Functional Divergence of the Glutathione S-Transferase Supergene Family in *Physcomitrella patens* Reveals Complex Patterns of Large Gene Family Evolution in Land Plants[1][W][OA]

Yan-Jing Liu[2], Xue-Min Han[2], Lin-Ling Ren, Hai-Ling Yang, and Qing-Yin Zeng*

State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China (Y.-J.L., L.-L.R., Q.-Y.Z.); College of Life Sciences and Biotechnology, Beijing Forestry University, Beijing 100083, China (X.-M.H., H.-L.Y.); and University of the Chinese Academy of Sciences, Beijing 100049, China (Y.-J.L.)

Plant glutathione S-transferases (GSTs) are multifunctional proteins encoded by a large gene family that play major roles in the detoxification of xenobiotics and oxidative stress metabolism. To date, studies on the GST gene family have focused mainly on vascular plants (particularly agricultural plants). In contrast, little information is available on the molecular characteristics of this large gene family in nonvascular plants. In addition, the evolutionary patterns of this family in land plants remain unclear. In this study, we identified 37 GST genes from the whole genome of the moss *Physcomitrella patens*, a nonvascular representative of early land plants. The 37 *P. patens* GSTs were divided into 10 classes, including two new classes (hemerythrin and iota). However, no tau GSTs were identified, which represent the largest class among vascular plants. *P. patens* GST gene family members showed extensive functional divergence in their gene structures, gene expression responses to abiotic stressors, enzymatic characteristics, and the subcellular locations of the encoded proteins. A joint phylogenetic analysis of GSTs from *P. patens* and other higher vascular plants showed that different class GSTs had distinct duplication patterns during the evolution of land plants. By examining multiple characteristics, this study revealed complex patterns of evolutionary divergence among the GST gene family in land plants.

Glutathione S-transferases (GSTs; EC 2.5.1.8) constitute a family of multifunctional proteins that catalyze the nucleophilic conjugation of reduced tripeptide glutathione (GSH; γ-Glu-Cys-Gly) into a wide variety of hydrophobic and electrophilic substrates. As phase II detoxification enzymes, GSTs are ubiquitous proteins found in the cells of microbes, animals, and plants. Based on sequence and structural similarities, bacterial and fungal GSTs have been mainly divided into beta, zeta, Ure2p, GTT, and γ-subunit classes of the eukaryotic translation elongation factor 1B (EF1Bγ) classes (Sheehan et al., 2001; McGoldrick et al., 2005). Plant GSTs have been divided into eight classes: tau, phi, theta, zeta, lambda, EF1Bγ, dehydroascorbate reductase (DHAR), and tetrachlorohydroquinone dehalogenase (TCHQD; Oakley, 2005; Lan et al., 2009; Dixon and Edwards, 2010a; Jain et al., 2010). Among these, phi, tau, DHAR, and lambda GSTs are specific to plants.

In plants, tau and phi GSTs are the most abundant and are involved mainly in xenobiotic metabolism (Frova, 2003, 2006). Genome-wide analysis of biochemical characteristics of Arabidopsis (*Arabidopsis thaliana*) and poplar (*Populus trichocarpa*) tau and phi GSTs found that these two classes of GSTs have broad substrate specificities (Dixon et al., 2009; Lan et al., 2009), which may be related to the high tolerance to abiotic stresses, especially to a broad spectrum of xenobiotics. Overexpression of tau or phi GSTs increases tolerance to herbicides, salt, and UV stressors in transgenic plants (Karavangeli et al., 2005; Benekos et al., 2010; Jha et al., 2011), indicating a protective role against abiotic stress. Recently, Arabidopsis GSTF2 was found to selectively bind the indole-derived phytoalexin camalexin and the flavonol quercetin-3-O-rhamnoside, suggesting a role in regulating the binding and transport of defense-related compounds in plants (Dixon et al., 2011). DHAR and lambda GSTs have no detectable GSH-conjugating activity toward standard xenobiotic GST substrates, such as 1-chloro-2,4-dinitrobenzene, 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole, or 1,2-dichloro-4-nitrobenzene, but they do function as thioltransferases (Dixon et al., 2002a).

DHAR is also a key enzyme in the ascorbate-glutathione cycle that maintains reduced pools of ascorbic acid, which serves as an important antioxidant (Moons, 2005). A recent study found that lambda GSTs used flavonols as high-affinity ligands and demonstrated a novel glutathione-dependent role for these enzymes in recycling oxidized quercetin (Dixon and Edwards, 2010b). Zeta and theta GSTs have very restricted activities toward xenobiotics. Theta GSTs are glutathione peroxidases and are involved in oxidative stress metabolism (Basantani and Srivastava, 2007). Zeta GSTs participate in Tyr catabolism and have GSH-dependent isomerase reaction activity (Edwards and Dixon, 2005). These studies showed that the plant GST gene family has a high degree of functional overlap and divergence both within and between classes.

Plant GSTs form a large gene family with more than 55 members in the Arabidopsis, poplar, and rice (*Oryza sativa*) genomes (Lan et al., 2009; Dixon and Edwards, 2010a; Jain et al., 2010). To date, studies on the GST gene family have focused mainly on vascular plants, especially on agricultural plants. In contrast, there is virtually no information on the molecular characteristics of this large gene family in nonvascular plants. Fortunately, the sequenced genome of the moss *Physcomitrella patens* allows us to examine the molecular characteristics of the whole GST gene family in a nonvascular plant. As a member of the bryophytes, *P. patens* has emerged as a new model organism because of its simple morphology, dominant haploid gametophyte stage, and highly efficient homologous recombination (Cove, 2005; Rensing et al., 2008). To date, *P. patens* is the only bryophyte whose genome has been sequenced. Bryophytes are the closest extant relatives of early land plants, which began to diverge about 450 million years ago (Rensing et al., 2008). Other model plants, such as Arabidopsis, poplar, and rice, have a much more recent lineage that emerged approximately 140 to 180 million years ago (Soltis et al., 2008). Thus, *P. patens* is an ideal model for plant evolutionary studies and can provide new insights into the evolution of the mechanisms behind the complexity of modern plants, such as rice and Arabidopsis. In addition, when combined with analyses of data from other higher plants, such as Arabidopsis, rice, and *Selaginella moellendorffii*, *P. patens* can be used to reveal evolutionary patterns in the large gene family of land plants.

In this study, a genome-wide analysis of *P. patens* GSTs was performed, and functional divergence was observed among *P. patens* GSTs by examining multiple characteristics, such as gene expression patterns, biochemical characteristics, and subcellular locations. In addition, based on joint phylogenetic analyses of GSTs from *P. patens* and other higher plants, this study revealed complex patterns of evolutionary divergence of the GST gene family in land plants.

RESULTS

P. patens Has a Large GST Gene Family

Thirty-seven full-length genes encoding putative GSTs were identified in the *P. patens* genome (Supplemental Table S1). Among them, three sequences were considered to be putative pseudogenes, based on the presence of a frame shift disrupting the coding region or a stop codon occurring prematurely, resulting in a truncated protein. After removing these stop codons or revising the frame shifts by deleting one or two nucleotides, the three sequences were included in the phylogenetic and gene expression analyses. Domain analysis using the National Center for Biotechnology Information (NCBI) conserved domain search indicated that all the predicted proteins encoded by the 37 genes contain typical GST N- and C-terminal domains, suggesting that all 37 genes were members of the GST family.

To define the classes of *P. patens* GSTs, the 37 full-length *P. patens* GST protein sequences were subjected to a joint phylogenetic analysis with 152 full-length GST protein sequences representing 36 GST classes from plants, animals, fungi, and bacteria (Supplemental Data Set S1). The phylogenetic tree showed that 28 *P. patens* GSTs were grouped with eight GST classes: the phi, DHAR, theta, zeta, lambda, EF1Bγ, Ure2p, and TCHQD classes (Fig. 1). The other nine *P. patens* GSTs were grouped into two separate clades with high bootstraps, indicating that these nine GSTs might belong to two new classes. Among the nine *P. patens* GSTs, eight contained both a GST and a hemerythrin domain and were grouped into one distinct clade; they were named as "hemerythrin class" GSTs. The other GST (PpGSTI1) was grouped as a distinct clade. To investigate whether PpGSTI1 represented a new GST class, we identified its homologs from *S. moellendorffii*, four green algae, and two diatoms species (listed in Supplemental Data Set S2) using TBLASTN searches of the NCBI nucleotide database. We did not find homologs of PpGSTI1 in bacteria, fungi, animals, or other plants. The phylogenetic tree showed that PpGSTI1 and its homologs were grouped into one distinct clade (Supplemental Fig. S1), confirming that PpGSTI1 and its homologs were a new class of GST. For notational convenience, PpGSTI1 and its homologs were thus named the "iota class" GSTs.

Among the 37 *P. patens* GSTs, phi and hemerythrin GSTs were the most numerous, with 10 and eight members, respectively. The TCHQD and EF1Bγ classes had five and four members, respectively; both the DHAR and theta classes had three members; and the zeta, lambda, Ure2p, and iota classes were represented by one member each.

