

# SCIENTIFIC REPORTS



OPEN

## The Glutathione Peroxidase Gene Family in *Gossypium hirsutum*: Genome-Wide Identification, Classification, Gene Expression and Functional Analysis

Mingyang Chen\*, Kun Li\*, Haipeng Li, Chun-Peng Song & Yuchen Miao

The plant glutathione peroxidase (GPX) family consists of multiple isoenzymes with distinct subcellular locations, tissue-specific expression patterns and environmental stress responses. In this study, 13 putative GPXs from the genome of *Gossypium hirsutum* (*GhGPXs*) were identified and a conserved pattern among plant GPXs were exhibited, besides this they also responded to multiple environmental stresses and we predicted that they had hormone responsive *cis*-elements in their promoter regions. Most of the *GhGPXs* on expression in yeast can scavenge H<sub>2</sub>O<sub>2</sub>. Our results showed that different members of the *GhGPX* gene family were co-ordinately regulated under specific environmental stress conditions, and suggested the importance of *GhGPXs* in hormone treatments and abiotic stress responses.

The generation of reactive oxygen species (ROS) is particularly enhanced when animals and plants are subjected to abiotic and/or biotic stresses. ROS are chemical species that are produced by incomplete reduction of oxygen; they include the superoxide anion (O<sub>2</sub><sup>-</sup>), hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>), singlet oxygen (<sup>1</sup>O<sub>2</sub>) and the hydroxyl radical (HO·). ROS participate in a wide range of important signaling processes including growth, development, the cell cycle, acclimation to stress and programmed cell death<sup>1-4</sup>. A set of H<sub>2</sub>O<sub>2</sub>-decomposing enzymes like, catalases (CATs) and peroxidases cope with cells to uncontrolled oxidation status<sup>1,5,6</sup>. Peroxidases may be heme or non-heme peroxidases, the heme peroxidases having a cofactor in their active site (such as ascorbate peroxidases, APXs) or, a redox active cysteine (Cys) or selenocysteine (Sec) residues respectively<sup>7</sup>. Thiol peroxidases such as thioredoxin (Trx) peroxidases or peroxiredoxins (Prxs) and glutathione peroxidases (GPXs) are belongs to family of non-heme peroxidases. Moreover, GPXs, glutathione S-transferases (GSTs) with GPX activity and Prxs are also able to decompose alkyl hydroperoxides in addition to H<sub>2</sub>O<sub>2</sub><sup>8,9</sup>. For decades, GPXs by using glutathione (GSH) or other reducing equivalents as a reductant, have been known to catalyse the reduction of H<sub>2</sub>O<sub>2</sub> or other organic hydroperoxides in to water or the corresponding alcohols<sup>10</sup>.

The GPXs were the first solenozyme that was discovered from mammals<sup>11,12</sup>. The reaction takes place at a single redox centre with Sec as the redox-active residue in selenocysteine-GPXs (Sec-GPXs). The catalytic centre of Sec-GPXs was first characterised as a triad composed of Sec or Cys, glutamine (Gln) and tryptophan (Trp)<sup>13</sup>, but later turned out to be a tetrad with an additional asparagine (Asn)<sup>14,15</sup>. In contrast, the peroxidative Sec was replaced by a Cys and function via a second redox centre that contains a resolving Cys in GPXs of most non-vertebrates. The former type of enzyme is more or less specific for GSH, while the latter is reduced by “redoxins”. The common denominator of the GPX family is the first redox centre comprising (seleno) Cys, Trp, Asn and Gln<sup>15,16</sup>; this kind of GPXs are shown to be reduced by redoxins, namely, thioredoxins or related proteins with a CXXC motif in plants, yeast, insects and protists<sup>15,17</sup>.

GPX genes from a range of plant species, such as *Nicotiana sylvestris*<sup>18</sup>, *Citrus sinensis*<sup>19,20</sup>, *Avena fatua*<sup>21</sup>, *Spinacia oleracea*<sup>22</sup>, *Brassica campestris*<sup>23</sup>, *Helianthus annuus*<sup>24</sup>, *Pisum sativum*<sup>25</sup>, *Lycopersicon esculentum*<sup>26</sup>,

Institute of Plant Stress Biology, State Key Laboratory of Cotton Biology, Department of Biology, Henan University, 85 Minglun Street, Kaifeng 475001, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to C.-P.S. (email: songcp@henu.edu.cn) or Y.M. (email: miaoych@henu.edu.cn)

*Arabidopsis thaliana*<sup>27–29</sup>, *Oryza sativa*<sup>30</sup>, *Triticum aestivum*<sup>31</sup>, *Panax ginseng*<sup>32</sup>, *Populus trichocarpa*<sup>33,34</sup> and *Thellungiella salsuginea*<sup>35</sup>, have been isolated and characterised. Unfortunately very few reports on genome-wide identification and characterisation of GPX family proteins to date have been reported. Furthermore it has been reported that the plant GPX family not only consists of multiple GPXs with distinct subcellular localisations and functions but also exhibited differential tissue-specific expression patterns, responses to environmental stress, functioning in co-ordination against ROS, immune responses, regulating the growth and development of plant root, and stomatal movement<sup>27,28,36,37</sup> based on studies of four GPX family proteins from *A. thaliana*<sup>36,37</sup>, *Lotus japonicus*<sup>38</sup>, *P. trichocarpa*<sup>5</sup> and *T. salsuginea*<sup>35</sup>. Therefore, these results also highlighting the importance of GPX study at the genome-wide level.

Cotton has a long cultivation history and is currently the main source of natural textile fibre and an important crop for oil and biofuel<sup>39</sup>. There are four commercially grown cotton species, *Gossypium hirsutum*, *Gossypium barbadense*, *Gossypium herbaceum* and *Gossypium arboreum*<sup>40</sup>. *G. hirsutum* and *G. barbadense* are the two most commonly cultivated species and producing ~98% of the textile fibre worldwide. Moreover, it requires to mention that there are no reports providing insight on the expression profiling of GPXs in *G. hirsutum* under abiotic stress conditions. In the present study, we have identified 13 GPX genes from *G. hirsutum* and not only their gene structure and promoter sequences were analysed, but also their potential subcellular locations were predicted. We examined the expression of *GhGPX* transcripts from the leaves and roots of *G. hirsutum* under short-term exposure to salt, osmotic and abscisic acid (ABA)-induced stresses to address their role in these stresses. Further exploring their role under abiotic stresses, *gpx3Δ* ( $H_2O_2$ -sensitive mutant) of *Saccharomyces cerevisiae* were complemented with the *GhGPXs*, and it suggests that *GhGPXs* have similar function to GPX3 in yeast, revealing their participation in the oxidative stress response.

## Results

**Identification and characterisation of *GhGPX* genes.** We searched the *GhGPX* sequences from the Cotton Genome Project (CGP) database (<http://cgp.genomics.org.cn/page/species/index.jsp>)<sup>39</sup>, which is the recent release of the first version of the *G. hirsutum* genome by using the coding sequences (CDSs) of *GPX1–8* from *A. thaliana* (*AtGPX*). 12 *GhGPX* genes had been identified and which were represented putative *GhGPX1*, *GhGPX2*, *GhGPX3*, *GhGPX4*, *GhGPX5*, *GhGPX6*, *GhGPX7*, *GhGPX8*, *GhGPX11*, *GhGPX12*, *GhGPX13* and *CotAD\_39521*. Moreover, we cloned CDSs of the 12 *GhGPXs* in at least three independent experiments and it was found that a nucleotide sequence of 45-bp inserted between nucleotides 373 and 374 bp of the *GhGPX2* CDS, leading to a 15-amino-acid (GFLGSRKWNFTKFL) insertion between 124S (Ser) and 125V (Val). However, this 45-bp nucleotide sequence was considered as part of an intron on CGP database<sup>39</sup>. Thus it could be possible that *GhGPX2* has two transcripts in young seedlings. We also found a 39-bp nucleotide sequence insertion behind the 441-bp of the CDS of *GhGPX4* (marked with yellow in the Supplementary Data 1), and this 39-bp nucleotide sequence was not exist in the intron of *GhGPX4* from the database. Therefore, we had cloned a part of the DNA sequence of *GhGPX4* and found an error in the sequence of *GhGPX4* that was reported on the database (the sequence between two yellow marked sequences of the cloned *GhGPX4* in Supplementary Data 1). 72% identity with *AtGPX6* to the 43–225-amino-acid sequence of the N terminus of *CotAD\_39521* realized it was a specific GPX, however, its amino-acid sequence (226–545) had 72% identity with a cysteinyl-tRNA synthetase (*AT5G38830*). Moreover, the cloned gene showed different CDS (Supplementary Data 1) to *CotAD\_39521*, when we cloned *CotAD\_39521* by using two sets of primers (Supplementary Table S1); *Gh\_A12G2084* and *Gh\_D12G2260* were found to have high homology with *CotAD\_39521*, when the protein sequence of *CotAD\_39521* was used for a Blast search with another genome database of the *G. hirsutum*<sup>41</sup>. This suggested us an error in the gene splicing of *CotAD\_39521*. The CDSs of *Gh\_A12G2084* and *Gh\_D12G2260* were also cloned by PCR and we found that there was a deletion of 13-bp nucleotide sequence behind the 675-bp nucleotide (from the start codon, marked with yellow in Supplementary Data 1) of *Gh\_A12G2084* and *Gh\_D12G2260*, this frame shift error evoked the emergence of a termination codon (marked with red in Supplementary Data 1). Thus it is possible that both *Gh\_A12G2084* and *Gh\_D12G2260* have two transcripts, *Gh\_A12G2084* and *Gh\_D12G2260* represented putative *GhGPX10* and *GhGPX9*, respectively. Therefore, the renewed *GhGPX2*, *GhGPX4*, *GhGPX9* and *GhGPX10* sequences were used in the following bioinformatic analysis of the *GhGPX* sequences. The CDSs of 13 *GhGPXs* were shown in the Supplementary data 2, and the genomic locations of 13 *GhGPXs* were shown in Table 1.

Notably, twice of some *GPX* genes were found in allotetraploid than those of diploid cotton. Three genes had one copy in the diploid, while five homologous genes, from the A and D subgenomes (*GhGPX2*, 4, *GhGPX3*, 5, *GhGPX7*, 11, *GhGPX8*, 12 and *GhGPX9*, 10) were found in the tetraploid *G. hirsutum*. The sequences of homologues genes were inconsistent, which might have been resulted from loss of gene during their individual evolutionary processes or from assembly error in partial chromosomal regions, although this requires confirmation. Although *GhGPX6* and *GhGPX9* were located in chromosome 8 and 12, respectively, the same part sequence of *GhGPX9* (190–702 bp) was shared by CDS of *GhGPX6* (513 bp) and the promoter sequence of *GhGPX6* was also similar to that of *GhGPX9*. This data vividly suggested that *GhGPX6* is another copy of *GhGPX9* and they may have two different transcripts.

