 2017, 56(29): 3708-3709. [CrossRef]

Andres-Mateos E, Perier C, Zhang L, et al. DJ-1 gene deletion reveals that DJ-1 is an atypical peroxiredoxin-like peroxidase[J]. Proceedings of the National Academy of Sciences, 2007, 104(37): 14807-14812. [CrossRef]

Baier M, Dietz K J. The plant 2‐Cys peroxiredoxin BAS1 is a nuclear‐encoded chloroplast protein: its expressional regulation, phylogenetic origin, and implications for its specific physiological function in plants[J]. The plant journal, 1997, 12(1): 179-190. [CrossRef]

Baier M, Ströher E, Dietz K J. The acceptor availability at photosystem I and ABA control nuclear expression of 2-Cys peroxiredoxin-A in Arabidopsis thaliana[J]. Plant and Cell Physiology, 2004, 45(8): 997-1006. [CrossRef]

Bailey T L, Williams N, Misleh C, et al. MEME: discovering and analyzing DNA and protein sequence motifs[J]. Nucleic acids research, 2006, 34(suppl_2): W369-W373. [CrossRef]

Barbazuk W B, Fu Y, McGinnis K M. Genome-wide analyses of alternative splicing in plants: opportunities and challenges[J]. Genome research, 2008, 18(9): 1381-1392. [CrossRef]

Berardini T Z, Mundodi S, Reiser L, et al. Functional annotation of the Arabidopsis genome using controlled vocabularies[J]. Plant physiology, 2004, 135(2): 745-755. [CrossRef]

Bhaduri A, Ravishankar R, Sowdhamini R. Conserved spatially interacting motifs of protein superfamilies: application to fold recognition and function annotation of genome data[J]. Proteins: Structure, Function, and Bioinformatics, 2004, 54(4): 657-670. [CrossRef]

Campbell M A, Haas B J, Hamilton J P, et al. Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis[J]. BMC genomics, 2006, 7(1): 327. [CrossRef]

Cannon S B, Mitra A, Baumgarten A, et al. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana[J]. BMC plant biology, 2004, 4(1): 10. [CrossRef]

614 Cerveau D, Henri P, Blanchard L, et al. Variability in the redox status of plant 2-Cys
615 peroxiredoxins in relation to species and light cycle[J]. Journal of experimental botany,
616 2019, 70(18): 5003-5016. [CrossRef]

617 Cerveau D, Ouahrani D, Marok M A, et al. Physiological relevance of plant 2‐Cys
618 peroxiredoxin overoxidation level and oligomerization status[J]. Plant, cell & environment,
619 2016, 39(1): 103-119. [CrossRef]

620 Chen Q, Han Y, Liu H, et al. Genome-wide association analyses reveal the importance of
621 alternative splicing in diversifying gene function and regulating phenotypic variation in
622 maize[J]. The Plant Cell, 2018, 30(7): 1404-1423. [CrossRef]

623 Cho C W, Chung E, Heo J E, et al. Molecular characterization of a 2-Cys peroxiredoxin induced
624 by abiotic stress in mungbean[J]. Plant Cell, Tissue and Organ Culture (PCTOC), 2012,
625 108(3): 473-484. [CrossRef]

626 Choi J, Choi S, Choi J, et al. Crystal Structure of Escherichia coli Thiol Peroxidase in the
627 Oxidized State INSIGHTS INTO INTRAMOLECULAR DISULFIDE FORMATION AND
628 SUBSTRATE BINDING IN ATYPICAL 2-CYS PEROXIREDOXINS[J]. Journal of
629 Biological Chemistry, 2003, 278(49): 49478-49486. [CrossRef]

630 Dietz K J, Jacob S, Oelze M L, et al. The function of peroxiredoxins in plant organelle redox
631 metabolism[J]. Journal of experimental botany, 2006, 57(8): 1697-1709. [CrossRef]

632 Ding M, Chen J, Jiang Y, et al. Genome-wide investigation and transcriptome analysis of the
633 WRKY gene family in Gossypium[J]. Molecular Genetics and Genomics, 2015, 290(1):
634 151-171. [CrossRef]

635 Du X, Huang G, He S, et al. Resequencing of 243 diploid cotton accessions based on an updated
636 A genome identifies the genetic basis of key agronomic traits[J]. Nature genetics, 2018,
637 50(6): 796. [CrossRef]

638 Emanuelsson O, Brunak S, Von Heijne G, et al. Locating proteins in the cell using TargetP,
639 SignalP and related tools[J]. Nature protocols, 2007, 2(4): 953. [CrossRef]

640 Fang J, Nakamura T, Cho D H, et al. S-nitrosylation of peroxiredoxin 2 promotes oxidative
641 stress-induced neuronal cell death in Parkinson's disease[J]. Proceedings of the National
642 Academy of Sciences, 2007, 104(47): 18742-18747. [CrossRef]

643 Fichman Y, Koncz Z, Reznik N, et al. SELENOPROTEIN O is a chloroplast protein involved in
644 ROS scavenging and its absence increases dehydration tolerance in Arabidopsis thaliana[J].
645 Plant science, 2018, 270: 278-291. [CrossRef]

646 Finkemeier I, Goodman M, Lamkemeyer P, et al. The mitochondrial type II peroxiredoxin F is
647 essential for redox homeostasis and root growth of Arabidopsis thaliana under stress[J].
648 Journal of Biological Chemistry, 2005, 280(13): 12168-12180. [CrossRef]

649 Finn R D, Bateman A, Clements J, et al. Pfam: the protein families database[J]. Nucleic acids
650 research, 2013, 42(D1): D222-D230. [CrossRef]

651 Finn R D, Mistry J, Tate J, et al. The Pfam protein families database[J]. Nucleic acids research,
652 2009, 38(suppl_1): D211-D222. [CrossRef]

653 Haddad R, Japelaghi R H. Isolation of grape peroxiredoxin gene responding to abiotic

654 stresses[J]. Russian Journal of Plant Physiology, 2015, 62(6): 856-865. [CrossRef]

655 Hampton M B, Vick K A, Skoko J J, et al. Peroxiredoxin involvement in the initiation and
656 progression of human cancer[J]. Antioxidants & redox signaling, 2018, 28(7): 591-
657 608..[CrossRef]

658 Haslekås C, Grini P E, Nordgard S H, et al. ABI3 mediates expression of the peroxiredoxin
659 antioxidant AtPER1 gene and induction by oxidative stress[J]. Plant molecular biology,
660 2003, 53(3): 313-326. [CrossRef]

661 Hofmann B, Hecht H J, Flohé L. Peroxiredoxins[J]. Biological chemistry, 2002, 383(3-4): 347-
662 364. [CrossRef]

663 Horling F , Baier M , Dietz K J . Redox-regulation of the expression of the peroxide-detoxifying
664 chloroplast 2-Cys peroxiredoxin in the liverwortRiccia fluitans[J]. Planta, 2001, 214(2):304-
665 313. [CrossRef]

666 Hu B, Jin J, Guo A Y, et al. GSDS 2.0: an upgraded gene feature visualization server[J].
667 Bioinformatics, 2014, 31(8): 1296-1297. [CrossRef]

668 Jaillon O, Aury J M, Noel B, et al. The grapevine genome sequence suggests ancestral
669 hexaploidization in major angiosperm phyla[J]. nature, 2007, 449(7161): 463. [CrossRef]

670 Kang Y, Udvardi M. Global regulation of reactive oxygen species scavenging genes in alfalfa
671 root and shoot under gradual drought stress and recovery[J]. Plant signaling & behavior,
672 2012, 7(5):539-543. [CrossRef]

673 Khan-Malek R, Wang Y. Statistical analysis of quantitative RT-PCR results[M]//Drug Safety
674 Evaluation. Humana Press, 2011: 227-241. [CrossRef]

675 Knoops B, Loumaye E É, Van der Eecken V. Evolution of the peroxiredoxins[M]//Peroxiredoxin
676 systems. Springer, Dordrecht, 2007: 27-40. [CrossRef]

677 Koch M A, Haubold B, Mitchell-Olds T. Comparative evolutionary analysis of chalcone
678 synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera
679 (Brassicaceae)[J]. Molecular biology and evolution, 2000, 17(10): 1483-1498. [CrossRef]

680 Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative
681 genomics[J]. Genome research, 2009, 19(9): 1639-1645. [CrossRef]

682 Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0
683 for bigger datasets[J]. Molecular biology and evolution, 2016, 33(7): 1870-1874. [CrossRef]

684 Lamkemeyer P, Laxa M, Collin V, et al. Peroxiredoxin Q of Arabidopsis thaliana is attached to
685 the thylakoids and functions in context of photosynthesis[J]. The Plant Journal, 2006, 45(6):
686 968-981. [CrossRef]

687 Lee E S, Kang C H, Park J H, et al. Physiological significance of plant peroxiredoxins and the
688 structure-related and multifunctional biochemistry of peroxiredoxin 1[J]. Antioxidants &
689 redox signaling, 2018, 28(7): 625-639. [CrossRef]

690 Lee T H, Kim S U, Yu S L, et al. Peroxiredoxin II is essential for sustaining life span of
691 erythrocytes in mice[J]. Blood, 2003, 101(12): 5033-5038. [CrossRef]

692 Lescot M, Déhais P, Thijs G, et al. PlantCARE, a database of plant cis-acting regulatory
693 elements and a portal to tools for in silico analysis of promoter sequences[J]. Nucleic acids

694 research, 2002, 30(1): 325-327. [CrossRef]

695 LI J, LI L, ZHANG G. A retrotransposon-like insertion——Characterization of Intron 1 in
696 peroxiredoxin V from bay scallop Argopecten irradians[J]. Marine Sciences, 2013 (2): 02.

