Research Article

Functional significance of asymmetrical retention of parental alleles in a hybrid pine species complex

Chang Qu[1] , Hong-Na Kao[1], Hui Xu[1], Bao-Sheng Wang[2], Zhi-Ling Yang[3], Qi Yang[4], Gui-Feng Liu[1], Xiao-Ru Wang[5] , Yan-Jing Liu[1]*, and Qing-Yin Zeng[1]*

[1]State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry and Northeast Forestry University, Beijing 100091, China
[2]Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
[3]Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan 666303, China
[4]State Key Laboratory of Subtropical Silviculture, College of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou, Zhejiang 311300, China
[5]Department of Ecology and Environmental Science, UPSC, Umeå University, SE-90187 Umeå, Sweden
*Authors for correspondence. Yan-Jing Liu. E-mail: yan.jing.liu@163.com; Qing-Yin Zeng. E-mail: qingyin.zeng@caf.ac.cn
Received 27 December 2022; Accepted 3 March 2023; Article first published online 10 March 2023

Abstract Hybrid genomes usually harbor asymmetrical parental contributions. However, it is challenging to infer the functional significance of asymmetrical retention of parental alleles in hybrid populations of conifer trees. Here we investigated the diversity in the glutathione *S*-transferase (GST) gene family in a hybrid pine *Pinus densata* and its parents (*Pinus tabuliformis* and *Pinus yunnanensis*). Plant GSTs play major roles in protecting plants against biotic and abiotic stresses. In this study, 19 orthologous groups of GST genes were identified and cloned from these three species. We examined their expression in different tissues, and then purified the corresponding proteins to characterize their enzymatic activities and specificities toward different substrates. We found that among the 19 GST orthologous groups, divergence in gene expression and in enzymatic activities toward different substrates was prevalent. *P. densata* preferentially retained *P. yunnanensis*-like GSTs for 17 out of the 19 gene loci. We determined the first GST crystal structure from conifer species at a resolution of 2.19 Å. Based on this structure, we performed site-directed mutagenesis to replace amino acid residuals in different wild-types of GSTs to understand their functional impacts. Reciprocal replacement of amino acid residuals in native GSTs of *P. densata* and *P. tabuliformis* demonstrated significant changes in enzyme functions and identified key sites controlling GSTs activities. This study illustrates an approach to evaluating the functional significance of sequence variations in conifer genomes. Our study also sheds light on plausible mechanisms for controlling the selective retention of parental alleles in the *P. densata* genome.

Key words: enzymatic function, functional divergence, gene expression, glutathione *S*-transferases, homoploid hybrid species.

1 Introduction

Homoploid hybrid speciation is a form of speciation without a change in chromosome number after hybridization (Rieseberg, 1997; Abbott et al., 2013). This mode of speciation has historically been considered rare, but large numbers of recent empirical studies have enforced the notion that it is common in both plants and animals (Mallet, 2007). In homoploid hybrids, each parent initially contributes one allele to the newly fused genome. During the subsequent evolution over generations, neutral and selective processes, such as genetic drift, introgression, and selection, may fix some alleles while losing others in the hybrid genome (Meier et al., 2017). Therefore, the parental components are usually unevenly represented in the hybrid, with some loci being homozygote of alternative parental alleles. One important

mechanism of homoploid hybrid speciation is the separation of ecological niche from the parents (Rieseberg et al., 2003; Mallet, 2007), which implies that hybrids experience novel selection resulting in the sorting of parental alleles in the hybrids that confer adaptation to a new environment. However, inferences of selection vs. neutral processes of allele sorting in natural hybrid populations are challenging without any knowledge about the functional implications of allelic variations (Zhao et al., 2020).

Pinus densata is a homoploid hybrid species between *Pinus tabuliformis* and *Pinus yunnanensis* (Wang & Szmidt, 1994; Mao & Wang, 2011). The current distribution of these three species forms a geographical succession, with *P. tabuliformis*, *P. densata*, and *P. yunnanensis* generally occupying the northern, middle, and southern latitudes, respectively (Mao & Wang, 2011). *Pinus densata* forms extensive forests that

regenerate well on the southeastern Tibetan Plateau at elevations ranging from 2700 to 4200 m above sea level (asl), making it an ecologically successful hybrid species in terms of the geographic scale of the establishment (Mao & Wang, 2011; Gao et al., 2012). Reciprocal transplant experiments revealed strong evidence of local adaptation and genotype-environment association analyses detected genome-wide signals of genetic adaptation (Zhao et al., 2014, 2020). However, the validation of DNA sequence variations for functional significance is difficult in conifers due to the lack of an effective gene transformation system and long generation times, leaving answers to why and how allelic frequencies differ among populations and species largely speculative.

Glutathione *S*-transferases (GSTs; EC 2.5.1.18) are ubiquitous proteins in plants, animals, and microorganisms. GSTs catalyze the nucleophilic attack of the sulfhydryl group of tripeptide glutathione (GSH, *γ*-Glu–Cys–Gly) on various electrophilic and hydrophobic toxic molecules. In plants, GSTs are encoded by a large gene family with many members. For example, *Populus trichocarpa* and *Arabidopsis thaliana* contain 81 and 55 members, respectively (Lan et al., 2009). Plant GSTs are divided into eight types, among which tau and phi GSTs are the most abundant (Liu et al., 2013). Plant tau and phi GSTs protect cells from a wide range of biotic and abiotic stresses, including pathogen attack, xenobiotic and heavy metal toxins, oxidative stress, and UV radiation (Loyall et al., 2000; Labrou et al., 2015). Different tau and phi GST members within a plant genome show extensive functional diversification (Dixon et al., 2009; Lan et al., 2009; Liu et al., 2013). Enzymatic proteins as biological catalysts are essential components of every biological system. The key functional characteristics of an enzyme are its catalytic activity toward different substrates and its substrate specificity. Isolation, purification, and biochemical examination of enzymatic proteins offer an opportunity of linking DNA variation to protein structure and function and thus putative adaptive implications.

In our previous study, we cloned 44 GST genes from *P. tabuliformis* and purified their corresponding proteins (Lan et al., 2013). We found that these GSTs are differentiated in gene expression and protein biochemical functions, and that positive selection played a role in their functional evolution (Lan et al., 2013). We became curious about the GST gene composition in the *P. densata* genome and whether sequence variations among species have functional implications. In this study, we cloned 43 GSTs from *P. densata* and 51 from *P. yunnanensis*. We examined the parental origins of these genes in *P. densata* and compared their expression in different tissues. We then purified these proteins to evaluate their differences in enzymatic activities. We produced the first GST crystal of conifer species to characterize its structure. After that, we selected two native GST proteins that differ in several amino acid sites and performed reciprocal site-directed mutagenesis to understand the impact of amino acid substitutions on protein structure and function. We illustrate an alternative approach to exploring the functional significances of sequence variation in conifer genomes.

2 Material and Methods

2.1 Cloning of GST genes in *Pinus*

The genome sequences of *Pinus densata*, *Pinus yunnanensis*, and *Pinus tabuliformis* were not available when we performed this study. The GST gene sequences of *P. tabuliformis* are reported in our previous studies (Kao et al., 2012; Lan et al., 2013). To amplify the GSTs from *P. densata* and *P. yunnanensis*, the GST genes of *P. tabuliformis* were used as reference sequences. To ensure the identification of GST genes from *P. densata* and *P. yunnanensis*, four strategies were used to amplify GST genes from the two pine species. First, in order to increase the possibility of amplifying genes that were specifically expressed in some specific tissues, we amplified the GST genes from different tissues. Thus, total RNAs were extracted from the root, stem, and leaf tissues of seedlings, and bud, leaf, and phloem tissues of mature trees. Second, it is possible that the target gene was not amplified due to the non-specificity primers. Thus, we designed two pairs of primers for the GST gene that had not been amplified from different tissues. The polymerase chain reaction (PCR) primers used for GST gene cloning are listed in Table S2. Third, for the GST genes not expressed in the tissues that we examined, we tried to amplify their genomic DNA using total genomic DNA as templates. Fourth, to verify the amplified sequences, the PCR products were cloned into the pGEM-T vector (Promega, Madison, WI, USA) and then transfected into JM109 competent cells. For each PCR product, we chose 5–15 positive single colonies and sequenced them in both directions.

