New Phytologist

Research

# Functional and structural profiles of GST gene family from three *Populus* species reveal the sequence–function decoupling of orthologous genes

**Qi Yang[1,2,3], Xue-Min Han[2], Jin-Ke Gu[4], Yan-Jing Liu[1,2], Mao-Jun Yang[4] and Qing-Yin Zeng[1,2,3]**

[1]State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing 100091, China; [2]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China; [3]University of Chinese Academy of Sciences, Beijing 100049, China; [4]State Key Laboratory of Biomembrane and Membrane Biotechnology, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China

Author for correspondence:
*Qing-Yin Zeng*
*Tel: +86 10 62824033*
*Email: qingyin.zeng@ibcas.ac.cn*

Received: *26 March 2018*
Accepted: *8 August 2018*

*New Phytologist* (2018)
**doi**: 10.1111/nph.15430

**Key words:** functional divergence, glutathione *S*-transferase (GST), orthologous gene, *Populus*, protein function.

## Summary

- A common assumption in comparative genomics is that orthologous genes are functionally more similar than paralogous genes. However, the validity of this assumption needs to be assessed using robust experimental data.
- We conducted tissue-specific gene expression and protein function analyses of orthologous groups within the glutathione *S*-transferase (GST) gene family in three closely related *Populus* species: *Populus trichocarpa*, *Populus euphratica* and *Populus yatungensis*.
- This study identified 21 GST orthologous groups in the three *Populus* species. Although the sequences of the GST orthologous groups were highly conserved, the divergence in enzymatic functions was prevalent. Through site-directed mutagenesis of orthologous proteins, this study revealed that nonsynonymous substitutions at key amino acid sites played an important role in the divergence of enzymatic functions. In particular, a single amino acid mutation (Arg39→Trp39) contributed to *P. euphratica* PeGSTU30 possessing high enzymatic activity via increasing the hydrophobicity of the active cavity.
- This study provided experimental evidence showing that orthologues belonging to the gene family have functional divergences. The nonsynonymous substitutions at a few amino acid sites resulted in functional divergence of the orthologous genes. Our findings provide new insights into the evolution of orthologous genes in closely related species.

## Introduction

Homologous genes are divided into orthologues and paralogues based on whether they arose by speciation or duplication respectively (Fitch, 1970). Orthologues, or orthologous genes, are genes in different species that are derived from a single gene in the last common ancestor of the species. Paralogues, or paralogous genes, are created by a duplication event within a species' genome. It is widely assumed that orthologous genes perform similar functions in different species, whereas paralogous genes tend to diverge in function (Tatusov *et al.*, 1997; Koonin, 2005). This belief has been termed as the orthologue conjecture hypothesis. In accordance with this hypothesis, biologists assign functional information to uncharacterized orthologous genes in newly sequenced genomes based on experimentally characterized genes in model organisms (Dolinski & Botstein, 2007). This hypothesis is the conceptual basis for the functional annotation of newly sequenced genomes, which is fundamental to experimental biology in the post-genomic era (Gabaldon & Koonin, 2013).

The orthologue conjecture hypothesis has been challenged by Nehrt *et al.* (2011), who compared the function and expression similarities of orthologues and paralogues in humans and mouse using experiment-based annotations in the Gene Ontology (GO) database and microarray gene expression data. They found that orthologues and between-species outparalogues (paralogous genes resulting from a duplication preceding a given speciation event) are much less functionally similar than within-species outparalogues, which contradicts the orthologue conjecture hypothesis (Nehrt *et al.*, 2011). Subsequently, several studies have proposed that GO annotations should be used very carefully or even are unsuitable for testing the orthologue conjecture hypothesis (Altenhoff *et al.*, 2012; Chen & Zhang, 2012; Thomas *et al.*, 2012). When controlling for biases in GO annotations or using RNA sequencing data, these studies found supports for the ortholog conjecture hypothesis (Altenhoff *et al.*, 2012; Chen & Zhang, 2012). For example, using a large RNA sequencing dataset of multiple tissues from eight mammals and chicken, Chen & Zhang (2012) found that the level of similarity in expression between orthologues is significantly higher than within-species paralogues. However, further assessments of the reliability and validity of the orthologue conjecture hypothesis are still

needed, particularly using more and better systematic biochemical data.

Single-copy genes, which lack paralogues within a genome, have clear one-to-one orthologous relationships between species. Functional change in a single-copy gene can be harmful, since such genes lack paralogues that may provide functional compensation (Dolinski & Botstein, 2007). Thus, single-copy orthologous genes are usually considered to perform similar functions in different organisms (Fulton *et al.*, 2002; Kojima *et al.*, 2002; Wu *et al.*, 2006). However, most structural and regulatory genes are not single-copy genes; rather, they are members of gene families. Particularly among plants, genes involved in secondary metabolism or in the response to biotic/abiotic stimuli tend to form large gene families, such as the resistance gene and glutathione *S*-transferase (GST) gene families (Meyers *et al.*, 2003; Lan *et al.*, 2009). After a speciation event, duplicate genes belonging to the same gene family persist within each lineage. A duplicate gene in one descendant species will therefore have a corresponding orthologue in another descendant species. When an ortholog within a large gene family in one species had functional changes, unlike the single-copy orthologs, organism had paralogous genes to compensate the loss of the original function. Thus, the orthologs in different descendent species would be functional diverged, which was not consistent with the ortholog conjecture hypothesis. However, to date, the lack of appropriate functional assays has hindered attempts to elucidate the molecular mechanisms for functional conservation or divergence of orthologous genes belonging to large gene family.

GSTs (EC 2.5.1.8) are ubiquitous genes encoding proteins that catalyze the conjugation of tripeptide glutathione (γ-glutamyl-cysteinylglycine; GSH) to various hydrophobic substrates (e.g. 1-chloro-2,4-dinitrobenzene). Plant GSTs form a large gene family of over 54 members in the *Arabidopsis thaliana*, *Populus trichocarpa*, and *Oryza sativa* genomes (Lan *et al.*, 2009; Dixon & Edwards, 2010; Jain *et al.*, 2010). Plant GSTs are divided into eight classes, in which the tau and phi classes are the most abundant and specific to plants (Liu *et al.*, 2013). In plants, tau and phi GSTs can protect cells from a wide range of biotic and abiotic stresses, including pathogen attacks, xenobiotics, heavy metals, toxins, oxidative stress, and UV radiation (Cummins *et al.*, 1999, 2013; Loyall *et al.*, 2000; Agrawal *et al.*, 2002). Our previous studies on this family found that tau and phi GSTs display extensive functional diversification in gene expression and enzymatic functions in *P. trichocarpa*, *Glycine max*, and *Larix kaempferi* (Lan *et al.*, 2009; Yang *et al.*, 2014; Liu *et al.*, 2015). These characteristics make the GST family an ideal model for elucidating the molecular mechanisms underlying the functional conservation or divergence of orthologous genes belonging to large gene families.

In this study, we focus on the GST gene family in three closely related *Populus* species: *P. trichocarpa*, *Populus yatungensis*, and *Populus euphratica*. *Populus* is a woody perennial plant that represents the most important currently available tree model system for plant genomics (Tuskan *et al.*, 2006). The genus *Populus* is divided into five sections: *Tacamahaca*, *Turanga*, *Populus*, *Leucoides*, and *Aigeiros* (Fang *et al.*, 1999). *P. trichocarpa* and *P. yatungensis* belong to sect. *Tacamahaca*, whereas *P. euphratica* belongs to sect. *Turanga* (Fang *et al.*, 1999). The phylogenetic tree of these three *Populus* species showed that *P. trichocarpa* was closer to *P. yatungensis* than to *P. euphratica* (Fig. 1). The *P. trichocarpa* and *P. euphratica* lineages diverged *c.* 8–11 Ma (Ma *et al.*, 2013), and the *P. trichocarpa* and *P. yatungensis* should have diverged more recently. *P. trichocarpa* (black cottonwood) is an ecologically and economically important forest tree species that is native to western North America and grows at altitudes of up to 2600 m (Argus *et al.*, 2010). *P. trichocarpa* is adapted to relatively humid, moist, and mild conditions in the west of the Rocky Mountains (Suarez-Gonzalez *et al.*, 2018). *P. yatungensis* is distributed only in the Tibetan Plateau; it grows on mountain slopes at an altitude of 2400–3600 m (Fang *et al.*, 1999) and exhibits unique ecological adaptation to the high plateau habitat. *P. euphratica*, also known as the desert poplar, grows in shelter belts along riversides in the natural desert forest ecosystems of China and Middle Eastern countries (Fang *et al.*, 1999). *P. euphratica* is highly adapted to salt stress, extreme temperatures, and drought (Ferreira *et al.*, 2006). Because the three *Populus* species grow in different habitats and diverged relatively recently, they provide an excellent system for studying the molecular mechanisms of functional conservation or divergence of orthologous genes. This study identified 21 orthologous GST groups from the three *Populus* species. By examining the gene expression responses to abiotic stresses, as well as the enzymatic and structural properties of GST proteins, this study showed that although the sequences of orthologous GST groups from the three *Populus* species were highly conserved, functional divergences were prevalent. Our findings provide new insights into the evolution of orthologous genes in closely related species.

## Materials and Methods

### Plant material and treatments

Cuttings of *P. trichocarpa* and *P. yatungensis* were cultivated in potting soil for *c.* 1 yr. Owing to the low survival rate of *P. euphratica* cuttings, this study used 2-yr-old *P. euphratica* seedlings germinated from seeds. The saplings of the three *Populus* species were cultivated in potting soil under the same growth conditions and were about breast height when sampled. The detailed methods of chemical treatments and total RNA extractions are described in Supporting Information Methods S1.

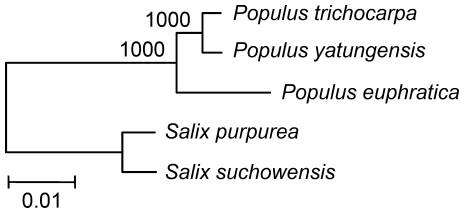



**Fig. 1** Phylogenetic relationships among *Populus trichocarpa*, *Populus euphratica*, and *Populus yatungensis*. The maximum likelihood tree was constructed with nucleotide sequences of 13 single-copy nuclear genes from each of the five species.

### Identification and nomenclature of *Populus* GST genes

Total DNA and complementary DNA (cDNA) from different tissues with different treatments of *P. yatungensis* and *P. euphratica* were used as templates for amplifying GST genes. All the primers were designed based on the GST sequences of *P. trichocarpa* (Table S1). Up to six primer pairs for each gene were used to perform PCR. This study only lists the working primer pairs. All PCR products were cloned into pEASY-T3 Vector (Transgen, Beijing, China). And for each PCR product, we selected 5–15 clones for sequencing. The cloned genes were analyzed by searching the National Center for Biotechnology Information conserved domain database to affirm the presence of typical GST domains.