697 Mastrangelo A M, Marone D, Laidò G, et al. Alternative splicing: enhancing ability to cope with
698 stress via transcriptome plasticity[J]. Plant Science, 2012, 185: 40-49. [CrossRef]

699 Motamayor J C, Mockaitis K, Schmutz J, et al. The genome sequence of the most widely
700 cultivated cacao type and its use to identify candidate genes regulating pod color[J].
701 Genome biology, 2013, 14(6): r53. [CrossRef]

702 Nelson K J, Knutson S T, Soito L, et al. Analysis of the peroxiredoxin family: using active‐site
703 structure and sequence information for global classification and residue analysis[J].
704 Proteins: Structure, Function, and Bioinformatics, 2011, 79(3): 947-964. [CrossRef]

705 Newick K, Cunniff B, Preston K, et al. Peroxiredoxin 3 is a redox-dependent target of
706 thiostrepton in malignant mesothelioma cells[J]. PloS one, 2012, 7(6): e39404. [CrossRef]

707 Nonn L, Berggren M, Powis G. Increased Expression of Mitochondrial Peroxiredoxin-3
708 (Thioredoxin Peroxidase-2) Protects Cancer Cells Against Hypoxia and Drug-Induced
709 Hydrogen Peroxide-Dependent Apoptosis11CA52995 and CA772049.[J]. Molecular Cancer
710 Research, 2003, 1(9):682-689. [CrossRef]

711 Ojeda V, Pérez-Ruiz J M, Cejudo F J. 2-Cys Peroxiredoxins participate in the oxidation of
712 chloroplast enzymes in the dark[J]. Molecular plant, 2018, 11(11): 1377-1388. [CrossRef]

713 Oláhová M, Taylor S R, Khazaipoul S, et al. A redox-sensitive peroxiredoxin that is important
714 for longevity has tissue-and stress-specific roles in stress resistance[J]. Proceedings of the
715 National Academy of Sciences, 2008, 105(50): 19839-19844. [CrossRef]

716 Ouyang S, Zhu W, Hamilton J, et al. The TIGR rice genome annotation resource: improvements
717 and new features[J]. Nucleic acids research, 2006, 35(suppl_1): D883-D887. [CrossRef]

718 Park J J, Chang H W, Jeong E J, et al. Peroxiredoxin IV protects cells from radiation-induced
719 apoptosis in head-and-neck squamous cell carcinoma[J]. International Journal of Radiation
720 Oncology* Biology* Physics, 2009, 73(4): 1196-1202. [CrossRef]

721 Paterson A H, Wendel J F, Gundlach H, et al. Repeated polyploidization of Gossypium genomes
722 and the evolution of spinnable cotton fibres[J]. Nature, 2012, 492(7429): 423. [CrossRef]

723 Pérez-Ruiz J M, Naranjo B, Ojeda V, et al. NTRC-dependent redox balance of 2-Cys
724 peroxiredoxins is needed for optimal function of the photosynthetic apparatus[J].
725 Proceedings of the National Academy of Sciences, 2017, 114(45): 12069-12074. [CrossRef]

726 Pérez-Ruiz, Juan Manuel, Naranjo, Belén, Ojeda V, Guinea M , Cejudo, FJ (2017). Ntrc-
727 dependent redox balance of 2-cys peroxiredoxins is needed for optimal function of the
728 photosynthetic apparatus. Proceedings of the National Academy of Sciences, 201706003.
729 [CrossRef]

730 Poulsen L R, Palmgren M G, López-Marqués R L. Transient expression of P-type ATPases in
731 tobacco epidermal cells[M]//P-Type ATPases. Humana Press, New York, NY, 2016: 383-
732 393. [CrossRef]

733 Pulido P, Spínola M C, Kirchsteiger K, et al. Functional analysis of the pathways for 2-Cys

peroxiredoxin reduction in Arabidopsis thaliana chloroplasts[J]. Journal of experimental botany, 2010, 61(14): 4043-4054. [CrossRef]

Rauch H B, Patrick T L, Klusman K M, et al. Discovery and expression analysis of alternative splicing events conserved among plant SR proteins[J]. Molecular biology and evolution, 2013, 31(3): 605-613. [CrossRef]

Reddy A S N, Marquez Y, Kalyna M, et al. Complexity of the alternative splicing landscape in plants[J]. The Plant Cell, 2013, 25(10): 3657-3683. [CrossRef]

Salazar M, González E, Casaretto J A, et al. The promoter of the TLC1. 1 retrotransposon from Solanum chilense is activated by multiple stress-related signaling molecules[J]. Plant cell reports, 2007, 26(10): 1861-1868. [CrossRef]

Sturn A, Quackenbush J, Trajanoski Z. Genesis: cluster analysis of microarray data[J]. Bioinformatics, 2002, 18(1): 207-208. [CrossRef]

Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments[J]. Nucleic acids research, 2006, 34(suppl_2): W609-W612. [CrossRef]

Syed N H, Kalyna M, Marquez Y, et al. Alternative splicing in plants–coming of age[J]. Trends in plant science, 2012, 17(10): 616-623. [CrossRef]

Tovar-Méndez A, Matamoros M A, Bustos-Sanmamed P, et al. Peroxiredoxins and NADPH-dependent thioredoxin systems in the model legume Lotus japonicus[J]. Plant physiology, 2011, 156(3): 1535-1547. [CrossRef]

Wang M, Tu L, Yuan D, et al. Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense[J]. Nature genetics, 2019, 51(2): 224. [CrossRef]

Wang Y, Li J, Paterson A H. MCScanX-transposed: detecting transposed gene duplications based on multiple colinearity scans[J]. Bioinformatics, 2013, 29(11): 1458-1460. [CrossRef]

Wang Y, Tang H, DeBarry J D, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity[J]. Nucleic acids research, 2012, 40(7): e49-e49. [CrossRef]

Wenders F, Chainiaux F, Remacle J, et al. Retrovirally mediated overexpression of peroxiredoxin VI increases the survival of WI-38 human diploid fibroblasts exposed to cytotoxic doses of tert-butylhydroperoxide and UVB[J]. Biogerontology, 2003, 4(3): 125-131. [CrossRef]

Xing D, Chen X, Wu Y, et al. Rapid prediction of the re-watering time point of Orychophragmus violaceus L. based on the online monitoring of electrophysiological indexes[J]. Scientia Horticulturae, 2019, 256: 108642. [CrossRef]

Xu J, Lan H, Fang H, et al. Quantitative proteomic analysis of the rice (Oryza sativa L.) salt response[J]. PLoS One, 2015, 10(3): e0120978. [CrossRef]

Xu Z, Wang M, Guo Z, et al. Identification of a 119-bp Promoter of the Maize Sulfite Oxidase Gene (ZmSO) That Confers High-Level Gene Expression and ABA or Drought Inducibility

774 in Transgenic Plants[J]. International journal of molecular sciences, 2019, 20(13): 3326.
775 [CrossRef]

776 Zhang D, Yang J F, Gao B, et al. Identification, evolution and alternative splicing profile
777 analysis of the splicing factor 30 (SPF30) in plant species[J]. Planta, 2019, 249(6): 1997-
778 2014. [CrossRef]

779 Zhang H, Ni Z, Chen Q, et al. Proteomic responses of drought-tolerant and drought-sensitive
780 cotton varieties to drought stress[J]. Molecular genetics and genomics, 2016, 291(3): 1293-
781 1303. [CrossRef]

782 Zhang T, Hu Y, Jiang W, et al. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc.
783 TM-1) provides a resource for fiber improvement[J]. Nature biotechnology, 2015, 33(5):
784 531. [CrossRef]

785 Zhang Z, Li J, Zhao X Q, et al. KaKs_Calculator: calculating Ka and Ks through model selection
786 and model averaging[J]. Genomics, proteomics & bioinformatics, 2006, 4(4): 259-263.
787 [CrossRef]

788 Zhiguo E, Wang L, Zhou J. Splicing and alternative splicing in rice and humans[J]. BMB
789 reports, 2013, 46(9): 439. [CrossRef]

790

Figure 1

The conserved active sites of the *PRX* protein sequence in cotton.

The multiple sequence alignment of 46 PRX proteins in cotton using ClustalW in MEGA7, and the active conserved sites were marked with columnar bars.