The DNAsecure Plant Kit was used to isolate DNA (Tiangen, Beijing, China). The RNAprep Pure Plant Plus Kit (Polysaccharides & Polyphenolics-rich) (Tiangen) was used to isolate total RNA and the PrimeScript™ RT reagent kit (Perfect Real Time) (TaKaRa, Dalian, China) was used for reverse transcription.

The sequences were then analyzed by the National Center for Biotechnology Information (NCBI) Conserved Structure Database (https://www.ncbi.nlm.nih.gov/cdd) with default parameters to confirm the presence of typical GST domains. The sequences containing the GST domains were identified as GSTs. All GST sequences of *P. densata* and *P. yunnanensis* are listed in Table S3.

2.2 Phylogenetic relationship and nomenclature of GST genes in three *Pinus* species

The GST amino acid sequences of *P. densata*, *P. yunnanensis*, and *P. tabuliformis* were aligned with the MAFFT v7 online service (https://mafft.cbrc.jp/alignment/software/) (Katoh et al., 2019). The amino acid substitution model was Jones, Taylor, and Thornton (JTT), which was selected using the ModelGenerator v.0.84 program (Keane et al., 2006). The phylogenetic tree of the GSTs in three *Pinus* species was constructed using the maximum likelihood (ML) algorithm in PhyML v2.4.4 (Guindon & Gascuel, 2003). The bootstrap value was 100 and the outgroup was *Escherichia coli* Glutaredoxin 2 (GRX2).

The nomenclature of *P. densata* and *P. yunnanensis* GST genes followed the criterion proposed by previous studies (Edwards et al., 2000; Lan et al., 2013). Each GST gene was named in three parts, including its abbreviated species name

(Pde for *P. densata*, Pyu for *P. yunnanensis*), capital letters next to the species name indicating the subfamily class (GSTU, GSTF, EF1BG, GSTL, GSTT, TCHQD, and GSTZ corresponding to tau, phi, EF1Bγ, lambda, theta, TCHQD, and zeta subfamily class, respectively), and the number was the serial number of the gene in the subfamily (e.g., PdeGSTU1).

2.3 Orthologous groups identification of *Pinus* GST genes

The tau and phi GST proteins of *P. densata*, *P. yunnanensis*, and *P. tabuliformis* were aligned using MAFFT v7 online software. The alignment was manually adjusted with BioEdit v7.0.9.0 (Tom Hall, Ibis Therapeutics, Carlsbad, CA, USA) (Hall, 1999). Then, the phylogenetic tree for *Pinus* GST proteins was constructed using the ML program in PhyML v2.4.4 with the JTT model that was selected by Model Generator v.0.84 software. The bootstrap values were set to 100. The phi GSTs were outgroups for tau GSTs.

The tau and phi GST genes of the three *Pinus* species were aligned with RevTrans v1.4 Server online software (http://www.cbs.dtu.dk/services/RevTrans/) (Wernersson & Pedersen, 2003). After manually adjusting the alignment with BioEdit v7.0.9.0, we selected the General Time Reversible (GTR) model using Model Generator v.0.84. We reconstructed the ML tree for *Pinus* GST genes in 1000 bootstrap replicates.

2.4 Tissue-specific gene expressions of *Pinus* GSTs

The expression pattern of *P. tabuliformis* GSTs has been reported in a previous study (Lan et al., 2013). To investigate the tissue-specific gene expression patterns of *P. densata* and *P. yunnanensis* GST genes under normal conditions, 3-year-old mature trees, 3-month-old seedlings, and sprouted seeds were sampled. The 3-year-old mature trees of *P. densata* and *P. yunnanensis* were collected from Nyingchi, Tibet, and Kunming, Yunnan, respectively. The 3-month-old seedlings of both *P. densata* and *P. yunnanensis* were cultivated in a greenhouse. Total RNA from the bud, leaf, phloem from the stem, phloem from the root, and root tip of mature trees, leaf, stem, and root of seedlings, and radicle of the sprouted seeds was isolated. Total RNA was then reverse transcribed into complementary DNA (cDNA) as the PCR templates. Based on the multiple sequence alignment of the tau and phi class GST genes from the three *Pinus* species, 78 pairs of primers were designed for the specific amplification (Table S4). The actin gene of *Pinus* was used as an internal control, and the primers for the actin gene were taken from a previous study (Lan et al., 2013).

2.5 Molecular cloning, expression, and purification of wild-type and mutant GST proteins

To measure the GST enzymatic activities of the three *Pinus* species, 19 GSTs from *P. densata*, 13 GSTs from *P. yunnanensis*, and 13 GSTs from *P. tabuliformis* were selected for protein expression and purification. The full-length GST sequences were subcloned into the pET-30a expression system (Novagen, Madison, WI, USA) to obtain a 6× His-tag at the N-terminal part of GST proteins. The primers used to construct the expression vectors are listed in Table S5. We performed the expression and purification of GSTs using the methods described by Liu et al. (2013). We performed

site-directed mutagenesis using the TRAN Fast Mutagenesis System kit (TransGen, Beijing, China). The primers used to construct the mutants of PdeGSTU1 and PtaGSTU1 are listed in Table S6.

2.6 Enzyme activities and kinetic studies

Enzymatic activities of the purified GST proteins were studied using five conventional substrates. The activities of GST proteins on 1-chloro-2,4-dinitrobenzene (CDNB; Sigma-Aldrich, Shanghai, China), 1,2-dichloro-4-nitrobenzene (DCNB; Sigma-Aldrich), and 4-nitrobenzyl chloride (NBC; Sigma-Aldrich) were determined according to Habig's description (Habig et al., 1974), and the activities on 7-chloro-4-nitrobenzo-2-oxa-1, 3-diazole (NBD-Cl; Sigma-Aldrich) were determined according to Ricci's method (Ricci et al., 1994), and the activities on diphenyl ethers (Fluorodifen) were determined according to the method of Edwards and Dixon (Edwards & Dixon, 2005). GST protein concentrations were determined by measuring the absorbance at 280 nm. All assays were measured using Evolution^{TM} 300 UV-visible spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) at room temperature, and each experiment was conducted in three replicates.

The steady-state kinetic parameters of the recombinant GST proteins on GSH and CDNB substrates were determined by the method described by Lan et al. (2009). To determine the K_m value of GSH, the concentration of CDNB was fixed at 1.0 mM, and the concentration of GSH was varied in the range 0.2–4.0 mM. The apparent K_m values for CDNB were determined using concentrations of CDNB ranging from 0.2 to 2.0 mM and a fixed GSH concentration of 1.0 mM. The kinetic parameters were obtained with nonlinear regression analysis in the HYPER32 program (https://hyper32.software.informer.com/).

2.7 Differential scanning calorimetry (DSC)

DSC experiments were performed using a MicroCal PEAQ-DSC (MicroCal, Northampton, MA, USA). All GST proteins were subjected to DSC scanning in 20 mM sodium HEPES and 200 mM NaCl (pH 7.4) buffer. All data were collected at a scan rate of 2 °C/min, and the DSC scans were run from 20 °C to 90 °C.

2.8 Crystallization, data collection, and processing

The PdeGSTU1 gene was subcloned into a modified vector ΔpET-30a (Yang et al., 2009), and transformed into *E. coli* BL21 (DE3). After initially being purified by affinity chromatography using a Nickel-Sepharose column, the protein was then subjected to size-exclusion chromatography through a 75 pg 16/600 GL column (GE Healthcare, Uppsala, Sweden) to desalt PdeGSTU1 into 10 mM Tris, pH 7.4, 10 mM GSH and obtain purer protein. The size-exclusion chromatography column was installed on the ÄKTA FPLC system (GE Healthcare Europe GmbH Succursale France, Orsay, France). Purified PdeGSTU1 protein concentration was adjusted to 10, 20, 30, 40, 50, 60, 70, and 80 mg/mL using an Amicon Ultra-15 centrifugal filter device (Merck Millipore, Country Cork, Ireland), respectively. These eight concentrations of proteins were used for crystallization with the Crystal Screen kit (Hampton Research, Laguna Beach, CA, USA) by hanging drop vapor diffusion at 16 °C. Each drop

consisted of a 2 μL protein solution mixed with a 2 μL reservoir solution in the Crystal Screen kit. After 6–10 days, platelet crystals could be found at 0.1 M MES monohydrate, pH 5.5, 25% v/v polyethylene glycol 400 (mother liquor), with a protein concentration of 42 mg/mL.