To confirm the potential functions of the 13 putative *GhGPXs*, we analysed the homology between *GhGPX* and *AtGPX* protein sequences using WU-BLAST2 (<http://www.arabidopsis.org/wublast/index2.jsp>). Using the protein sequences of *GhGPXs*, we found the highest homology of *AtGPX* to each *GhGPX* protein (Table 2), for example, the amino acid sequences of *GhGPX1* and *AtGPX1* had 79% identity, while *GhGPX6* and *AtGPX6* had 78% identity. This suggested that *GhGPXs* may have similar functions to those of their homologues in *Arabidopsis*.

**Identification of intron-exon from *GhGPX* family genes.** Open reading frames (ORFs) of all 13 *GhGPX* genes consisted of six exons and five introns (Fig. 1, and the sequences were shown in Supplementary data 3). Similar

Number	Sequence ID	Name	Location in genome
1	CotAD_10469	<i>GhGPX1</i>	Dt_chr1 42211918-42213928 (-)
2	CotAD_76066	<i>GhGPX2</i>	Dt_chr4 19846771-19848886 (-)
3	CotAD_22672	<i>GhGPX3</i>	Dt_chr4 21898946-21900968 (-)
4	CotAD_54994	<i>GhGPX4</i>	At_chr4 82962124-82964255 (-)
5	CotAD_58608	<i>GhGPX5</i>	At_chr4 89986498-89988499 (+)
6	CotAD_39880	<i>GhGPX6</i>	Dt_chr8 52575640-52576901 (+)
7	CotAD_39878	<i>GhGPX7</i>	Dt_chr8 52591449-52593319 (+)
8	CotAD_59095	<i>GhGPX8</i>	At_chr10 93484710-93486200 (+)
9	Gh_D12G2260	<i>GhGPX9</i>	D12 55725155-55726643 (+)
10	Gh_A12G2084	<i>GhGPX10</i>	A12 83692525-83694015 (+)
11	CotAD_39520	<i>GhGPX11</i>	At_chr13 80456382-80458243 (+)
12	CotAD_51884	<i>GhGPX12</i>	scaffold1637.1 220256-221724 (+)
13	CotAD_36707	<i>GhGPX13</i>	scaffold1048.1 502185-503584 (+)

**Table 1.** The genomic locations of *GhGPX* genes.

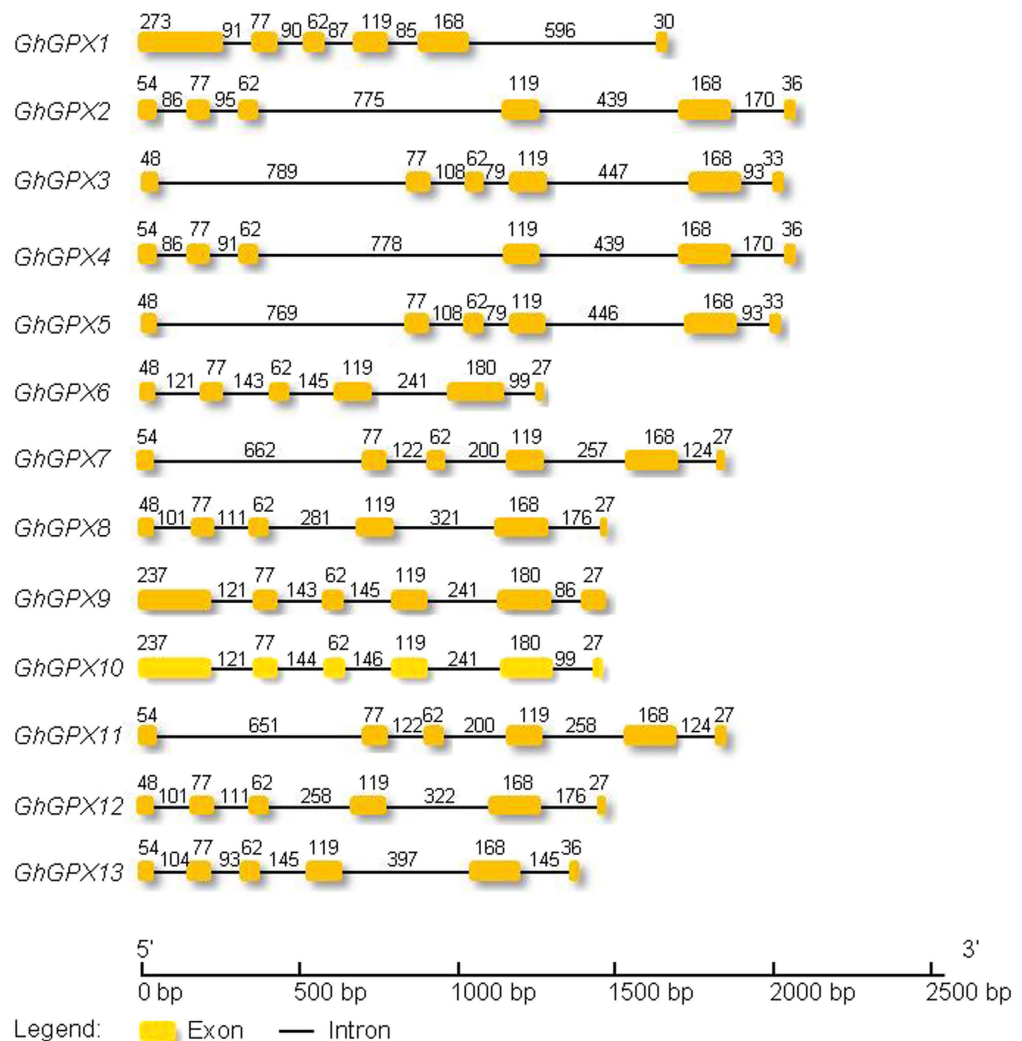
GHGPX	ATGPX	Identities	Positives
GHGPX1	ATGPX1	79%	88%
GHGPX2	ATGPX4	64%	77%
GHGPX3	ATGPX2	73%	85%
GHGPX4	ATGPX4	73%	86%
GHGPX5	ATGPX2	73%	85%
GHGPX6	ATGPX6	78%	85%
GHGPX7	ATGPX8	73%	89%
GHGPX8	ATGPX6	84%	93%
GHGPX9	ATGPX6	67%	76%
GHGPX10	ATGPX6	72%	80%
GHGPX11	ATGPX8	74%	90%
GHGPX12	ATGPX6	83%	93%
GHGPX13	ATGPX5	68%	82%

**Table 2.** Homology analysis between *GhGPXs* and *AtGPXs*. <http://www.arabidopsis.org/wublast/index2.jsp>.

exon-intron organisations in GPX genes were also reported from various plants like eight *AtGPXs*<sup>36</sup>, the six GPX genes from *P. trichocarpa* (*PtGPX*)<sup>5</sup>, the five GPX genes from *L. japonicus* (*LjGPX*)<sup>38</sup> and the eight GPX genes from *T. salsuginea* (*TsGPX*)<sup>35</sup>, as well as some GPX genes from citrus<sup>20</sup> and wheat<sup>31</sup>, suggesting a high degree of conservation in the exon-intron organisation among plant species. The second, third and fourth exons showed same length were of 77, 62 and 119-bp respectively, showed a same length in all 13 *GhGPXs*. The length of the fifth exon of 13 *GhGPXs* was also similar, with a length of 180 bp in *GhGPX6*, *GhGPX9* and *GhGPX10*, and other *GhGPXs* were 168 bp.

**Analysis of putative *cis*-acting regulatory elements from *GhGPXs* promoters.** The plant GPX genes are well known to be involved in responses to various environmental stresses<sup>29,31,32,36</sup>. Therefore, we analysed the promoter regions of all 13 *GhGPX* genes to find their putative *cis*-acting regulatory elements involved in stress and hormone signaling (Table 3) by using PlantCARE software<sup>42</sup>. Multiple different *cis*-acting regulatory elements were found in each *GhGPX* promoter region and their number varied with the individual promoter (Table 3 and Supplementary data 4). Almost all *GhGPX* promoters contained *cis*-acting elements that were responsive to hypoxia (*GhGPX1–3, 5–11*), methyl jasmonate (MeJA) (*GhGPX1–5, 7, 8, 11, 12*) and drought (*GhGPX1–3, 5–12*); many promoters of *GhGPX* contained elements that were responsive to salicylic acid (SA) (*GhGPX1, 3, 5, 7, 8, 10–13*), gibberellin (*GhGPX1, 2, 5, 6, 9–11, 13*) and auxin (*GhGPX1, 4–6, 8–10*); whereas fewer *GhGPX* promoters contained elements that were responsive to ABA (*GhGPX6, 7, 9, 10, 11*), ethylene (*GhGPX3, 8, 12, 13*) and low temperature (*GhGPX2, 4, 8, 13*). These findings indicate that *GhGPX* genes may also be involved in responses to various environmental stresses and plant hormones.

**Predicted properties of *GhGPX* proteins.** To analyse the properties of all 13 *GhGPX* proteins, the amino acid sequences of *GhGPXs* were subjected to alignment. The alignment result revealed that all the protein sequences contained three conserved Cys residues (Fig. 2, marked with triangles) present in most of the plant GPXs, as well as three conserved domains: GPX signature 1, GPX signature 2 and a domain WNF(S/T)KF (Fig. 2, marked with yellow boxes) found in GPXs of most plant and mammalian<sup>43</sup>, besides this the other conserved sequences were also presented, such as (L/I)Y(E/D/N)(K/Q)YK(D/N)(Q/K)G(F/L), C(T/S/K)(R/C/M)



**Figure 1. Genomic organisation of the GhGPX gene family.** The exons are depicted in yellow boxes and introns by black lines. All exon and intron lengths are drawn to scale.

(F/Y)(K/R)(A/S)E(Y/F)PIF and P(V/L/I)(Y/F)(K/Q)(Y/F)LK (Fig. 2, marked with red boxes). The comparing between the amino acid sequences of GhGPXs and previously reported GPXs of other plant species revealed that redox centre embraced with a Cys residue (known to be a catalytic site) as well as Trp, Asn and Gln conserved residues<sup>15</sup> (Fig. 2, marked with red triangles).