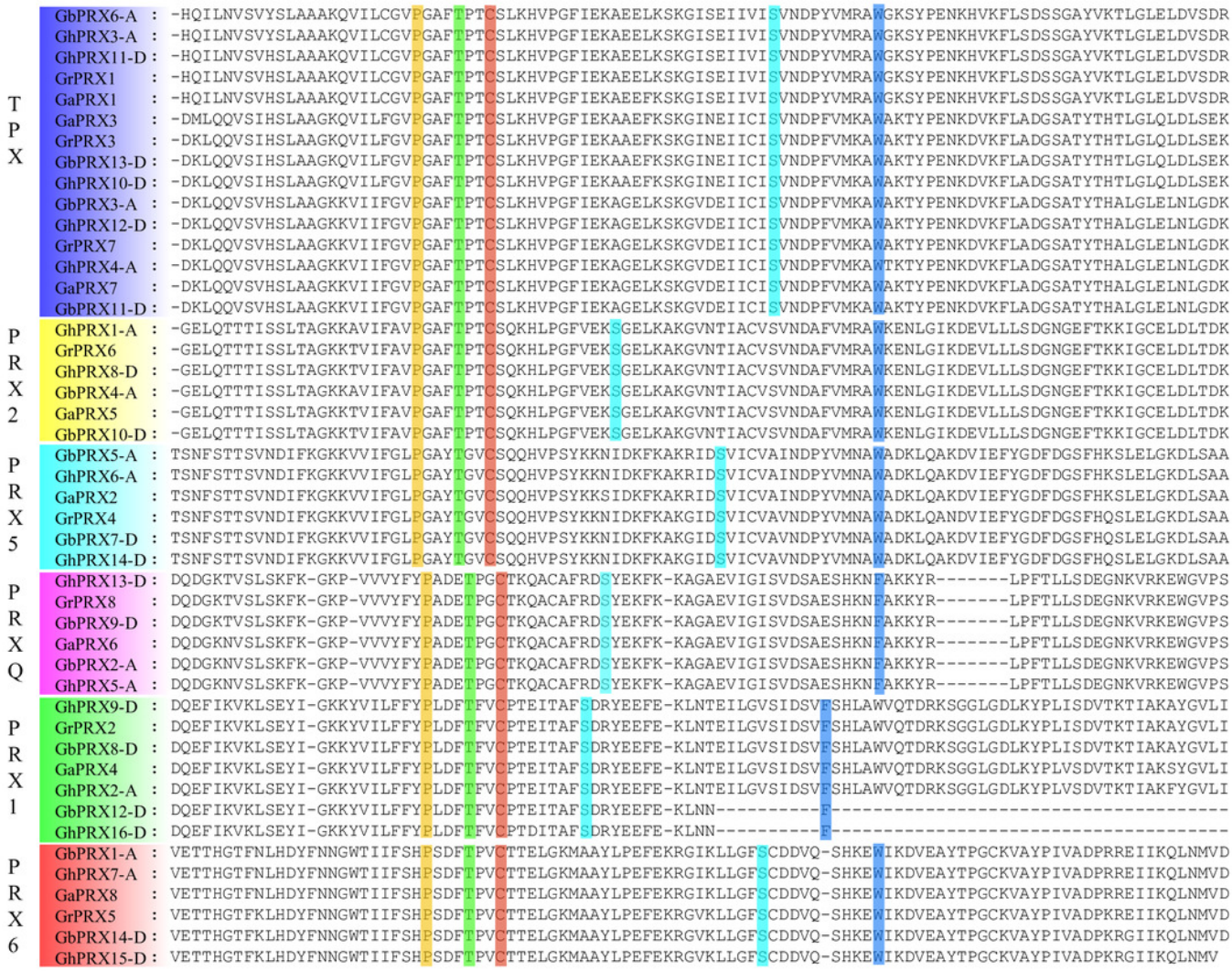


Figure 2

Phylogenetic analysis of *PRX* family members.

A maximum likelihood phylogenetic tree was constructed with *PRX* protein sequences from *G. hirsutum*, *G. barbadense*, *G. arboreum*, *G. raimondii*, *Arabidopsis*, *rice*, *cacao*, and *grapevine.*

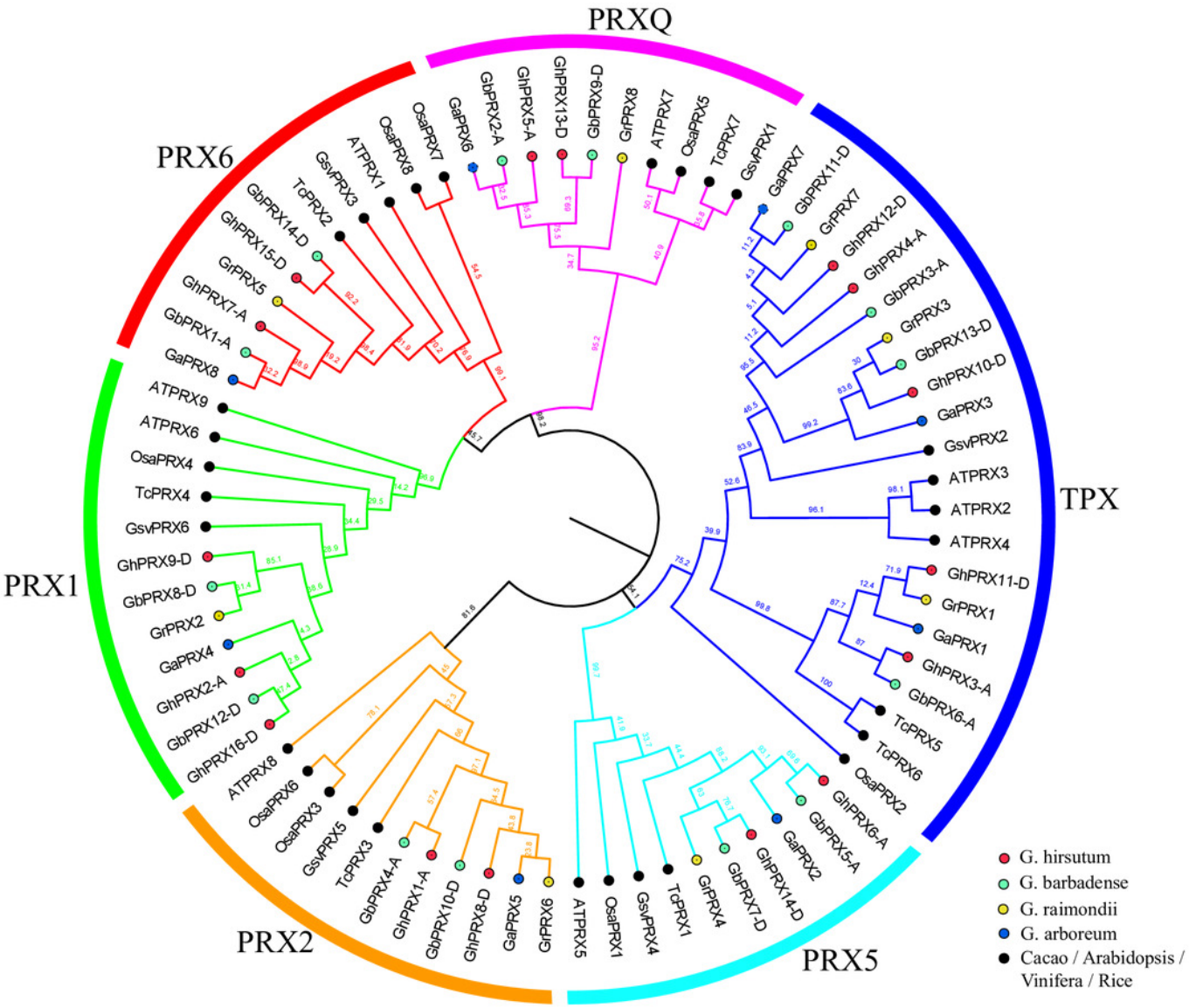


Figure 3

Phylogenetic tree, gene structure and conserved motif of *PRXs* in cotton.

(a) Phylogenetic analysis of *PRX* proteins in cotton. **(b)** The exon/intron distribution of *PRX* genes. **(c)** The distribution of domains in *PRX* proteins. The conserved active site motif of the PTC4 PRX subclass is circled in red, and the conserved active site motif of the PTC2 PRX subclass is circled in blue.