The protein crystals were taken out from the mother liqueur into the cryoprotectant, which was the mother liqueur containing 20% (v/v) glycerol using a Mounted CryoLoop (0.05–0.1 mm). And then the crystals encased in cryoprotectant were stored in liquid nitrogen for X-ray data collection. The X-ray reflection data were collected from the BL19U beamline of the Shanghai Synchrotron Radiation Facility (SSRF) and preprocessed using the HKL3000 software package (Otwinowski & Minor, 1997). The best template for molecular replacement of PdeGSTU1 was GmGSTU4-4 (PDB code 4TOP), which was selected by the SWISS-MODEL online website (https://swissmodel.expasy.org/interactive) with the default parameters. The CCP4 v6.5.0 suite (Collaborative Computational Project 1994) (Potterton et al., 2003) was used to solve the phase problem. Then we used the Coot v0.8.1 program (Emsley & Cowtan, 2004) and the PHENIX v1.9-1692 program (Adams et al., 2002) to build, manually adjust, and refine the model. High-quality crystal data were difficult to obtain. In this study, the R-free factor of the PdeGSTU1 protein crystal was 30.21%, slightly larger than 30%. Even so, the backbone structure of the protein crystal was relatively accurate and did not affect the subsequent analysis. The PYMOL v1.0r2 (DeLano Scientific LLC, San Carlos, CA, USA) was used to compare and analyze the three-dimensional structure of the PdeGSTU1 protein, and generate the final images.

3 Results

3.1 Identification of GST orthologues from the three pine species

This study cloned 43 and 51 GST genes from *Pinus densata* and *Pinus yunnanensis*, respectively, and 44 GST genes from *Pinus tabuliformis* had been identified previously (Lan et al., 2013). We analyzed the genetic relationships among these genes (Fig. S1) and classified 30 tau and 6 phi GST genes from *P. densata*, and 30 tau and 12 phi GST genes from *P. yunnanensis*.

To identify GST orthologues in the three species, we constructed a phylogenetic tree using 86 tau and 25 phi GST protein sequences from *P. densata*, *P. yunnanensis*, and *P. tabuliformis* (Fig. 1A). Because of the hybrid origin, we expected to see both parental copies of genes in the hybrid. We defined the orthologous group using two criteria: (i) the genes from each of the three pine species group together; (ii) the genes of *P. densata* should group either with *P. yunnanensis* or *P. tabuliformis* (Model A or B in Fig. 1B). In cases in which genes from *P. tabuliformis* and *P. yunnanensis* group together, the gene in *P. densata* is then regarded as not reflecting progenitor-derivative relationship, thus not considered as an orthologous group with the parental species (Model C in Fig. 1B). Based on the phylogenetic tree of protein sequences, we identified 19 orthologous groups (named OG1–OG19), each with three genes from each of the three species (Fig. 1). All 19 groups had high bootstrap support (>60%). We also analyzed the genetic relationships of these genes using their cDNA sequences and found the identification of the 19 groups was also supported (Fig. S2). Among the 19 orthologous groups, *P. densata* retained GST genes from *P. yunnanensis* in 17 cases and two from *P. tabuliformis*.

### 3.2 Expression of the GST orthologues in different tissues

We investigated the expression patterns of 19 GST orthologues in nine types of tissues of the three species. In total, 513 tissue-expression profiles were generated. We found that 12 groups (OG1, 3, 4, 6, 7, 8, 12, 14, 16, 17, 18, and 19) were expressed in all tissues examined in each respective species, while the other seven groups (OG2, 5, 9, 10, 11, 13, and 15) showed tissue-specific patterns among the three members of each orthologous group (OG) (Fig. 2). Interestingly, in cases in which differential expression was observed among the three members of an OG, *P. densata* genes showed expression similar to that of *P. yunnanensis* rather than to *P. tabuliformis*, in accordance with their sequence similarities.

3.3 Expression and purification of GST proteins from the three pine species

As an enzyme, a key characteristic of the catalytic activity of GSTs is its substrate activity and specificity. In this study, we analyzed substrate activities and specificities of the 57 GST proteins from the 19 GST OGs using five substrates, CDNB, NBD-Cl, DCNB, NBC, and Fluorodifen. We expressed these genes in *Escherichia coli* and purified the corresponding proteins for biochemical assays. The enzyme activities of six *P. tabuliformis* GSTs (PtaGSTU6, 7, 10, 12, 17, and 26) toward these five substrates were evaluated in our previous studies (Lan et al., 2013). Among the 19 OGs, six pairs of proteins in six OGs (PdeGSTU2/PyuGSTU2, PdeGSTU10/PyuGSTU10, PdeGSTU25/PyuGSTU25, PdeGSTF1/PyuGSTF1, PdeGSTF2/PyuGSTF2, and PdeGSTF3/PyuGSTF3) had the same amino acid sequence. For these six pairs, we only selected the GST proteins from *P. densata* in each pair for expression and purification. Thus, a total of 45 GST proteins (13 from *P. tabuliformis*, 19 from *P. densata*, and 13 from *P. yunnanensis*) were expressed and purified in this study. We found that nine proteins (PdeGSTU6, PdeGSTF4, PyuGSTU12, PyuG-STUF4, PtaGSTU18, PtaGSTU20, PtaGSTU25, PtaGSTF1, PtaGSTF4) were expressed as inclusion bodies in *E. coli*, the remaining 36 GSTs were expressed as soluble proteins. After purification through the Nickel-Sepharose High-Performance column, these 36 soluble GST proteins were stable in the enzyme assay buffer and were used for subsequent analysis of substrate activities and specificities.