The nomenclature of all GSTs followed the system suggested by Edwards *et al.* (2000) for plant GSTs. The name of each cloned gene began with two letters denoting the source organism: Pt was used for *P. trichocarpa*, Py for *P. yatungensis*, and Pe for *P. euphratica*. The number in each gene's name corresponded to the most similar GST of *P. trichocarpa*; this similarity was evaluated using sequence identities and phylogenetic relationships.

### Phylogenetic relationship among three *Populus* species

To determine the species relationship among *P. trichocarpa*, *P. yatungensis*, and *P. euphratica*, we used *Salix purpurea* and *Salix suchowensis* as outgroup species. Thirteen nuclear genes were used to construct a phylogenetic tree (Table S2). These nuclear genes are single-copy genes shared by *Arabidopsis*, *Populus*, *Vitis*, and *Oryza* (Duarte *et al.*, 2010), and previously proved to be good markers for constructing a phylogenetic tree of *Populus* species (Wang *et al.*, 2014). These 13 genes from *P. trichocarpa* were used as templates to retrieve orthologues from the *P. euphratica*, *S. purpurea*, and *S. suchowensis* genome databases using BLASTN searches. The sequences of *P. yatungensis* orthologues were acquired using BLASTN searches in the *P. yatungensis* whole-genome resequencing data (Q.Y. Zeng, Q. Yang & Y.J. Liu, unpublished).

Nucleotide sequence alignments were generated using the MUSCLE program (https://www.ebi.ac.uk/Tools/msa/muscle/). The phylogenetic analyses were performed for the combined dataset of these single-copy nuclear genes. The maximum likelihood tree was constructed using the PHYML program (Guindon & Gascuel, 2003) with 1000 bootstrap replicates. The optimal substitution model of nucleotide substitution was selected using the MODELGENERATOR v.0.84 program (Keane *et al.*, 2006).

### Orthologous groups identification of *Populus* GST genes

The GSTs cloned from *P. yatungensis* and *P. euphratica* were aligned with the tau and phi GSTs of *P. trichocarpa* using the MUSCLE program. The alignments were adjusted manually using BIOEDIT (Tom Hall, Ibis Therapeutics, Carlsbad, CA, USA). Subsequently, we constructed phylogenetic trees for *Populus* GSTs using a maximum likelihood procedure in PHYML. The phylogenetic trees of protein and cDNA sequences were respectively constructed using JTT and GTR models, which were selected using the MODELGENERATOR program. One hundred and 1000 bootstrap replicates were performed for phylogenetic trees of protein and cDNA sequences respectively. The Lambda GSTs of *P. trichocarpa* were used as outgroups.

### Tissue-specific gene expressions of *Populus* GSTs under normal conditions and abiotic stress

To investigate the tissue-specific gene expression patterns of *P. trichocarpa*, *P. yatungensis*, and *P. euphratica* GST genes under normal conditions and abiotic stresses, total RNA was isolated from root, shoot, leaf, phloem, and bud tissues under normal growth conditions and abiotic stress. Based on multiple sequence alignments of all 101 GST genes of *P. yatungensis* and *P. euphratica*, 42 specific primer pairs were designed (Table S3). The primers used for detecting GSTs expression from *P. trichocarpa* were from previous work (Lan *et al.*, 2009). The *Populus* actin gene (GenBank no. XM_002316253) was used as an internal control. The PCR products from each sample were analyzed by 1% agarose gel electrophoresis, and then validated by DNA sequencing. More details for investigating the tissue-specific gene expression of *Populus* GSTs are described in Methods S2.

### Molecular cloning, expression, and purification of GST proteins

Twenty-five GSTs from *P. yatungensis* and *P. euphratica* were selected for protein expression and purification. Each coding sequence was sub-cloned into a pET-30a expression vector (Novagen, Madison, WI, USA). The primers used to construct the expression vectors are listed in Table S4. To investigate the effect of nonsynonymous substitutions on enzymatic activity, we performed site-directed mutagenesis using methods described by Zeng & Wang (2005). The mutagenesis primers are listed in Table S5. GST proteins were expressed and purified as described by Liu *et al.* (2015). More details are given in Methods S3.

### Enzyme activity and kinetic studies

The substrate specificities of the purified GSTs were investigated using four substrates: 1-chloro-2,4-dinitrobenzene (CDNB; Sigma-Aldrich), 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl; Sigma-Aldrich), 1,2-dichloro-4-nitrobenzene (DCNB; Sigma-Aldrich), and 4-nitrobenzyl chloride (NBC; Sigma-Aldrich). The activities of recombinant proteins toward CDNB, DCNB, and NBC were measured according to description by Habig *et al.* (1974), and activities toward NBD-Cl were measured as described by Ricci *et al.* (1994). All assays were determined at 25°C using an Evolution™ 300 UV–visible spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA) with three replicates.

The steady-state kinetic parameters of recombinant GST proteins were determined by the methods described by Lan *et al.* (2009). The GSH concentrations ranging from 0.1 to 0.9 mM with a fixed CDNB concentration of 1.0 mM were used to determine the $K_m$ value for GSH. To determine the apparent $K_m$ values for CDNB, concentrations ranging from 0.2 to 2.0 mM with

a fixed GSH concentration of 1.0 mM were used. The kinetic parameters were derived using nonlinear regression as implemented in the HYPER32 program (http://hyper32.software.informer.com/).

### Viscosity experiments

The kinetic parameters were assayed at 25°C with various concentrations of glycerol to analyze the influence of viscosity. The reaction buffer contained 0.05 M potassium phosphate (pH 6.5) with variable glycerol concentrations (from 5% v/v to 30% v/v). Values of viscosity $\eta$ at 25°C were calculated as described by Wolf *et al.* (1985).

### Crystallization, data collection and processing

The genes encoding PtGSTU30 and R39W mutant proteins were sub-cloned into a modified ΔpET-30a expression vector (Yang *et al.*, 2009), and transformed into *Escherichia coli* Tuner (DE3). After the purification on the Nickel-Sepharose column (GE Healthcare, Chicago, IL, USA), the protein was concentrated to *c.* 10 mg ml$^{-1}$ and desalted in 10 mM Tris, pH 7.0. Subsequently, the protein was loaded onto a Superdex 75 10/300 GL column (GE Healthcare, Chicago, IL, USA). Then the protein was used for crystal growth. Initial crystallization screens were performed with Crystal Screen kits (Hampton Research, Laguna Beach, CA, USA). Crystals of the recombinant PtGSTU30 and R39W mutant were grown by the hanging-drop vapor diffusion method at 18°C. The detailed methods of crystal growth are described in Methods S4.

X-ray data were collected from small loop-mounted crystals (0.05–0.1 mm). Diffraction datasets were collected at beamline BL17U of the Shanghai Synchrotron Radiation Facility. The X-ray reflection data were integrated and scaled using the HKL2000 package (Otwinowski & Minor, 1997). The structure of GmGSTU4-4 was used as a model to solve the phase problem in CCP4 suits (Collaborative Computational Project, 1994). Then a model was built by using the COOT program and manually adjusted (Emsley & Cowtan, 2004). The final structure was refined with PHENIX (Adams *et al.*, 2002). All structure comparison and analysis were performed by PYMOL software (DeLano Scientific LLC, San Carlos, CA, USA). The structural figures were produced using PYMOL and Accelrys Discovery Studio (Accelrys Software Inc., San Diego, CA, USA). The detailed methods of data collection and processing are described in Methods S4.

### 1-Anilino-8-naphthalene sulfonic acid binding assays

1-Anilino-8-naphthalene sulfonic acid (ANS) is a fluorescent dye. The fluorescence intensity of ANS increases when it binds to hydrophobic regions of proteins, but it is strongly quenched in water (Schonbrunn *et al.*, 2000). To detect the differences in surface hydrophobicity between PtGSTU30 and R39W mutant, we performed the fluorescence measurement using an Hitachi F-7000 fluorescence spectrophotometer (Hitachi High-Technologies, Tokyo, Japan). The protein sample contained 0.2 mM ANS and 0.05 mg ml$^{-1}$ enzyme in 10 mM Tris buffer (pH 7.0). The control contained 0.2 mM ANS in 10 mM Tris buffer (pH 7.0). Each sample and control was scanned for three replications.

### Statistical analyses

Statistical analyses of the data were performed using SPSS 16.0 software (SPSS Inc., Chicago, IL, USA). The difference was considered significant at $P < 0.05$.

## Results

### Identification of GST orthologous genes from three *Populus* species

Our previous study identified 58 tau and nine phi GSTs in the *P. trichocarpa* genome (Lan *et al.*, 2009). In this study, we cloned 50 tau and nine phi GSTs from *P. yatungensis*, and 33 tau and nine phi GSTs from *P. euphratica* (Table S6). To identify GST orthologues, we used the 141 tau and 27 phi GST cDNA sequences from *P. trichocarpa*, *P. yatungensis*, and *P. euphratica* to construct a phylogenetic tree (Fig. 2). Based on the phylogenetic tree of cDNA sequences, we identified 21 groups (named OG1, OG2, OG3, etc.), and the three genes from three *Populus* species in each group were grouped into a single clade (Fig. 2). Among these 21 groups, 20 had >80% bootstrap support, and one (OG20) had 68% bootstrap support. We also constructed a phylogenetic tree with the protein sequences of these GSTs (Fig. S1). Based on the phylogenetic tree of the protein sequences, the three genes in each of the 21 groups were still grouped into a single clade (Fig. S1). Among the 21 groups, the gene tree topology in 15 groups (OG1–OG15) was consistent with the species tree (Fig. 2). In this study, these 15 groups (OG1–OG15) were considered strictly as orthologous groups. In the other six groups (OG16–OG21), the gene tree topology in each group was not consistent with the species tree (Fig. 2). Because we could not completely exclude the possibility that the genes within these six groups were not one-to-one GST orthologues, these six groups were included in the gene expression and enzymatic activity analyses. Protein sequence identities within each orthologous group were much higher than those of paralogous GSTs (independent sample *t*-test, $P < 0.0001$, Fig. S2), which further supported the assignment of the 21 groups as orthologous.