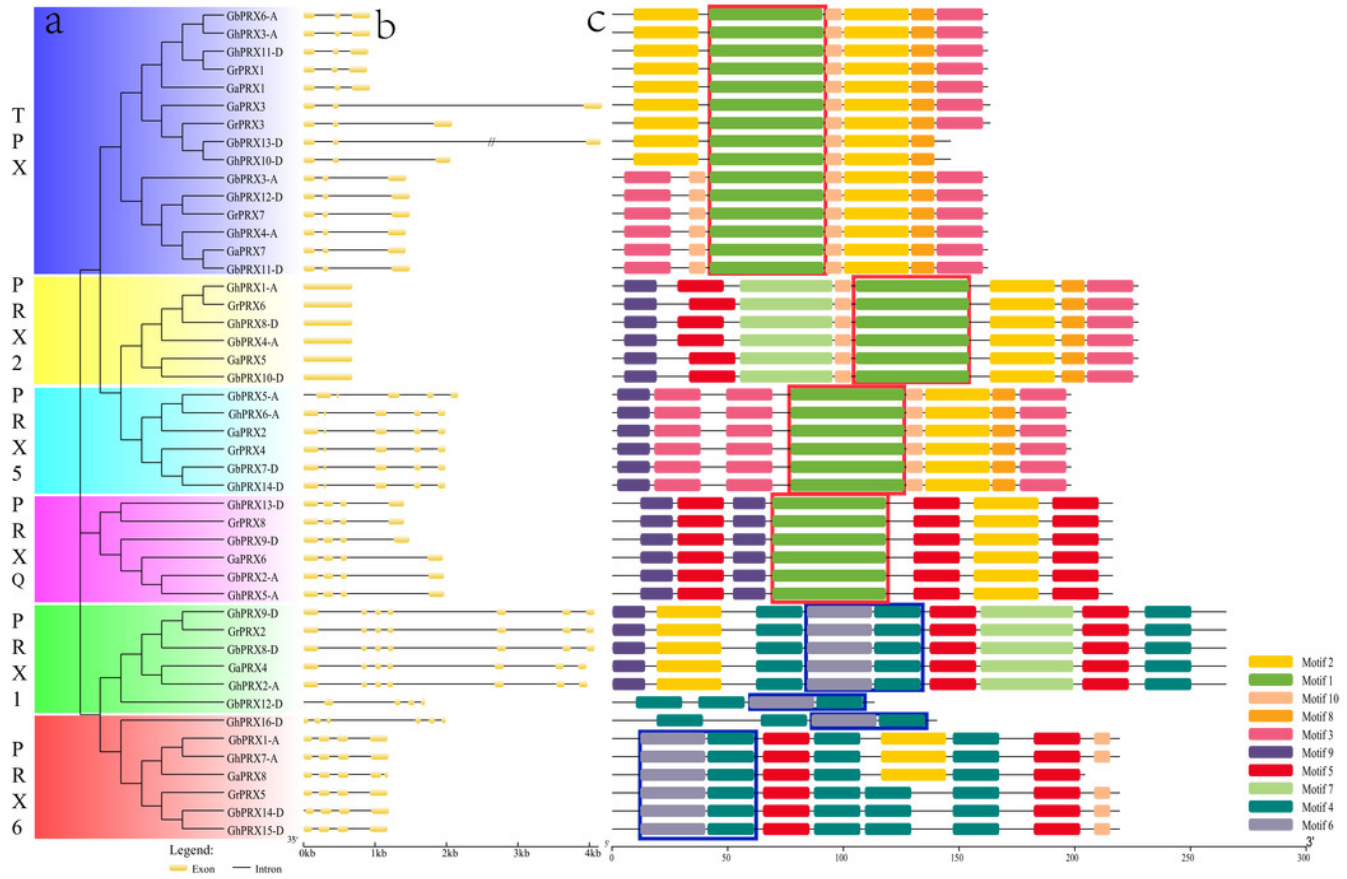


Figure 4

Chromosome physical mapping and duplication events of *PRX* genes in the genomes of four cotton species.

Different color blocks represent different chromosomes, and the minimum scale represents 1000000 chromosome units. red links mark the physical site of genes on chromosomes. The homologous genes of the genes marked in black have been lost during the evolution of the species. The segmental duplication event occurred in the genes marked in blue . The homologous genes produced by segmental duplication events are connected by a blue curve, and the homologous genes produced by genomic polyploidization events are connected by a red curve. (a) Chromosomal physical mapping and duplication events of PRX genes in the *G. hirsutum* genome. (b) Chromosomal physical mapping and duplication events of PRX genes in the *G. barbadense* genome. (c) Chromosomal physical mapping and duplication events of PRX genes in the *G. raimondii* genome. (d) Chromosomal physical mapping and duplication events of PRX genes in the *G. arboreum* genome.

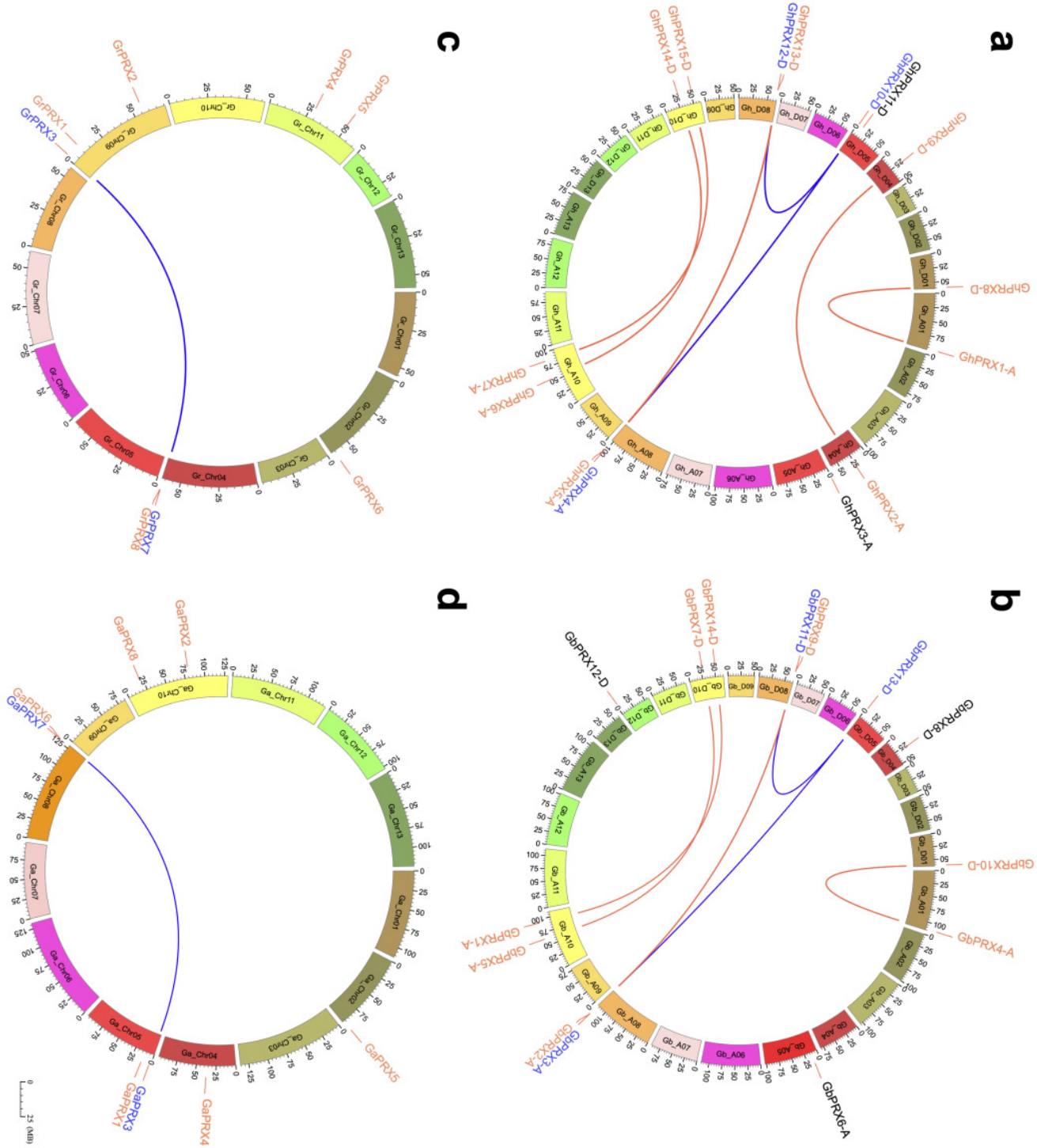
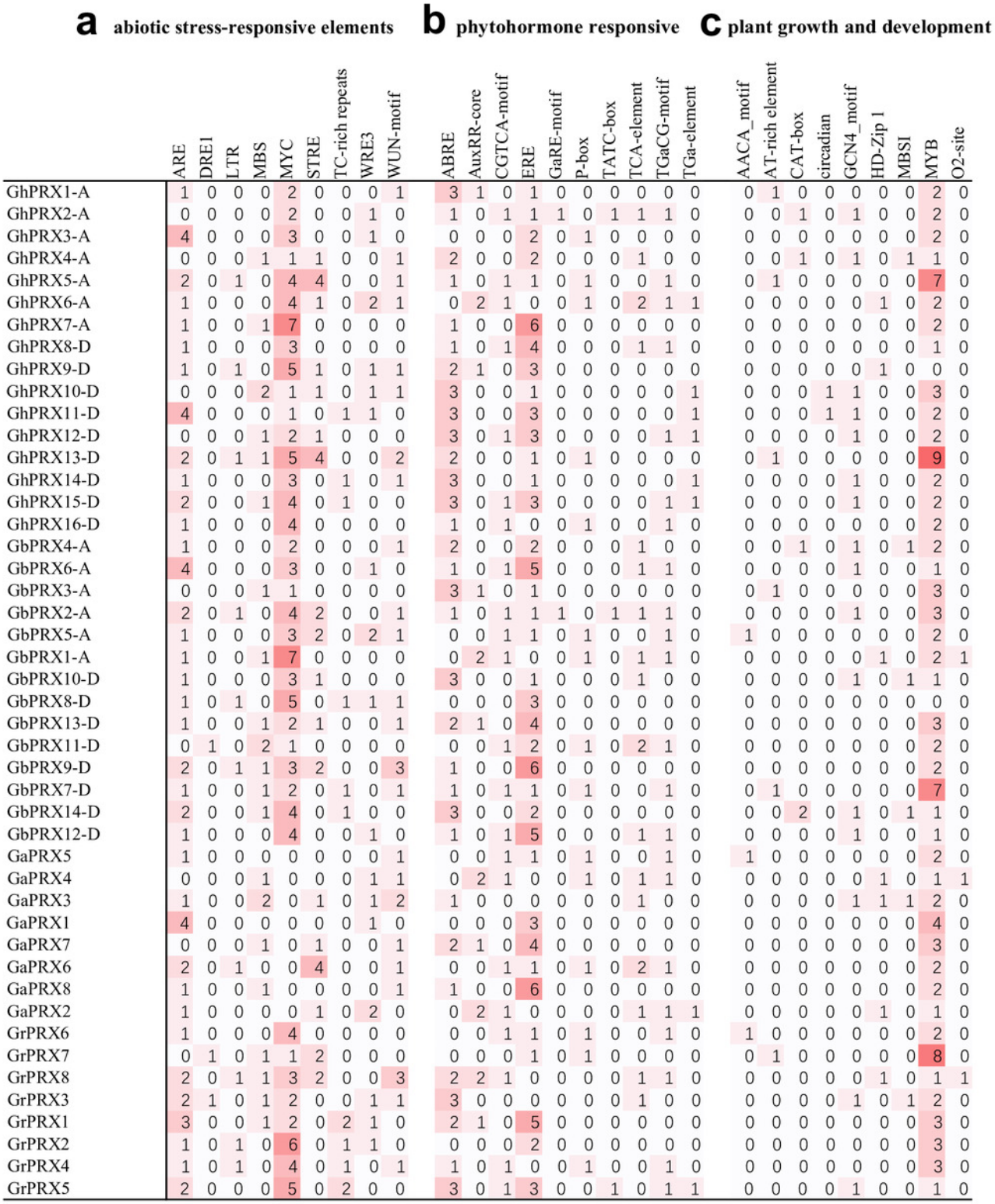