Divergent Gene Structures of *P. patens* GST Family Genes

The exon-intron organization of the 37 *P. patens* GST genes was examined. The eight hemerythrin class GSTs have highly conserved intron positions and numbers.
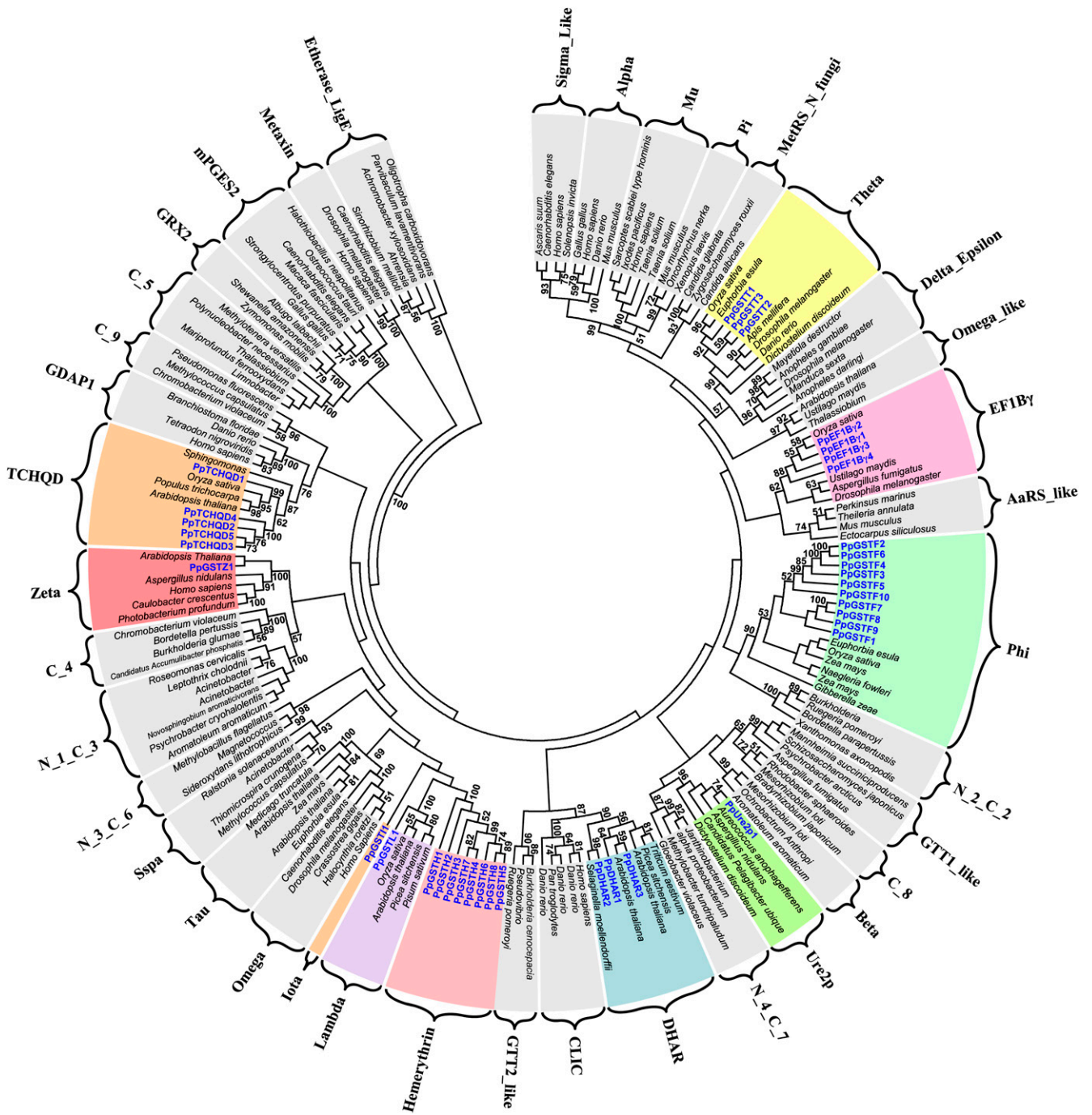
Figure 1. Phylogenetic relationships among 37 *P. patens* GSTs and 152 full-length proteins representing 36 GST classes from plants, animals, fungi, and bacteria. Phylogenetic trees were constructed using RAxML software with the WAG+G model and 100 bootstrap replicates. Numbers at each node in the phylogenetic tree represent bootstrap values, and only bootstrap values greater than 50% are shown. Blue letters represent *P. patens* GSTs. The 10 clades harboring *P. patens* GSTs are shaded in different colors, and the remaining 28 clades without *P. patens* GSTs are shaded in gray.

However, intron numbers were variable within each of the phi, theta, EF1Bγ, TCHQD, and DHAR classes (Fig. 2). Among the 10 *P. patens* phi GST genes, four (*PpGSTF1* and *PpGSTF7* to *PpGSTF9*) had two introns, while the other six (*PpGSTF2* to *PpGSTF6* and *PpGSTF10*) contained three introns. Three theta GSTs (*PpGSTT1* to

PpGSTT3) contained six, four, and five introns, respectively. However, intron positions and numbers were conserved in the N-terminal domain of the three theta GSTs (Fig. 2, theta class). Four *PpEF1Bγ*s contained both a GST and an EF1Bγ domain; intron positions and numbers were conserved in the EF1Bγ

Figure 2. Gene structures of *P. patens* GSTs. The GST N-terminal domain, C-terminal domain, EF1Bγ domain, and hemerythrin domain are highlighted in blue, purple, gray, and green boxes, respectively. Introns are indicated as lines. Conserved intron positions within each class are indicated by dashed lines. Phylogenetic relationships among these classes are shown in Figure 4. Bootstrap values are given at each node.
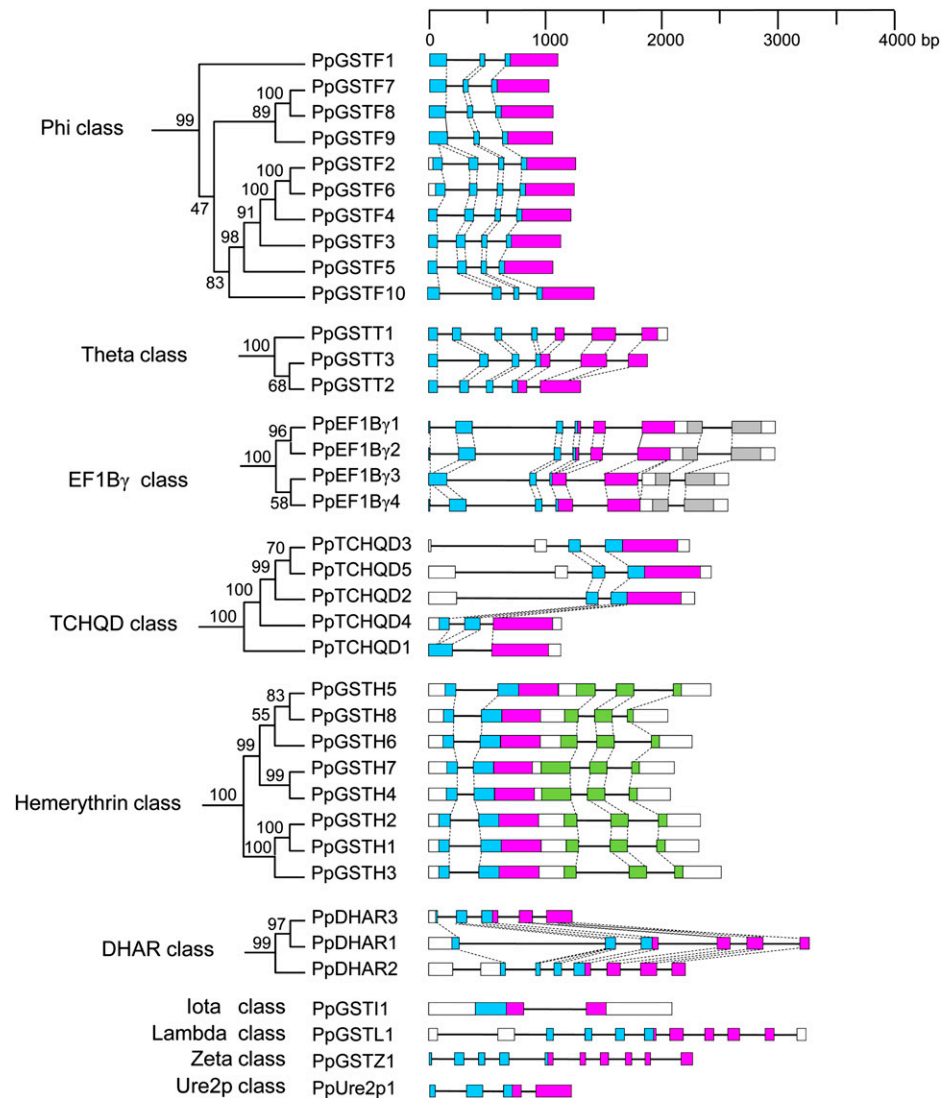
domain, while variable intron numbers were observed in the GST domain (Fig. 2, EF1Bγ class). Among the five *P. patens* TCHQD GSTs, *PpTCHQD3* and *PpTCHQD5* contained three introns, *PpTCHQD2* and *PpTCHQD4* had two introns, and *PpTCHQD1* had only one. Intron positions and numbers were conserved in the GST domain regions of *PpTCHQD2*, *PpTCHQD3*, and *PpTCHQD5*. Highly variable gene structures were observed in three *PpDHAR* genes. *PpDHAR1* to *PpDHAR3* contained five, seven, and four introns, respectively. The extensive divergence of gene structures in *P. patens* phi, theta, EF1Bγ, TCHQD, and DHAR classes indicated high rates of intron gain/loss.

Sequence and Structural Features of *P. patens* Hemerythrin and Iota Class GSTs

Two new GST classes (hemerythrin and iota classes) were identified in the *P. patens* genome. The *P. patens*

genome contained eight hemerythrin GSTs and one iota GST. Pairwise comparisons of GST domain protein sequences of 92 plant GSTs representing 10 GST classes (Supplemental Data Set S3) revealed some notable features (Supplemental Fig. S2). Eight *P. patens* hemerythrin GSTs showed 38.0% to 94.7% pairwise sequence identity. However, the GST domain protein sequences between the eight hemerythrin GSTs and the nine other GST classes were significantly different (independent sample Student's *t* test, $P < 0.0001$) and had less than 23.8% pairwise sequence identity (Supplemental Fig. S2). *P. patens* iota GST (PpGSTI1) showed 67.6% protein sequence identity with *S. moellendorffii* iota GST (SmGSTI1), while PpGSTI1 showed less than 27.7% protein sequence identity with the other nine GST classes (Supplemental Fig. S2). The protein sequence analyses indicate that the hemerythrin and iota GSTs represent two novel GST classes.

The three-dimensional structures of the GST domains from eight hemerythrin GSTs (PpGSTH1 to PpGSTH8) and one iota GST (PpGSTI1) were modeled based on the x-ray structure of wheat (*Triticum aestivum*) tau GST (TaGSTU4; Protein Data Bank [PDB] code 1GWC) and human omega GST (PDB code 1EEM), respectively. Nine proteins showed typical GST structures (Fig. 3B; Supplemental Fig. S3) with two distinct domains, namely, a smaller thioredoxin-like N-terminal domain (purple domain in Fig. 3B) and a larger helical C-terminal domain (gray domain in Fig. 3B). The N-terminal domain consisted of a $\beta\alpha\beta\alpha\beta\beta\alpha$ structural motif in which $\beta 3$ was antiparallel with respect to the other β-strands and the C-terminal domain was composed entirely of helices. The GST domains of eight hemerythrin GSTs shared the same conformation of the structural elements of α-helices and β-sheets, but structural modifications were present in the loop regions (Supplemental Fig. S3).

Despite the low protein sequence identity among plant GSTs, certain structural conservation was evident. For example, the GSH-binding N-terminal domain was highly conserved (Frova, 2003). The x-ray structure of TaGSTU4 indicated that Ser-15, Lys-42, Ile-56, Glu-68, and Ser-69 (alignment positions 10, 40, 55, 68, and 69 in Fig. 3A) are GSH-binding site (G-site) residues (Thom et al., 2002) and form H bonds with GSH. Based on the crystal structure of TaGSTU4 and multiple sequence alignment of the N-terminal protein sequence of plant GSTs (Fig. 3), Cys-46, Tyr-76, Tyr-90, Gly-102, and Ser-103 of PpGSTH3 and Cys-140, Lys-170, Leu-183, Glu-195, and Ser-196 of PpGSTI1 (alignment positions 10, 40, 55, 68, and 69 in Fig. 3A) were predicted to be G-site residues.