3.4 Divergence in substrate specificity and activity of the GST orthologues

The 36 purified soluble proteins were evaluated for substrate activity and specificity (Figs. 3, S3). We observed three divergent modes of enzyme specificity: (i) the three orthologues in each OG shared a similar substrate spectrum, for example, OG1, 2, 4, and 5 groups. (ii) GST proteins of *P. densata* and *P. yunnanensis* had similar substrate spectrums but were different from those of *P. tabuliformis*, for example, OG6, 13, 17, and 18 groups. (iii) the three orthologues had a partially overlapping substrate spectrum, for example, OG8, 11, 12, and 15 groups.
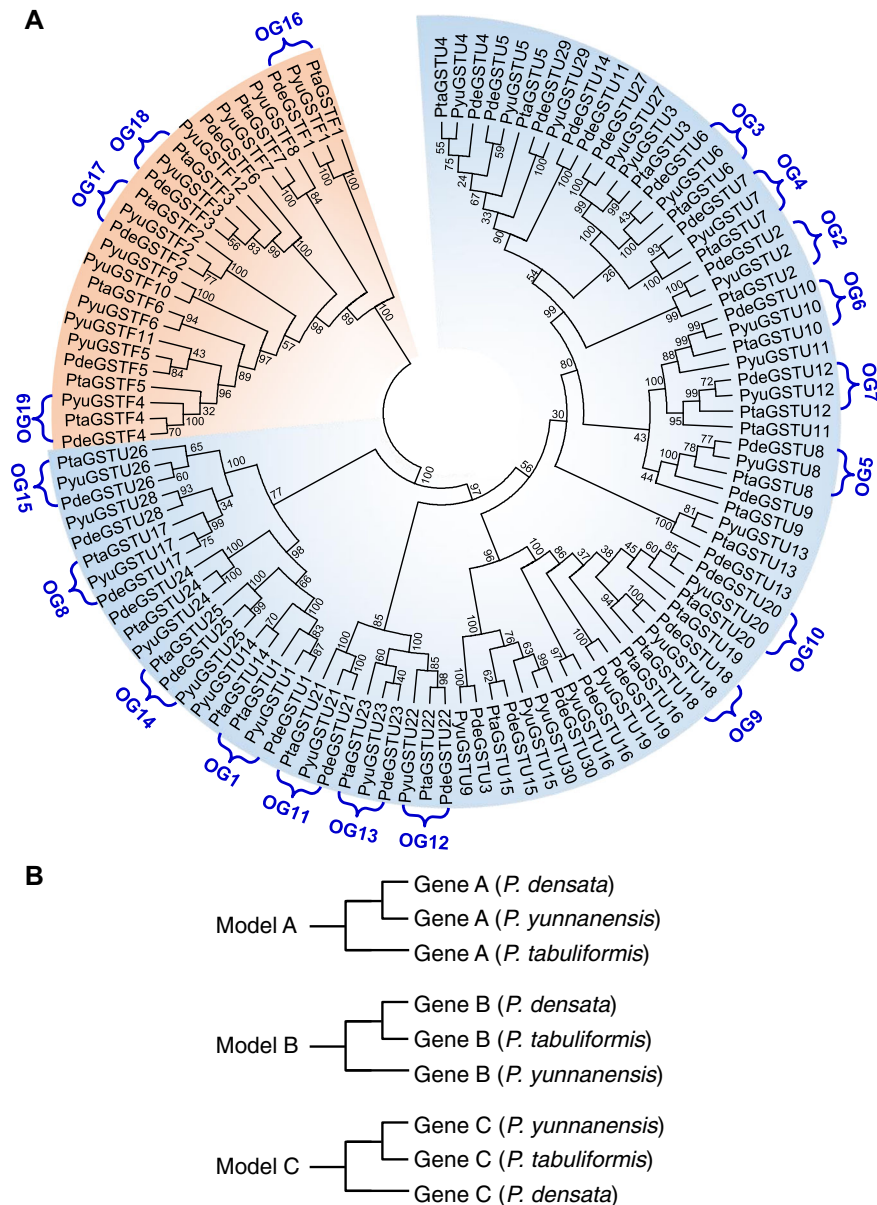
Fig. 1. Phylogenetic relationships of the tau and phi glutathione *S*-transferase (GST) proteins from *Pinus densata*, *Pinus yunnanensis*, and *Pinus tabuliformis* (**A**), and three models of clustering GSTs in three *Pinus* species (**B**). **A,** Numbers on branches indicate the bootstrap values calculated from 100 bootstrap replicates. The orthologous groups (OG1–OG19) identified in this study are marked with curly brackets. The tau and phi GSTs are colored in blue and orange, respectively. **B,** The GST genes belonging to Model A or Model B are defined as an orthologous group in this study.

For OG3, 7, 9, 10, 14, and 16 groups, we only obtained the enzymatic activities of two orthologues in each orthologous group. For OG3 and 10 groups, the two orthologues in each group shared a partially overlapping substrate spectrum. We classified these two groups into the third model. For OG7, 9, 14, and 16 groups, the two orthologues in each group shared a similar substrate spectrum, which was classified into the first model (Fig. S3).

We measured the substrate activities of the GSTs in 12 OGs (Fig. 3). Among these OGs, only in OG12 *P. densata* retained the genes from *P. tabuliformis*, while in the other 11 OGs it

retained genes from *P. yunnanensis*. Enzymatic activities of the GST proteins from *P. densata* reflected their relationships with the two parental species, with enzymatic properties toward the closest copy. For example, PdeGSTU22 and PtaGSTU22 had enzymatic activities toward the substrates CDNB and Fluorodifen, but PyuGSTU22 did not.

We compared pairwise differences in enzymatic activities of the three proteins in each OG toward each of the five substrates. Comparisons were made when all three proteins in each OG had activities toward the substrate. This generated 108 data points, which were divided into three
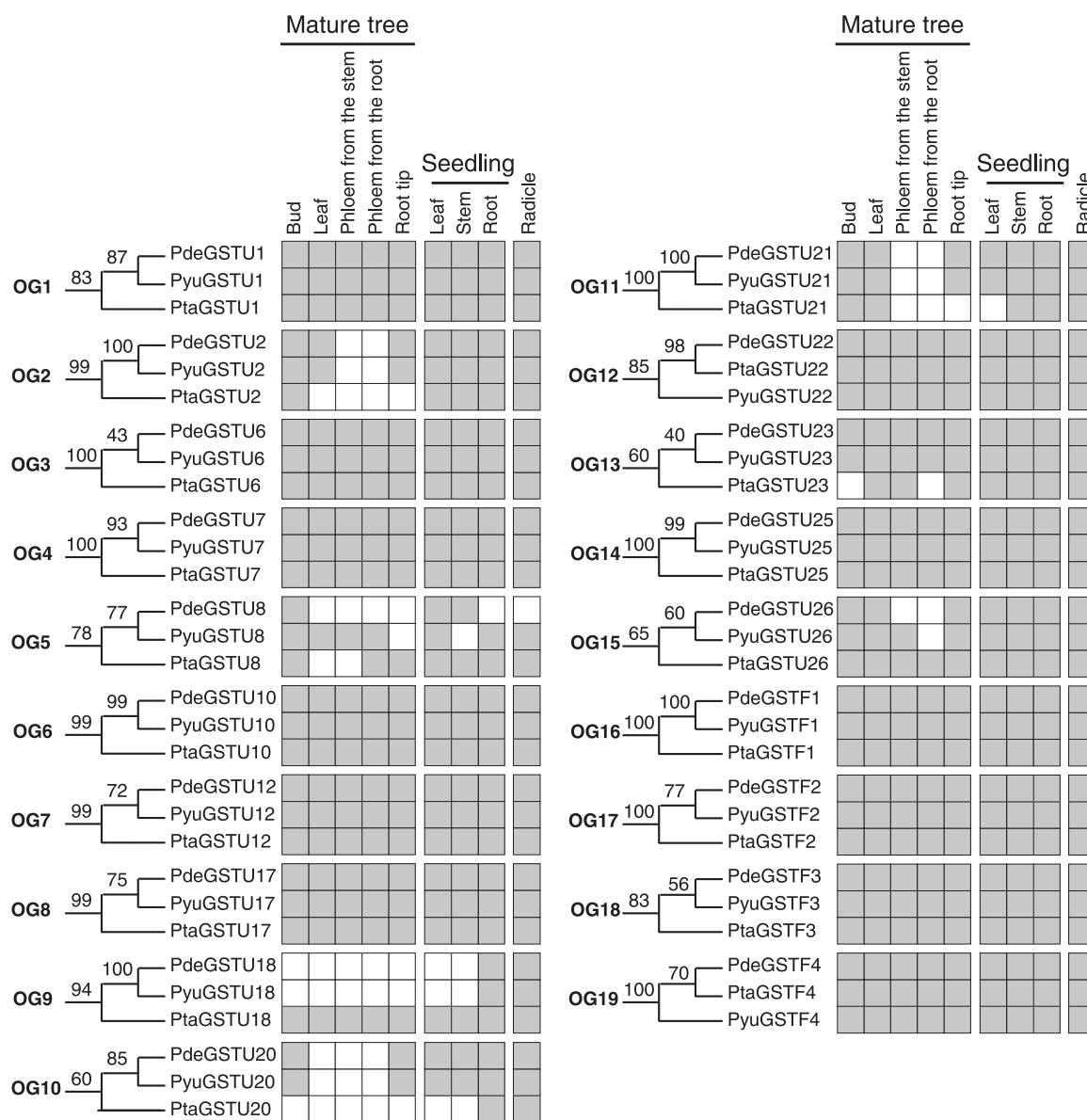
Fig. 2. The expression patterns of members of 19 glutathione *S*-transferase orthologous groups in three *Pinus* species. The phylogenetic trees are from Fig. 1. The gray boxes indicated the genes that are expressed in the corresponding tissue under normal growth conditions. The expression patterns of *Pinus tabuliformis* are taken from our previous study (Lan et al., 2013).

sets: PdeGST and PyuGST, PdeGST and PtaGST, and PyuGST and PtaGST. Each set contained 36 data points (Fig. S4). The PdeGST–PyuGST set showed significant differences from that of the PdeGST–PtaGST set or the PyuGST–PtaGST set (Mann–Whitney *U*-test, $P < 0.001$), while the difference between PdeGST–PtaGST and PyuGST–PtaGST was not significant (Mann–Whitney *U*-test, $P = 0.444$). These results illustrated the fact that the enzymatic activities of the GST proteins in *P. densata* were more similar to that of *P. yunnanensis*.