a
GhPRX15-D
GhPRX14-D
GhPRX13-D
GhPRX12-D
GhPRX11-D
GhPRX10-D
GhPRX9-D
GhPRX8-D
GhPRX1-A
GhPRX2-A
GhPRX3-A
GhPRX4-A
GhPRX5-A
GhPRX6-A
GhPRX7-A
b
GbPRX14-D
GbPRX13-D
GbPRX12-D
GbPRX11-D
GbPRX10-D
GbPRX8-D
GbPRX4-A
GbPRX6-A
GbPRX1-A
GbPRX2-A
GbPRX3-A
GbPRX5-A
GbPRX7-D
c
GrPRX1
GrPRX2
GrPRX3
GrPRX4
GrPRX5
GrPRX6
GrPRX7
d
GaPRX1
GaPRX2
GaPRX3
GaPRX4
GaPRX5
GaPRX6
GaPRX7
GaPRX8

Figure 5

Cis-acting elements on the promoter of *PRX* genes in cotton.

(a) Abiotic stress-related cis-acting elements. (b) Plant hormone-induced cis-acting elements.

(c) Cis-acting elements related to the growth and development of plants.



	a abiotic stress-responsive elements									b phytohormone responsive										c plant growth and development									
	ARE	DRE1	LTR	MBS	MYC	STRE	TC-rich repeats	WRE3	WUN-motif	ABRE	AuxRR-core	CGTCA-motif	ERE	GaRE-motif	P-box	TATC-box	TCA-element	TGaCG-motif	TGa-element	AACA_motif	AT-rich element	CAT-box	circadian	GCN4_motif	HD-Zip 1	MBSI	MYB	O2-site
GhPRX1-A	1	0	0	0	2	0	0	0	1	3	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0
GhPRX2-A	0	0	0	0	2	0	0	1	0	1	0	1	1	1	0	1	1	1	0	0	0	1	0	1	0	0	2	0
GhPRX3-A	4	0	0	0	3	0	0	1	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0
GhPRX4-A	0	0	0	1	1	1	0	0	1	2	0	0	2	0	0	0	1	0	0	0	0	1	0	1	0	1	1	0
GhPRX5-A	2	0	1	0	4	4	0	0	1	1	0	1	1	0	1	0	0	1	0	0	1	0	0	0	0	0	7	0
GhPRX6-A	1	0	0	0	4	1	0	2	1	0	2	1	0	0	1	0	2	1	1	0	0	0	0	0	1	0	2	0
GhPRX7-A	1	0	0	1	7	0	0	0	0	1	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
GhPRX8-D	1	0	0	0	3	0	0	0	0	1	0	1	4	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0
GhPRX9-D	1	0	1	0	5	1	0	1	1	2	1	0	3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
GhPRX10-D	0	0	0	2	1	1	0	1	1	3	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	3	0
GhPRX11-D	4	0	0	0	1	0	1	1	0	3	0	0	3	0	0	0	0	0	1	0	0	0	1	1	0	0	2	0
GhPRX12-D	0	0	0	1	2	1	0	0	0	3	0	1	3	0	0	0	0	1	1	0	0	0	0	1	0	0	2	0
GhPRX13-D	2	0	1	1	5	4	0	0	2	2	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	9	0
GhPRX14-D	1	0	0	0	3	0	1	0	1	3	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	2	0
GhPRX15-D	2	0	0	1	4	0	1	0	0	3	0	1	3	0	0	0	0	1	1	0	0	0	0	1	0	0	2	0
GhPRX16-D	1	0	0	0	4	0	0	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	2	0
GbPRX4-A	1	0	0	0	2	0	0	0	1	2	0	0	2	0	0	0	1	0	0	0	0	1	0	1	0	1	2	0
GbPRX6-A	4	0	0	0	3	0	0	1	0	1	0	1	5	0	0	0	1	1	0	0	0	0	0	1	0	0	1	0
GbPRX3-A	0	0	0	1	1	0	0	0	0	3	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	3	0
GbPRX2-A	2	0	1	0	4	2	0	0	1	1	0	1	1	1	0	1	1	1	0	0	0	0	0	1	0	0	3	0
GbPRX5-A	1	0	0	0	3	2	0	2	1	0	0	1	1	0	1	0	0	1	0	1	0	0	0	0	0	0	2	0
GbPRX1-A	1	0	0	1	7	0	0	0	0	0	2	1	0	0	1	0	1	1	0	0	0	0	0	0	1	0	2	1
GbPRX10-D	1	0	0	0	3	1	0	0	0	3	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	1	1	0
GbPRX8-D	1	0	1	0	5	0	1	1	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GbPRX13-D	1	0	0	1	2	1	0	0	1	2	1	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
GbPRX11-D	0	1	0	2	1	0	0	0	0	0	0	1	2	0	1	0	2	1	0	0	0	0	0	0	0	0	2	0
GbPRX9-D	2	0	1	1	3	2	0	0	3	1	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
GbPRX7-D	1	0	0	1	2	0	1	0	1	1	0	1	1	0	1	0	0	1	0	0	1	0	0	0	0	0	7	0
GbPRX14-D	2	0	0	1	4	0	1	0	0	3	0	0	2	0	0	0	0	0	0	0	0	2	0	1	0	1	1	0
GbPRX12-D	1	0	0	0	4	0	0	1	0	1	0	1	5	0	0	0	1	1	0	0	0	0	0	1	0	0	1	0
GaPRX5	1	0	0	0	0	0	0	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	0	0	0	0	2	0
GaPRX4	0	0	0	1	0	0	0	1	1	0	2	1	0	0	1	0	1	1	0	0	0	0	0	0	1	0	1	1
GaPRX3	1	0	0	2	0	1	0	1	2	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	2	0
GaPRX1	4	0	0	0	0	0	0	1	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0
GaPRX7	0	0	0	1	0	1	0	0	1	2	1	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
GaPRX6	2	0	1	0	0	4	0	0	1	0	0	1	1	0	1	0	2	1	0	0	0	0	0	0	0	0	2	0
GaPRX8	1	0	0	1	0	0	0	0	1	1	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
GaPRX2	1	0	0	0	0	1	0	2	0	0	2	1	0	0	0	0	1	1	1	0	0	0	0	0	1	0	1	0
GrPRX6	1	0	0	0	4	0	0	0	0	0	0	1	1	0	1	0	0	1	0	1	0	0	0	0	0	0	2	0
GrPRX7	0	1	0	1	1	2	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	8	0
GrPRX8	2	0	1	1	3	2	0	0	3	2	2	1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	1	1
GrPRX3	2	1	0	1	2	0	0	1	1	3	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	2	0
GrPRX1	3	0	0	1	2	0	2	1	0	2	1	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
GrPRX2	1	0	1	0	6	0	1	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
GrPRX4	1	0	1	0	4	0	1	0	1	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	3	0
GrPRX5	2	0	0	0	5	0	2	0	0	3	0	1	3	0	0	1	0	1	1	0	0	0	0	1	0	0	1	0

Figure 6

Analysis of cis-acting elements in the promoter region of *Gh-PRX* genes.