The 13 GhGPX proteins can be divided into four major groups: GhGPX6, 8–10, 12 were classified into group 1; GhGPX1, 3, 5 into group 2; GhGPX7, 11 into group 3 and GhGPX2, 4, 13 into group 4 (Fig. 3a). Moreover, based on the phylogenetic tree analysis of GPXs from *A. thaliana*, *G. hirsutum*, *O. sativa* and *P. trichocarpa*, the GPXs could also be divided into four clades. Specifically, the GPXs of *A. thaliana*, *G. hirsutum*, *O. sativa* and *P. trichocarpa* were divided into each clade, which indicated that each GPX has a homologue in the other three kinds of plant, and thus they might have similar functions in different kinds of plant. These data suggested that the conserved nature of evolution of GPX family protein among the four plant species, it is consistent with the same number of exons and introns found in GPXs of different plants species (Fig. 1). GhGPX6, 8–10, 12 were grouped into clade I, GhGPX1, 3, 5 into clade II and the GPXs in sub-classification of clade II, including GhGPX1, PtGPX1, AtGPX1, 7 and OsGPX6, 8 were all predicted to have the subcellular localisation of the chloroplast or mitochondria (Table 4). Clade III composed of GhGPX7, 11, while GhGPX2, 4, 13 were in clade IV (Fig. 3b). These results suggested that GhGPXs have similar characteristics like AtGPXs, OsGPXs and PtGPXs.

The GhGPX proteins can be divided into two categories based on the difference between the length of amino acid sequences: the first category comprised GhGPX1 (242 amino acids, 26.65 kDa), GhGPX9 (233 amino acids, 25.49 kDa), and GhGPX10 (233 amino acids, 25.56 kDa); while the second category comprised with the other GhGPXs having 166–171 amino acids (Table 4). Protein subcellular localisation prediction programmes suggested most of the GhGPX proteins localised in the cytoplasm, while GhGPX1 in chloroplasts and GhGPX9, 10 in chloroplasts or mitochondria (Table 4). The putative homologues of GhGPXs in *Arabidopsis* and poplar are also shown in Table 4.

Environmental stress or hormone	Cis-acting regulator elements	Sequence	GhGPX
Abscisic acid (ABA)	ABRE	TACGTG CACGTG	GhGPX6,7,9,10 GhGPX6,7,9,11
Auxin	TGA-element AuxRR-core	AACGAC GGTCCAT	GhGPX4,5,6,8,9,10 GhGPX1
Salicylic acid (SA)	TCA-element	CAGAAAAGGA CCATCTTTT GAGAAGAATA TCAGAAGAGG	GhGPX1,7,8,11,12 GhGPX1,8,10,13 GhGPX3,5,11,13 GhGPX11
Gibberellins (GA)	GARE-motif P-box	TCTGTTG AAACAGA GCCTTTGAGT CCTTTTG	GhGPX2,5,13 GhGPX1,2,6,9,10,11 GhGPX2,6,10 GhGPX2,6,9,10
Methyl Jasmonate (MeJA)	CGTCA/ TGACG-motif	CGTCA/ TGACG	GhGPX1,2,3,4,5,7,8,11,12
Ethylene	ERE	ATTTCAAA	GhGPX3,8,12,13
Drought inducibility	MBS	TAACTG CAACTG CGGTCA	GhGPX2,3,5,6,8,9,10,11,12 GhGPX1,6,7,8,9,10,12 GhGPX8
Low-temperature	LTR	CCGAAA	GhGPX2,4,8,13
Anaerobic induction	ARE	TGGTTT	GhGPX1,2,3,5,6,7,8,9,10,11

**Table 3. Putative cis-acting regulatory elements related to stress and hormone response in GhGPXs.** The tool to analyse cis-elements in the promoter of GhGPXs: Plant CARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html>).

Although GhGPX9 and GhGPX10 (consisted of ~233 amino acids) were clustered in clade I along with GhGPX6, 8, 12, which each had about 166–170 amino acids. The GhGPX1 comprised of 242 amino acids and a chloroplast transit peptide signal (1–30 amino acids) was at the N-terminal; in contrast, the 1–30 amino acids residues of GhGPX9, 10 may be a mitochondrial or chloroplastic targeting peptide, as predicted by iPSORT and TargetP, which were used to predict the subcellular localisation of eukaryotic proteins (<http://ipsort.hgc.jp/index.html#predict>; <http://www.cbs.dtu.dk/services/TargetP/>).

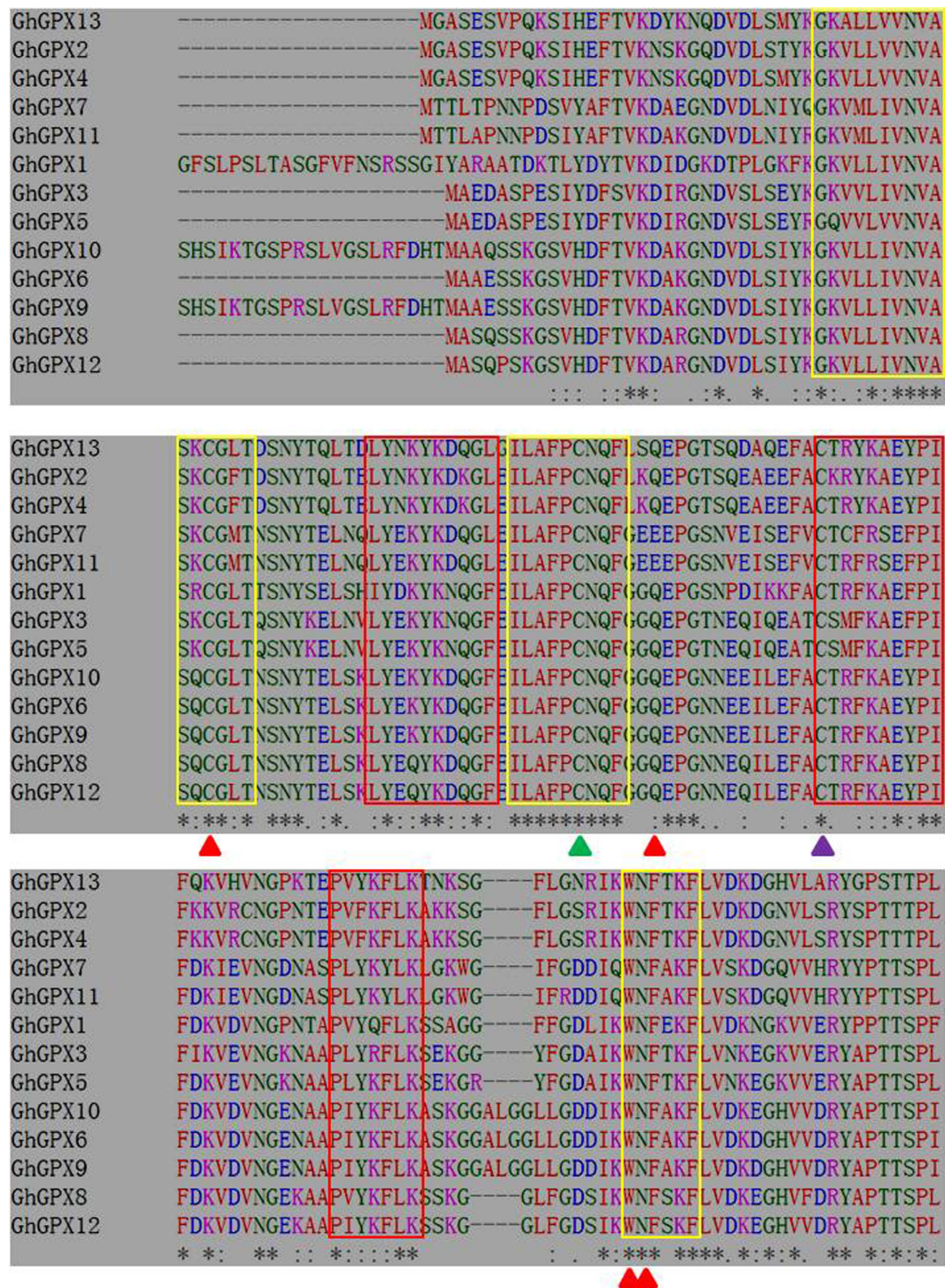
**Differential expression pattern and subcellular localisation of GhGPX products.** To determine which tissues normally express GhGPX genes, gene-specific primers were used for transcription analysis in roots, stems, flowers, cotyledons and leaves of *G. hirsutum*. We had designed six specific primers for real-time quantitative PCR (qRT-PCR): GhGPX1–3, 5, 8, 13, owing to the high homology between GhGPX genes.

As shown in Fig. 4a, the GhGPX1 gene encoding a putative chloroplast protein not only showed expression pattern in roots, stems, flowers, leaves and cotyledons, but was also with most strongly expression level in flowers, and high transcript level in leaves. However, it was nearly absent in roots; the possible explanation to this may be due to the predicted subcellular localisation of GhGPX1 being the chloroplast. The *Arabidopsis* chloroplastic GPXs (AtGPX1 and AtGPX7) play a role in scavenging of active oxygen species under photo-oxidative stress<sup>28</sup>, and AtGPX7 also shows the highest expression level in flowers (<http://bar.utoronto.ca/>). In addition, the evolutionary relationship among AtGPX1, AtGPX7 and GhGPX1 is close (Fig. 3b), indicating the role of GhGPX1 in protection against ROS in response to photosynthesis. GhGPX2 transcript was high in roots, and with the lowest expression level in leaves (Fig. 4b). GhGPX3 and GhGPX5 were homologous from the A and D subgenomes, the transcripts of them showed similar patterns, with low levels in stems and high expression level in flowers and leaves (Fig. 4c,d), which is consistent with their close evolutionary relationship (Fig. 3a). GhGPX8 transcripts were present at high levels in flowers and stems, and the region of the promoter had many cis-elements, including SA, gibberellins and ethylene-responsive elements. This suggested that GhGPX8 may regulate the seedling growth and development (Fig. 4e). The level of GhGPX13 transcript was high in flowers and low in other tissues; compared with roots, the transcript in flowers was found to be nearly 30 times (Fig. 4f). This data indicates that GhGPXs play roles in specific tissues and in different growth periods.