Expression Divergence of *P. patens* GST Genes under Normal and Abiotic Stress Conditions

To examine functional diversification in *P. patens* GST genes, the patterns of transcript accumulation of the 37 genes were examined by semiquantitative PCR under normal conditions and in response to stress treatments (hydrogen peroxide [H_2O_2], salt, salicylic acid, and atrazine applications). Some *P. patens* GST genes (e.g. *PpGSTF3* and *PpDHAR1*) could be detected by PCR that underwent 24, 26, 28, or 30 amplification cycles, while some genes (e.g. *PpGSTF6*) were only detected by PCR that underwent 30 amplification cycles. Thus, this study focused on the expression patterns of *P. patens* GST genes under 30 PCR amplification cycles. Among the 37 *P. patens* GST genes, 23 were expressed under all growth conditions, six (*PpGSTF2* and *PpGSTF5* to *PpGSTF7*, *PpGSTT3*, and *PpDHAR2*) were selectively expressed in response to specific treatments, and the expression of eight genes (*PpGSTF8*, *PpGSTT2*, *PpEF1Bγ3*, *PpGSTH1*, *PpGSTH5*, and *PpGSTH8*, *PpTCHQD4*, and *PpDHAR3*) was not detected by PCR in any of the samples examined (Fig. 4). To further examine the expression patterns of these eight *P. patens* GST genes not detected by PCR, BLASTN searches of the NCBI *P. patens* EST database (containing 382,587 *P. patens* EST sequences) were performed. Except for *PpGSTH1*, we did not find any EST sequence for the other seven genes in the *P. patens* EST database. The *PpGSTH1* gene had three EST sequences (GenBank accession nos. BJ592106, BY971487, and BJ184722) that were expressed in *P. patens* cultivated under continuous light conditions. Thus, except for *PpGSTH1*, the other seven *P. patens* GST genes may be
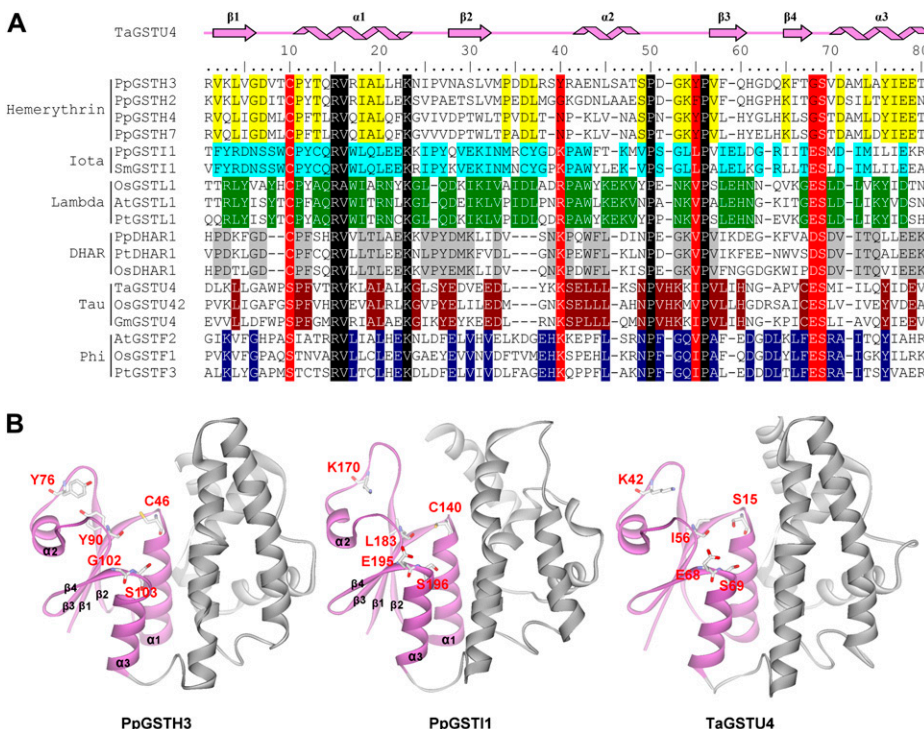

Figure 3. Sequence and structural analysis of plant GST proteins. A, Protein sequence alignment of the N-terminal domains of plant hemerythrin, iota, lambda, DHAR, tau, and phi GSTs. α-Helices and β-strands are represented as helices and arrows, respectively. Highly conserved residues in all plant GSTs are marked in black. Predicted G-sites are marked in red. Residues marked in yellow, light blue, green, gray, carmine, and navy blue indicate conserved residues in hemerythrin, iota, lambda, DHAR, tau, and phi GSTs, respectively. The protein sequences used to create this sequence alignment are available as Supplemental Data Set S3. B, Structural comparison between the GST domains of PpGSTH3, PpGSTI1, and TaGSTU4. The N- and C-terminal domains of three GST proteins are illustrated in purple and gray, respectively. The PDB code and reference for the x-ray structure of TaGSTU4 is 1GWC (Thom et al., 2002).

Figure 4. Phylogenetic relationships among *P. patens* GSTs (A) and their expression patterns (B). The maximum-likelihood procedure with the WAG+I+G model and 100 bootstrap replicates was used for phylogeny reconstruction. Numbers at each node in the phylogenetic tree are bootstrap values, and only bootstrap values greater than 50% are shown. Putative pseudogenes are indicated with asterisks. GST genes designated as GSTF, GSTT, GSTZ, GSTL, GSTH, and GSTI correspond to phi, theta, zeta, lambda, hemerythrin, and iota class GSTs, respectively. Different GST classes are shaded with different colors. The expression patterns of *P. patens* GST genes were examined using semiquantitative PCR under normal conditions (NC) and in response to H_2O_2 (HO), salt (ST), salicylic acid (SA), and atrazine (AT) stress treatments.
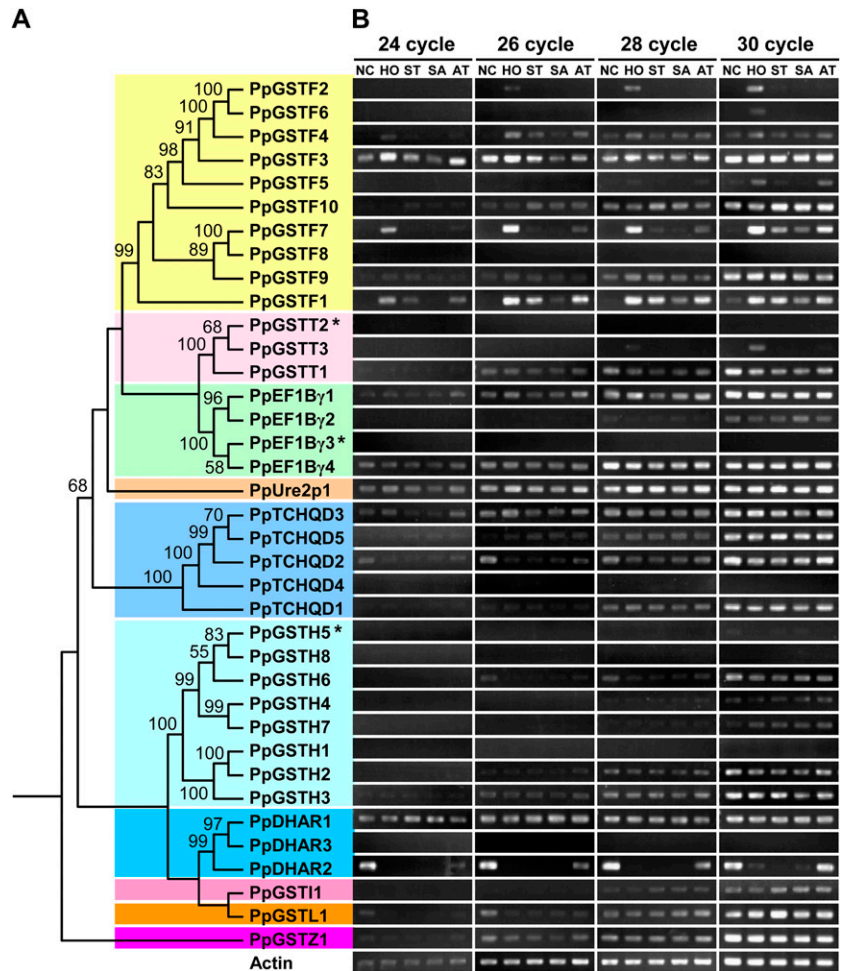
expressed at subdetectable levels, or they may only be induced in response to a specific treatment not tested here, or they may be pseudogenes.

Among the six *P. patens* GST genes selectively expressed in response to specific treatments, expression of the *PpDHAR2* gene was repressed by H_2O_2, salt, and salicylic acid treatments, while the other five genes (*PpGSTF2* and *PpGSTF5* to *PpGSTF7* and *PpGSTT3*) were induced by specific treatments. Three genes (*PpGSTF2*, *PpGSTF6*, and *PpGSTT3*) were expressed only in response to H_2O_2 treatment, suggesting that they may play roles in oxidative tolerance. Expression of *PpGSTF7* was not detected under normal conditions, but it was induced by all the stress treatments performed here. *PpGSTF5*, which was expressed at a low level under normal conditions, was induced by H_2O_2, salt, and atrazine treatments. Thus, *PpGSTF7* and *PpGSTF5* may play roles in tolerance to abiotic stresses.