The promoter region has 1500 bp before the gene translation initiation site.

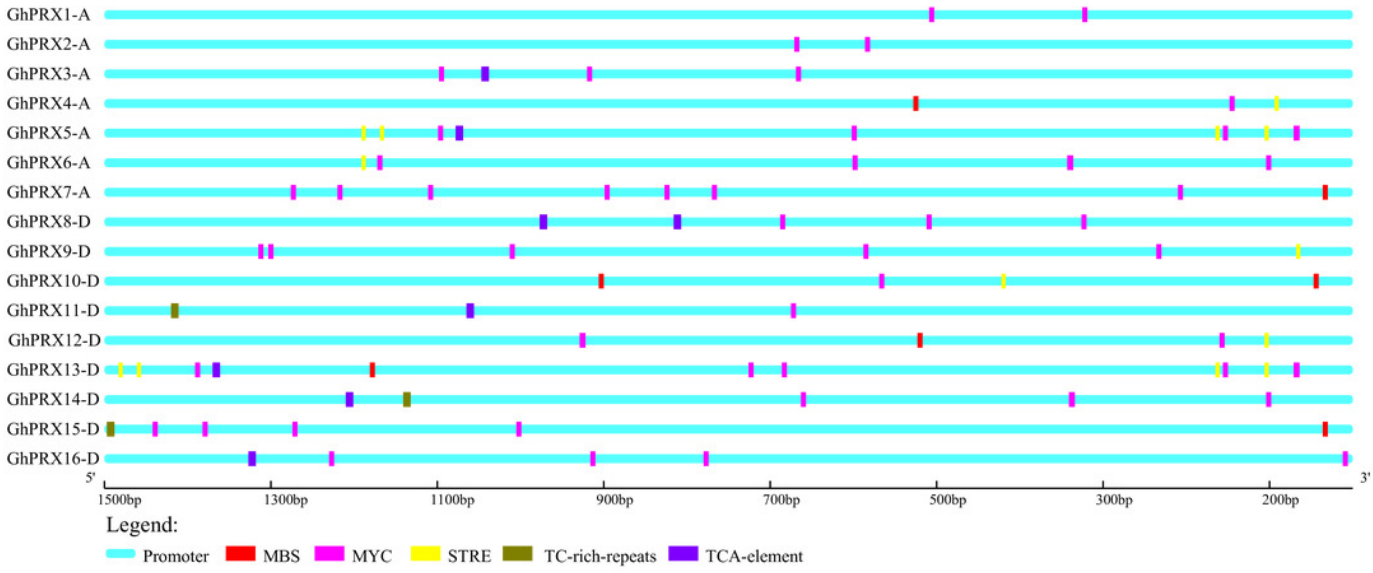


Figure 7

Expression patterns of *Gh-PRX* genes in *G. hirsutum* roots.

The $2^{-\Delta\Delta CT}$ method was used to analyze the expression patterns of *Gh-PRX* genes in *G. hirsutum* roots. *G. hirsutum* seedlings were treated with drought, salt stress and SA hormone at the trefoil stage, and the expression patterns of *Gh-PRX* genes were identified by qRT-PCR. The blue strip column is the control, the orange strip column is NaCl stress, the gray bar column is SA stress, and the yellow bar column is drought stress. Using the cotton ubiquitin gene as an internal reference control, the statistical significant differences of the expression levels were show using the fold change values, all the values at all times were compared with the corresponding values at 0 h. Results are the average of three replicates, and the error bars indicating standard deviations.

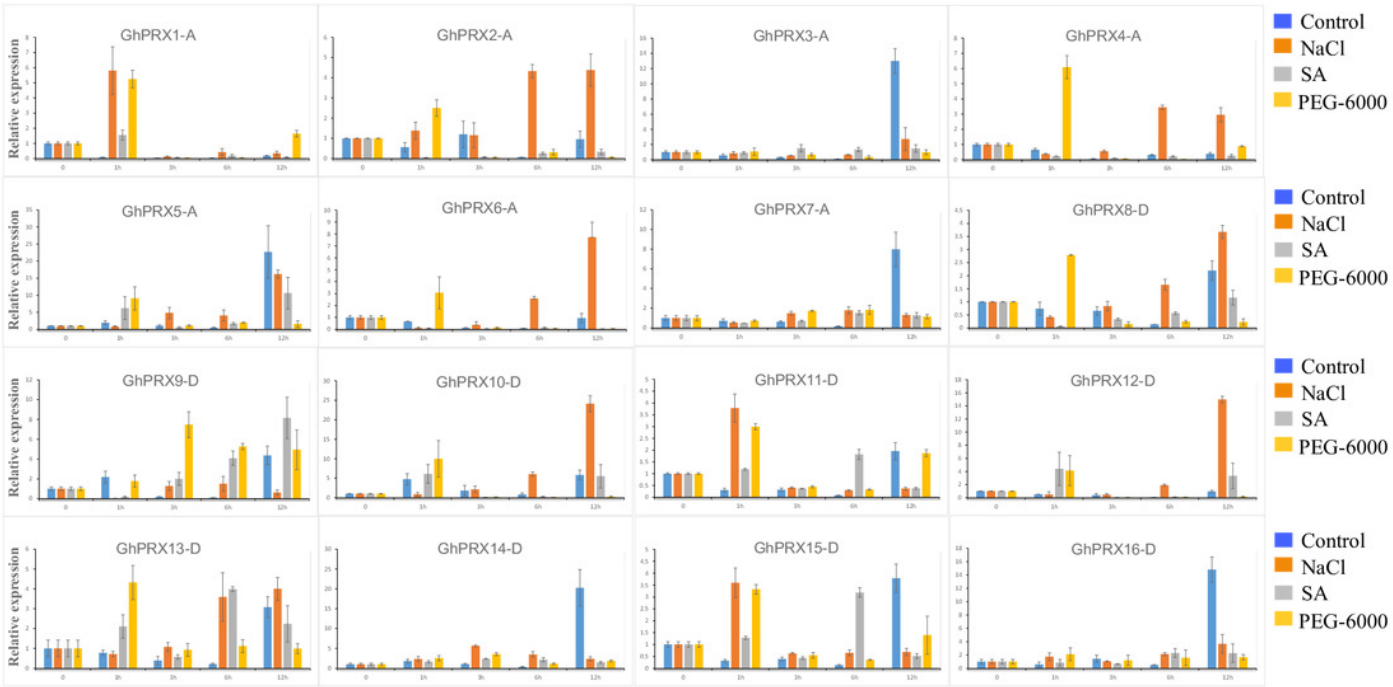


Figure 8

Expression patterns of *Gh-PRX* genes in *G. hirsutum* leaves.

The $2^{-\Delta\Delta CT}$ method was used to analyze the expression patterns of *Gh-PRX* genes in *G. hirsutum* leaves. *G. hirsutum* seedlings were treated with drought, salt stress and SA hormone at the trefoil stage, and the expression patterns of *Gh-PRX* genes were identified by qRT-PCR. The blue strip column is the control, the orange strip column is NaCl stress, the gray bar column is SA stress, and the yellow bar column is drought stress. Using the cotton ubiquitin gene as an internal reference control, the statistical significant differences of the expression levels were show using the fold change values, all the values at all times were compared with the corresponding values at 0 h. Results are the average of three replicates, and the error bars indicating standard deviations.