3.5 Site-specific mutagenesis analysis

To understand whether there were key amino acid sites that affected substrate activities among GST orthologues, we conducted a site-specific mutagenesis analysis in OG1. Among

the 19 orthologous groups, the three GSTs of OG1 showed a broad substrate spectrum and high enzymatic activities toward the five substrates. Of the three proteins in this group, PdeGSTU1 and PyuGSTU1 differed by only one amino acid site, and PdeGSTU1 and PtaGSTU1 differed by five sites (Fig. 4). We chose PdeGSTU1 and PtaGSTU1 to construct two groups of mutants.

First, we mutated the amino acid residues in PdeGSTU1 to the corresponding amino acid residues in PtaGSTU1. We constructed and purified five PdeGSTU1 mutants, A38P, L106I, G111S, M220I, and R222Q, and examined their enzymatic activities toward substrates CDNB, NBD-Cl, and Fluorodifen (Fig. 5). When the Ala38 residue was replaced by a Pro residue, and Gly111 was replaced by Ser, the A38P and G111S mutant activities toward the three substrates

GST (analysis)	Specific activity (µmol min⁻¹ mg⁻¹)				
	CDNB	NBD-Cl	DCNB	NBC	Fluorodifen
OG1 (87, 83)					
PdeGSTU1	49.82 ± 4.45	18.44 ± 1.88	0.22 ± 0.02	0.45 ± 0.21	297.11 ± 23.30
PyuGSTU1	59.46 ± 5.64	19.30 ± 0.64	0.20 ± 0.02	0.65 ± 0.20	301.57 ± 15.97
PtaGSTU1	22.50 ± 1.43	5.27 ± 0.33	0.18 ± 0.01	0.52 ± 0.26	169.14 ± 10.99
Significance	•	•	•	•	•
OG2 (100, 99)					
PdeGSTU2	1.41 ± 0.02	0.93 ± 0.03	0.05 ± 0.01	2.65 ± 0.16	0.06 ± 0.01
PyuGSTU2	1.41 ± 0.02	0.93 ± 0.03	0.05 ± 0.01	2.65 ± 0.16	0.06 ± 0.01
PtaGSTU2	4.17 ± 0.03	2.69 ± 0.05	0.07 ± 0.01	0.96 ± 0.06	0.12 ± 0.01
Significance	ns	ns	ns	ns	ns
OG4 (93, 100)					
PdeGSTU7	11.59 ± 0.03	50.55 ± 0.78	0.72 ± 0.04	3.12 ± 0.12	4.05 ± 0.08
PyuGSTU7	10.30 ± 0.25	48.32 ± 3.63	0.86 ± 0.02	3.02 ± 0.15	3.26 ± 0.09
PtaGSTU7*	12.17 ± 0.01	10.66 ± 0.10	0.07 ± 0.01	0.55 ± 0.09	2.38 ± 1.00
Significance	•	ns	•	ns	ns
OG5 (77, 78)					
PdeGSTU8	8.19 ± 0.22	1.26 ± 0.02	0.05 ± 0.01	2.74 ± 0.08	12.00 ± 0.14
PyuGSTU8	16.68 ± 0.28	12.64 ± 0.42	0.07 ± 0.01	3.54 ± 0.11	9.17 ± 0.18
PtaGSTU8	3.80 ± 0.04	1.64 ± 0.04	0.01 ± 0.01	3.49 ± 0.11	3.61 ± 0.01
Significance	•	•	ns	ns	•
OG6 (99, 99)					
PdeGSTU10	0.49 ± 0.01	3.81 ± 0.32	nd	0.68 ± 0.02	3.12 ± 0.08
PyuGSTU10	0.49 ± 0.01	3.81 ± 0.32	nd	0.68 ± 0.02	3.12 ± 0.08
PtaGSTU10*	nd	0.04 ± 0.01	nd	0.16 ± 0.02	0.56 ± 0.01
Significance	•	•	ns	•	ns
OG8 (75, 99)					
PdeGSTU17	5.32 ± 0.01	0.17 ± 0.01	nd	0.33 ± 0.04	nd
PyuGSTU17	14.78 ± 0.22	0.69 ± 0.01	0.03 ± 0.01	0.91 ± 0.02	nd
PtaGSTU17*	6.17 ± 0.04	1.58 ± 0.01	0.03 ± 0.01	0.51 ± 0.05	0.35 ± 0.01
Significance	•	•	•	•	•
OG11 (100, 100)					
PdeGSTU21	nd	0.64 ± 0.01	nd	1.37 ± 0.07	0.29 ± 0.02
PyuGSTU21	nd	0.77 ± 0.01	nd	0.91 ± 0.04	nd
PtaGSTU21	0.02 ± 0.01	0.52 ± 0.01	nd	1.77 ± 0.09	0.81 ± 0.06
Significance	•	•	ns	•	•
OG12 (98, 85)					
PdeGSTU22	0.13 ± 0.01	0.11 ± 0.01	0.02 ± 0.01	2.01 ± 0.09	1.66 ± 0.01
PtaGSTU22	0.03 ± 0.01	0.03 ± 0.01	nd	0.58 ± 0.03	0.35 ± 0.04
PyuGSTU22	nd	0.02 ± 0.01	nd	0.17 ± 0.01	nd
Significance	•	•	•	•	ns
OG13 (40, 60)					
PdeGSTU23	nd	nd	nd	1.52 ± 0.06	0.29 ± 0.04
PyuGSTU23	nd	nd	nd	1.16 ± 0.05	0.16 ± 0.02
PtaGSTU23	nd	nd	nd	1.75 ± 0.06	nd
Significance	ns	ns	ns	•	ns
OG15 (60, 65)					
PdeGSTU26	0.57 ± 0.01	0.33 ± 0.01	nd	0.35 ± 0.04	nd
PyuGSTU26	nd	0.01 ± 0.01	nd	nd	nd
PtaGSTU26*	0.35 ± 0.01	0.49 ± 0.01	nd	nd	0.45 ± 0.07
Significance	•	•	ns	•	•
OG17 (77, 100)					
PdeGSTF2	0.13 ± 0.01	nd	0.02 ± 0.01	0.73 ± 0.05	0.12 ± 0.03
PyuGSTF2	0.13 ± 0.01	nd	0.02 ± 0.01	0.73 ± 0.05	0.12 ± 0.03
PtaGSTF2	0.14 ± 0.01	nd	nd	2.00 ± 0.06	0.19 ± 0.01
Significance	ns	ns	ns	ns	ns
OG18 (56, 83)					
PdeGSTF3	0.17 ± 0.01	0.83 ± 0.02	nd	3.89 ± 0.06	nd
PyuGSTF3	0.17 ± 0.01	0.83 ± 0.02	nd	3.89 ± 0.06	nd
PtaGSTF3	0.16 ± 0.01	0.92 ± 0.02	nd	4.20 ± 0.01	0.03 ± 0.01
Significance	ns	ns	ns	ns	•