Subcellular Localization of *P. patens* GSTs

Eukaryotic cells are highly compartmentalized, and the correct localization of proteins is essential for their function (Boruc et al., 2010). Subcellular divergence

may play a significant role in the functional divergence of gene families (Marques et al., 2008; Qian and Zhang, 2009). To investigate the subcellular divergence of *P. patens* GSTs, 21 GST proteins representing 10 classes were tested for subcellular localization using C-terminal GFP fusions and visualization by confocal microscopy after transient expression of the fusions in *Nicotiana benthamiana*. GST-GFP fusions of 16 *P. patens* GSTs showed typical cytosolic and nuclear localizations, similar to GFP alone, as illustrated by PpGSTF1-GFP (Fig. 5). The 16 proteins consisted of four phi (PpGSTF1, PpGSTF4, PpGSTF9, and PpGSTF10), four hemerythrin (PpGSTH1 to PpGSTH3 and PpGSTH7), three EF1Bγ (PpEF1Bγ1, PpEF1Bγ2, and PpEF1Bγ4), two theta (PpGSTT1 and PpGSTT3), one TCHQD (PpTCHQD2), one Ure2p (PpUre2p1), and one zeta (PpGSTZ1) GSTs. PpTCHQD3-GFP was only found localized in the cytosol. Two GST-GFP fusions (PpTCHQD5 and PpGSTL1) localized to discrete structures; they colocalized with the red autofluorescence of chlorophyll (Fig. 5), a hallmark of chloroplasts. Fluorescent spots were not observed in other subcellular locations. Thus, PpTCHQD5 and PpGSTL1 localized to chloroplasts. Using the same
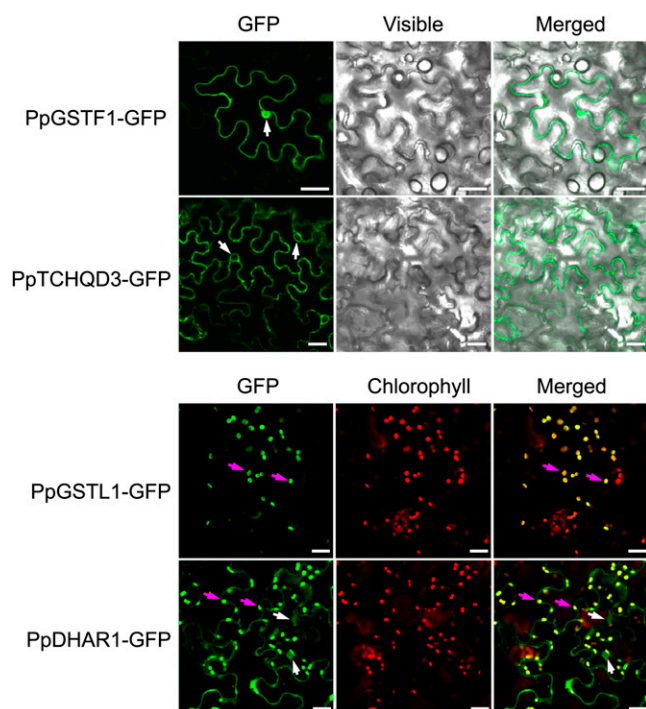
Figure 5. Subcellular localizations of *P. patens* GST proteins. GFP signal (green) and chlorophyll autofluorescence (red) were observed using confocal laser-scanning microscopy, and an overlay is shown in yellow. Chloroplasts and nuclei are indicated with purple and white arrows, respectively. Bars = 20 μm.

method, PpDHAR1 and PpGSTI1 proteins were shown to localize to both the cytosol and chloroplasts (Fig. 5).

Expression and Purification of *P. patens* GST Proteins

P. patens GST proteins were examined for their catalytic activities, which may be related to their biological functions. Fifteen *P. patens* GSTs from six classes were selected for protein expression and purification: eight phi (PpGSTF1 to PpGSTF5, PpGSTF7, PpGSTF9, and PpGSTF10), two DHAR (PpDHAR1 and PpDHAR2),

two hemerythrin (PpGSTH3 and PpGSTH7), one lambda (PpGSTL1), one Ure2p (PpUre2p1), and one iota (PpGSTI1) GSTs. Nine *P. patens* GST proteins (PpGSTF1, PpGSTF3, PpGSTF7, PpGSTF9, and PpGSTF10, PpDHAR1 and PpDHAR2, PpUre2p1, and PpGSTL1) were expressed as soluble proteins in *Escherichia coli* BL21 (DE3) cells. However, six proteins (PpGSTF2, PpGSTF4, and PpGSTF5, PpGSTH3 and PpGSTH7, and PpGSTI1) were expressed as inclusion bodies in *E. coli* BL21 (DE3) and in *E. coli* Tuner (DE3) cells, respectively. In addition, the GST domains of PpGSTH3 and PpGSTH7 proteins were expressed as inclusion bodies in these cell lines. To increase the solubility of PpGSTH3 and PpGSTH7 and PpGSTI1 proteins, the PpGSTH3 and PpGSTH7 and PpGSTI1 GST domains were fused to the thioredoxin tag, which is a highly soluble polypeptide that can increase the solubility of target proteins (Kern et al., 2003). However, the recombinant proteins fused to the thioredoxin tag were still expressed as inclusion bodies in *E. coli* BL21 (DE3) and *E. coli* Tuner (DE3) cells. Thus, in this study, only nine *P. patens* GSTs (PpGSTF1, PpGSTF3, PpGSTF7, PpGSTF9, and PpGSTF10, PpDHAR1 and PpDHAR2, PpUre2p1, and PpGSTL1) expressed as soluble proteins were subsequently analyzed for enzymatic characteristics.

Substrate Specificity

Nine soluble *P. patens* GST proteins were assayed for activities toward eight substrates: 1-chloro-2,4-dinitrobenzene (CDNB), 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl), 1,2-dichloro-4-nitrobenzene (DCNB), 4-nitrobenzyl chloride (NBC), fluorodifen, cumene hydroperoxide (Cum-OOH), 2-hydroxyethyl disulfide (HED), and dehydroascorbic acid (DHA; Table I). None of the five purified phi GSTs and PpUre2p1 showed activity toward DHA; in contrast, two DHAR GSTs (PpDHAR1 and PpDHAR2) and one lambda GST (PpGSTL1) did, suggesting that different GST classes showed functional divergence. PpUre2p1 protein showed enzymatic activities toward CDNB, NBD-Cl, and Cum-OOH. Two DHAR proteins (PpDHAR1

Table I. *Specific activities of P. patens GSTs toward eight substrates*

Values shown are means ± SD, calculated from three replicates. n.d., No activity detected.

GST	Specific Activities							
	CDNB	NBD-Cl	DCNB	NBC	Fluorodifen	Cum-OOH	DHA	HED
	$\mu mol\ min^{-1}\ mg^{-1}$							
PpGSTF1	1.21 ± 0.02	0.46 ± 0.10	n.d.	0.62 ± 0.10	n.d.	0.21 ± 0.01	n.d.	n.d.
PpGSTF3	7.45 ± 0.23	3.16 ± 0.06	0.04 ± 0.01	1.23 ± 0.23	n.d.	0.13 ± 0.01	n.d.	n.d.
PpGSTF7	0.85 ± 0.02	0.24 ± 0.01	0.01 ± 0.01	0.61 ± 0.05	0.43 ± 0.03	n.d.	n.d.	n.d.
PpGSTF9	n.d.	n.d.	n.d.	n.d.	n.d.	0.01 ± 0.01	n.d.	n.d.
PpGSTF10	4.89 ± 0.07	0.34 ± 0.01	0.04 ± 0.01	0.25 ± 0.05	n.d.	0.14 ± 0.01	n.d.	n.d.
PpUre2p1	0.29 ± 0.02	1.55 ± 0.07	n.d.	n.d.	n.d.	0.04 ± 0.01	n.d.	n.d.
PpDHAR1	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	54.10 ± 0.87	n.d.
PpDHAR2	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.03 ± 0.05	n.d.
PpGSTL1	n.d.	n.d.	n.d.	n.d.	n.d.	0.02 ± 0.01	0.01 ± 0.01	n.d.

and PpDHAR2) only showed activity toward the DHA substrate, while PpDHAR1 had 52.5-fold higher enzyme activity than PpDHAR2. PpGSTL1 showed low enzyme activities to Cum-OOH and DHA.

The enzyme assays showed that five purified phi GSTs had large variation in substrate spectra. For example, PpGSTF9 only had activity to substrate Cum-OOH, while PpGSTF3 and PpGSTF10 showed activities to five substrates: CDNB, NBD-Cl, DCNB, NBC, and Cum-OOH. Although PpGSTF3 and PpGSTF10 shared a similar spectrum, their enzymatic activities varied 1.5- to 9.3-fold toward substrates CDNB, NBD-Cl, and NBC. Of five purified phi GSTs, only PpGSTF7 showed specific activity to the herbicide fluorodifen. These results indicate that *P. patens* phi GSTs may have extensive functional divergence.

Kinetics of the Conjugation Reaction of *P. patens* GSTs

GST enzymes can catalyze the conjugation of GSH to a variety of electrophilic substrates. In this study, the apparent kinetic constants of four *P. patens* phi GSTs (PpGSTF1, PpGSTF3, PpGSTF7, and PpGSTF10), PpUre2p1, and two DHARs (PpDHAR1 and PpDHAR2) were determined using CDNB and DHA as substrates (Table II). Because PpGSTF9 had only weak activity to substrate Cum-OOH and PpGSTL1 had low activities to Cum-OOH and DHA, we did not examine the kinetics of the two proteins. For four phi GSTs and PpUre2p1, the apparent K_m^{GSH} values fell within the range of 0.11 to 0.71 mM, indicating that they had high affinities for GSH. The catalytic efficiency (k_{cat}/K_m) for GSH, however, varied significantly among the four phi GSTs. Among the four *P. patens* phi GSTs, PpGSTF1 had the highest affinity to CDNB but lower k_{cat}/K_m due to a lower turnover rate (k_{cat}). Similar to PpGSTF1, PpDHAR2 had higher affinity for DHA than PpDHAR1, but a lower k_{cat} resulted in a 958-fold lower k_{cat}/K_m than PpDHAR1.

Temperature and pH Profiles of *P. patens* Phi GSTs

Four *P. patens* phi GSTs appeared to have a broad temperature optimum (Fig. 6A). PpGSTF1 and PpGSTF7

showed a similar optimum temperature spectrum, retaining more than 50% of its maximum enzymatic activity at a temperature range of 20°C to 55°C. Compared with PpGSTF1, PpGSTF3, and PpGSTF7, PpGSTF10 showed a distinct optimum temperature spectrum, losing all activity at 55°C, whereas PpGSTF1, PpGSTF3, and PpGSTF7 retained 89%, 32%, and 65% of their maximum enzymatic activity, respectively.

Four *P. patens* phi GSTs showed broad pH optima (Fig. 6B), retaining more than 65% of their maximum enzymatic activity over a pH range of 7.0 to 9.5. At pH 6.0, PpGSTF1 still retained more than 71% of its maximum enzymatic activity, while PpGSTF3, PpGSTF7, and PpGSTF10 retained only 43%, 10%, and 12% of their maximum enzymatic activity, respectively. At pH 10.0, PpGSTF1, PpGSTF3, and PpGSTF7 retained more than 64% of their maximum enzymatic activity, while PpGSTF10 retained only 18% of its maximum enzymatic activity. These data indicated that, among the four proteins, PpGSTF1 and PpGSTF10 had a broader and a narrower pH optima, respectively.