### Tissue-specific gene expression divergence among *Populus* GST orthologues

We characterized the tissue-specific gene expression patterns of GST genes belonging to the 21 orthologous GST groups from *P. trichocarpa*, *P. yatungensis*, and *P. euphratica* under normal growth and abiotic stress conditions (Fig. 3b). Among the 21 orthologous groups, the GST genes in 10 groups (OG1, OG2, OG5, OG11, OG13, OG14 and OG18–OG21) were expressed in all tissues in all treatments examined. However, two tissue-specific gene expression divergence patterns were observed in the

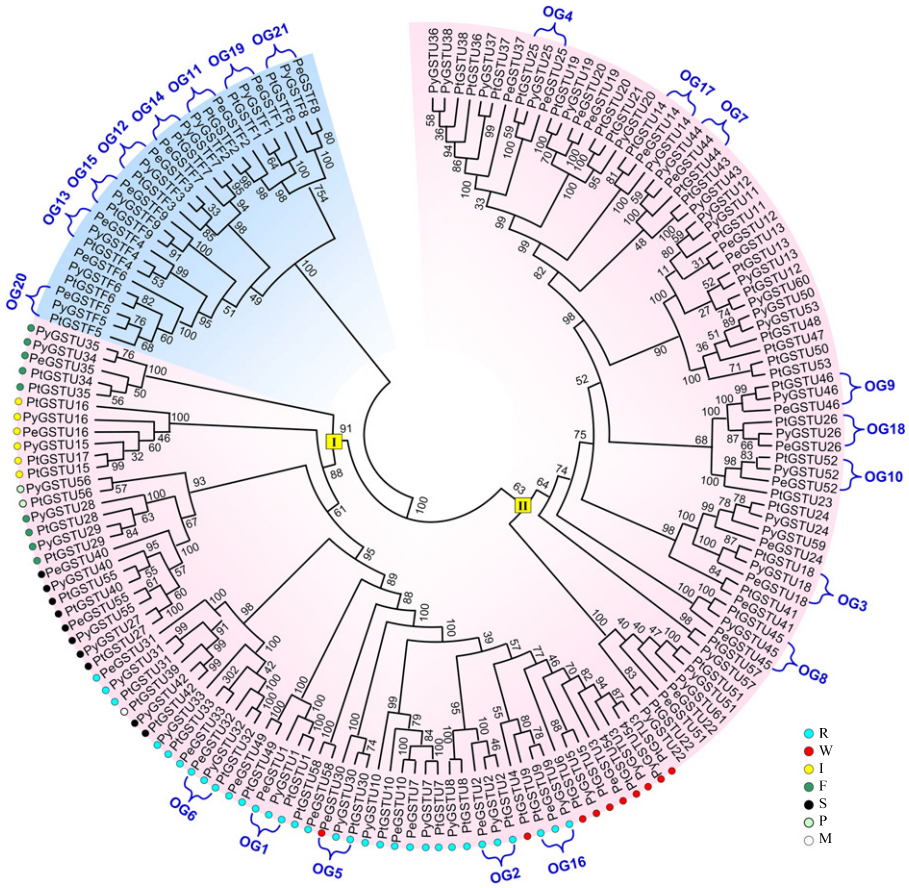



**Fig. 2** Phylogenetic relationships of the tau and phi glutathione *S*-transferase (GST) complementary DNAs in *Populus trichocarpa*, *Populus euphratica*, and *Populus yatungensis*. Numbers on branches indicate the bootstrap percentage values calculated from 1000 bootstrap replicates. The orthologous groups identified in this study are labeled with square extensions. Tau and phi GSTs are shaded in red and blue. Clade I and clade II are represented by yellow boxes. In clade I, the amino acid sites corresponding to Trp39 in PeGSTU30 are represented by colored circles.

other 11 orthologous groups. The first category applied to six groups (OG3, OG4, OG10, and OG15–OG17). In this category, among the three orthologous genes in each group, two had similar tissue-specific gene expression patterns and the other had a divergent pattern. For example, PtGSTU52 and PeGSTU52 in the OG10 group were expressed in all tissues examined, whereas PyGSTU52 was expressed only in bud tissues under the normal growth condition and the CDNB and hydrogen peroxide stress conditions. In the second category (applied to OG6–OG9 and OG12), three orthologous genes in each group had different tissue-specific gene expression patterns. For example, among the three genes in the OG12 group, PyGSTF3 was expressed in all tissues examined, PeGSTF3 was expressed only in bud tissue under normal conditions, and PtGSTF3 was expressed in shoot tissue under normal conditions and in root and bud tissues under normal conditions, CDNB treatments, and hydrogen peroxide treatments.

Of the 11 groups that showed tissue-specific gene expression divergence patterns, only OG15 did not show significant variation in tissue-specific gene expression patterns (nonparametric Kendall's *W* test, $P = 0.82$), while the other 10 groups showed significant variation in at least one tissue (nonparametric Kendall's *W* test, $P < 0.05$). For example, the three orthologous genes in OG9 showed significant variation in tissue-specific gene expression only in phloem tissue (nonparametric Kendall's *W* test, $P = 0.05$).

## Divergence in enzymatic activities among *Populus* GST orthologues

Fourteen GST orthologous groups (OG1–OG3, OG5, OG6, OG11–OG14 and OG16–OG20), which include eight tau and six phi GST groups, were selected for protein expression and purification. Among the 14 orthologous groups, nine (OG1, OG2, OG5, OG11, OG13, OG14 and OG18–OG20) did not show divergent patterns of tissue-specific gene expression, whereas the other five groups (OG3, OG6, OG12, OG16 and OG17) did show divergent patterns. Our previous study assessed the enzymatic activities of *P. trichocarpa* GSTs (Lan *et al.*, 2009). In this study, we investigated enzymatic activities for GSTs belonging to the 14 orthologous groups from other two *Populus* species (Fig. 4). Within OG1 and OG12, PtGSTU1 and PyGSTU1 shared identical protein sequences, as did PtGSTF3 and PyGSTF3. And PyGSTU14 of OG17 was a pseudogene. Therefore, we did not select PyGSTU1, PyGSTU14, and PyGSTF3 for protein expression and purification. All GSTs were expressed as soluble proteins in *E. coli*, except for two proteins (PeGSTU9 and PyGSTF5) which were expressed as inclusion bodies. The purified PeGSTF7 protein was unstable and precipitated easily in the assay buffer. Thus, we used 22 purified GST proteins to determine substrate specificity.

For these 14 orthologous groups, we observed three patterns of differentiation in enzyme specificity. In the first pattern (applied

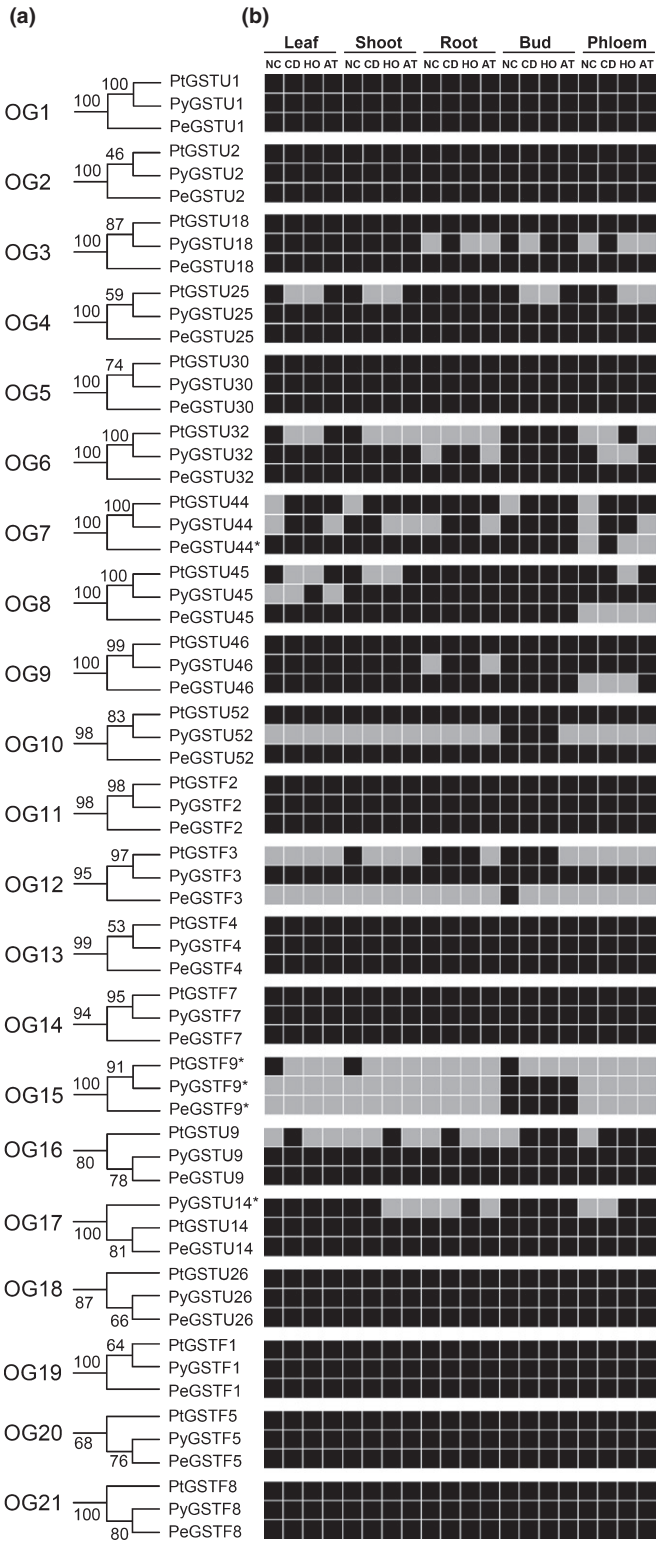



**Fig. 3** Phylogenetic relationships among the members of orthologous groups (a) and their expression patterns (b). The phylogenetic trees are from Fig. 2. Putative pseudogenes are indicated with asterisks. The black box indicates positive detection of gene expression in the corresponding tissue under normal growth conditions (NC) and following 1-chloro-2,4-dinitrobenzene (CD), hydrogen peroxide (HO), and atrazine (AT) treatments.