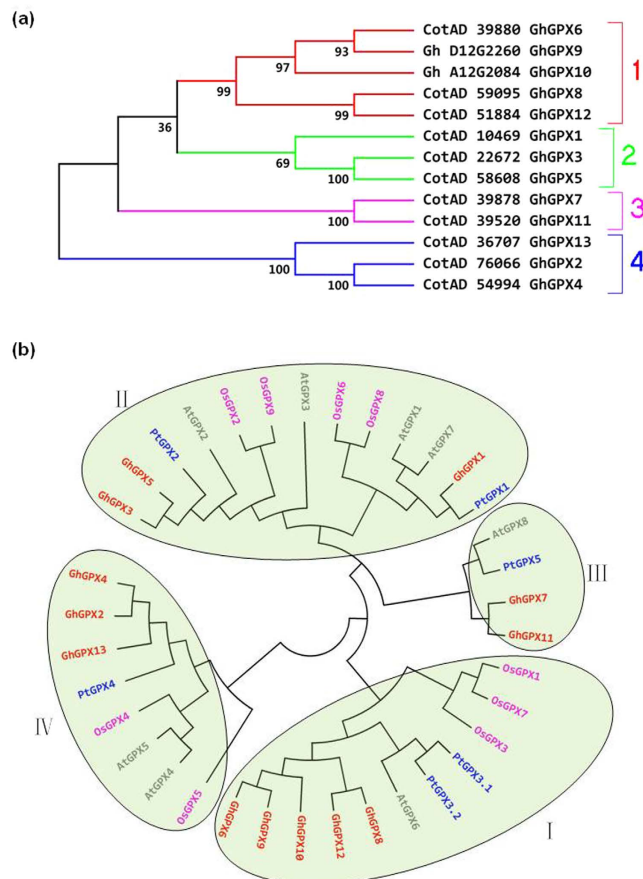
To test the subcellular localisations of GhGPXs, N-terminal GFP-fused GhGPX proteins were expressed in the protoplast of *Arabidopsis* leaves. As shown in Fig. 5, GhGPX1 was expressed in chloroplasts, while GhGPX3, 5–8 were located in the cytoplasm. This result is consistent with the prediction of the subcellular localisation of GhGPXs as shown in Table 4.

**Expression analyses of GhGPX genes in response to various stress conditions.** It is widely accepted that diverse environmental conditions, such as exposure to high concentrations of salt and metals, or high temperatures, induce oxidative stress in plants. Many plant species have been shown to respond to environmental stress by increasing the expression of GPXs, which might protect the cells from damage<sup>19,27–29,35,44</sup>. However, there were no studies of whether GhGPXs could respond to the multiple environmental stresses. Considering that multiple stress-response cis-acting elements were predicted in the promoters of GhGPX genes (Table 2), we used qRT-PCR to investigate the expression profiles of GhGPXs when the plants were under high-temperature, high-salinity, osmotic and heavy-metal stress.





**Figure 2. Amino acid sequences of GHGPX proteins.** Multiple GhGPX sequences alignment using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). The three conserved domains found in most animal and plant GPXs are marked in yellow boxes; the amino acid residues that form part of the proposed catalytic site of the GPXs are marked with red triangles, including the strictly conserved Cys (C), Gln (Q), Asn (N) and Trp (W); other conserved domains are marked with red boxes. The cysteine residue marked with a red triangle is replaced by Sec in mammalian phospholipid hydroperoxide GPXs; the other two conserved Cys residues are marked with green and purple triangles. Amino acid residues that are identical (or have similar properties) in all 13 sequences are marked with a black asterisk below the sequences. Amino acid residues that have similar properties are marked with a point below the sequences. Amino acid residues that are identical in at least five sequences are marked with a point below the sequences.



**Figure 3. Phylogenetic analysis of GPX proteins in plants.** Only complete sequences were considered. (a) Phylogenetic analysis of GhGPXs. (b) Phylogenetic analysis of GPX in *G. hirsutum*, *A. thaliana*, *O. sativa* and *P. trichocarpa*. The tree was constructed using the neighbour-joining method of ClustalW, with 1000 bootstraps, and the bar indicates 0.1 substitutions per site. Different species of GPX proteins are denoted by different colours: red, *G. hirsutum*; blue, *P. trichocarpa*; pink, *O. sativa*; grey, *A. thaliana*. Abbreviations of plant species: Gh, *G. hirsutum*; At, *A. thaliana*; Os, *O. sativa*; Pt, *P. trichocarpa*. Accession numbers for *G. hirsutum* genes are indicated in Table 1 and those of other plant species are as follows (in brackets): AtGPX1 (At2g25080); AtGPX2 (At2g31570); AtGPX3 (At2g43350); AtGPX4 (At2g48150); AtGPX5 (At3g63080); AtGPX6 (At4g11600); AtGPX7 (A4g31870); AtGPX8 (At1g63460); OsGPX1 (Os04g0556300); OsGPX2 (Os06g0185900); OsGPX3 (Os02g0664000); OsGPX4 (Os03g0358100); OsGPX5 (Os11g0284900); OsGPX6 (P0568D10.7); OsGPX7 (Os04g0556300); OsGPX8 (Os06g0185900); OsGPX9 (LOC\_Os11g18170); PtGPX1 (ABK96776); PtGPX2 (DT518382); PtGPX3.1 (ABK96047); PtGPX3.2 (ABK94488); PtGPX4 (ABK95195); PtGPX5 (2P5Q\_A).

Short-term exposure of *G. hirsutum* plants to heavy metal stress (400 mM ferrous sulphate) induced or suppressed the expression of some members of the *GhGPX* genes at different processing times and the effect varied with the tissue, such as leaves and roots (Fig. 6a,b). Compared with the control samples, the *GhGPX1–3*, 5, 8, 13 expression levels changed not more than four times. In leaves, the increased of transcript level could be seen within 3–6 h of treatment for *GhGPX1*, 3, 5, 8; however, within 12–24 h of treatment, the expression level decreased. However, the transcript level of *GhGPX2* decreased when the seedlings were treated with ferrous sulphate for 6 h but increased after 12 h of treatment. In addition, the transcript level of *GhGPX13* increased with ferrous sulphate treatment for 3 h, and then decreased after 6 h. In roots, the increased in transcript level could be seen within 3 h of treatment for *GhGPX1–3*, 5, 8, 13, with a decline after 6 h of treatment and a further increased after 12 h of treatment (Fig. 6b).

The expression patterns of *GhGPX* genes were also analysed in detail following the treatments of salt (Fig. 6c,d) and high-temperature (Fig. 6e,f). Salt stress with 400 mM NaCl treatment for 24 h increased the transcript levels for *GhGPX1–3*, 5, 8, 13 in roots. The expression level of *GhGPX2* was induced in leaves by 24 h treatment with 400 mM NaCl, and the transcript levels of *GhGPX1*, 3, 5, 8, 13 decreased in leaves after treatment with 400 mM NaCl for 3 h (Fig. 6c). A high temperature (35 °C) induced expression of *GhGPX1–3*, 5, 13 after treatment for 24 h in leaves and roots (Fig. 6e,f); however, the transcript level of *GhGPX8* increased in roots after 24 h treatment and in leaves at 3 h treatment (Fig. 6f).

**Induced expression patterns of GPX genes under multiple phytohormone treatments in *G. hirsutum*.** Previous results indicated that plant hormones play a direct role in regulation of a rice GPX



Protein	Length (aa)	MW (Da)	pI	Subcellular localisation	Putative homologues
<i>GhGPX1</i>	242	26645.51	9.29	Chlo	<i>AtGPX1/7, PtGPX1</i>
<i>GhGPX2</i>	171	19167.99	9.24	Cyt	<i>AtGPX4, PtGPX4</i>
<i>GhGPX3</i>	168	18847.42	5.66	Cyt	<i>AtGPX2, PtGPX2</i>
<i>GhGPX4</i>	171	19171.01	9.16	Cyt	<i>AtGPX4, PtGPX4</i>
<i>GhGPX5</i>	168	18948.44	5.36	Cyt	<i>AtGPX2, PtGPX2</i>
<i>GhGPX6</i>	170	18667.19	5.46	Cyt	<i>AtGPX6, PtGPX3.1/3.2</i>
<i>GhGPX7</i>	168	19129.61	4.59	Cyt	<i>AtGPX8, PtGPX5</i>
<i>GhGPX8</i>	166	18525.00	6.73	Cyt	<i>AtGPX6, PtGPX3.1/3.2</i>
<i>GhGPX9</i>	233	25489.14	8.62	Mito/Chlo	<i>AtGPX6, PtGPX3.1/3.2</i>
<i>GhGPX10</i>	233	25561.26	8.98	Mito/Chlo	<i>AtGPX6, PtGPX3.1/3.2</i>
<i>GhGPX11</i>	168	19292.91	5.10	Cyt	<i>AtGPX8, PtGPX5</i>
<i>GhGPX12</i>	166	18501.02	6.73	Cyt	<i>AtGPX6, PtGPX3.1/3.2</i>
<i>GhGPX13</i>	171	19187.91	8.84	Cyt	<i>AtGPX5, PtGPX4</i>

**Table 4. Predicted properties of GhGPXs.** The tool to analyse the isoelectric point (pI) and molecular weight (MW) of *GhGPXs*: EXPASY ([http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/)). The tools to predict the subcellular localisation of *GhGPXs*: TargetP (<http://www.cbs.dtu.dk/services/TargetP/>) and SubLoc (<http://www.bioinfo.tsinghua.edu.cn/SubLoc/>).