Molecular Evolution of the GST Supergene Family in Land Plants

With some new plant genomes now available (e.g. the lycophyte *S. moellendorffii*), we are able to gain new insights into the evolution of the GST gene family in land plants. In this study, we identified GST genes from the *S. moellendorffii* genome. GST gene families in the poplar, Arabidopsis, and rice genomes were identified in previous studies (Lan et al., 2009; Dixon and Edwards, 2010a; Jain et al., 2010). No whole-genome sequence is currently available for any gymnosperm species; the most comprehensive genomic resource is the EST collection for loblolly pine (*Pinus taeda*). In the database of 328,662 loblolly pine ESTs, 56 GST genes were identified using TBLASTN searches (Table III). Tau GSTs were the most numerous in the poplar, Arabidopsis, rice, loblolly pine, and *S. moellendorffii* genomes, with 58, 28, 52, 35, and 47 members, respectively. However, the *P. patens* genome contained no tau GST. Ure2p, hemerythrin, and iota GSTs were found only in the *S. moellendorffii* and *P. patens* genomes. Lambda GST was absent in *S. moellendorffii*.

Table II. *Kinetic constants of P. patens GSTs for CDNB and DHA conjugation reactions*

Values shown are means ± SD, calculated from three replicates.

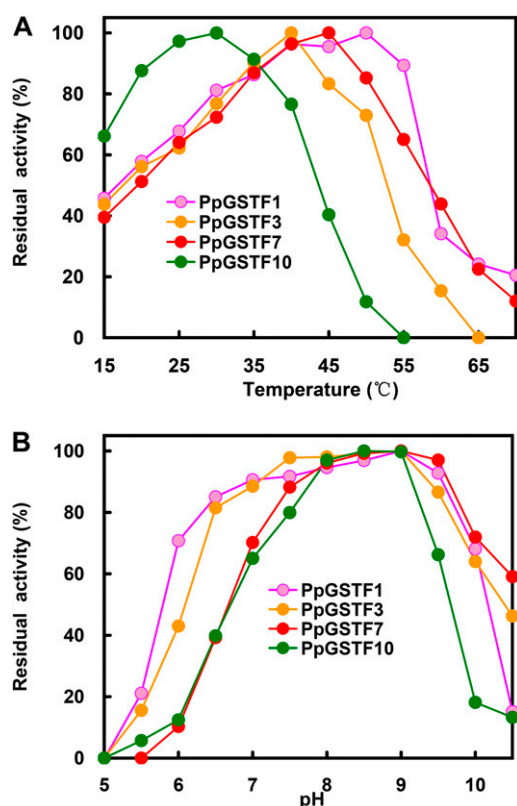| GST	K_m^{GSH}	k_{cat}^{GSH}	$(k_{cat}/K_m)^{GSH}$	K_m^{CDNB}	k_{cat}^{CDNB}	$(k_{cat}/K_m)^{CDNB}$	K_m^{DHA}	k_{cat}^{DHA}	$(k_{cat}/K_m)^{DHA}$
	mM	s^{-1}	m$M^{-1}\,s^{-1}$	mM	s^{-1}	m$M^{-1}\,s^{-1}$	mM	s^{-1}	m$M^{-1}\,s^{-1}$
PpGSTF1	0.16 ± 0.01	1.35	8.44	1.35 ± 0.08	2.03	1.50			
PpGSTF3	0.11 ± 0.01	27.93	253.91	3.69 ± 0.45	83.31	22.58			
PpGSTF7	0.55 ± 0.04	1.02	1.85	2.74 ± 0.35	2.89	1.05			
PpGSTF10	0.71 ± 0.11	53.59	75.48	1.73 ± 0.23	82.19	47.51			
PpUre2p1	0.41 ± 0.04	0.33	0.80	3.46 ± 0.47	0.96	0.28			
PpDHAR1	3.64 ± 0.34	50,123.99	13,770.33				0.17 ± 0.01	8,105.09	47,677.00
PpDHAR2	1.08 ± 0.10	12.85	11.90				0.13 ± 0.02	6.47	49.77

Figure 6. Effects of temperature (A) and pH (B) on *P. patens* phi GSTs.

We separately reconstructed phylogenetic relationships among the phi, zeta, theta, lambda, DHAR, TCHQD, and EF1Bγ GST classes from poplar, Arabidopsis, rice, loblolly pine, *S. moellendorffii*, and *P. patens* (Supplemental Fig. S4; Supplemental Data Set S4). Among the 59 phi GSTs from the six species, 10 *P. patens* phi GSTs were grouped into three distinct clades (Supplemental Fig. S4A). Two *S. moellendorffii* phi GSTs were grouped into one clade, while angiosperm and gymnosperm phi GSTs were grouped together.

Fifteen lambda GSTs from poplar, Arabidopsis, rice, loblolly pine, and *P. patens* were grouped into two clades (clades A and B in Supplemental Fig. S4B). *P. patens* only had one lambda GST, which was grouped into clade B. The other 14 lambda GSTs were grouped into a seed plant-specific clade (clade A).

Sixteen EF1Bγ GSTs from the six species were divided into two clades (clades A and B in Supplemental Fig. S4C). Two *S. moellendorffii* EF1Bγ GSTs were grouped into clade B, and the other 14 were grouped into clade A. Clade A was divided into seed plant-specific subclade A1 and *P. patens*-specific subclade A2. Based on the phylogenetic tree, we predicted that two ancestral EF1Bγ GSTs might occur in the ancestor of land plants. One ancestral copy was retained only in lycophytes, while the other copy was lost only in lycophytes (Fig. 7).

P. patens had five TCHQD GSTs, while each of the other five species had only one. Four *P. patens* TCHQD GSTs (PpTCHQD2 to PpTCHQD5) were grouped into one clade with 100% bootstrap support (clade B in Supplemental Fig. S4D). PpTCHQD1 and five TCHQD GSTs from poplar, Arabidopsis, rice, loblolly pine, and *S. moellendorffii* were grouped into one clade with 99% bootstrap support (clade A in Supplemental Fig. S4D). In clade A, the gene tree was consistent with the species tree, indicating that a duplication event might not have occurred in this clade. Thus, the six genes in clade A might be orthologous. Based on the phylogenetic tree, the ancestor of the land plants may contain two TCHQD genes (Fig. 7). Among the two ancestral TCHQD genes, one (*PpTCHQD1* in *P. patens*) was retained in all land plants and did not expand during the evolution of land plants, while the other one existed only in bryophytes and underwent rapid expansion in the *P. patens* genome.

Eighteen DHARs from six species were divided into three clades (clades A, B, and C in Supplemental Fig. S4E). Clade A contained 13 DHARs from six species, while clade B contained four DHARs from loblolly pine, *S. moellendorffii*, and *P. patens*. Clade C contained only one *P. patens* DHAR (PpDHAR3). The ancestor of the land plants might contain three ancestral DHARs (Fig. 7). The first ancestral DHAR (*PpDHAR1* in *P. patens*) was retained in all land plants, the second (*PpDHAR2* in *P. patens*) was lost in angiosperms, and the third (*PpDHAR3* in *P. patens*) was only found in bryophytes (Fig. 7).

Among the 13 zeta GSTs from six species, PpGSTZ1 and SmGSTZ1 and SmGSTZ2 were grouped into one *P. patens* + *S. moellendorffii*-specific clade (clade B in Supplemental Fig. S4F). The other 10 zeta GSTs from poplar, Arabidopsis, rice, and loblolly pine were grouped into a seed plant-specific clade (clade A in Supplemental Fig. S4F). Based on a phylogenetic tree of 13 zeta GSTs, we predicted that two ancestral zeta GSTs might occur in the ancestor of land plants. One

Table III. *GST gene family in land plants*

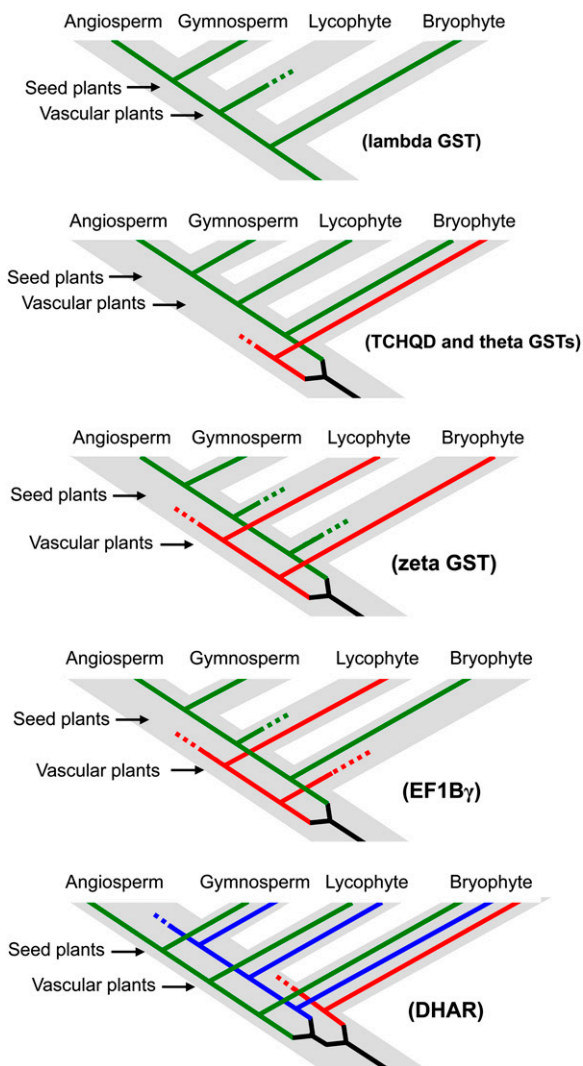| Organism | GST Class | | | | | | | | | | | Total GSTs |
	Tau	Phi	Zeta	Theta	Lambda	DHAR	TCHQD	EF1Bγ	Ure2p	Hemerythrin	Iota	
Poplar	58	9	2	2	3	3	1	3	0	0	0	81
Arabidopsis	28	13	2	3	3	5	1	2	0	0	0	57
Rice	52	17	4	1	3	2	1	3	0	0	0	83
Loblolly pine	35	8	2	1	5	2	1	2	0	0	0	56
S. moellendorffii	47	2	2	3	0	3	1	2	2	2	1	65
P. patens	0	10	1	3	1	3	5	4	1	8	1	37

Figure 7. Schematic representation of the duplication histories of each class of GSTs in land plants. Gene loss events are indicated by dashed lines.

ancestral copy was only retained in lycophytes and bryophytes, while the other copy was only retained in seed plants (Fig. 7).

P. patens contained three theta GSTs, two (PpGSTT2 and PpGSTT3) of which were grouped into one clade (clade B in Supplemental Fig. S4G). PpGSTT1 and 10 theta GSTs from poplar, Arabidopsis, rice, loblolly pine, and *S. moellendorffii* were grouped into another clade with high bootstrap support (clade A in Supplemental Fig. S4G). The ancestor of land plants might contain two ancestral theta GSTs. One ancestral copy was retained in all land plants, while the other copy was lost in vascular plants (Fig. 7).