Fig. 3. Enzymatic activities of orthologous group proteins. The phylogenetic trees are taken from Fig. 1. Values shown are mean ± SD, as calculated from three replicates. Data cited from Lan et al. (2013) are indicated by asterisks. To detect whether there were significant differences among the three proteins of each orthologous group for each substrate, the nonparametric Kruskal–Wallis *H* test was performed. ns, not significantly different (*P* > 0.05); •, significantly different (*P* < 0.05).
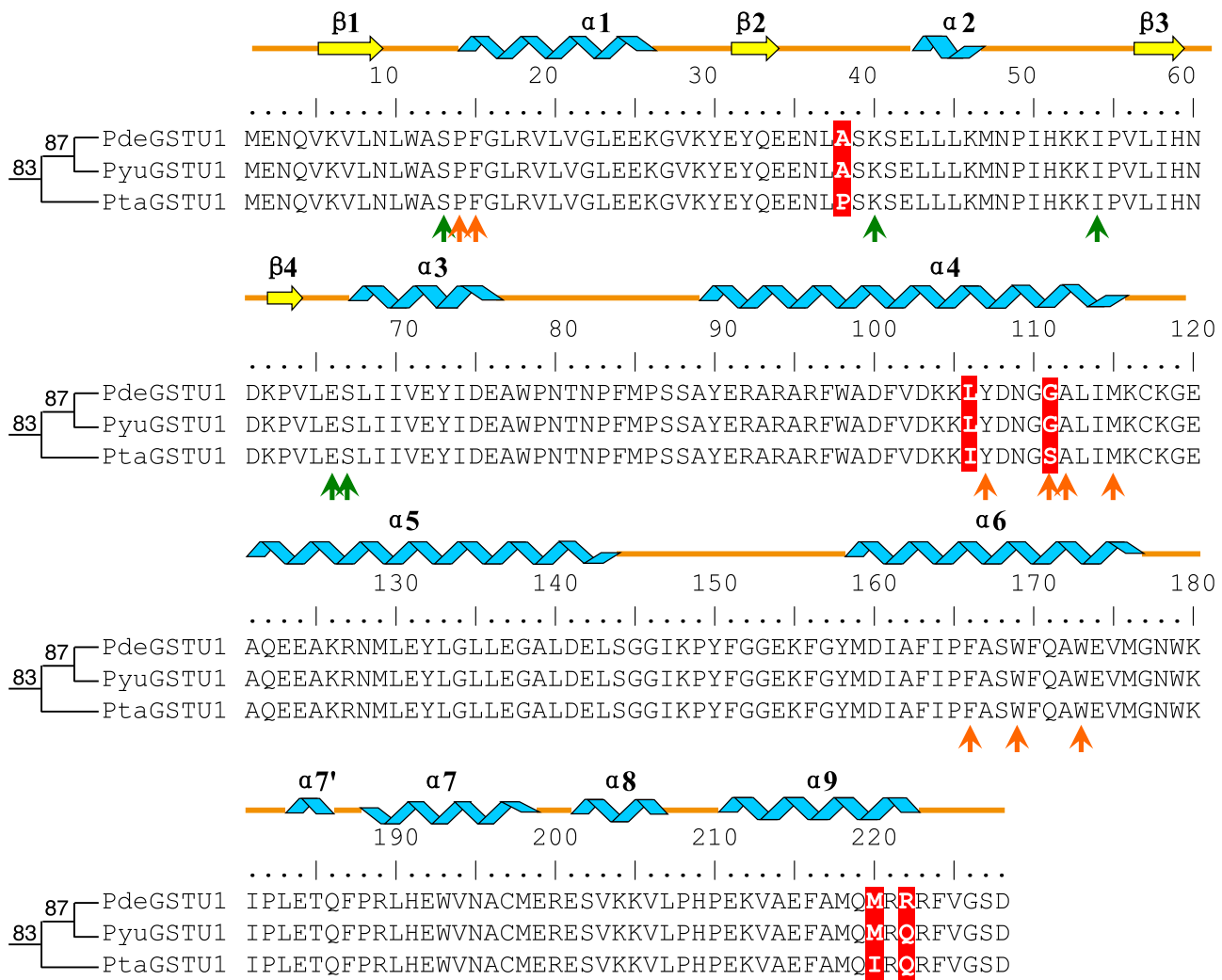
Fig. 4. Sequence alignment of three glutathione *S*-transferase (GST) proteins in OG1 orthologous group. The blue helices and yellow arrows represent α-helices and β-strands, respectively. Different residues among three GST proteins are shaded in red. The green and orange arrows indicate the G-site and H-site residues, respectively.

decreased by 0.92- and 0.93-fold, 0.72- and 0.81-fold, and 0.60- and 0.30-fold relative to the wild-type PdeGSTU1, respectively.

Next, we mutated the amino acid residues in PtaGSTU1 to the corresponding divergent amino acid residues in PdeGSTU1. We constructed and purified five PtaGSTU1 mutants (P38A, I106L, S111G, I220M, and Q222R) (Fig. 5). When the Pro38 residue was replaced by an Ala residue, and Ser111 was replaced by Gly, the resulting P38A and S111G mutants showed 1.35- and 1.65-fold much higher enzymatic activities toward substrate CDNB, 1.92- and 1.65-fold higher enzymatic activities toward substrate NBD-Cl, and 1.23- and 3.05-fold higher enzymatic activities toward substrate Fluorodifen than the wild-type protein PtaGSTU1. These results support the important roles of the two residues in the functional divergence of GST orthologues.

To study the thermal unfolding of wild-type and mutant proteins, DSC studies were conducted (Fig. 6). The transition melting temperature (T_m) of PdeGSTU1 and PtaGSTU1 was 72.4 °C and 66.5 °C, respectively. Compared with wild-type

PdeGSTU1, T_m values of the A38P and G111S mutants decreased significantly. In contrast, the T_m values of the P38A and S111G mutants were significantly increased compared with wild-type PtaGSTU1. These results indicated that the 38th and 111th amino acid sites play an important role in the thermal stability of proteins.

3.6 Kinetic characteristics of PdeGSTU1, PtaGSTU1, and the mutant proteins

To further understand the functional significance of the 38th and 111th amino acid sites, we examined the steady-state kinetic characteristics of PdeGSTU1, PtaGSTU1, and their mutant proteins (Table 1). We observed significant differences in steady-state kinetic constant values for the substrates GSH and CDNB between wild-type PdeGSTU1 and PtaGSTU1. The affinity (1/K_m) and catalytic efficiency (k_{cat}/K_m) for the substrates GSH and CDNB were significantly higher in PdeGSTU1 than that of PtaGSTU1 (Mann–Whitney *U*-test, $P < 0.05$).
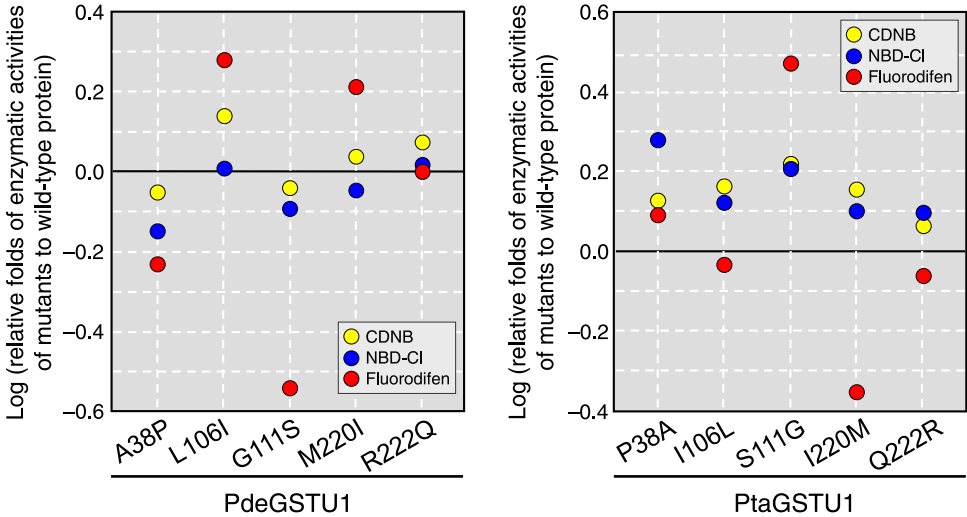
Fig. 5. Changes in enzymatic activities induced by mutations at five residues to the wild-type proteins. The enzymatic activities of PdeGSTU1 and PtaGSTU1 toward each substrate are set as the baselines for comparison with the mutants specified on the X-axis.
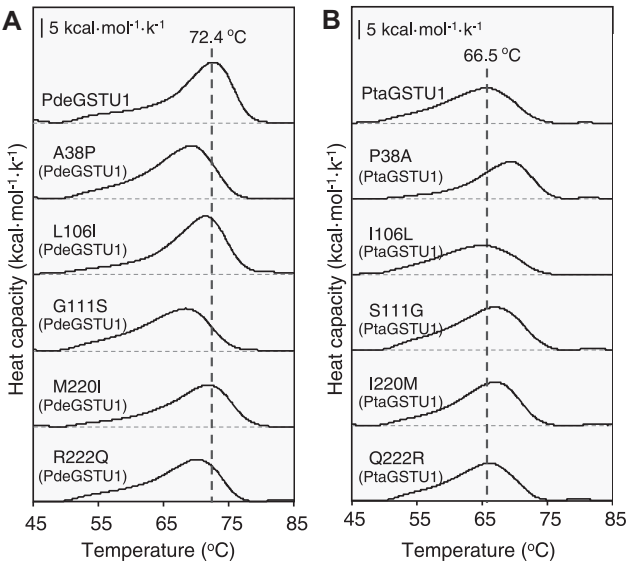
Fig. 6. Thermal denaturation profiles of PdeGSTU1 (**A**) and PtaGSTU1 (**B**) and their mutants.