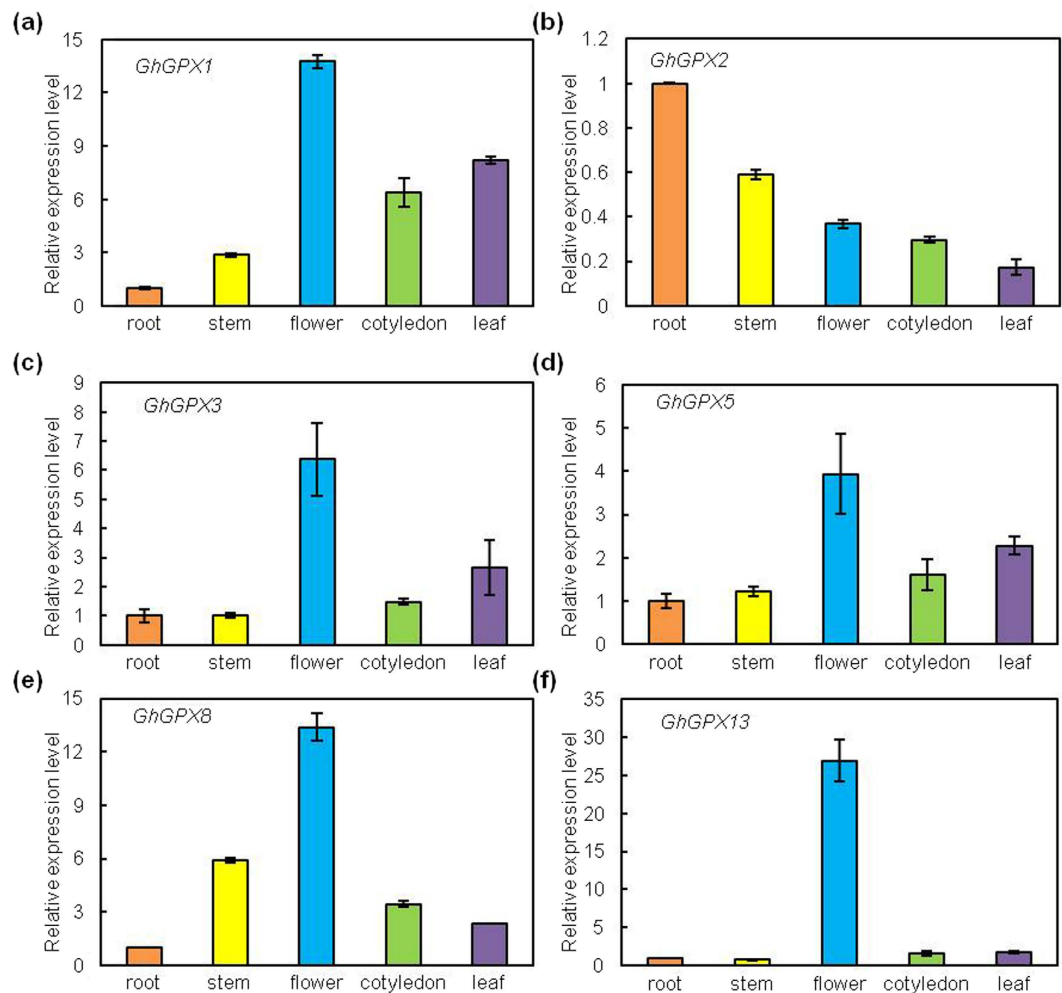
gene<sup>45</sup>. AtGPX3 might mediate ABA and oxidative signaling through physical interaction with the 2C-type protein phosphatase ABA INSENSITIVE2 (ABI2) and, to a lesser extent, with ABI1<sup>27</sup>. AtGPX1 and AtGPX7 were also shown to regulate ROS and SA pathways, which affect leaf development, light acclimation, basal defence and cell death programmes<sup>28</sup>. RNA was isolated from roots and leaves of 10-day-old *G. hirsutum* plants for further study of the effects of signaling molecules on regulation of the expression level of the *GhGPX* gene family. The leaves were then sprayed with SA, jasmonic acid (JA) and ABA, and the roots were dipped in ABA. We performed qRT-PCR analyses to examine the expression levels of the *GhGPX* genes in response to the three phytohormones. ABA induced the expression of *GhGPX2* in both leaves and roots with respect to controls (Fig. 7a,b). A slight increase in expression of *GhGPX3* and *GhGPX5* were observed in leaves and roots, while *GhGPX3* and *GhGPX5* showed similar expression patterns under ABA treatment. The expression of *GhGPX1*, *GhGPX8* and *GhGPX13* were also induced by ABA in leaves; however, in roots, ABA induced the expression of *GhGPX1* and *GhGPX13*, while no obvious change was found for *GhGPX8*. We previously reported that the enhanced production of H<sub>2</sub>O<sub>2</sub> in guard cells on treatment with ABA, resulted stomatal closure<sup>46</sup>. In addition, it was shown that AtGPX3 might play dual and distinctive roles in H<sub>2</sub>O<sub>2</sub> homeostasis, acting as a general scavenger and specifically relaying the H<sub>2</sub>O<sub>2</sub> signal as an oxidative signal transducer in ABA and drought stress signaling<sup>27</sup>. Hence, ABA induced expression of *GhGPX* genes may involve enzymatic responses to oxidative stress. JA treatment specifically increased the expression levels of *GhGPX8* among the six analysed *GhGPX* genes (Fig. 7c). SA treatment induced slight up-regulation of the expression level of *GhGPX1* and *GhGPX8*, and down-regulation of *GhGPX3* (Fig. 7d). These results indicated that individual members of the *GhGPX* gene family respond differentially against abiotic stress and plant hormones, and these findings may provide a clue to explain the differential regulation of transcripts under stress conditions.

**Functional complementation of the H<sub>2</sub>O<sub>2</sub>-sensitive yeast growth by transforming GhGPXs in *gpx3Δ* mutant yeasts.** The GPX homolog YIR037W/HYR1/GPX3 was found in the *Saccharomyces* Genome Database, and a previous study indicated that *gpx3Δ* mutant was hypersensitive to peroxides, and with decreased ~57% GPX activity in *gpx3Δ* mutants with respect to wild type (BY4741)<sup>47</sup>. Plant GPXs contain Cys instead of Sec at their active site and some of them have both GPX and thioredoxin peroxidase activities, and the thioredoxin regenerating system is much more efficient *in vitro* than the GSH system. At present, the functions of these enzymes in plants are not addressed. To test whether the GhGPXs can detoxify H<sub>2</sub>O<sub>2</sub>, the CDSs of *GhGPX1*, *GhGPX2*, *GhGPX3*, *GhGPX5*, *GhGPX6*, *GhGPX7* and *GhGPX8* were cloned into the *pYES2* vector and the recombinant plasmids were transformed into *gpx3Δ* cells. As shown in Fig. 8, the *gpx3Δ* cells were sensitive to H<sub>2</sub>O<sub>2</sub> when OD<sub>600</sub> was <0.01, and few yeasts were seen under 1 mM H<sub>2</sub>O<sub>2</sub> treatment compared with that of controls, in terms of the number of wild-type BY4741 yeast cells. The *gpx3Δ* cells, which had been transformed with *GhGPX1*, *GhGPX3*, *GhGPX5*, *GhGPX6*, *GhGPX7* and *GhGPX8*, were insensitive to H<sub>2</sub>O<sub>2</sub> compared with the *gpx3Δ* cells when OD<sub>600</sub> was 0.1; however, the *GhGPX2*-transformed *gpx3Δ* was found to be sensitive to H<sub>2</sub>O<sub>2</sub>. This suggested that GhGPX1, GhGPX3, GhGPX5, GhGPX6, GhGPX7 and GhGPX8 can scavenge H<sub>2</sub>O<sub>2</sub> in yeast, with functions similar to that of GPX3 in yeast. The *gpx3Δ* cells, which had been transformed with *GhGPX6*, *GhGPX7* and *GhGPX8*, could even grow in medium containing 1 mM H<sub>2</sub>O<sub>2</sub>, although their OD<sub>600</sub> were 0.001 (Fig. 8). This indicated that GhGPX6, GhGPX7 and GhGPX8 could be more effective scavengers of H<sub>2</sub>O<sub>2</sub> than other GhGPXs, which could not grow in medium containing 1 mM H<sub>2</sub>O<sub>2</sub> when the OD<sub>600</sub> was 0.001.

## Discussion

Most plant defence proteins are encoded by families of closely related genes that usually display specific structural and regulatory features<sup>48</sup>. In this study, we performed genome-wide identification of the GPX family genes in *G. hirsutum*. The gene expression profiles of the leaves and roots of *GhGPXs* were investigated under high-salinity