DISCUSSION

GSTs are multifunctional proteins encoded by a large gene family found in Arabidopsis, poplar,

soybean (*Glycine max*), rice, and maize (*Zea mays*). Plant GSTs are divided into eight distinct classes: tau, phi, theta, zeta, lambda, DHAR, TCHQD, and EF1Bγ (Oakley, 2005; Dixon and Edwards, 2010a). Among the eight classes, tau GSTs were the most numerous in the poplar, Arabidopsis, rice, loblolly pine, and *S. moellendorffii* genomes but are absent from the *P. patens* genome, indicating that tau GSTs might have originated in vascular plants. Compared with nonvascular plants, vascular plants include a root system, shoot system, and vascular system, which can allow them to conserve and conduct water and to grow to larger sizes. These characteristics allow vascular plants to adapt well to land living. Tau GSTs play important roles in plant detoxification metabolism, stress tolerance, and light signaling (Loyall et al., 2000; Thom et al., 2002; Benekos et al., 2010; Jha et al., 2011). The origin and expansion of tau GSTs in vascular plants might be correlated with their broad adaptations to land living.

In this study, we identified two new GST classes, namely, the iota and hemerythrin classes. Iota class GSTs were found only in *P. patens* and *S. moellendorffii* genomes. These two species only contained one iota GST each. *P. patens* iota GST (*PpGSTI1*) was expressed under all growth conditions, indicating that this gene might be functional in vivo. Based on sequence and structural comparisons, four predicted G-site residues, Lys-170, Leu-183, Glu-195, and Ser-196 of PpGSTI1 (alignment positions 40, 55, 68, and 69 in Fig. 3A), were conserved in the iota, tau, and phi GSTs, suggesting that PpGSTI1 may play a role in binding GSH. The Ser-15 residue of TaGSTU4 (alignment position 10 in Fig. 3A) is highly conserved in tau and phi GSTs and was thought to be the catalytic residue responsible for stabilization of the thiolate anion of enzyme-bound GSH (Armstrong, 1997; Thom et al., 2002). However, this critical residue was replaced by the Cys in iota, DHAR, and lambda GSTs. In DHAR and lambda GSTs, this Cys residue was thought to catalyze the formation of mixed disulfides with GSH instead of stabilizing the GSH thiolate anion (Dixon et al., 2002a). Mutation of this Cys residue in Arabidopsis AtDHAR1 abolished enzymatic activity (Dixon et al., 2002). Thus, similar to DHAR and lambda GSTs, iota class GSTs may function as oxidoreductases rather than conjugating GSTs.

Hemerythrin is a nonheme iron protein found in metazoans, prokaryotes, protozoans, and fungi (Bailly et al., 2008; French et al., 2008). In a phylogenetic tree of plant, fungi, and bacteria hemerythrin domains (Supplemental Data Set S5), all plant hemerythrin domains were grouped into one distinct clade (Supplemental Fig. S5), indicating that hemerythrin occurred in the ancestor of the land plants. In the green algae (*Micromonas pusilla* and *Volvox carteri*) and other angiosperm species listed in Supplemental Figure S5, the hemerythrin proteins were fused to zinc finger proteins. However, the hemerythrin proteins were fused to GST proteins in *P. patens*. The phylogenetic

trees showed that the GST domains of the eight *P. patens* hemerythrin GST fusion proteins were grouped into one distinct clade with 100% support (Fig. 1; Supplemental Fig. S1), indicating that these eight proteins form a new GST class. Based on the sequence and structural comparisons, two G-site residues, Lys-42 and Glu-68 of TaGSTU4 (alignment positions 40 and 68 in Fig. 3A), were highly conserved in tau, phi, and iota GSTs. In tau and phi GSTs, these two G-site residues were important components that contributed to the enzyme's affinities for GSH (Zeng and Wang, 2005; Wei et al., 2012). However, these two G-site residues were replaced by Tyr/Cys/Gly/Asn and Gly/Ser/His in PpGSTHs, respectively, which suggests that PpGSTHs may have a low affinity for GSH. Hemerythrin class GSTs contained GST and hemerythrin domains. Hemerythrin domain proteins can bind heavy metal ions such as iron and cadmium (French et al., 2008), and GST domains can bind GSH. One possible function of hemerythrin class GSTs may be to detoxify heavy metals by catalyzing the conjugation of GSH with metal ions through the metal-thiolate bond. In plants, phytochelatin synthase (PCS) plays an important role in heavy metal detoxification (Vivares et al., 2005). However, we did not identify a PCS gene in the *P. patens* genome using TBLASTN searches with Arabidopsis, rice, or *S. moellendorffii* PCS protein sequences. The functional absence of PCS genes in *P. patens* might be compensated by the hemerythrin class GSTs. However, understanding the biological function of hemerythrin class GSTs in *P. patens* requires further study.

Ure2p GSTs were found mainly in bacteria and fungi genomes. The *P. patens* and *S. moellendorffii* genomes contained one and two Ure2p GSTs, respectively, while other seed plants did not have any (Table III). The alga *M. pusilla* has a Ure2p gene (Supplemental Data Set S6). One possibility is that the Ure2p gene in *P. patens* was transferred from bacteria or fungi by horizontal gene transfer events. To test this, we constructed a phylogenetic tree of plant, fungi, and bacteria Ure2p proteins (Supplemental Fig. S6). The phylogenetic tree showed that *P. patens* and *S. moellendorffii* Ure2p proteins were grouped together, indicating that the Ure2p genes occurred in the ancestor of the land plants. Plant and some bacterial Ure2p proteins were grouped into one clade (green circle in Supplemental Fig. S6), indicating that the Ure2p genes might have been transferred from the bacteria by horizontal gene transfer events. During the evolution of land plants, this gene might have been kept in *S. moellendorffii* but deleted from gymnosperms and angiosperms. *P. patens* Ure2p GST (PpUre2p1) showed high protein sequence identities (39.2%–65.0%) with the bacterial Ure2p proteins listed in Supplemental Figure S7A. Compared with bacterial Ure2p proteins, PpUre2p1 had lower protein sequence identities (26.1%–33.9%) with the fungal Ure2p proteins listed in Supplemental Figure S7A. Based on the x-ray structure of *E. coli* Ure2p protein (YfcG; PDB code 3GX0), we predicted the three-dimensional

structure of PpUre2p1 (Supplemental Fig. S7B). We found that PpUre2p1 and YfcG shared the same conformation of α-helices and β-sheets. The x-ray structure of the YfcG protein indicated that Thr-9, Gln-38, Ile-52, Glu-71, and Ser-72 (alignment positions 11, 41, 55, 74, and 75 in Supplemental Fig. S7A) are G-site residues (Wadington et al., 2009). Based on the crystal structure data of YfcG and multiple sequence alignment of the N-terminal protein sequences of Ure2p proteins, Thr-11, Gln-40, Ile-54, Glu-72, and Ser-73 of PpUre2p1 were predicted to be G-site residues (Supplemental Fig. S7). These five G-site residues were highly conserved in PpUre2p1 and bacterial Ure2p proteins (Supplemental Fig. S7A). Fungal and bacterial Ure2p proteins were known to be active as dimers (Wadington et al., 2009; Zhang and Perrett, 2009). Fungal Ure2p proteins did not show detectable GST activity toward the standard substrate CDNB, but they had GSH-dependent peroxidase activity (Zhang et al., 2008). However, bacterial Ure2p proteins had GSH-conjugating activity toward CDNB and GSH-dependent peroxidase activity toward Cum-OOH (Kanai et al., 2006). Similar to bacterial Ure2p proteins, *P. patens* PpUre2p1 showed activities toward CDNB and Cum-OOH. PpUre2p1 and bacterial Ure2p proteins had highly conserved G-site residues and shared similar protein structures and biochemical functions, implying that the *P. patens* Ure2p genes might have been transferred from bacteria. However, the biological function of PpUre2p1 in *P. patens* is in need of further study.

In poplar and Arabidopsis, the phi GST has a three-exon/two-intron structure (Dixon et al., 2002b; Lan et al., 2009). Four *P. patens* phi GST genes (*PpGSTF1* and *PpGSTF7* to *PpGSTF9*) had similar gene structures to poplar and Arabidopsis phi GSTs, but six *P. patens* phi GST genes (*PpGSTF2* to *PpGSTF6* and *PpGSTF10*) contained three introns. A joint analysis of gene structures and phylogenetic trees showed that the ancestor of *P. patens* phi GSTs might contain a three-exon/two-intron structure (Fig. 2, phi class). The first exon of *PpGSTF2* to *PpGSTF6* and *PpGSTF10* might have undergone an intron insertion event, resulting in the four-exon/three-intron gene structures of the six genes. Apart from this variation in gene structure, the 10 *P. patens* phi GSTs also showed extensive functional diversification. Six phi GSTs (*PpGSTF2* to *PpGSTF6* and *PpGSTF10*) were grouped into one clade (clade B in Supplemental Fig. S4A). *PpGSTF3*, *PpGSTF4*, and *PpGSTF10* genes were expressed under all growth conditions, while *PpGSTF2* and *PpGSTF6* genes were expressed only in response to H_2O_2 treatment. Although *PpGSTF3*, *PpGSTF4*, and *PpGSTF10* had similar expression patterns, the proteins encoded by these three genes showed different biochemical characteristics. For example, the PpGSTF3 protein had 9.3-fold higher activity for NBD-Cl than PpGSTF10, and PpGSTF10 showed a distinct optimum temperature spectrum versus PpGSTF3. This extensive functional diversification of *P. patens* phi GSTs might facilitate the

retention of duplicate genes, which can result in a large class with a distinct gene expression pattern, a broad substrate spectrum, and a wide range of reactivity toward different substrates.

DHAR can reduce DHA to ascorbic acid in the ascorbate-glutathione cycle. Most ascorbic acid has been reported to be localized in the cytoplasm and chloroplasts (Horemans et al., 2000; Pignocchi et al., 2003), indicating that the first DHAR protein (PpDHAR1 in *P. patens*) might localize to cytosol and chloroplast. This was supported by the result that the PpDHAR1 protein was localized to the cytosol and chloroplasts. Interestingly, a previous study found that the Arabidopsis genome contained four DHAR proteins: two were predicted to be cytosolic, while the other two were chloroplast proteins (Dixon et al., 2002a). As *PpDHAR1* and the most recent ancestor of Arabidopsis *DHAR*s are orthologous genes (Supplemental Fig. S4E), divergence in the subcellular localization of Arabidopsis DHAR proteins may indicate subfunctionalization (each duplicate copy partitions the ancestral gene function).