Table 1 Steady-state kinetic parameters of wild-type proteins (PdeGSTU1 and PtaGSTU1) and their mutants toward substrates GSH and CDNB

Proteins	$1/K_m^{GSH}$ (mM^{-1})	k_{cat}^{GSH} (s^{-1})	$(k_{cat}/K_m)^{GSH}$ (mM^{-1}s^{-1})	$1/K_m^{CDNB}$ (mM^{-1})	k_{cat}^{CDNB} (s^{-1})	$(k_{cat}/K_m)^{CDNB}$ (mM^{-1}s^{-1})
PdeGSTU1	1.14 ± 0.13	12.44	14.32	1.88 ± 0.23	10.31	19.46
A38P (PdeGSTU1)	0.97 ± 0.04	10.62	10.30	1.47 ± 0.18	9.16	13.56
G111S (PdeGSTU1)	1.19 ± 0.17	9.06	10.85	2.28 ± 0.40	6.88	15.89
PtaGSTU1	0.84 ± 0.06	6.90	5.80	1.53 ± 0.20	5.12	7.87
P38A (PtaGSTU1)	1.10 ± 0.04	9.09	9.96	3.79 ± 0.18	5.79	21.94
S111G (PtaGSTU1)	0.90 ± 0.07	8.14	7.34	2.43 ± 0.36	5.46	13.38

CDNB, 1-chloro-2,4-dinitrobenzene; GSH, glutathione.

Compared with the wild-type PdeGSTU1, the affinity and catalytic efficiency of the A38P mutant toward GSH and CDNB were significantly lower (Mann–Whitney U-test, $P < 0.05$). The affinity of the G111S mutant for GSH and CDNB was not significantly different from that of the wild-type (Mann–Whitney U-test, $P > 0.05$), but its catalytic efficiency for the two substrates was 0.76- and 0.82-fold lower, respectively (Mann–Whitney U-test, $P < 0.05$).

The P38A mutant of PtaGSTU1 had an increased affinity and catalytic efficiency for GSH and CDNB (Mann–Whitney U-test, $P < 0.05$), while the S111G mutant showed higher affinity only toward CDNB (Mann–Whitney U-test, $P < 0.05$) but not GSH relative to the wild-type (Mann–Whitney U-test, $P > 0.05$). However, the catalytic efficiency of S111G toward GSH and CDNB was significantly higher (Mann–Whitney U-test, $P < 0.05$).

3.7 Crystal structure of PdeGSTU1 protein
To understand how key amino acid sites affect substrate activities, we investigated the structures of the PdeGSTU1 protein through protein crystallization. Using the co-crystallization method, we obtained crystals of PdeGSTU1 in complex with its substrate GSH. The structure of PdeGSTU1 bound to GSH was determined at 2.19 Å resolution (Table S1). PdeGSTU1 adopted a characteristic GST fold, including two spatially distinct domains: a smaller N-terminal domain (residues 1–78) and a larger C-terminal domain (residues 89–228) (Fig. S5). These two domains were connected by a nine-residue linker (residues 79–88). The N-terminal domain had a topology similar to that of the thioredoxin fold ($\beta\alpha\beta\alpha\beta\beta\alpha$ structure motif), which contained a GSH binding site (G-site). The C-terminal domain consisted only of α-helices, and contained the hydrophobic substrate-binding site (H-site).

Each GST monomer combines one molecule of GSH in the G-site, which is formed by conserved amino acid residues. The correct position and orientation of GSH in the G-site are essential for activating GSH. The crystal structure of PdeGSTU1 showed that the G-site amino acids were Ser13, Lys40, Ile54, Glu66, and Ser67 (Fig. 7A). These five amino acids formed five hydrogen bonds with GSH. The amino acid sites of Ala38 and Arg223 were located at the top of the GSH
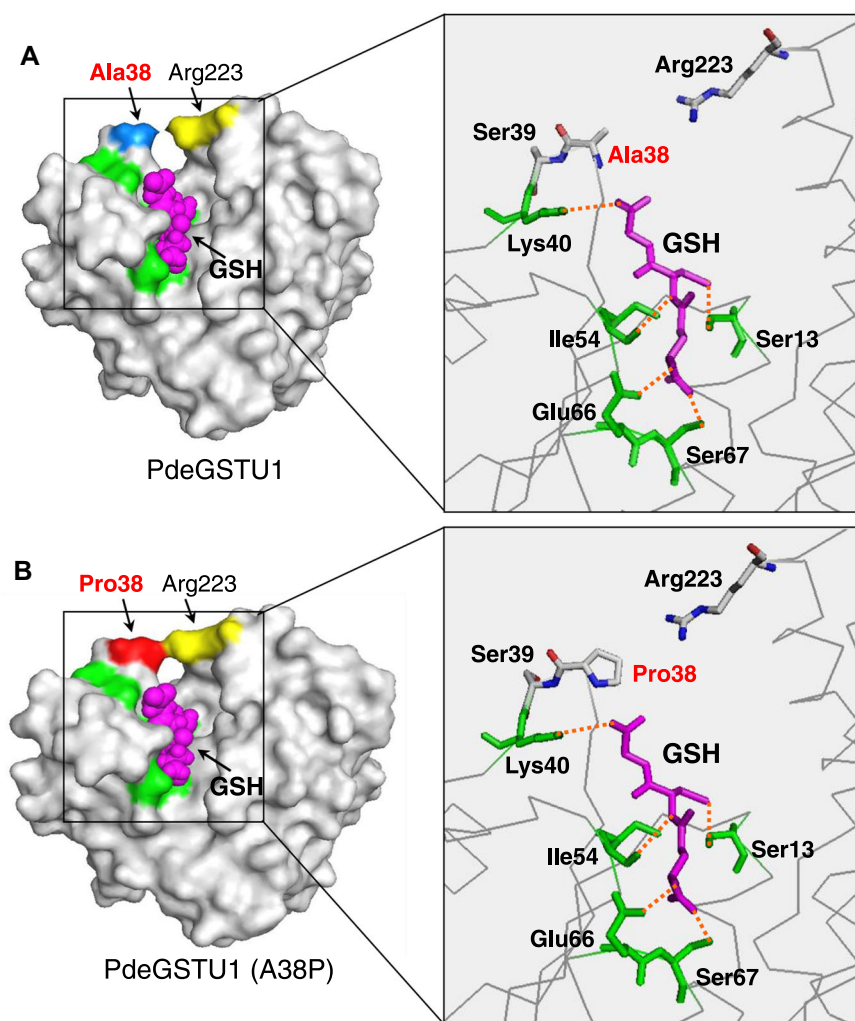
Fig. 7. Protein surface and G-sites of PdeGSTU1 (**A**) and its A38P mutant (**B**). The surface of G-site residues is marked with green and the molecules of G-site residues are indicated by green sticks. Hydrogen bonds are indicated by orange dashed lines.

binding pocket (Fig. 7A). Using the structure of PdeGSTU1 as a template, we modeled the three-dimensional structure of the A38P mutant of PdeGSTU1. Compared with the wild-type, the surfaces of Pro38 and Arg223 in A38P were closer (Fig. 7B), which changed the structure of the GSH binding pocket.