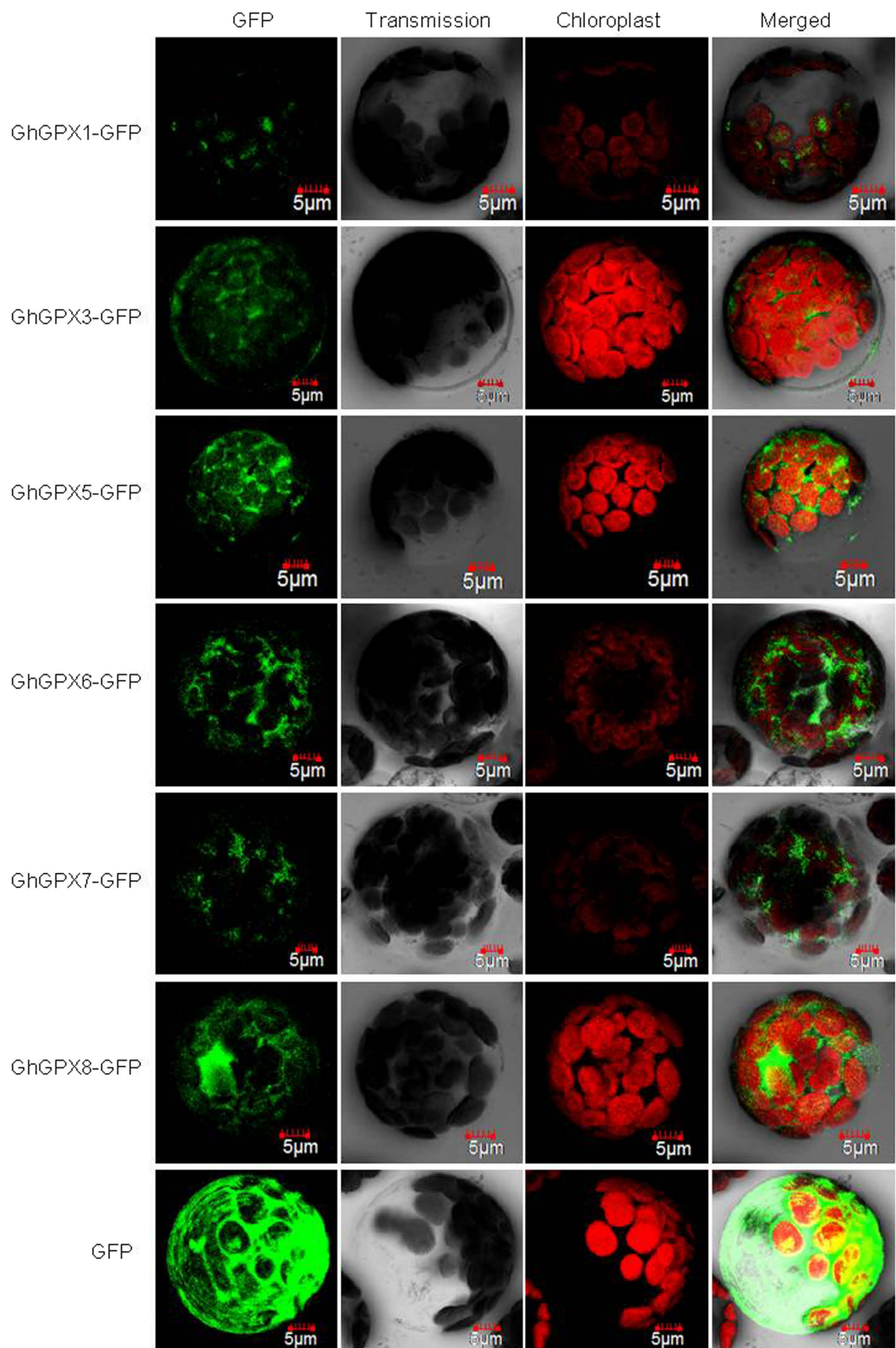




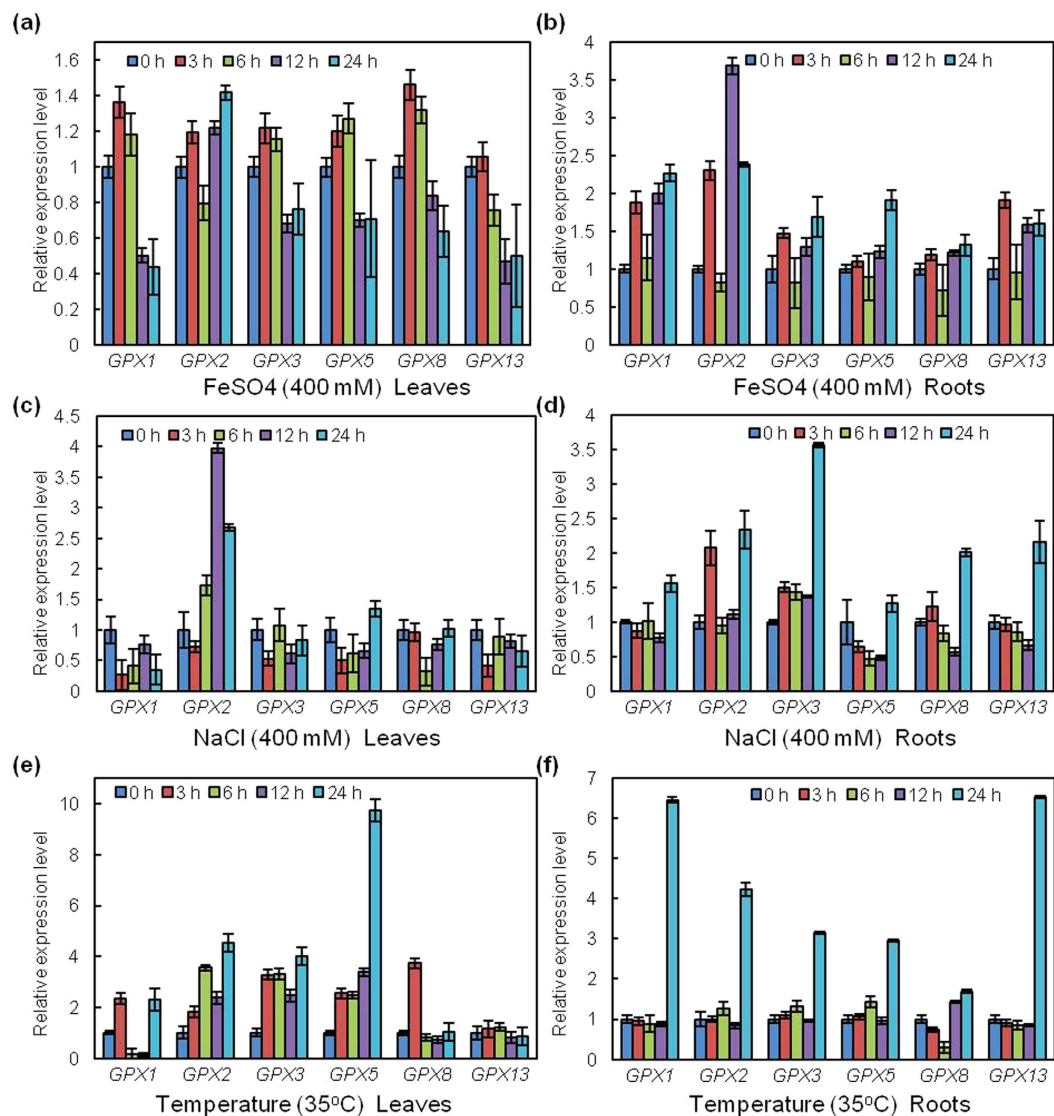
**Figure 4. Expression of GhGPX genes in leaves, flowers, roots, stems and cotyledons.** Relative expression of each gene in organs with respect to leaves, flowers, roots, stems and cotyledons. (a) *GhGPX1*, (b) *GhGPX2*, (c) *GhGPX3*, (d) *GhGPX5*, (e) *GhGPX8*, (f) *GhGPX13*. Cotton seedlings were grown in sterile culture pots under long-day conditions (16-h photoperiod) with 26/20°C day/night temperatures for 25 days. The leaves were the first true leaves. Steady-state mRNA levels were normalised with respect to *Ubiquitin7* (UBQ7, DQ116441) and are expressed relative to the values found in the roots, which were given an arbitrary value of 1. All data are means  $\pm$  SD of six replicates.

and hormone treatments to analyse their differences. Our results showed that the different members of the GPX gene family were co-ordinately regulated in different tissues under specific environmental stress conditions. The complementation analysis using the *gpx3* $\Delta$  cells in this study therefore provided evidence that the GhGPX1, GhGPX3, GhGPX5, GhGPX6, GhGPX7 and GhGPX8 play crucial role in scavenging of H<sub>2</sub>O<sub>2</sub> in response to different stresses.

**Characterisation of GPXs from *G. hirsutum* and their evolution.** Based on the scans of several plant genomes, GPX family genes have been systematically investigated from *A. thaliana*<sup>36,37</sup>, *L. japonicas*<sup>38</sup>, *P. trichocarpa*<sup>5</sup> and *T. salsuginea*<sup>35</sup>. In this study, a total of 13 GPXs from *G. hirsutum* were identified and classified into four groups based on their phylogenetic clades (Fig. 3a). These 13 GhGPX proteins contained not only the three motifs present in most plant GPXs, but also three conserved cysteines (Fig. 2). GhGPX1 was shown to belong to clade 2, and is an enzyme localised to plastids (Table 4, Fig. 5), with 88% identity in its amino acid sequence compared with chloroplast-localised AtGPX1. The expression level of *GhGPX1* was also high in flowers and leaves, and it is consistent with the high concentrations of GSH in chloroplasts<sup>1</sup> and its low levels in roots and stems, which suggested an important role for GPXs. This matches with previous report describing the involvement of AtGPX1 and AtGPX7 in the protection against photosynthesis-induced oxidative damage<sup>28</sup>. The plant GPXs are ubiquitous enzymes that have been shown to be present in different plant tissues, subcellular compartments and developmental stages<sup>5,35–38</sup>. ROS were produced at high levels in endoplasmic reticulum and peroxisome, surprisingly, no GhGPX was reported to target in these cellular compartments, it could be a possibility that the APX, CAT and SOD are the main antioxidant enzymes in these organs. Most of the 13 GhGPX proteins in terms of the



**Figure 5. The subcellular localisation of GhGPX proteins in mesophyll protoplast of *Arabidopsis*.** Localisation of GFP signals from GhGPX proteins fused with GFP. Epifluorescence, chloroplast autofluorescence, bright field and merged images of *Arabidopsis* mesophyll protoplasts transfected with constructs expressing different fusion proteins. Bars = 5  $\mu$ m.

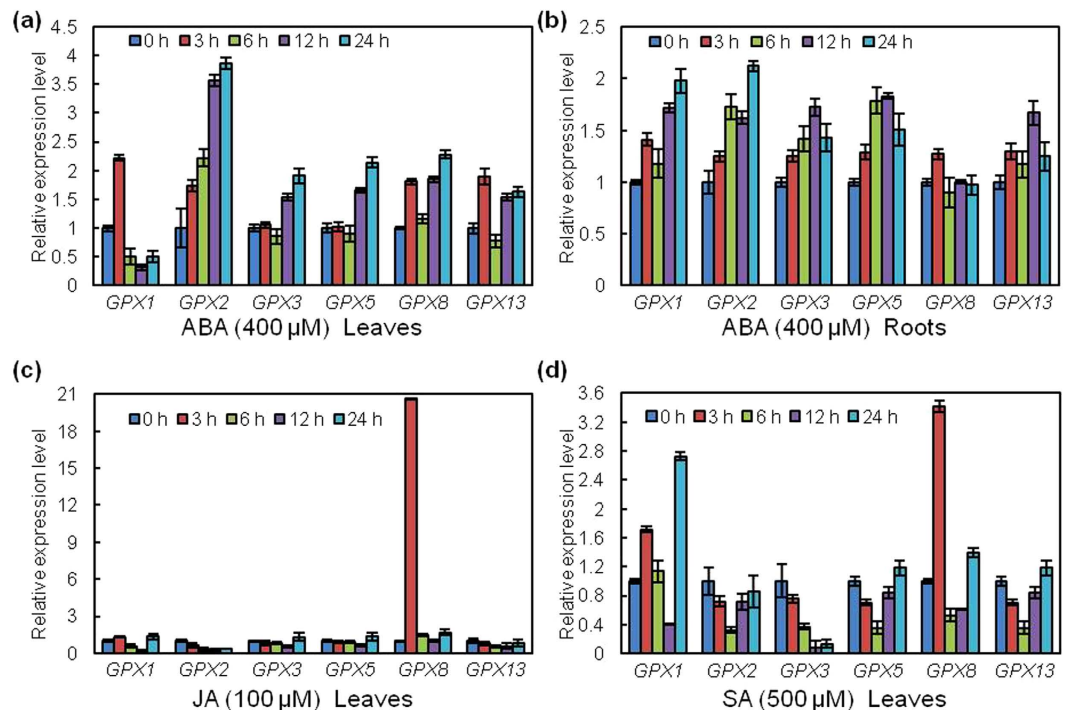


**Figure 6.** Effect of various stresses on the expression of *GhGPX* genes in leaves and roots. *GhGPX* gene transcripts in response to various stress conditions were analysed by qRT-PCR. Cotton seedlings were grown in sterile culture pots under long-day conditions (16-h photoperiod) with 26/20 °C day/night temperatures for 25 days and then treated with 400 mM FeSO<sub>4</sub> (a,b), 400 mM NaCl (c,d) and a high temperature (35 °C) (e,f). mRNA levels were normalised with respect to *GhUBQ7* and are expressed relative to the values at 0 h (control), which were given an arbitrary value of 1 for each gene. Data represent the means  $\pm$  SD of at least three replicates.

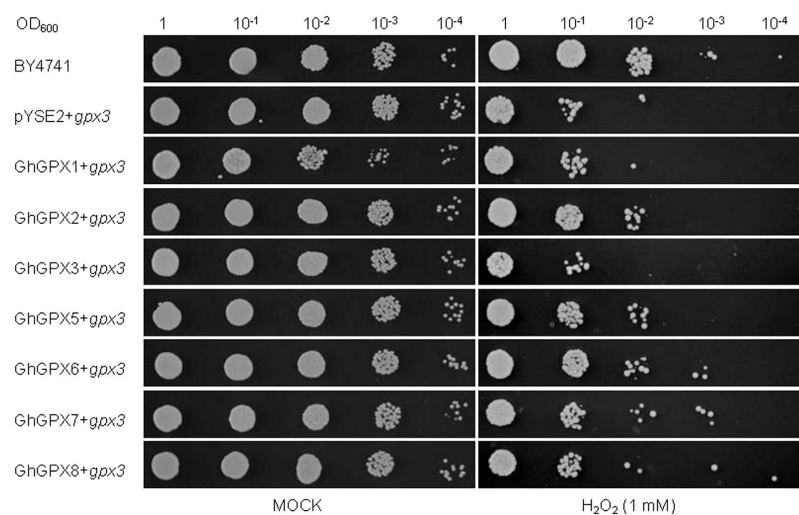
evolutionary relationships were found to be closer to those of *P. trichocarpa*, revealed by phylogenetic analysis and in the same clade, each GhGPX was closer to PtGPX (Fig. 3b, marked in red and blue).