The *P. patens* genome has only one lambda GST, while at least three were found in the poplar, Arabidopsis, rice, and loblolly pine genomes. As all plant lambda GSTs form a monophyletic group (Fig. 7), *PpGSTL1* and the most recent ancestor of Arabidopsis *GSTL*s are orthologous genes. The PpGSTL1 protein localized to the chloroplast, while among the three Arabidopsis GSTLs, one was cytosolic and one was chloroplast targeted (Dixon et al., 2002a). The new cytosolic localization of Arabidopsis GSTL proteins indicated that the Arabidopsis lambda GST duplicate genes might have experienced neofunctionalization (one duplicate copy acquires a new function, whereas another duplicate copy retains the original function).

In conclusion, this study identified 37 GST genes from the moss *P. patens* genome. They were divided into 10 classes, including two new classes (hemerythrin and iota). Although tau GST is the largest class in vascular plants, the *P. patens* genome lacks any tau GST gene. Among the 10 *P. patens* GST classes, phi GSTs were the most numerous, with 10 members. *P. patens* phi GSTs showed extensive functional divergence in their gene expression responses to abiotic stressors and enzymatic characteristics of the encoded proteins. The vascular plants contained only one TCHQD gene, but *P. patens* had five copies that showed divergence in their gene structure and subcellular locations of the encoded proteins. *P. patens* DHARs were active as thiol transferases but had no GSH-conjugating activity. Other *P. patens* GST classes also showed extensive divergence in their gene structure and/or gene expression responses to abiotic stressors or subcellular locations of the encoded proteins. Phylogenetic analyses of GSTs from *P. patens* and other higher plants showed that different class GSTs had distinct duplication patterns during the evolution of land plants. Thus, by examining multiple characteristics, this study revealed the complex patterns of evolutionary divergence of the GST gene family in land plants.

MATERIALS AND METHODS

GST Gene Identification and Nomenclature

Physcomitrella patens GST genes were identified by TBLASTN searches of the *P. patens* genome database using 221 full-length GST proteins from rice (*Oryza sativa*; Jain et al., 2010), Arabidopsis (*Arabidopsis thaliana*; Dixon and Edwards, 2010a), poplar (*Populus trichocarpa*; Lan et al., 2009), and 152 of other plants, animals, fungi, and bacteria (Supplemental Data Sets S1, S4, and S7). These 152 full-length GSTs represent 36 classes defined by the NCBI conserved domain database (Marchler-Bauer et al., 2005). *P. patens* GST candidates were primarily analyzed using the NCBI conserved domain search to identify typical GST N- and C-terminal domains in their protein structures. When a *P. patens* GST candidate was identified by the NCBI conserved domain search to contain typical GST N- and C-terminal domains in its protein structure, it was considered as a GST gene. The predicted *P. patens* GST genes were amplified from *P. patens* complementary DNA (cDNA), cloned into the pEASY-T3 vector (TransGen), and sequenced in both directions to verify the gene sequence. The coding sequences of *P. patens* GSTs were mapped to the *P. patens* genome to verify the gene structure. For genes that PCR did not detect (eight out of 37 in this study), their structures were assumed to be identical to those of their closest relatives on the phylogenetic tree. This approach was adapted from other studies (Meyers et al., 2003).

The nomenclature for *P. patens* GSTs follows the system suggested by Dixon et al. (2002b) for plant GSTs; a univocal name was assigned to each *P. patens* GST gene consisting of two italic letters denoting the source organism, the family name (e.g. *PpGSTF*, *PpGSTT*, *PpGSTZ*, *PpGSTL*, *PpDHAR*, *PpEF1Bγ*, *PpUre2p*, *PpTCHQD*, *PpGSTH*, and *PpGSTI*, corresponding to phi, theta, zeta, lambda, DHAR, EF1Bγ, Ure2p, TCHQD, hemerythrin, and iota classes), and a progressive number for each gene (e.g. *PpGSTF1*).

Phylogenetic Analysis

The GST domain sequences were aligned using MUSCLE software (Edgar, 2004). The protein alignment was further adjusted manually using BioEdit (Hall, 1999). The optimal substitution model of amino acid substitution was selected using the modelGenerator version 0.84 program (Keane et al., 2006). The phylogenetic tree in Figure 1 was constructed using the maximum likelihood procedure with RAxML software (Stamatakis, 2006). All other trees were constructed using the maximum likelihood method with PhyML software (Guindon and Gascuel, 2003). One hundred bootstrap replicates were performed in each analysis to obtain the confidence support. Cytosolic GSTs are thought to be derived from the GRX2 protein (Oakley, 2005). Thus, GRX2 protein was used as an outgroup for phylogenetic analysis of the *P. patens* GST family.

Homology Modeling

The crystal structures of wheat (*Triticum aestivum*) TaGSTU4 (PDB code 1GWC) and human GST (PDB code 1EEM) were used as templates to construct the structure models of GST domains of *P. patens* hemerythrin and iota GSTs, respectively. Sequences were aligned using the Align 2D structure alignment program (homology module; InsightII; Accelrys). Structures were built using the modeler module of InsightII. All structures were verified using the profile-3D program of InsightII. The structure models were selected according to model evaluation score calculated by Profile-3D.

Expression of *P. patens* GST Genes under Normal Conditions and Abiotic Stress

To investigate the expression patterns of the *P. patens* GST genes under normal conditions and abiotic stress, *P. patens* was cultured in BCDATG solidified medium (Nishiyama et al., 2000) under an 18-h/6-h light/dark cycle for about 60 d. Then, the *P. patens* plants were separately sprayed with 0.5% H_2O_2, 150 mM NaCl, 1 mM salicylic acid, and 0.1% atrazine solutions. Each treatment consisted of three replicates. Then, 12 h after the H_2O_2 and atrazine treatments and 24 h after the NaCl and salicylic acid treatments, total RNA

was isolated from the whole plant of *P. patens* using Aurum Total RNA Kits (Bio-Rad Laboratories). Total RNA was treated with RNase-free DNase I (Promega) and reverse transcribed into cDNA using the RNA PCR Kit (AMV) version 3.0 (TaKaRa). Based on multiple sequence alignment of all *P. patens* GST sequences, 37 specific primer pairs were designed (Supplemental Table S2). The *P. patens* actin gene (GenBank accession no. XM_001783849) was used as an internal control. PCR was performed in a volume of 25 μL containing 3 μL of first-strand cDNA, 2.5 μL of TaKaRa 10$\times$PCR buffer, 0.125 μL of TaKaRa ExTaq (5 units μL^{-1}), 2 μL of deoxyribonucleotide triphosphate (2.5 mM each), and 10 pmol of each primer. PCR conditions were 94°C for 3 min, followed by cycles of 94°C for 30 s, 60°C annealing for 40 s, and 72°C for 1 min, with a final extension at 72°C for 10 min. To be in the linear range, PCR was performed at 24, 26, 28, and 30 cycles. The PCR products were analyzed by 1% agarose gel electrophoresis. PCR products from each sample were validated by DNA sequencing.

Subcellular Localization

To investigate the subcellular divergence of *P. patens* GSTs, 21 *P. patens* GST proteins representing 10 classes were tested for subcellular localization using C-terminal GFP tagging. The *P. patens* GST genes were subcloned into modified pCAMBIA1302 vectors (Supplemental Fig. S8). The primers used to construct the GST subcellular localization vectors are listed in Supplemental Table S3. Colonies containing the appropriate insert were identified by sequencing. The modified pCAMBIA1302 vectors containing GST sequences were transformed into *Agrobacterium tumefaciens* LBA4404. Cultures were infiltrated into epidermal cells of tobacco (*Nicotiana tabacum*) leaves as described by Sparkes et al. (2006). Transformed tissues were harvested 3 to 6 d later and immediately observed with a confocal laser microscope (Olympus FV1000 MPE). GFP fusion fluorescence was excited with a 488-nm laser, while autofluorescence of chlorophyll was imaged using a 543-nm laser.

Expression and Purification of Recombinant *P. patens* GST Proteins

To investigate the enzymatic functions of *P. patens* GST proteins, the *P. patens* GST genes were subcloned into pET-30a expression vectors (Novagen) to obtain an N-terminal 6$\times$His tag. The primers used to construct the GST expression vectors are listed in Supplemental Table S4. Colonies containing appropriate inserts were identified by sequencing.

Escherichia coli BL21 (DE3) cells harboring pET-30a/GST plasmids were cultured overnight, diluted 1:100, and grown until the optical density (A_{600}) reached 0.5. A final concentration of 0.1 mM isopropyl-β-D-thiogalactopyranoside was added to induce synthesis of the recombinant GST proteins. Then, 10 h after induction, cells were harvested by centrifugation (8,000g, 3 min, 4°C), resuspended in binding buffer (20 mM sodium phosphate, 0.5 M NaCl, and 20 mM imidazole, pH 7.4), and disrupted by cold sonication. In each case, the homogenate was then subjected to centrifugation (10,000g, 10 min, 4°C). The resulting particulate material and a small portion of the supernatant were analyzed by SDS-PAGE. The rest of the supernatant was loaded onto a Nickel-Sepharose High Performance column (GE Healthcare Bio-Sciences) preequilibrated with binding buffer. The GST proteins that bound to the Nickel-Sepharose High Performance column were eluted with elution buffer (20 mM sodium phosphate, 0.5 M NaCl, and 500 mM imidazole, pH 7.4).

Enzyme Assays and Kinetic Studies

GST activity toward CDNB, DCNB, and NBC was measured as described by Habig et al. (1974); that toward NBD-Cl was measured as described by Ricci et al. (1994); and that toward HED, DHA, fluorodifen, and Cum-OOH was measured as described by Edwards and Dixon (2005). All assays were carried out at 25°C. SDS-PAGE was performed on 10% separating gels and 5% stacking gels. The dependence of GST activity on pH was measured as reported by Yuen and Ho (2001). The dependence of GST activity on temperature was measured using CDNB as the substrate at various temperatures from 15°C to 70°C at 5°C intervals. Protein concentrations were determined by measuring A_{280} (Layne, 1957).

The apparent kinetic constants of *P. patens* GSTs were studied in assays with various concentrations of GSH and CDNB or DHA. The apparent K_m value for GSH was determined using a GSH range from 0.1 to 1.0 mM at a

fixed CDNB concentration of 1.0 mM for phi GSTs and at a fixed DHA concentration of 0.2 mM for DHARs. The apparent K_m value for CDNB was determined using a CDNB range from 0.2 to 2.0 mM at a fixed GSH concentration of 1.0 mM. The apparent K_m value for DHA was determined using a DHA range from 0.02 to 0.2 mM at a fixed GSH concentration of 0.5 mM. The kinetic parameters were derived using nonlinear regression as implemented in the Hyper32 program, available at http://homepage.ntlworld.com/john.easterby/hyper32.html.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Phylogenetic relationships among iota and 37 *P. patens* GSTs.