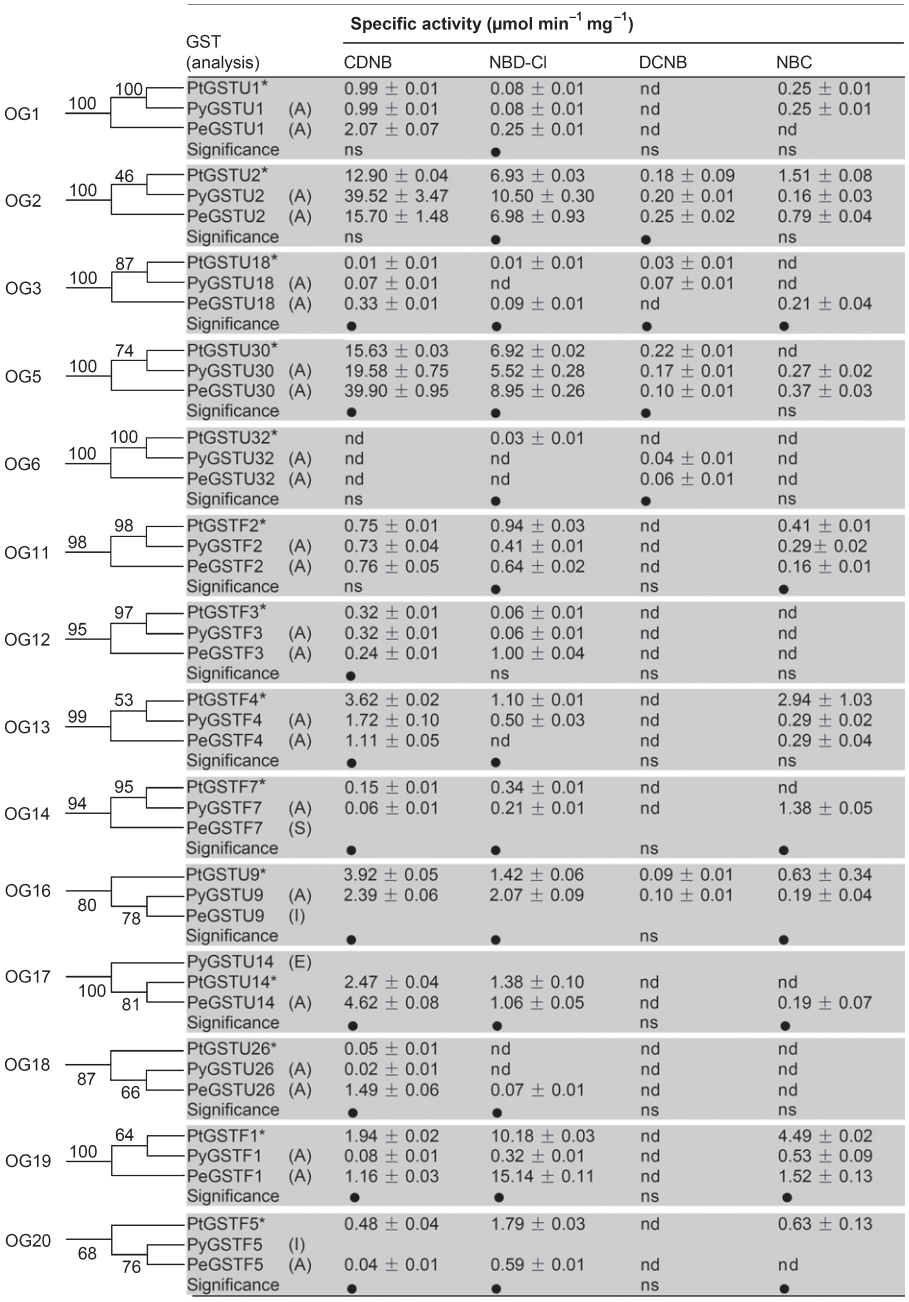
to OG2, OG11, OG12 and OG19), three orthologues of each orthologous group shared similar substrate spectra but differed in their enzymatic activity toward at least one substrate. In the second pattern (applied to OG1, OG5, OG6, OG13 and OG18), two orthologues in each group had a similar substrate spectrum and the other had a different substrate spectrum. In the third pattern, three orthologues in each group had partially overlapping substrate spectra. One orthologous group (OG3) showed this pattern. For OG14, OG16, OG17 and OG20 we obtained the enzymatic activities of only two orthologues in each orthologous group. Because PtGSTU9 and PyGSTU9 in OG16 shared similar substrate spectra, but differed in their enzymatic activities toward substrates CDNB, NBD-Cl, and NBC, OG16 was classified into the first pattern. For OG14, OG17 and OG20, two orthologues in each orthologous group had partially overlapping substrate spectra. We classified these three groups into the third pattern.

To test whether there were significant differences between the three proteins of each orthologous group for each substrate, this study used the nonparametric Kruskal–Wallis $H$ test (Fig. 4). The three orthologues in OG3 showed significant variation in enzymatic activities toward four substrates ($P < 0.04$). For the other 13 orthologous groups (OG1, OG2, OG5, OG6, OG11–OG14 and OG16–OG20), the three proteins of each orthologous group showed significant variation in enzymatic activities toward at least one substrate ($P < 0.05$). For example, the three orthologues in OG2 showed significant variation in enzymatic activities toward the substrates NBD-Cl and DCNB ($P < 0.05$), and no significant variation in enzymatic activities toward the substrates CDNB and NBC ($P > 0.097$).

## Site-directed mutagenesis analysis

To investigate the roles of nonsynonymous substitutions accumulating in different species, we selected GSTs within OG5 to perform site-directed mutagenesis analysis. Among the three GSTs within OG5, PeGSTU30 showed the highest activity toward CDNB and NBD-Cl. Compared with PtGSTU30 and PyGSTU30 respectively, PeGSTU30 had 2.55- and 2.03-fold higher activity toward CDNB, and 1.29- and 1.62-fold higher activity toward NBD-Cl. Compared with PeGSTU30, PtGSTU30 and PyGSTU30 exhibited more similar activity to CDNB and NBD-Cl. PtGSTU30 and PyGSTU30 had only two divergent sites in their protein sequences. PeGSTU30 had 10 different sites compared with the protein sequences of PtGSTU30 and PyGSTU30 (Fig. 5). Thus, we selected PtGSTU30 and PeGSTU30 to construct two sets of mutants for biochemical assays. In this study, the enzyme activities of wild-type and its mutant proteins were determined simultaneously (Table S7).

First, we mutated the amino acid residues in PeGSTU30 to the corresponding divergent amino acid residues in PtGSTU30. We constructed and purified 10 PeGSTU30 mutants (G3S, G4D, R34S, W39R, D61N, A96S, T124A, L183M, I212M and +220E). Among these 10 mutants, +220E indicated a Glu insert



| | | GST (analysis) | Specific activity (μmol min$^{-1}$ mg$^{-1}$) CDNB | NBD-Cl | DCNB | NBC |
|---|---|---|---|---|---|---|
| OG1 | 100 | PtGSTU1* | 0.99 ± 0.01 | 0.08 ± 0.01 | nd | 0.25 ± 0.01 |
| | 100 | PyGSTU1 (A) | 0.99 ± 0.01 | 0.08 ± 0.01 | nd | 0.25 ± 0.01 |
| | | PeGSTU1 (A) | 2.07 ± 0.07 | 0.25 ± 0.01 | nd | nd |
| | | Significance | ns | ● | ns | ns |
| OG2 | 46 | PtGSTU2* | 12.90 ± 0.04 | 6.93 ± 0.03 | 0.18 ± 0.09 | 1.51 ± 0.08 |
| | 100 | PyGSTU2 (A) | 39.52 ± 3.47 | 10.50 ± 0.30 | 0.20 ± 0.01 | 0.16 ± 0.03 |
| | | PeGSTU2 (A) | 15.70 ± 1.48 | 6.98 ± 0.93 | 0.25 ± 0.02 | 0.79 ± 0.04 |
| | | Significance | ns | ● | ● | ns |
| OG3 | 87 | PtGSTU18* | 0.01 ± 0.01 | 0.01 ± 0.01 | 0.03 ± 0.01 | nd |
| | 100 | PyGSTU18 (A) | 0.07 ± 0.01 | nd | 0.07 ± 0.01 | nd |
| | | PeGSTU18 (A) | 0.33 ± 0.01 | 0.09 ± 0.01 | nd | 0.21 ± 0.04 |
| | | Significance | ● | ● | ● | ● |
| OG5 | 74 | PtGSTU30* | 15.63 ± 0.03 | 6.92 ± 0.02 | 0.22 ± 0.01 | nd |
| | 100 | PyGSTU30 (A) | 19.58 ± 0.75 | 5.52 ± 0.28 | 0.17 ± 0.01 | 0.27 ± 0.02 |
| | | PeGSTU30 (A) | 39.90 ± 0.95 | 8.95 ± 0.26 | 0.10 ± 0.01 | 0.37 ± 0.03 |
| | | Significance | ● | ● | ● | ns |
| OG6 | 100 | PtGSTU32* | nd | 0.03 ± 0.01 | nd | nd |
| | 100 | PyGSTU32 (A) | nd | nd | 0.04 ± 0.01 | nd |
| | | PeGSTU32 (A) | nd | nd | 0.06 ± 0.01 | nd |
| | | Significance | ns | ● | ● | ns |
| OG11 | 98 | PtGSTF2* | 0.75 ± 0.01 | 0.94 ± 0.03 | nd | 0.41 ± 0.01 |
| | 98 | PyGSTF2 (A) | 0.73 ± 0.04 | 0.41 ± 0.01 | nd | 0.29± 0.02 |
| | | PeGSTF2 (A) | 0.76 ± 0.05 | 0.64 ± 0.02 | nd | 0.16 ± 0.01 |
| | | Significance | ns | ● | ns | ● |
| OG12 | 97 | PtGSTF3* | 0.32 ± 0.01 | 0.06 ± 0.01 | nd | nd |
| | 95 | PyGSTF3 (A) | 0.32 ± 0.01 | 0.06 ± 0.01 | nd | nd |
| | | PeGSTF3 (A) | 0.24 ± 0.01 | 1.00 ± 0.04 | nd | nd |
| | | Significance | ● | ns | ns | ns |
| OG13 | 53 | PtGSTF4* | 3.62 ± 0.02 | 1.10 ± 0.01 | nd | 2.94 ± 1.03 |
| | 99 | PyGSTF4 (A) | 1.72 ± 0.10 | 0.50 ± 0.03 | nd | 0.29 ± 0.02 |
| | | PeGSTF4 (A) | 1.11 ± 0.05 | nd | nd | 0.29 ± 0.04 |
| | | Significance | ● | ● | ns | ns |
| OG14 | 95 | PtGSTF7* | 0.15 ± 0.01 | 0.34 ± 0.01 | nd | nd |
| | 94 | PyGSTF7 (A) | 0.06 ± 0.01 | 0.21 ± 0.01 | nd | 1.38 ± 0.05 |
| | | PeGSTF7 (S) | | | | |
| | | Significance | ● | ● | ns | ● |
| OG16 | | PtGSTU9* | 3.92 ± 0.05 | 1.42 ± 0.06 | 0.09 ± 0.01 | 0.63 ± 0.34 |
| | 80 | PyGSTU9 (A) | 2.39 ± 0.06 | 2.07 ± 0.09 | 0.10 ± 0.01 | 0.19 ± 0.04 |
| | 78 | PeGSTU9 (I) | | | | |
| | | Significance | ● | ● | ns | ● |
| OG17 | | PyGSTU14 (E) | | | | |
| | 100 | PtGSTU14* | 2.47 ± 0.04 | 1.38 ± 0.10 | nd | nd |
| | 81 | PeGSTU14 (A) | 4.62 ± 0.08 | 1.06 ± 0.05 | nd | 0.19 ± 0.07 |
| | | Significance | ● | ● | ns | ● |
| OG18 | | PtGSTU26* | 0.05 ± 0.01 | nd | nd | nd |
| | 87 | PyGSTU26 (A) | 0.02 ± 0.01 | nd | nd | nd |
| | 66 | PeGSTU26 (A) | 1.49 ± 0.06 | 0.07 ± 0.01 | nd | nd |
| | | Significance | ● | ● | ns | ns |
| OG19 | 64 | PtGSTF1* | 1.94 ± 0.02 | 10.18 ± 0.03 | nd | 4.49 ± 0.02 |
| | 100 | PyGSTF1 (A) | 0.08 ± 0.01 | 0.32 ± 0.01 | nd | 0.53 ± 0.09 |
| | | PeGSTF1 (A) | 1.16 ± 0.03 | 15.14 ± 0.11 | nd | 1.52 ± 0.13 |
| | | Significance | ● | ● | ns | ● |
| OG20 | | PtGSTF5* | 0.48 ± 0.04 | 1.79 ± 0.03 | nd | 0.63 ± 0.13 |
| | 68 | PyGSTF5 (I) | | | | |
| | 76 | PeGSTF5 (A) | 0.04 ± 0.01 | 0.59 ± 0.01 | nd | nd |
| | | Significance | ● | ● | ns | ● |