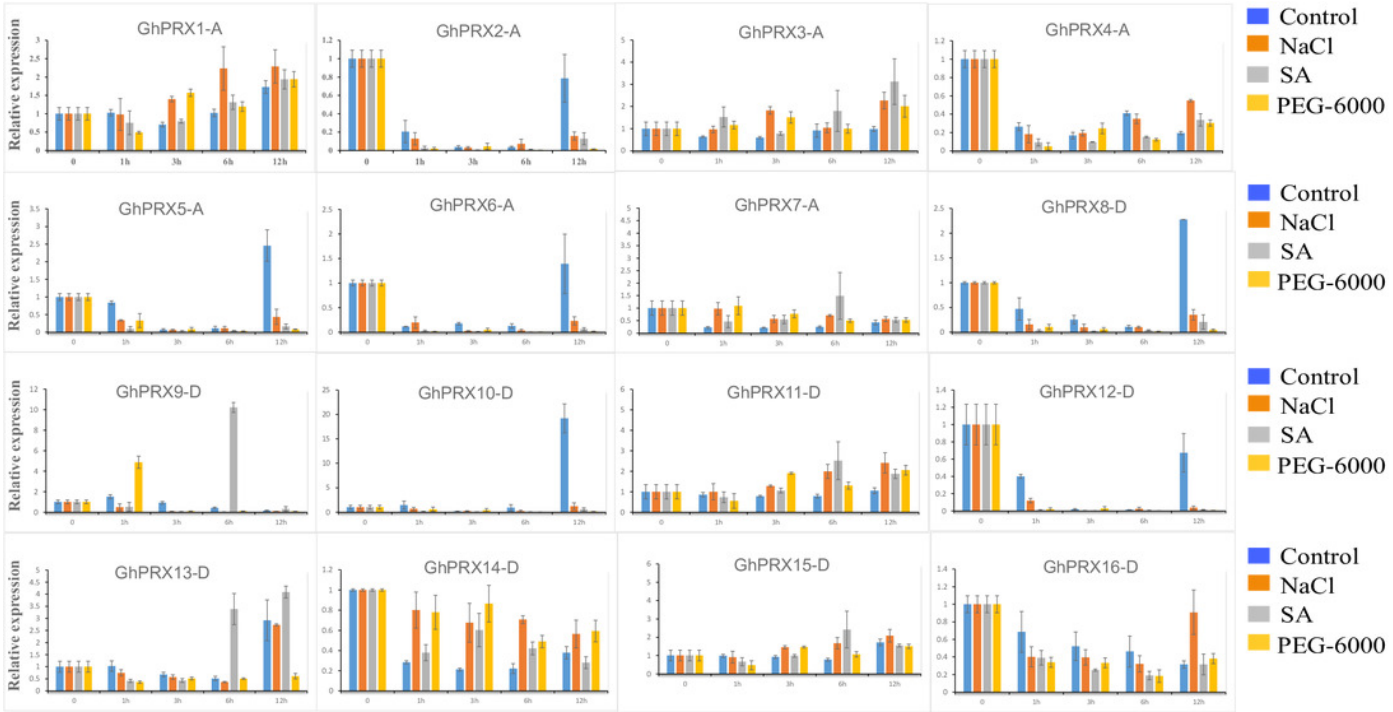


Figure 9

Analysis of tissue expression and alternative shearing profiles of *PRX* genes.

(a) Expression patterns of *PRX* genes in different tissues and organs of *G. hirsutum* TM-1. The color represents *Gh-PRX* expression levels: Log2 (FPKM). (b) Agarose gel electrophoresis map of PRX genes in *G. hirsutum* sGK9708 leaves and roots. (c) Structural diagram of GhPRX14 alternative shearing transcripts. GhPRX14-D is the reference transcript of *G. hirsutum* TM-1. The DNA marker is the Trans2K DNA marker from TransGen.

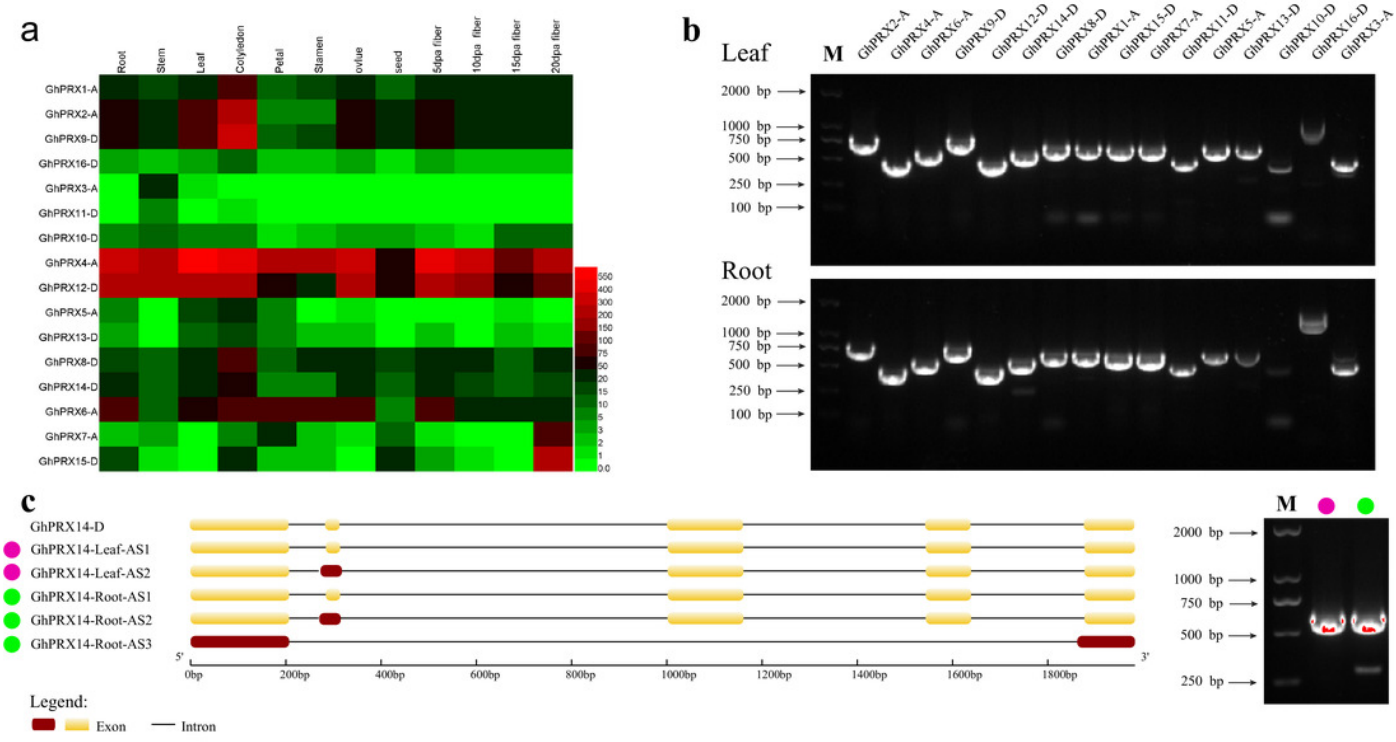


Figure 10

Subcellular localization map of *PRX* proteins.

Subcellular localization map of *PRX* proteins. (**a**) Localization of fluorescence signals of eGFP fluorescent control and homologous gene pair 1 (*GhPRX4-A* and *GhPRX12-D*) protein on the cell membrane ; (**b**) Localization of fluorescence signals of homologous gene pair 2 (*GhPRX7-A* and *GhPRX15-D*) protein on the cell membrane and nucleus; (**c**) Localization of the fluorescence signals of homologous gene pair 3 (*GhPRX1-A* and *GhPRX8-D*), homologous gene pair 4 (*GhPRX2-A* and *GhPRX9-D*) and homologous gene pair 5 (*GhPRX6-A* and *GhPRX14-D*) protein chloroplasts. GFP is a fluorescence signal excited by recombinant pcambia2300-eGFP (CAMBIA) transient expression vector, DAPI stains the nucleus, and chloroplast autofluorescence was obtained at 635 nm fluorescence receiving wavelength.