**The role of GPXs during plant growth and development.** It had been reported that GPX genes were involved in not only plant growth and development, but also play very crucial role in the control of plant responses to abiotic stresses, biotic stresses and phytohormones<sup>37,49</sup>. For example, AtGPX1 and AtGPX7 contribute to cross-talk between photo-oxidative stress and immune responses, therefore they might be involved in defence against virulent pathogen infection<sup>28</sup>. AtGPX2 was also found to be part of a cytosolic protein complex involved in stress defence<sup>50</sup>. In addition, AtGPX3 plays a dual role in response to stresses: on one hand, it contributes to the maintenance of H<sub>2</sub>O<sub>2</sub> homeostasis by the elimination of H<sub>2</sub>O<sub>2</sub> and organic hydroperoxides; while on the other hand it works as a switch in the ABA and H<sub>2</sub>O<sub>2</sub> signaling pathway in guard cells<sup>27</sup>; AtGPX8 plays an important role in the protection of cellular components including nuclear DNA against oxidative stress<sup>29</sup>. Moreover, *Atgpx1*, *Atgpx4*, *Atgpx6*, *Atgpx7* and *Atgpx8* mutants have been shown to have significantly larger lateral root density than the wild type, while the *Atgpx2* and *Atgpx3* mutants were embraced with significantly lower. The role of GPXs in the hormone-mediated (auxin, strigolactone and ABA) control of lateral root development has been documented<sup>37</sup>. In this study, we observed similar characters between GhGPXs and AtGPXs. First, the same number of exon-intron organisation, conserved Cys residues and GPX domains of GhGPXs and





**Figure 7.** The effects of different plant hormones on the steady-state transcript level of *GhGPX* genes by qRT-PCR. Leaves or roots of *G. hirsutum* exposed to 400  $\mu$ M ABA (a,b), 100  $\mu$ M JA (c) and 500  $\mu$ M SA (d). mRNA levels were normalised with respect to *GhUBQ7* and are expressed relative to the values at 0 h (control), which were given an arbitrary value of 1. Data represent the means  $\pm$  SD of at least three replicates.



**Figure 8.** Some *GhGPXs* can functionally complement the  $H_2O_2$ -sensitive morphology in *gpx3* $\Delta$ . Growth of yeast strains on solid minimal medium containing 1 mM  $H_2O_2$  (MOCK represents equal volume of  $H_2O$ ). Strain BY4741 (*gpx3* $\Delta$ , WT) transformed with plasmid *pYSE2* and *pYSE2-GhGPXs* was incubated at 30  $^{\circ}C$ . Pictures were taken 4 d later.

AtGPXs. Second, the different expression patterns of GPXs were also observed in cotton. The genes *GhGPX1*, 3, 5, 8, 13 were expressed at high levels in leaves, of which the expression of *GhGPX1* being the highest; in addition, *GhGPX8* was expressed at a high level in stems and cotyledons. High expression levels of *GhGPX2*, 3, 13, were also found in roots. The tissue- or organ-specific GPX expression patterns indicated their functional divergence during plant development and growth. Moreover, the expression of all *GhGPX* genes induced by stress- or phytohormone-related signal treatment (Figs 6 and 7). Taken together, these results showed that *GhGPXs* are generally responsive to stress-related signals and phytohormone treatments, and play an important role in the responses to environmental stresses and the regulation of plant growth and development.

**The level of GhGPXs has antioxidant function.** The plant GPX isoenzymes were demonstrated to detoxify  $H_2O_2$  and organic hydroperoxides, and could be involved in maintenance of the thiol/disulphide or NADPH/NADP<sup>+</sup> balance. Our results showed that the amount of GhGPXs could functionally complement a yeast mutant defective of *GPX3* gene (Fig. 8). Besides GhGPX2, the other GhGPXs such as GhGPX1, GhGPX3, GhGPX5, GhGPX6, GhGPX7 and GhGPX8 could functionally complement the  $H_2O_2$ -sensitive morphology in *gpx3Δ* yeast. AtGPX isoenzymes were able to utilize Trx as a sole electron donor for the reduction of  $H_2O_2$  or hydroperoxides<sup>31</sup>, thiol-dependent redox regulation is more diverse in plants than in animals, bacteria or fungi<sup>52</sup>. These findings are consistent with the function of GPXs in catalysing reduction of  $H_2O_2$ . Therefore, further work to clone Trxs in *G. hirsutum*, which may act as downstream components of GhGPXs to regulate redox homostasis, is crucial and of great interest.

## Materials and Methods

**Plant materials and stress treatments.** The seeds of cotton were wrapped in moist absorbent cotton and placed in a Petri dish, which were placed in an incubator for germinate with illumination at 25 °C for 3 days to germinate. The seedlings were grown in a sterile culture pots under long-day conditions (16 h/8 h light/dark photoperiod) with 26/20 °C day/night temperatures until the second true leaf appeared. Cotton seedlings were inoculated using the root dip method<sup>53</sup>. The root tissues were harvested at 24 h post-inoculation and >10 plants were collected per treatment.

**Isolation and cloning of GhGPX genes.** We searched most of the *G. hirsutum* GPX gene sequences in the Cotton Genome Project database (<http://cgp.genomics.org.cn/page/species/index.jsp>) using the CDS of *GPX1–8* from *A. thaliana* with the BLAST tool, and searched GhGPX9, 10 from another database (<http://mascotton.njau.edu.cn/html/Data/Genomefhsequence/2015/05/05/16ab0945-19e9-49f7-a09e-8e956ec866bf.html>). The GhGPX genes were amplified by PCR (Fast Pfu DNA polymerase, Novoprotein) and the primers used for cloning and analysing the full-length CDS of *GhGPX* genes were designed by using Primer Premier 5 and their sequences were listed in Supplementary Table S1.

**Sequence alignments, phylogenetic analysis and other bioinformatic analyses.** Isoelectric point (pI) and molecular weight (MW) predictions for each deduced *GhGPX* were performed using the Compute pI/MW tool: EXPASY ([http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/)). The deduced sequences of 13 GhGPXs were aligned with the DNAMAN program (version 4.0; Lynnon Corporation, Pointe-Claire, Quebec, Canada) to identify conserved domains. These GhGPXs sequences were also used together with the sequences of the homologous proteins from *A. thaliana*, *O. sativa* and *P. trichocarpa*, to construct an unrooted phylogenetic tree by the neighbour-joining method using the ClustalW2.0 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) and MEGA 6.0.5<sup>54</sup> programs. For promoter analysis, 2 kb upstream region from translation start code ATG of each *GhGPXs* was from database of CGP<sup>39</sup> and another database<sup>41</sup>, and then predicted *cis*-acting elements in (+) strand and (–) strand of each promoter of *GhGPXs* was found using the database of PlantCARE<sup>42</sup> (Plant *Cis*-Acting Regulatory Element, <http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>).

**Gene expression analysis.** Total RNA was extracted by using TRIzol reagent (Invitrogen) from *G. hirsutum* seedlings. The cDNA was synthesised using M-MLV Reverse Transcriptase (Promega) and was 30-fold diluted with deionised water before use as a template in qRT-PCR. The quantitative reaction was performed on a 7500 Fast Real-Time PCR System (Applied Biosystems) using GoTaq<sup>®</sup> qPCR Master Mix kit (Promega). Three independent biological and technical replicates were subjected to qRT-PCR analysis for each sample. The expression levels of *GhGPXs* were normalised against that of an internal reference gene, *GhUBQ7* (*ubiquitin 7*, accession number: DQ116441). All primers used in this study are listed in Supplementary Table S1.

**Subcellular localisation of GhGPXs.** For green fluorescent protein (GFP) analysis, the final construct *pHBT-GFP-GhGPXs* and empty vector *pHBT-GFP* were transiently expressed in *Arabidopsis* mesophyll protoplasts<sup>55</sup>. Transgenic cells were examined under an FV1000 (Olympus) confocal microscope.

**Fluorescence microscopy.** Fluorescence was observed by using the laser scanning confocal microscope LSM FV1000 (Olympus). The fluorescence was excited at 488 nm and detected at 500–560 nm for GFP, while autofluorescence of chloroplasts was detected at 650–750 nm. Transmitted fields were collected simultaneously for use in merging images.

**Yeast complementation analysis.** Full-length complementary DNA of *GhGPXs* was cloned into the vector *pYES2* and transformed into the wild type and *gpx3Δ* of *Saccharomyces cerevisiae* (genotype BY4741). Successfully transformed colonies were screened by complementation of leucine autotrophy. The presence of the correct plasmid was confirmed by PCR. The cell density of cultures was measured during exponential growth and was adjusted to OD<sub>600</sub> 1.0, was analysed with the Biophotometer No. 6131–21283 (Eppendorf, Germany). The growth assays were conducted at 30 °C. Yeast strains were precultured in liquid and solid yeast peptone dextrose medium. Yeast colonies grew on medium that containing 10% galactose and 10% raffinose. Primer sequences used in the PCR are presented in Supplementary Table S1.

## References

- Mittler, R. Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci.* **7**, 405–410 (2002).
- Mittler, R., Vanderauwera, S., Gollery, M. & Van Breusegem, F. Reactive oxygen gene network of plants. *Trends Plant Sci.* **9**, 490–498 (2004).
- Wang, P. & Song, C. P. Guard-cell signalling for hydrogen peroxide and abscisic acid. *New Phytol.* **178**, 703–718 (2008).