Supplemental Figure S2. Pairwise protein sequence identities of the GST domains of plant GSTs.

Supplemental Figure S3. Structural comparison of the GST domains of eight *P. patens* hemerythrin class GSTs.

Supplemental Figure S4. Phylogenetic relationships of phi (A), lambda (B), EF1Bγ (C), TCHQD (D), DHAR (E), zeta (F), and theta (G) GSTs in land plants.

Supplemental Figure S5. Phylogenetic relationships among the hemerythrin domains from bacteria, fungi, and plants.

Supplemental Figure S6. Phylogenetic relationships among the full-length Ure2p proteins from bacteria, fungi, algae, and plants.

Supplemental Figure S7. Sequence and structural analyses of Ure2p proteins.

Supplemental Figure S8. Modified pCAMBIA1302 vector.

Supplemental Table S1. Full-length GST genes identified in the *P. patens* genome.

Supplemental Table S2. PCR primers used to detect the expression of *P. patens* GST genes.

Supplemental Table S3. Primers used to construct the *P. patens* GST subcellular localization vectors.

Supplemental Table S4. Primers used to construct the *P. patens* GST protein expression vector.

Supplemental Data Set S1. Full-length GST protein sequences used to construct the phylogenetic tree in Figure 1.

Supplemental Data Set S2. Protein sequences used to construct the phylogenetic tree in Supplemental Figure S1.

Supplemental Data Set S3. Protein sequence alignment of the GST domains of hemerythrin, TCHQD, iota, lambda, EF1Bγ, DHAR, tau, theta, zeta, and phi GSTs.

Supplemental Data Set S4. Protein sequences of *P. patens*, *S. moellendorffii*, loblolly pine, poplar, Arabidopsis, and rice GSTs used to construct the phylogenetic trees in Supplemental Figure S4.

Supplemental Data Set S5. Protein sequences of plant, fungi, and bacteria hemerythrin domains used to construct the phylogenetic tree in Supplemental Figure S5.

Supplemental Data Set S6. Ure2p protein sequences used to construct the phylogenetic tree in Supplemental Figure S6.

Supplemental Data Set S7. A FASTA file containing all protein sequences used in this study.

LITERATURE CITED

Armstrong RN (1997) Structure, catalytic mechanism, and evolution of the glutathione transferases. Chem Res Toxicol 10: 2–18

Bailly X, Vanin S, Chabasse C, Mizuguchi K, Vinogradov SN (2008) A phylogenomic profile of hemerythrins, the nonheme diiron binding respiratory proteins. BMC Evol Biol 8: 244

Basantani M, Srivastava A (2007) Plant glutathione transferases: a decade falls short. Can J Bot 85: 443–456

Benekos K, Kissoudis C, Nianiou-Obeidat I, Labrou N, Madesis P, Kalamaki M, Makris A, Tsaftaris A (2010) Overexpression of a specific soybean GmGSTU4 isoenzyme improves diphenyl ether and chloroacetanilide herbicide tolerance of transgenic tobacco plants. J Biotechnol 150: 195–201

Boruc J, Mylle E, Duda M, De Clercq R, Rombauts S, Geelen D, Hilson P, Inzé D, Van Damme D, Russinova E (2010) Systematic localization of the Arabidopsis core cell cycle proteins reveals novel cell division complexes. Plant Physiol 152: 553–565

Cove D (2005) The moss Physcomitrella patens. Annu Rev Genet 39: 339–358

Dixon DP, Davis BG, Edwards R (2002a) Functional divergence in the glutathione transferase superfamily in plants: identification of two classes with putative functions in redox homeostasis in Arabidopsis thaliana. J Biol Chem 277: 30859–30869

Dixon DP, Edwards R (2010a) Glutathione transferases. The Arabidopsis Book 8: e0131 10.1199/tab.0131

Dixon DP, Edwards R (2010b) Roles for stress-inducible lambda glutathione transferases in flavonoid metabolism in plants as identified by ligand fishing. J Biol Chem 285: 36322–36329

Dixon DP, Hawkins T, Hussey PJ, Edwards R (2009) Enzyme activities and subcellular localization of members of the Arabidopsis glutathione transferase superfamily. J Exp Bot 60: 1207–1218

Dixon DP, Lapthorn A, Edwards R (2002b) Plant glutathione transferases. Genome Biol 3: reviews 3004.3001–reviews 3004.3010

Dixon DP, Sellars JD, Edwards R (2011) The Arabidopsis phi class glutathione transferase AtGSTF2: binding and regulation by biologically active heterocyclic ligands. Biochem J 438: 63–70

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797

Edwards R, Dixon DP (2005) Plant glutathione transferases. Methods Enzymol 401: 169–186

French CE, Bell JML, Ward FB (2008) Diversity and distribution of hemerythrin-like proteins in prokaryotes. FEMS Microbiol Lett 279: 131–145

Frova C (2003) The plant glutathione transferase gene family: genomic structure, functions, expression and evolution. Physiol Plant 119: 469–479

Frova C (2006) Glutathione transferases in the genomics era: new insights and perspectives. Biomol Eng 23: 149–169

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704

Habig WH, Pabst MJ, Jakoby WB (1974) Glutathione S-transferases: the first enzymatic step in mercapturic acid formation. J Biol Chem 249: 7130–7139

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95–98

Horemans N, Foyer CH, Potters G, Asard H (2000) Ascorbate function and associated transport systems in plants. Plant Physiol Biochem 38: 531–540

Jain M, Ghanashyam C, Bhattacharjee A (2010) Comprehensive expression analysis suggests overlapping and specific roles of rice glutathione S-transferase genes during development and stress responses. BMC Genomics 11: 73

Jha B, Sharma A, Mishra A (2011) Expression of SbGSTU (tau class glutathione S-transferase) gene isolated from Salicornia brachiata in tobacco for salt tolerance. Mol Biol Rep 38: 4823–4832

Kanai T, Takahashi K, Inoue H (2006) Three distinct-type glutathione S-transferases from Escherichia coli important for defense against oxidative stress. J Biochem 140: 703–711

Karavangeli M, Labrou NE, Clonis YD, Tsaftaris A (2005) Development of transgenic tobacco plants overexpressing maize glutathione S-transferase I for chloroacetanilide herbicides phytoremediation. Biomol Eng 22: 121–128

Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol 6: 29

Kern R, Malki A, Holmgren A, Richarme G (2003) Chaperone properties of Escherichia coli thioredoxin and thioredoxin reductase. Biochem J 371: 965–972

Lan T, Yang ZL, Yang X, Liu YJ, Wang XR, Zeng QY (2009) Extensive functional diversification of the Populus glutathione S-transferase supergene family. Plant Cell 21: 3749–3766

Layne E (1957) Spectrophotometric and turbidimetric methods for measuring proteins. Methods Enzymol 3: 447–455

Loyall L, Uchida K, Braun S, Furuya M, Frohnmeyer H (2000) Glutathione and a UV light-induced glutathione S-transferase are involved in signaling to chalcone synthase in cell cultures. Plant Cell 12: 1939–1950

Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, et al (2005) CDD: a Conserved Domain Database for protein classification. Nucleic Acids Res 33: D192–D196

Marques AC, Vinckenbosch N, Brawand D, Kaessmann H (2008) Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. Genome Biol 9: R54

McGoldrick S, O'Sullivan SM, Sheehan D (2005) Glutathione transferase-like proteins encoded in genomes of yeasts and fungi: insights into evolution of a multifunctional protein superfamily. FEMS Microbiol Lett 242: 1–12

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell 15: 809–834

Moons A (2005) Regulatory and functional interactions of plant growth regulators and plant glutathione S-transferases (GSTs). Vitam Horm 72: 155–202

Nishiyama T, Hiwatashi Y, Sakakibara I, Kato M, Hasebe M (2000) Tagged mutagenesis and gene-trap in the moss, Physcomitrella patens by shuttle mutagenesis. DNA Res 7: 9–17

Oakley AJ (2005) Glutathione transferases: new functions. Curr Opin Struct Biol 15: 716–723

Pignocchi C, Fletcher JM, Wilkinson JE, Barnes JD, Foyer CH (2003) The function of ascorbate oxidase in tobacco. Plant Physiol 132: 1631–1641

Qian W, Zhang J (2009) Protein subcellular relocalization in the evolution of yeast singleton and duplicate genes. Genome Biol Evol 1: 198–204

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, et al (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319: 64–69

Ricci G, Caccuri AM, Lo Bello M, Pastore A, Piemonte F, Federici G (1994) Colorimetric and fluorometric assays of glutathione transferase based on 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole. Anal Biochem 218: 463–465

Sheehan D, Meade G, Foley VM, Dowd CA (2001) Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. Biochem J 360: 1–16

Soltis DE, Bell CD, Kim S, Soltis PS (2008) Origin and early evolution of angiosperms. Ann N Y Acad Sci 1133: 3–25

Sparkes IA, Runions J, Kearns A, Hawes C (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc 1: 2019–2025

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690

Thom R, Cummins I, Dixon DP, Edwards R, Cole DJ, Lapthorn AJ (2002) Structure of a tau class glutathione S-transferase from wheat active in herbicide detoxification. Biochemistry 41: 7008–7020

Vivares D, Arnoux P, Pignol D (2005) A papain-like enzyme at work: native and acyl-enzyme intermediate structures in phytochelatin synthesis. Proc Natl Acad Sci USA 102: 18848–18853

Wadington MC, Ladner JE, Stourman NV, Harp JM, Armstrong RN (2009) Analysis of the structure and function of YfcG from Escherichia coli reveals an efficient and unique disulfide bond reductase. Biochemistry 48: 6559–6561

Wei T, Wang CL, Kao HN, Yang HL, Wang XR, Zeng QY (2012) Molecular and catalytic characterization of a phi class glutathione transferase from Cathaya argyrophylla. Biochem Syst Ecol 40: 75–85

Yuen WK, Ho JW (2001) Purification and characterization of multiple glutathione S-transferase isozymes from Chironomidae larvae. Comp Biochem Physiol A Mol Integr Physiol 129: 631–640

Zeng QY, Wang XR (2005) Catalytic properties of glutathione-binding residues in a tau class glutathione transferase (PtGSTU1) from Pinus tabulaeformis. FEBS Lett 579: 2657–2662

Zhang ZR, Bai M, Wang XY, Zhou JM, Perrett S (2008) "Restoration" of glutathione transferase activity by single-site mutation of the yeast prion protein Ure2. J Mol Biol 384: 641–651

Zhang ZR, Perrett S (2009) Novel glutaredoxin activity of the yeast prion protein Ure2 reveals a native-like dimer within fibrils. J Biol Chem 284: 14058–14067