**Fig. 4** Enzymatic activities of orthologous group proteins. The phylogenetic trees are from Fig. 2. Values shown are means ± SD, calculated from three replicates. nd, no activity detected. Each glutathione *S*-transferase (GST) name is followed by a code in parentheses describing the associated analysis: A, purified GST protein assayed; I, recombinant protein totally insoluble; S, purified recombinant protein instable in buffer; E, analysis not performed. CDNB, 1-chloro-2,4-dinitrobenzene; NBD-Cl, 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole; DCNB, 1,2-dichloro-4-nitrobenzene; NBC, 4-nitrobenzyl chloride. Data cited from Lan *et al*. (2009) are indicated with asterisks. Nonparametric Kruskal–Wallis *H* test was used to test whether there are significant differences among the three proteins of each orthologous group for each substrate. ns, not significant ($P > 0.05$); ●, significant difference ($P < 0.05$).

mutation at the end of PeGSTU30 protein. Compared with wild-type PeGSTU30, the G4D and A96S mutants showed slight increases in enzymatic activities toward CDNB and NBD-Cl (Fig. 6). Conversely, four PeGSTU30 mutants showed decreased enzymatic activities toward these two substrates. Among these mutants, W39R showed the largest decline in enzymatic activity toward CDNB and NBD-Cl (Fig. 6). Next, we mutated the amino acid residues in PtGSTU30 to the corresponding divergent amino acid residues in PeGSTU30. We constructed and purified 10 PtGSTU30 mutants (S3G, D4G, S34R, R39W, N61D, S96A, A124T, M183L, M212I and ΔE220). Compared with wild-type PtGSTU30 protein, except for the R39W mutant, all PtGSTU30 mutants showed decreased enzymatic activities toward CDNB and NBD-Cl (Fig. 6). The R39W mutant exhibited 1.39- and 1.51-fold higher enzymatic activities toward CDNB and NBD-Cl respectively. These results indicate that the Trp39 residue contributed to the high enzymatic activity of PeGSTU30 toward CDNB and NBD-Cl. The phylogenetic tree divided tau GSTs into two clades (clades I and II), and PeGSTU30 was grouped into clade I (Fig. 2). In clade I, the amino acid corresponding to the Trp39 site in PeGSTU30 was an Arg residue in sister clades of the OG5 group (Fig. 2). In the 39$^{th}$ residue site of PeGSTU30, Arg was mutated to Trp through nonsynonymous substitution (AGG→TGG codon change). This mutation plays an important role for PeGSTU30 possessing high enzymatic activity toward substrates CDNB and NBD-Cl.

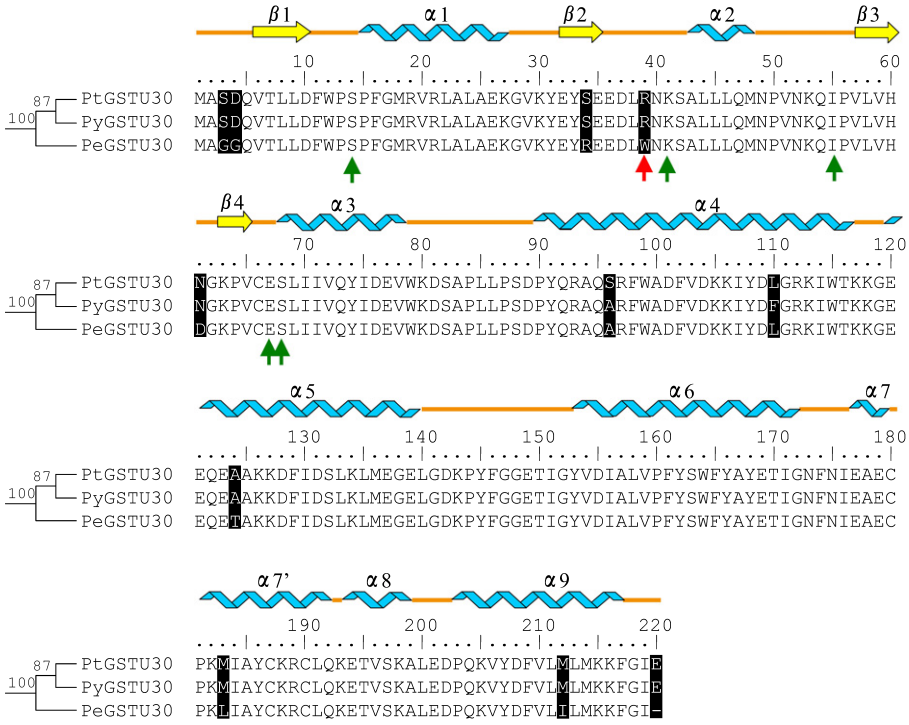



**Fig. 5** Sequence alignment of three orthologous proteins. α-Helices and β-strands are represented by blue helices and yellow arrows, respectively. Black shading, residue substitutions in members of OG5; green arrows, residues involved in binding glutathione; red arrow, the 39th amino acid.

## Steady-state kinetic characteristics of PtGSTU30 and the R39W mutant

To understand the functional significance of nonsynonymous substitutions in the 39th amino acid sites of OG5 orthologues, we investigated the steady-state kinetics of the PtGSTU30 and R39W mutant (Table 1). Significant differences in steady-state kinetic constant values toward substrates GSH and CDNB were observed between the PtGSTU30 and the R39W mutant (Wilcoxon nonparametric test, $P = 0.028$). Compared with the wild-type, the R39W mutant showed 1.11-fold higher affinity (higher $1/K_m$) toward GSH and 1.75-fold higher affinity toward CDNB (Mann–Whitney $U$ test, $P = 0.05$). R39W showed much higher catalytic efficiency ($k_{cat}/K_m$) toward GSH and CDNB compared with the wild-type (Mann–Whitney $U$ test, $P = 0.05$).

To analyze the rate-limiting step of the catalytic reaction in PtGSTU30 and R39W mutant, we determined the effects of viscosity on the kinetic parameters. The dependence of the relative steady-state turnover rate constant ($k_{cat}°/k_{cat}$) vs the relative solvent viscosity ($\eta/\eta°$) will have a slope of 1.0 when a physical step of the reaction (e.g. product release) is fully rate limiting, whereas a slope of zero indicates that the chemical reaction step is rate limiting (Nilsson *et al.*, 2002). For both PtGSTU30 and R39W, their $k_{cat}$ values were sensitive to the viscosity of the buffer (Fig. 7). The effects of viscosity on $k_{cat}$ for PtGSTU30 and R39W yielded slopes of 0.45 and 0.69 respectively, indicating that a physical step and chemical reaction affect the rate-limiting step. The $k_{cat}$ was more sensitive to changes in viscosity in R39W than in wild-type PtGSTU30. Compared with PtGSTU30, the higher slope of R39W mutant indicated that the reaction rate of R39W mutant was more significantly limited by a physical step.

## Structural variation between PtGSTU30 and the R39W mutant

To further characterize the structural significance of nonsynonymous substitutions in the 39th amino acid sites of OG5 orthologues, we investigated the crystal structures of PtGSTU30 and the R39W mutant. We obtained crystals of PtGSTU30 and R39W co-crystallized with GSH but failed to obtain crystals of PtGSTU30 and R39W co-crystallized with the model inhibitor *S*-hexylglutathione. Thus, we only analyzed the crystal structures of protein–GSH complexes. The resolution of the PtGSTU30 and R39W crystal structures were determined at 1.25 Å and 2.65 Å respectively. Parameters for data collection and refinement are provided in Table S8.

PtGSTU30 displayed a typical GST folding structure (Fig. 8a). The monomer of PtGSTU30 consisted of two spatially distinct domains: a small thioredoxin-like N-terminal domain (residues 1–78) and a large helical C-terminal domain (residues 90–220). The N-terminal domain contained the GSH binding site (G-site), and the C-terminal domain contained the hydrophobic substrate binding site (H-site). The R39W mutant had a structure similar to that of PtGSTU30, except for the loop region where the mutation site is located (red arrow in Fig. 8b).