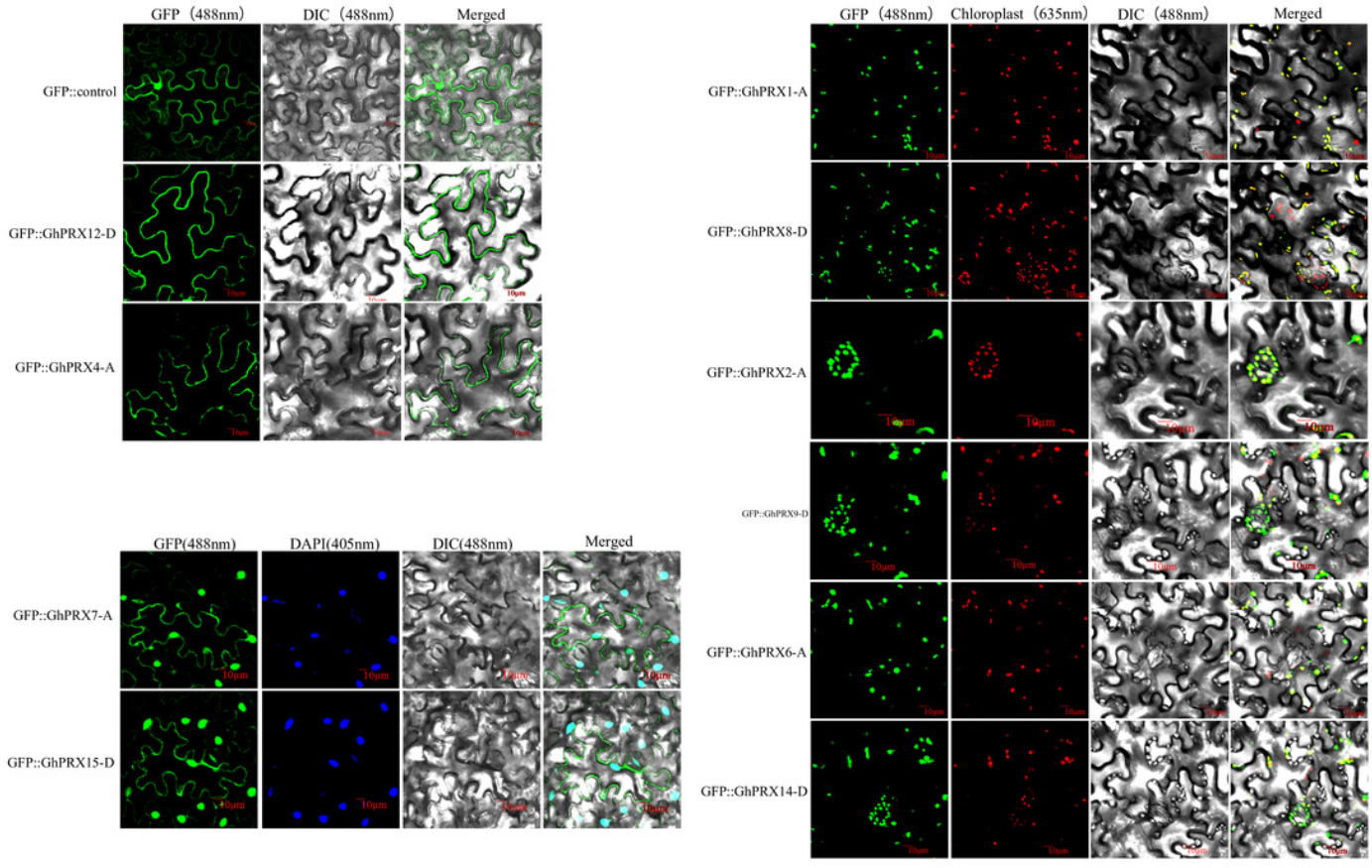


Table 1(on next page)

The characteristics of the *PRX* family genes in cotton.

Table 1 The characteristics of the PRX family genes in cotton

Gene name	Gene ID	Subcalade	Exon	Protein (aa)	pI	Mw(Dalton)	Chromosome location
GhPRX1-A	Gh_A01G1793	PRX2	1	228	9.1	24095.74	ChrA01:97705624-97706307
GhPRX2-A	Gh_A04G0493	PRX1	7	266	6.19	29149.42	ChrA04:26311787-26315757
GhPRX3-A	Gh_A05G3849	TPX	3	163	5.6	17466.1	ChrA05:104118-105049
GhPRX4-A	Gh_A08G2151	TPX	3	163	5.58	17226.79	ChrA08:102615361-102616794
GhPRX5-A	Gh_A08G2202	PRXQ	4	217	9.74	23887.43	ChrA08:103031734-103033703
GhPRX6-A	Gh_A10G1104	PRX5	5	199	8.98	21446.44	ChrA10:54462313-54464296
GhPRX7-A	Gh_A10G1567	PRX6	4	220	5.98	24555.88	ChrA10:85538436-85539624
GhPRX8-D	Gh_D01G2034	PRX2	1	228	9.1	24111.83	ChrD01:59334370-59335053
GhPRX9-D	Gh_D04G0917	PRX1	7	266	6.19	29043.3	ChrD04:26446572-26450644
GhPRX10-D	Gh_D05G0244	TPX	3	147	6.82	15755.16	ChrD05:2202510-2204565
GhPRX11-D	Gh_D05G1251	TPX	3	163	5.78	17454.09	ChrD05:10748631-10749537
GhPRX12-D	Gh_D08G2518	TPX	3	163	5.58	17196.76	ChrD08:64930986-64932473
GhPRX13-D	Gh_D08G2567	PRXQ	4	217	9.7	23793.27	ChrD08:65350134-65351542
GhPRX14-D	Gh_D10G1403	PRX5	5	199	8.45	21361.29	ChrD10:28600472-28602454
GhPRX15-D	Gh_D10G1825	PRX6	4	220	6.22	24392.79	ChrD10:51342185-51343360
GhPRX16-D	Gh_D12G2827	PRX1	6	141	4.9	16074.44	ChrD12:15048-17035
GbPRX1-A	Gbar_A10G018830.1	PRX6	4	219	5.98	24555.88	ChrA10:93866990-93868166
GbPRX2-A	Gbar_A08G026850.1	PRXQ	4	216	9.74	23887.43	ChrA08:119243921-119246359
GbPRX3-A	Gbar_A08G026350.1	TPX	3	162	5.58	17196.76	ChrA08:118790855-118792723
GbPRX4-A	Gbar_A01G020700.1	PRX2	1	227	9.1	24095.74	ChrA01: 113395150-113395833
GbPRX5-A	Gbar_A10G013	PRX5	5	198	8.98	21446.44	ChrA10:62896091-

	540.1						62898528
	Gbar_A05G012						ChrA05:11767776-
GbPRX6-A	820.1	TPX	3	162	5.6	17466.1	11768707
	Gbar_D10G014						ChrD10:28146247-
GbPRX7-D	760.1	PRX5	5	198	8.45	21333.23	28148956
	Gbar_D04G010						ChrD04:28395391-
GbPRX8-D	400.1	PRX1	7	265	6.19	29103.4	28400177
	Gbar_D08G027						ChrD08:65405467-
GbPRX9-D	530.1	PRXQ	4	216	9.7	23793.27	65407385
	Gbar_D01G021						ChrD01:60905018-
GbPRX10-D	840.1	PRX2	1	227	9.34	24091.84	60905701
	Gbar_D08G027						ChrD08:65009751-
GbPRX11-D	000.1	TPX	3	162	5.58	17196.76	65011737
	Gbar_D12G002						ChrD12:2619109-
GbPRX12-D	260.1	PRX1	5	113	4.16	13050.82	2620809
	Gbar_D05G002						ChrD05:2189964-
GbPRX13-D	520.1	TPX	3	146	6.82	15755.16	2198116
	Gbar_D10G019						ChrD10:51564063-
GbPRX14-D	020.1	PRX6	4	219	6.22	24392.79	51565371
	Gorai.009G135						Chr9:10193477-
GrPRX1	600.1	TPX	3	162	5.78	17468.12	10194845
	Gorai.009G397						Chr9:56100514-
GrPRX2	800.1	PRX1	7	265	6.19	29043.3	56105218
	Gorai.009G026						
GrPRX3	300.1	TPX	3	163	5.9	17491.13	Chr9:2005154-2007919
	Gorai.011G158						Chr11:28114223-
GrPRX4	200.1	PRX5	10	198	7.71	21319.16	28116745
	Gorai.011G204						Chr11:49579270-
GrPRX5	600.1	PRX6	4	219	5.98	24498.87	49580743
	Gorai.002G243						Chr2:60733892-
GrPRX6	200.1	PRX2	1	227	9.24	24170.9	60734923
	Gorai.004G280						Chr4:61267661-
GrPRX7	800.1	TPX	3	162	5.58	17196.76	61269611
	Gorai.004G285						Chr4:61634457-
GrPRX8	400.1	PRXQ	4	216	9.7	23793.27	61636274
							Chr05:12175377-
GaPRX1	Ga05G1380	TPX	3	162	5.78	17500.12	12176307
							Chr10:80207013-
GaPRX2	Ga10G1482	PRX5	5	198	8.98	21445.49	80208997
GaPRX3	Ga05G0262	TPX	3	163	5.58	17538.2	Chr05:2267143-2271319
GaPRX4	Ga04G0982	PRX1	7	265	6.19	29045.27	Chr04:36972413-

							36976371
							Chr02:97187166-
GaPRX5	Ga02G1606	PRX2	1	227	9.1	24135.81	97187849
							Chr08:128777936-
GaPRX6	Ga08G2895	PRXQ	4	216	9.7	23846.33	128779888
							Chr08:128370617-
GaPRX7	Ga08G2842	TPX	3	162	5.58	17196.76	128372047
							Chr10:20870875-
GaPRX8	Ga10G0941	PRX6	5	219	5.96	22794.86	20872058

2

Table 2(on next page)

Alternative splicing profile analysis of *PRX* genes.

1 **Table 2.** Alternative splicing profile analysis of *PRX* genes

Transcript	alternative splicing mode	alternative splicing site	alternative splicing boundary	Number of clones	ORF length(bp)	Amino acid length (aa)
GhPRX14-Leaf-AS1	—	—	—	24	597	199
GhPRX14-Leaf-AS2	IR	The first intron 3' 15-bp base	AG/AA	3	612	204
GhPRX14-Root-AS1	—	—	—	24	597	199
GhPRX14-Root-AS3	IR	The first intron 3' 15-bp base	AG/AA	2	612	204
GhPRX14-Root-AS2	ES and A3SS	The fourth intron 3' 14-bp base, the second, third, and fourth exons were skipped.	AG/AT, GG/TT	2	**327**	109

2 Differential shearing of *GhPRX14* in root and leaf.