4. Song, Y., Miao, Y. & Song, C. P. Behind the scenes: the roles of reactive oxygen species in guard cells. *New Phytol.* **201**, 1121–1140 (2013).
5. Navrot, N. *et al.* Plant glutathione peroxidases are functional peroxiredoxins distributed in several subcellular compartments and regulated during biotic and abiotic stresses. *Plant Physiol.* **142**, 1364–1379 (2006).
6. Caverzan, A. *et al.* Plant responses to stresses: role of ascorbate peroxidase in the antioxidant protection. *Genet. Mol. Biol.* **35**, 1011–1019 (2012).
7. Koua, D. *et al.* PeroxiBase: a database with new tools for peroxidase family classification. *Nucleic Acids Res.* **37**, D261–266 (2009).
8. Dietz, K. J. *et al.* The function of peroxiredoxins in plant organelle redox metabolism. *J. Exp. Bot.* **57**, 1697–1709 (2006).
9. Laborde, E. Glutathione transferases as mediators of signaling pathways involved in cell proliferation and cell death. *Cell Death Differ.* **17**, 1373–1380 (2010).
10. Ursini, F. *et al.* Diversity of glutathione peroxidases. *Methods Enzymol.* **252**, 38–53 (1995).
11. Rotruck, J. T. *et al.* Relationship of selenium to GSH peroxidase. *Fed. Proc.* **31**, 691–694 (1972).
12. Ursini, F., Maiorino, M. & Gregolin, C. The selenoenzyme phospholipid hydroperoxide glutathione peroxidase. *Biochim. Biophys. Acta* **839**, 62–70 (1985).
13. Epp, O., Ladenstein, R. & Wendel, A. The refined structure of the selenoenzyme glutathione peroxidase at 0.2-nm resolution. *Eur. J. Biochem.* **133**, 51–69 (1983).
14. Tosatto, S. C. *et al.* The catalytic site of glutathione peroxidases. *Antioxid. Redox Signal.* **10**, 1515–1526 (2008).
15. Toppo, S., Flohé, L., Ursini, F., Vanin, S. & Maiorino, M. Catalytic mechanisms and specificities of glutathione peroxidases: variations of a basic scheme. *Biochim. Biophys. Acta* **1790**, 1486–1500 (2009).
16. Brigelius-Flohé, R. & Maiorino, M. Glutathione peroxidases. *Biochim. Biophys. Acta* **1830**, 3289–3303 (2013).
17. Maiorino, M. *et al.* The thioredoxin specificity of *Drosophila* GPX: a paradigm for a peroxiredoxin-like mechanism of many glutathione peroxidases. *J. Mol. Biol.* **365**, 1033–1046 (2007).
18. Criqui, M. C. *et al.* Isolation and characterization of a plant cDNA showing homology to animal glutathione peroxidases. *Plant Mol. Biol.* **18**, 623–627 (1992).
19. Holland, D. *et al.* Molecular characterization of salt-stress associated protein in citrus: protein and cDNA sequence homology to mammalian glutathione peroxidases. *Plant Mol. Biol.* **21**, 923–927 (1993).
20. Avsian-Kretchmer, O., Gueta-Dahan, Y., Lev-Yadun, S., Gollop, R. & Ben-Hayyim, G. The salt-stress signal transduction pathway that activates the *gpx1* promoter is mediated by intracellular H<sub>2</sub>O<sub>2</sub>, different from the pathway induced by extracellular H<sub>2</sub>O<sub>2</sub>. *Plant Physiol.* **135**, 1685–1696 (2004).
21. Johnson, R. R., Cranston, H. J., Chaverra, M. E. & Dyer, W. E. Characterization of cDNA clones for differentially expressed genes in embryos of dormant and non-dormant *Avena fatua* (L.) caryopses. *Plant Mol. Biol.* **28**, 113–122 (1995).
22. Sugimoto, M., Furui, S. & Suzuki, Y. Molecular cloning and characterization of a cDNA encoding putative phospholipid hydroperoxide glutathione from spinach. *Biosci. Biotechnol. Biochem.* **61**, 1379–1381 (1997).
23. Eshdat, Y., Holland, D., Faltin, Z. & Ben-Hayyim, G. Plant glutathione peroxidases. *Plant Physiol.* **100**, 234–240 (1997).
24. Roeckel-Drevet, P. *et al.* Molecular characterization, organ distribution and stress-mediated induction of two glutathione peroxidase-encoding mRNAs in sunflower (*Helianthus annuus*). *Physiol. Plant.* **103**, 385–394 (1998).
25. Mullineaux, P. M. *et al.* Identification of cDNAs encoding plastid-targeted glutathione peroxidase. *Plant J.* **13**, 375–379 (1998).
26. Depege, D., Drevet, J. & Boyer, N. Molecular cloning and characterization of tomato cDNA encoding glutathione peroxidase-like proteins. *Eur. J. Biochem.* **253**, 445–451 (1998).
27. Miao, Y. *et al.* An Arabidopsis glutathione peroxidase functions as both a redox transducer and a scavenger in abscisic acid and drought stress responses. *Plant Cell* **18**, 2749–2766 (2006).
28. Chang, C. C. *et al.* Arabidopsis chloroplastic glutathione peroxidases play a role in cross talk between photooxidative stress and immune responses. *Plant Physiol.* **150**, 670–683 (2009).
29. Gaber, A. *et al.* The involvement of Arabidopsis glutathione peroxidase 8 in the suppression of oxidative damage in the nucleus and cytosol. *Plant Cell Physiol.* **53**, 1596–1606 (2012).
30. Passaia, G. *et al.* The mitochondrial glutathione peroxidase GPX3 is essential for H<sub>2</sub>O<sub>2</sub> homeostasis and root and shoot development in rice. *Plant Sci.* **208**, 93–101 (2013).
31. Zhai, C. Z. *et al.* Two wheat glutathione peroxidase genes whose products are located in chloroplasts improve salt and H<sub>2</sub>O<sub>2</sub> tolerances in Arabidopsis. *PLoS One* **8**, e73989 (2013).
32. Kim, Y. J. *et al.* Molecular characterization of two glutathione peroxidase genes of *Panax ginseng* and their expression analysis against environmental stresses. *Gene* **535**, 33–41 (2014).
33. Koh, C. S. *et al.* Crystal structures of a poplar thioredoxin peroxidase that exhibits the structure of glutathione peroxidases: insights into redox-driven conformational changes. *J. Mol. Biol.* **370**, 512–529 (2007).
34. Selles, B. *et al.* Hydroperoxide and peroxynitrite reductase activity of poplar thioredoxin-dependent glutathione peroxidase 5: kinetics, catalytic mechanism and oxidative inactivation. *Biochem. J.* **442**, 369–380 (2012).
35. Fei, G. *et al.* The glutathione peroxidase gene family in *Thellungiella salsuginea*: genome-wide identification, classification, and gene and protein expression analysis under stress conditions. *Int. J. Mol. Sci.* **15**, 3319–3335 (2014).
36. Rodriguez Milla, M. A., Maure, A., Rodriguez Huete, A. & Gustafson, J. P. Glutathione peroxidase genes in Arabidopsis are ubiquitous and regulated by abiotic stresses through diverse signaling pathways. *Plant J.* **36**, 602–615 (2003).
37. Passaia, G., Queval, G., Bai, J., Margis-Pinheiro, M. & Foyer, C. H. The effects of redox controls mediated by glutathione peroxidases on root architecture in *Arabidopsis thaliana*. *J. Exp. Bot.* **65**, 1403–1413 (2014).
38. Romas, J. *et al.* The glutathione peroxidase gene family of *Lotus japonicus*: Characterization of genomic clones, expression analyses and immunolocalization in legumes. *New Phytol.* **181**, 103–114 (2009).
39. Li, F. *et al.* Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.* **33**, 524–530 (2015).
40. Bao, Y. *et al.* Parallel up-regulation of the profilin gene family following independent domestication of diploid and allopolyploid cotton (*Gossypium*). *Proc. Natl. Acad. Sci. USA* **108**, 21152–21157 (2011).
41. Zhang, T. *et al.* Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **33**, 531–537 (2015).
42. Lescot, M. *et al.* PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **30**, 325–327 (2002).
43. Churin, Y., Schilling, S. & Börner, T. A gene family encoding glutathione peroxidase homologues in *Hordeum vulgare* (barley). *FEBS Lett.* **459**, 33–38 (1999).
44. Hernández, J. A., Jiménez, A., Mullineaux, P. & Sevilla, F. Tolerance of pea (*Pisum sativum* L.) to long-term salt stress is associated with induction of antioxidant defences. *Plant Cell Environ.* **23**, 853–862 (2000).
45. Agrawal, G. K., Rakwal, R., Jwa, N. S. & Agrawal, V. P. Effects of signaling molecules, protein phosphatase inhibitors and blast pathogen (*Magnaporthe grisea*) on the mRNA level of a rice (*Oryza sativa* L.) phospholipid hydroperoxide glutathione peroxidase (*OsPHGPX*) gene in seedling leaves. *Gene* **283**, 227–236 (2002).
46. Miao, Y. C., Song, C. P., Dong, F. C. & Wang, X. C. ABA-induced hydrogen peroxide generation in guard cells of *Vicia faba*. *Acta Phytophysiol. Sin.* **26**, 53–58 (2000).



47. Inoue, Y., Matsuda, T., Sugiyama, K., Izawa, S. & Kimura, A. Genetic analysis of glutathione peroxidase in oxidative stress response of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **274**, 27002–27009 (1999).
48. Ferrari, S., Vairo, D., Ausubel, F. M., Cervone, F. & De Lorenzo, G. Tandemly duplicated Arabidopsis genes that encode polygalacturonase-inhibiting proteins are regulated coordinately by different signal transduction pathways in response to fungal infection. *Plant Cell* **15**, 93–106 (2003).
49. Bela, K. *et al.* Plant glutathione peroxidases: emerging role of the antioxidant enzymes in plant development and stress responses. *J. Plant Physiol.* **176**, 192–201 (2015).
50. Xu, X. M. *et al.* The Arabidopsis DJ-1a protein confers stress protection through cytosolic SOD activation. *J. Cell Sci.* **123**, 1644–1651 (2010).
51. Iqbal, A., Yabuta, Y., Takeda, T., Nakano, Y. & Shigeoka, S. Hydroperoxide reduction by thioredoxin-specific glutathione peroxidase isoenzymes of *Arabidopsis thaliana*. *FEBS J.* **273**, 5589–5597 (2006).
52. Dietz, K. J. *et al.* The function of peroxiredoxins in plant organelle redox metabolism. *J. Exp. Bot.* **57**, 1697–1709 (2006).
53. Zhang, W. W. *et al.* Cotton gene expression profiles in resistant *Gossypium hirsutum* cv. Zhongzhimian KV<sub>1</sub> responding to *Verticillium dahliae* strain V<sub>991</sub> infection. *Mol. Biol. Rep.* **39**, 9765–9774 (2012).
54. Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
55. Yoo, S. D., Cho, Y. H. & Sheen, J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572 (2007).

## Acknowledgements

We thank Dr. Prashant Kumar Singh for reading of the manuscript. This research was financially supported in part by grants from the National Natural Science Foundation of China (31370332), Plan for Scientific Innovation Talent of Henan Province (144200510017), National Key Research and Development Program (2016YFD0101006) and the Ministry of Agriculture of China (2016ZX08009-003).

## Author Contributions

Y.M. and C.-P.S. conceived and supervised the study; Y.M. designed the experiments; M.C., K.L. and H.L. performed the experiments; Y.M. provided new tools and reagents; M.C., K.L. and Y.M. analysed the data; K.L. and Y.M. wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Chen, M. *et al.* The Glutathione Peroxidase Gene Family in *Gossypium hirsutum*: Genome-Wide Identification, Classification, Gene Expression and Functional Analysis. *Sci. Rep.* **7**, 44743; doi: 10.1038/srep44743 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017