**GSH binding site (G-site)** The G-site residues were important for initial binding of GSH and enzyme catalysis. Each GST monomer combined with one molecule of GSH in the N-terminal domain. Based on crystal structure data, Ser14, Lys41, Ile55, Glu67, and Ser68 of PtGSTU30 and R39W mutant are critical active G-site residues. The conformations of these five amino acid residues were consistent within the structures of PtGSTU30 and R39W mutant (Fig. 8c,d). However, in

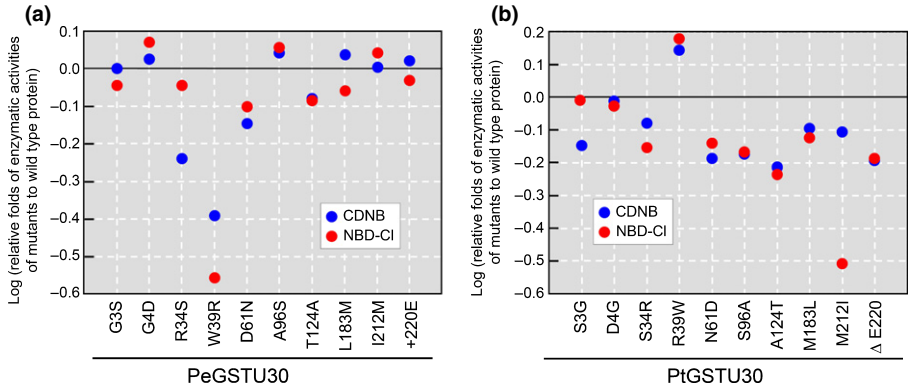



**Fig. 6** Changes in enzymatic activity induced by mutations at the 10 substitution sites relative to the wild-type proteins. The two substrates examined are each denoted by a distinct color: blue, 1-chloro-2,4-dinitrobenzene (CDNB); red, 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl). The enzymatic activities are means as calculated from three replicates. The enzymatic activities of (a) PeGSTU30 and (b) PtGSTU30 toward each substrate are set as the baselines for comparison with the mutants specified on the *x*-axis. The enzymatic activities of wild-type PtGSTU30 protein were cited from Lan *et al.* (2009).

**Table 1** Steady-state kinetic parameters of PtGSTU30 and the R39W mutant.

| | γ-Glutamyl-cysteinylglycine | | | 1-Chloro-2,4-dinitrobenzene | | |
|---|---|---|---|---|---|---|
| Proteins | $1/K_m$ ($mM^{-1}$) | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_m$ ($mM^{-1}\ s^{-1}$) | $1/K_m$ ($mM^{-1}$) | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_m$ ($mM^{-1}\ s^{-1}$) |
| PtGSTU30* | 1.39 | 543.56 | 754.94 | 4.76 | 341.21 | 1624.81 |
| R39W | 1.54 | 705.5 | 1085.38 | 8.33 | 444.83 | 3706.92 |

*Data from Lan *et al.* (2009).

R39W the side chain of Leu38 was in a different position compared with PtGSTU30; it pointed to the hydrophobic pocket in which GST bound the hydrophobic substrates. In R39W, the replacement of hydrophilic Arg39 by Trp with a larger hydrophobic indole (benzopyrrole) side chain resulted in a new hydrophobic interaction between Trp39 and its neighboring Leu38. This hydrophobic interaction dragged the side chain of Leu38 to move and form a new hydrophobic environment between the side chains of Trp39 and Leu38. This new hydrophobic environment resulted in the hydrophilic head of GSH moving 0.7 Å away from Trp39, which might contribute to the higher affinity of R39W for GSH (Table 1).

**Hydrophobic substrate binding site (H-site)** Since the structure of the H-site governs the substrate specificity and activity of a GST protein, structural variations in H-site give the GST the ability to catalyze a wide range of hydrophobic substrates. The side chains of H-site amino acids constitute a hydrophobic pocket in which GST binds hydrophobic substrates. The H-site of GST protein is situated next to the G-site and formed by hydrophobic residues derived mainly from the C-terminal domain. Since we did not obtain crystals of PtGSTU30 or R39W co-crystallized with the model inhibitor *S*-hexylglutathione, we superimposed each of the structures of PtGSTU30 and R39W on the structure of soybean GST (PDB code 2VO4) to infer the H-site amino acid. Next, we mimicked the hydrophobic substrate binding status and inferred the location of the hydrophobic pocket (Fig. 9a,b). Our results showed that the main part of the

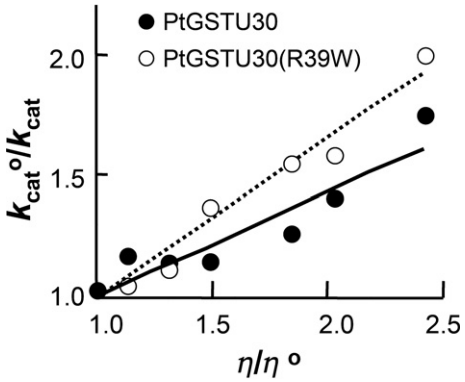



**Fig. 7** The effect of viscosity on $k_{cat}$ for the 1-chloro-2,4-dinitrobenzene conjugation reaction catalyzed by PtGSTU30 and its R39W mutant. A plot of the reciprocal of the relative turnover number ($k_{cat}°/k_{cat}$) as a function of relative viscosity ($\eta/\eta°$) with glycerol as a co-solvent for PtGSTU30 (closed circles) and for the R39W mutant (open circles). Lines were calculated using least-squares regression analysis.

H-site was formed by three α-helices (α4, α6 and α9). Hydrophobic Leu213 and Phe217 were located at the top of the hydrophobic pocket. The aromatic amino acids Tyr108, Trp115, Trp164, and Phe209 formed the bulk of the hydrophobic surface of the H-site pocket. Phe161 and Tyr168 were located at the bottom of the hydrophobic pocket. The conformations of the aforementioned H-site amino acid residues varied slightly between PtGSTU30 and R39W mutant (Fig. 9c,d). In PtGSTU30, a crack between the N-terminal and C-terminal domains was

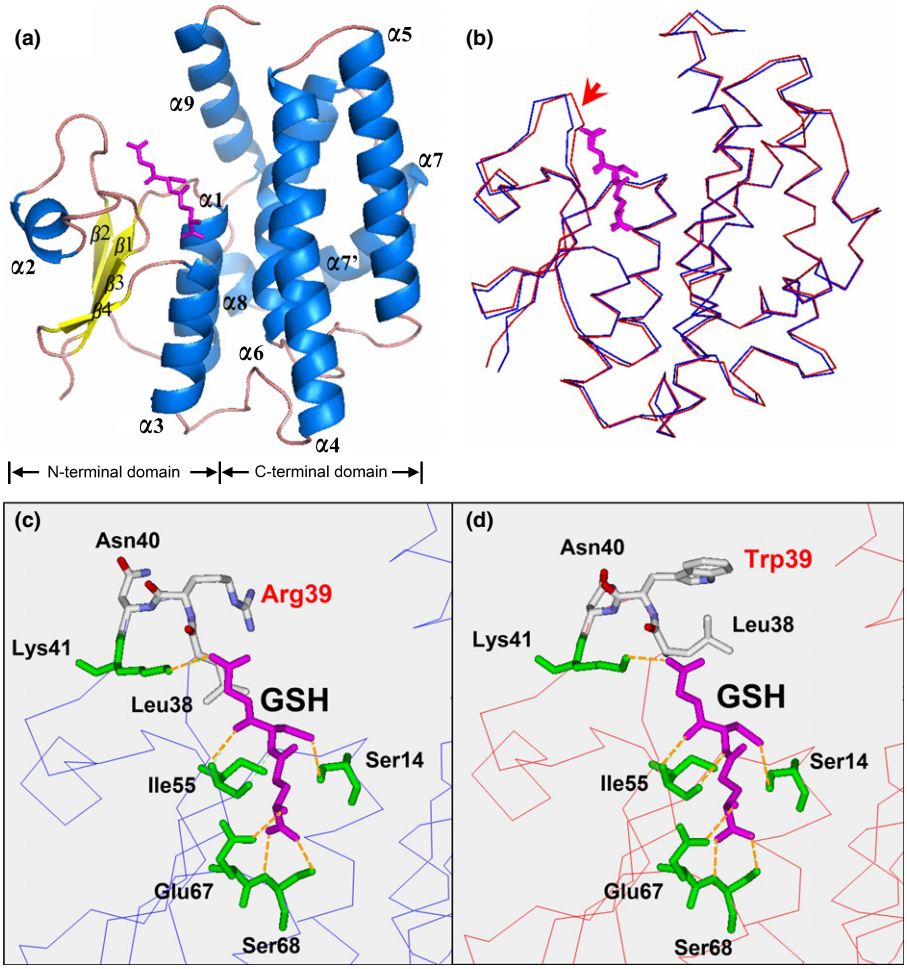



**Fig. 8** A monomer structure of PtGSTU30 (a) and structural comparison (b) and G-sites (c, d) of PtGSTU30 and its R39W mutant. In (a), α-helices and β-strands are represented as blue helices and yellow arrows, respectively. Glutathione (GSH) is represented by purple sticks. In (b), blue and red represent the PtGSTU30 and R39W structures, respectively, and structural difference is indicated by the red arrow. (c, d) Residues involved in GSH binding are labeled and colored in green, and the mutant site and adjacent residues are colored according to atom type. H-bonds are colored in orange.

located at the top of the hydrophobic pocket (Figs 9a, 10). When Arg39 was replaced by Trp, the larger hydrophobic indole side chain of Trp interacted with Phe217 in the C-terminal domain and filled this crack in R39W mutant (Fig. 10). The fill of this crack provided the R39W mutant with a more closed hydrophobic substrate binding pocket and increased the hydrophobic surface of the protein. To confirm this observation, we performed ANS fluorescence probe binding analysis to investigate the hydrophobic surfaces of the proteins. The R39W mutant showed marked increases in fluorescence intensity compared with the wild-type enzyme PtGSTU30 (Fig. S3), confirming the enhanced exposure of hydrophobic surface in R39W mutant protein. In short, the Trp39 substitution provided the R39W mutant with a more closed hydrophobic pocket, resulting in higher enzymatic activity and affinity toward hydrophobic substrates than PtGSTU30.

## Discussion

Currently, the molecular mechanisms underlying the functional conservation or divergence of orthologous genes in closely related species are poorly understood. In this study, we identified 21 orthologous GST groups from three *Populus* species. Based on high-quality whole-genome sequence data of *P. trichocarpa*, our previous study identified 58 tau and nine phi GSTs from the *P. trichocarpa* (Lan *et al.*, 2009). Although genome sequence data were available for *P. euphratica*, the data had not yet been assembled into chromosomes and had numerous gaps in the scaffolds. Up to now, no whole-genome sequence data were available for *P. yatungensis*. To ensure identification of GST genes from *P. yatungensis* and *P. euphratica*, this study used several strategies to amplify GSTs from these two *Populus* species. First, using the full-length *P. trichocarpa* tau and phi GSTs as templates, we designed up to six primer pairs for each gene to perform PCR. Second, we amplified GST genes from different tissues. Total RNAs were extracted from root, shoot, leaf, bud, and phloem tissues under conditions of normal growth and abiotic stresses. This approach was used to increase the possibility of amplifying genes that were specifically expressed under stress conditions. Third, for the GST genes not expressed in the tissues and treatments that we examined, this study tried to amplify their genomic DNA using total genomic DNA from each of the species as a template. Fourth, all PCR products amplified from RNA or genomic DNA were cloned into pEASY-T3 vector and fully sequenced. This approach was used to reduce the possibility of losing the GST fragments. These strategies were used in combination to improve the

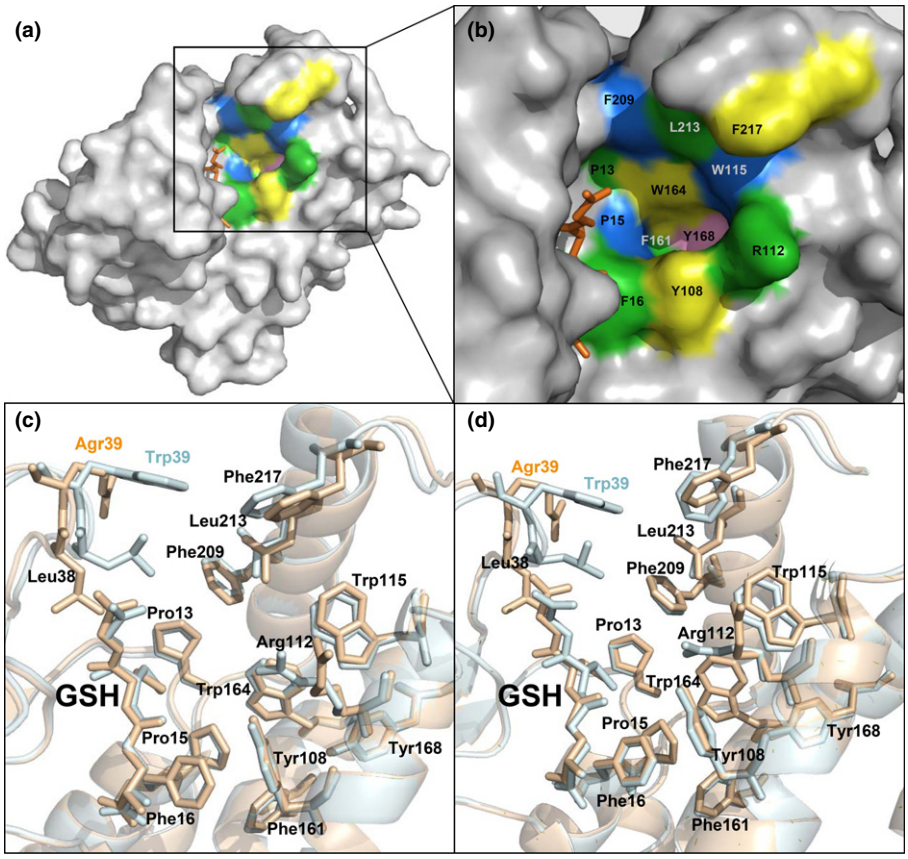



**Fig. 9** The hydrophobic cavity of PtGSTU30 and the R39W mutant. (a, b) The hydrophobic cavity of PtGSTU30, and the residues that constitute the surface of this cavity are labeled and colored. (c, d) The superposition of the PtGSTU30 H-site residues (light orange) with two chains of the R39W mutant (light blue). All residues are labeled with sticks. GSH, glutathione.

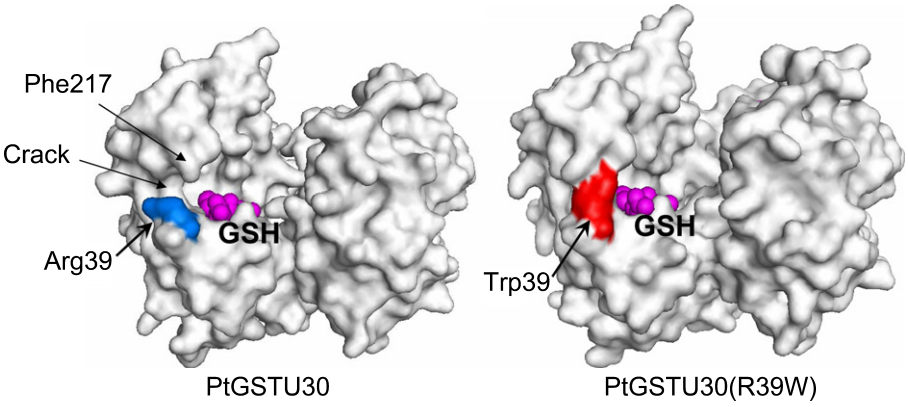



**Fig. 10** Protein surfaces of the PtGSTU30 dimer and R39W mutant dimer. The surfaces of Arg39 and Trp39 are colored in blue and red, respectively. Glutathiones (GSHs) are shown as purple balls.

likelihood of identifying GST genes in these two *Populus* species. Using these strategies, 50 tau and nine phi GSTs were identified from *P. yatungensis*, and 33 tau and nine phi GSTs were identified from *P. euphratica*. Of course, it is possible that some GST genes in *P. yatungensis* and *P. euphratica* were not identified. Another possibility is that there are fewer GSTs in the *P. yatungensis* and *P. euphratica* genomes than in the genome of *P. trichocarpa*.

This study used conventional PCR to investigate tissue-specific gene expression divergence of GST genes in three *Populus* species. Among the 21 orthologous groups, the GST genes in 10 groups (OG1, OG2, OG5, OG11, OG13, OG14 and OG18–OG21) were expressed in all tissues and treatments examined. It is possible that these genes have tissue-specific gene expression divergences in the tissues not examined in this study. Another possibility is that these genes had differences in gene expression levels that were not detected by conventional PCR, but they may be detectable by a more accurate quantitative method. Although conventional PCR only could detect whether a gene was expressed or not, this study still showed that that 11 of 21 GST orthologous groups had tissue-specific gene expression divergence patterns. The gene expression is the first step in creating ecologically and evolutionarily relevant phenotypes (Landry *et al.*, 2007). The three *Populus* species investigated in this study grow in significantly different habitats. The divergence in tissue-specific expression of orthologous genes (particularly those

involved in biotic/abiotic stimuli, such as tau and phi GSTs) may play a role in ecological adaptation of the three *Populus* species. Tissue-specific gene expression divergence is often the first step in functional divergence of duplicate genes (Ferris & Whitt, 1979; Force *et al.*, 1999). As with duplicate genes, divergence in tissue-specific gene expression might also be a key step in the functional divergence of orthologous genes within large gene families in closely related species.

The biological functions of different species have been optimized according to the conditions under which they evolved. When environmental conditions change, existing traits need to be fine-tuned, or new traits need to arise (Camps *et al.*, 2007). The three *Populus* lineages examined in this study diverged *c.* 8–11 Ma (Ma *et al.*, 2013). After speciation, these species occupied different ecological niches. These shifting environmental conditions required fine-tuning of the existing activities of the proteins. Plant tau and phi GSTs play important roles in defense responses against both biotic and abiotic stresses, by detoxifying xenobiotics and combating oxidative stress (Cummins *et al.*, 1999, 2013; Loyall *et al.*, 2000; Agrawal *et al.*, 2002). Because of the diversity of potential xenobiotics and stressors, functional divergence within the GST family has major adaptive significance (Frova, 2003). In this study, we investigated the activities and specificities of 14 orthologous GST groups from three *Populus* species. In five groups, the three orthologues within each group shared similar substrate spectra but differed in their enzymatic activity toward at least one substrate. For the other nine orthologous groups, we observed divergence in substrate spectra. Thus, all orthologous GST groups examined in this study showed divergence in either enzymatic activities or substrate specificities, indicating differences in their biochemical properties. It is possible that some orthologous genes retain the same functions in closely related species. However, some orthologous genes within large gene families in closely related species have undergone functional divergence. In this study, although the sequences of orthologous GST groups from the three *Populus* species were highly conserved, divergence in either enzymatic activities or substrate specificities was prevalent. The selection of different ecological niches for the three *Populus* species may have resulted in the divergence in enzymatic function of the GST orthologous genes in the three *Populus* lineages.

Mutations can directly affect protein function. To understand the molecular mechanisms of protein evolution, this study used a site-directed mutagenesis strategy to investigate how mutation affects protein function. By investigating the biochemical activities of 20 mutant proteins, we found that one mutation site (Arg39→Trp39) contributed to PeGSTU30 possessing high enzymatic activities. A crystallography study showed that this Trp39 residue was situated near the activity site and could affect the shape of the active cavity (Fig. 9). Trp39 residue could increase the hydrophobicity of the active cavity via hydrophobic interactions with adjacent residues. The conformational changes caused by the Trp39 mutant could increase the affinity with electrophilic substrates; on the other hand, they may also have contributed to the dissociation of product. For a given protein, evolvability increases with mutation robustness (the ability of proteins to tolerate mutations) (Camps *et al.*, 2007). Although each amino acid substitution potentially affects protein function, proteins can endure many substitutions with little change in structure, stability, or function (Taverna & Goldstein, 2002; Bloom *et al.*, 2006). This property may be intrinsic to a given protein. Many studies showed that GST proteins can tolerate a large number of mutations that affected enzymatic activity, kinetic characteristics, and thermal stability (Dixon *et al.*, 2003; Zeng & Wang, 2005; Lan *et al.*, 2013). Once an advantageous mutation occurs, positive selection can drive the fixation of this advantageous mutation. Our previous studies found that the plant GST gene family is under positive selection (Lan *et al.*, 2009, 2013; Liu *et al.*, 2015). In particular, in the *P. trichocarpa* GST gene family, the site corresponding to the Trp39 residue site in PeGSTU30 is under positive selection (Lan *et al.*, 2009). Many studies also show that positive selection can drive the evolution of large gene families involved in resistance to biotic and abiotic stresses (Mondragon-Palomino *et al.*, 2002; Gingerich *et al.*, 2007). We inferred that positive selection may contribute to the functional diversification of orthologous genes in closely related species.

## Acknowledgements

We thank the three anonymous reviewers for valuable comments and suggestions that helped to improve the manuscript. This study was supported by the National Science Fund for Distinguished Young Scholars (31425006) and the Chinese Academy of Forestry (CAFYBB2018ZX001). We would like to thank the staff at beamline BL17U of the Shanghai Synchrotron Radiation Facility for their assistance with data collection.

## Author contributions

Q-YZ designed research; QY, X-MH and J-KG performed the experiments; QY and Y-JL analyzed data; M-JY provided technical assistance to QY; QY and Q-YZ wrote the article.

## ORCID

Qi Yang http://orcid.org/0000-0002-7492-9619
Qing-Yin Zeng http://orcid.org/0000-0002-5730-3016

## References

**Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. 2002.** *PHENIX*: building new software for automated crystallographic structure determination. *Acta Crystallographica Section D: Biological Crystallography* **58**: 1948–1954.

**Agrawal GK, Jwa NS, Rakwal R. 2002.** A pathogen-induced novel rice (*Oryza sativa* L.) gene encodes a putative protein homologous to type II glutathione *S*-transferases. *Plant Science* **163**: 1153–1160.

**Altenhoff AM, Studer RA, Robinson-Rechavi M, Dessimoz C. 2012.** Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs. *PLoS Computational Biology* **8**: e1002514.

**Argus WG, Eckenwalder JE, Kiger RW. 2010.** Salicaceae. In: Flora of North America Editorial Committee, eds. *Flora of North America, vol. 7*. New York, NY, USA: Oxford University Press, 13.

**Bloom JD, Labthavikul ST, Otey CR, Arnold FH. 2006.** Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences, USA* **103**: 5869–5874.

**Camps M, Herman A, Loh E, Loeb LA. 2007.** Genetic constraints on protein evolution. *Critical Reviews in Biochemistry and Molecular Biology* **42**: 313–326.

**Chen X, Zhang J. 2012.** The ortholog conjecture is untestable by the current Gene Ontology but is supported by RNA sequencing data. *PLoS Computational Biology* **8**: e1002784.

**Collaborative Computational Project, Number 4. 1994.** The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D: Biological Crystallography* **50**: 760–763.

**Cummins I, Cole DJ, Edwards R. 1999.** A role for glutathione transferases functioning as glutathione peroxidases in resistance to multiple herbicides in black-grass. *Plant Journal* **18**: 285–292.

**Cummins I, Wortley DJ, Sabbadin F, He Z, Coxon CR, Straker HE, Sellars JD, Knight K, Edwards L, Hughes D *et al.* 2013.** Key role for a glutathione transferase in multiple-herbicide resistance in grass weeds. *Proceedings of the National Academy of Sciences, USA* **110**: 5812–5817.

**Dixon DP, Edwards R. 2010.** Glutathione transferases. *The Arabidopsis Book* **8**: e0131.

**Dixon DP, McEwen AG, Lapthorn AJ, Edwards R. 2003.** Forced evolution of a herbicide detoxifying glutathione transferase. *The Journal of Biological Chemistry* **278**: 23930–23935.

**Dolinski K, Botstein D. 2007.** Orthology and functional conservation in eukaryotes. *Annual Review of Genetics* **41**: 465–507.

**Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, Leebens-Mack J, dePamphilis CW. 2010.** Identification of shared single copy nuclear genes in *Arabidopsis, Populus, Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* **10**: 1471–2148.

**Edwards R, Dixon DP, Walbot V. 2000.** Plant glutathione *S*-transferases: enzymes with multiple functions in sickness and in health. *Trends in Plant Sciences* **5**: 193–198.

**Emsley P, Cowtan K. 2004.** *Coot*: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60**: 2126–2132.

**Fang C, Zhao S, Skvortsov AK. 1999.** Salicaceae. In: Wu ZY, Raven PH, Hong DY, eds. *Flora of China, vol. 4*. Beijing, China/St. Louis, MO, USA: Science Press/Missouri Botanical Garden Press 158.

**Ferreira S, Hjerno K, Larsen M, Wingsle G, Larsen P, Fey S, Roepstorff P, Salome Pais M. 2006.** Proteome profiling of *Populus euphratica* Oliv. upon heat stress. *Annals of Botany* **98**: 361–377.

**Ferris SD, Whitt GS. 1979.** Evolution of the differential regulation of duplicate genes after polyploidization. *Journal of Molecular Evolution* **12**: 267–317.

**Fitch WM. 1970.** Distinguishing homologous from analogous proteins. *Systematic Zoology* **19**: 99–113.

**Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999.** Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.

**Frova C. 2003.** The plant glutathione transferase gene family: genomic structure, functions, expression and evolution. *Physiologia Plantarum* **119**: 469–479.

**Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD. 2002.** Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *The Plant Cell* **14**: 1457–1467.

**Gabaldon T, Koonin EV. 2013.** Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics* **14**: 360–366.

**Guindon S, Gascuel O. 2003.** A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**: 696–704.

**Gingerich DJ, Hanada K, Shiu SH, Vierstra RD. 2007.** Large-scale, lineage-specific expansion of a bric-a-brac/tramtrack/broad complex ubiquitin-ligase gene family in rice. *The Plant Cell* **19**: 2329–2348.

**Habig WH, Pabst MJ, Jakoby WB. 1974.** Glutathione *S*-transferases. The first enzymatic step in mercapturic acid formation. *The Journal of Biological Chemistry* **249**: 7130–7139.

**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**: e73.

**Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney 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 Evolutionary Biology* **6**: e29.

**Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M. 2002.** *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. *Plant and Cell Physiology* **43**: 1096–1105.

**Koonin EV. 2005.** Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics* **39**: 309–338.

**Lan T, Wang XR, Zeng QY. 2013.** Structural and functional evolution of positively selected sites in pine glutathione *S*-transferase enzyme family. *The Journal of Biological Chemistry* **288**: 24441–24451.

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

**Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. 2007.** Genetic properties influencing the evolvability of gene expression. *Science* **317**: 118–121.

**Liu HJ, Tang ZX, Han XM, Yang ZL, Zhang FM, Yang HL, Liu YJ, Zeng QY. 2015.** Divergence in enzymatic activities in the soybean GST supergene family provides new insight into the evolutionary dynamics of whole-genome duplicates. *Molecular Biology and Evolution* **32**: 2844–2859.

**Liu YJ, Han XM, Ren LL, Yang HL, Zeng QY. 2013.** Functional divergence of the GST supergene family in *Physcomitrella patens* reveals complex patterns of large gene family evolution in land plants. *Plant Physiology* **161**: 773–786.

**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. *The Plant Cell* **12**: 1939–1950.

**Ma T, Wang J, Zhou G, Yue Z, Hu Q, Chen Y, Liu B, Qiu Q, Wang Z, Zhang J *et al.* 2013.** Genomic insights into salt adaptation in a desert poplar. *Nature Communications* **4**: e2797.

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

**Mondragon-Palomino M, Meyers BC, Michelmore RW, Gaut BS. 2002.** Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Research* **12**: 1305–1315.

**Nehrt NL, Clark WT, Radivojac P, Hahn MW. 2011.** Testing the ortholog conjecture with comparative functional genomic data from mammals. *PLoS Computational Biology* **7**: e9.

**Nilsson LO, Edalat M, Pettersson PL, Mannervik B. 2002.** Aromatic residues in the C-terminal region of glutathione transferase A1-1 influence rate-determining steps in the catalytic mechanism. *Biochimica et Biophysica Acta* **1597**: 157–163.

**Otwinowski Z, Minor W. 1997.** Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **276**: 307–326.

**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. *Analytical Biochemistry* **218**: 463–465.

**Schonbrunn E, Eschenburg S, Luger K, Kabsch W, Amrhein N. 2000.** Structural basis for the interaction of the fluorescence probe 8-anilino-1-naphthalene sulfonate (ANS) with the antibiotic target MurA. *Proceedings of the National Academy of Sciences, USA* **97**: 6345–6349.

**Suarez-Gonzalez A, Hefer CA, Lexer C, Douglas CJ, Cronk QCB. 2018.** Introgression from *Populus balsamifera* underlies adaptively significant variation and range boundaries in *P. trichocarpa*. *New Phytologist* **217**: 416–427.

**Tatusov RL, Koonin EV, Lipman DJ. 1997.** A genomic perspective on protein families. *Science* **278**: 631–637.

**Taverna DM, Goldstein RA. 2002.** Why are proteins so robust to site mutations? *Journal of Molecular Biology* **18**: 479–484.

**Thomas PD, Wood V, Mungall CJ, Lewis SE, Blake JA. 2012.** On the use of Gene Ontology annotations to assess functional similarity among orthologs and paralogs: a short report. *PLoS Computational Biology* **8**: e1002386.

**Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al.* 2006.** The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.

**Wang Z, Du S, Dayanandan S, Wang D, Zeng Y, Zhang J. 2014.** Phylogeny reconstruction and hybrid analysis of *Populus* (Salicaceae) based on nucleotide sequences of multiple single-copy nuclear genes and plastid fragments. *PLoS ONE* **9**: e103645.

**Wolf AV, Brown MG, Prentiss PG. 1985.** Concentrative properties of aqueous solutions: conversion tables. In: Weast RC, ed. *CRC handbook of chemistry and physics*. Boca Raton, FL, USA: CRC Press Inc, D222.

**Wu F, Mueller LA, Crouzillat D, Petiard V, Tanksley SD. 2006.** Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics* **174**: 1407–1420.

**Yang Q, Liu YJ, Zeng QY. 2014.** Biochemical functions of the glutathione transferase supergene family of *Larix kaempferi*. *Plant Physiology and Biochemistry* **77**: 99–107.

**Yang X, Sun W, Liu JP, Liu YJ, Zeng QY. 2009.** Biochemical and physiological characterization of a tau class glutathione transferase from rice (*Oryza sativa*). *Plant Physiology and Biochemistry* **47**: 1061–1068.

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

## Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article:

**Fig. S1** Phylogenetic relationships of the tau and phi classes GST proteins in *Populus trichocarpa*, *Populus euphratica* and *Populus yatungensis*.

**Fig. S2** Protein sequence identities of orthologous GSTs (blue box) and of paralogous GSTs (red box).

**Fig. S3** Fluorescence emission spectra of ANS binding to wild-type PtGSTU30 and the R39W mutant.

**Methods S1** Plant material and treatments.

**Methods S2** Tissue-specific gene expressions of *Populus* GSTs under normal conditions and abiotic stresses.

**Methods S3** Molecular cloning, expression and purification of GST proteins.

**Methods S4** Crystallization, data collection and processing.

**Table S1** PCR primers used to clone *Populus euphratica* and *Populus yatungensis* GST genes.

**Table S2** Sequences used for the reconstruction of a species tree.

**Table S3** PCR primers used to detect the expression of *Populus euphratica* and *Populus yatungensis* GST genes.

**Table S4** Primers used to construct the orthologous GST protein expression vectors.

**Table S5** PCR primers used to construct the expression vectors of mutant proteins.

**Table S6** Full-length GST genes from *Populus euphratica* and *Populus yatungensis*.

**Table S7** Enzymatic activities of wild type and single.-site mutants.

**Table S8** Data collection and structure refinement statistics.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.