GSTs bind a wide range of hydrophobic substrates in the H-site. The H-site is formed by elements of both N-terminal and C-terminal domains, but most of the H-site amino acids are composed of the C-terminal domain. The H-site region is usually hydrophobic and consists of hydrophobic amino acids. In this study, we did not obtain crystals of PdeGSTU1 co-crystallized with model hydrophobic substrate *S*-hexylglutathione. We, therefore, superimposed the structure of PdeGSTU1 onto the structure of wheat TaGSTU4 (PDB code: 1GWC) to infer the H-site and the location of the hydrophobic pocket of the H-site in PdeGSTU1 (Fig. 8). We found that the main part of the H-site was formed by Pro14, Phe15, Tyr107, Gly111, Ala112, Met115, Phe166, Trp169, and Trp173, with Gly111, Phe166, and Trp173 located at the bottom of the H-site pocket, and Pro14, Phe15, Tyr107, Ala112, Met115, and Trp169 formed most of the walls of the H-site pocket. The Gly111 site of PdeGSTU1 was replaced by Ser111 in PtaGSTU1. The side chain of Ser contained a hydroxyl group, which might reduce the hydrophobicity of the H-site pocket.

4 Discussion

In this study, we identified 19 GST OGs from *Pinus densata* and its putative parental species *Pinus tabuliformis* and *Pinus yunnanensis*. Among these groups, *P. densata* inherited GST genes from *P. yunnanensis* in 17 cases and only two were from *P. tabuliformis*. The biased representation of *P. yunnanensis*-like GSTs in *P. densata* might be due to predominant introgression from *P. yunnanensis* and selective retention driven by ecological adaptation. Previous studies have shown that *P. densata* originated from the eastern margin of its current distribution range and colonized the Tibetan Plateau by westward migrations (Wang et al., 2011; Zhao et al., 2020). During this process, *P. densata* populations were predominantly introgressed with *P. yunnanensis*, resulting in a similar genomic composition between *P. yunnanensis* and central and western populations of *P. densata* (Wang et al., 2011; Gao et al., 2012; Zhao et al., 2020). The population of *P. densata* sampled in this study was from the western range, thus, the detection of *P. yunnanensis*-like copies of *GST* in *P. densata* reflects the generally similar genomic composition between these two species.

Selective inheritance of alleles from parental species can confer the hybrid species' ecotopic biological features (Rieseberg et al., 2003). For example, the homoploid hybrid species *Ostryopsis intermedia* inherited the iron tolerance-related alleles from the parent *Ostryopsis nobilis*, enabling the hybrid to appear on iron-rich soils (Wang et al., 2021). Plant GSTs play major roles in protecting plants against biotic and abiotic stresses and could contribute to the adaptation of plants to different environmental stresses (Nianiou-Obeidat et al., 2017). It is a possibility that *P. yunnanensis*-like GST copies are more favored by natural selection in *P. densata* due to functional advantages. By analyzing the enzymatic characters of 12 groups of orthologous GSTs, we found that the enzymatic activities of *P. densata* GSTs were more similar to those of *P. yunnanensis* with higher enzymatic activities than that of *P. tabuliformis* GSTs. The higher activities of GSTs might confer *Pinus* trees' adaptations to environmental stresses. Many studies have found that natural selection and recombination shaped the evolution of hybrid genomes (Schumer et al., 2018; Wersebe et al., 2022). Our previous studies on the GST gene family of *Populus* and *P. tabuliformis* found that natural selection played an important role in the retention and functional divergence of GST genes (Lan et al., 2009, 2013). Our previous study also detected recombination events among *P. tabuliformis* tau GST genes using Recombination Detection Program (Lan et al., 2013). Thus, it is likely that both natural selection and recombination play an important role in the retention and functional divergence of the three pine GST family.

Sequence variation among orthologues is common, however, inference of functional significance of sequence divergence is challenging in conifers due to the lack of gene manipulation and transformation systems, and their long
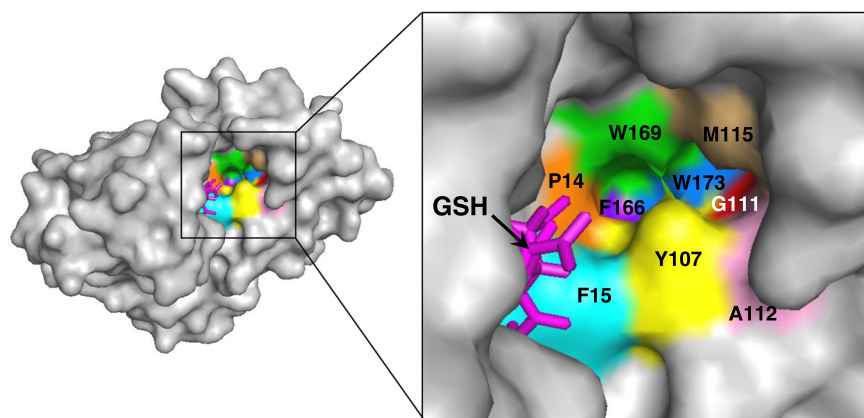
Fig. 8. Hydrophobic cavity of the PdeGSTU1 protein. Residues that constituted the hydrophobic cavity are marked with different colors. Purple sticks represent glutathione.

generation times. In this study, we explored the method of biochemical function determination to study the functional divergence among orthologues. We observed clear divergence in substrate activities among the GST orthologues from three pine species. To verify the causal relationships, we used a site-directed mutagenesis strategy to replace a few selected sites to evaluate the subsequent impact on enzyme activities. All mutants modified the enzymes' properties toward the selected substrates. Particularly the mutation of Ala38 or Gly111 in PdeGSTU1 to the corresponding Pro38 or Ser111 in PtaGSTU1 resulted in a dramatic decrease in enzymatic activity toward CDNB, NBD-Cl, and Fluorodifen. By contrast, mutation of Pro38 or Ser111 in PtaGSTU1 to the corresponding Ala38 or Gly111 in PdeGSTU1 resulted in a significant increase in enzymatic activity (Fig. 5). Our results illustrate that divergence at key amino acid sites plays an important role in the divergence of enzymatic functions.

To understand the GST protein structure, we obtained the crystal structure of PdeGSTU1 in complex with its substrate GSH with a resolution of 2.19 Å. This is the first GST crystal structure in conifer species. Based on the crystal structure of PdeGSTU1, we found that the Ala38 residue was located at the top of the GSH binding pocket (Fig. 7A). When the Ala38 residue of PdeGSTU1 was mutated to Pro38, its methyl side chain was replaced by a pyrrolidine loop. Compared with the wild-type, the surfaces of Pro38 and Arg223 in the A38P mutant were in contact, thus increasing both the steric hindrance and hydrophobicity of the G-site. This conformational change in the G-site might decrease the mutant protein binding ability to hydrophilic GSH. In fact, compared with the wild-type PdeGSTU1, the affinity and catalytic efficiency of the A38P mutant toward GSH were significantly lower. By contrast, the P38A mutant of PtaGSTU1 showed increased affinity and catalytic efficiency for GSH. Thus, steady-state kinetic parameters of wild-type proteins (PdeGSTU1 and PtaGSTU1) and their mutants toward GSH supported our inference about the key function of the Ala38 residue.

The amino acid sequence variation in the H-site region in the GST family is much greater than that in the G-site region (Thom et al., 2002; Lan et al., 2009). This high sequence variation at the H-site enables the GSTs to acquire novel biochemical functions (Frova, 2003). An important feature of the H-site is the presence of a hydrophobic pocket responsible for binding secondary substrates. The Gly111 residue of PdeGSTU1 is located at the bottom of the H-site pocket (Fig. 8). In PtaGSTU1 this site is a Ser instead. The Ser111 residue of PtaGSTU1 had a side chain composed of a hydroxyl group, which might reduce the hydrophobicity of the H-site pocket, resulting in a decreased ability of the H-site pocket to bind hydrophobic secondary substrates. Replacement of Ser111 by Gly resulted in a higher affinity toward CDNB and affected the steady-state kinetic parameters of the protein, illustrating the functional role of this key site.

Acknowledgements

Conflict of Interests

No author had any conflicts of interest associated with any aspect of the manuscript.

Author Contributions

Q.-Y.Z., X.-R.W., and Y.-J.L. designed the research; C.Q. and H.-N.K. performed the experiments; H.X. B.-S.W., Z.-L.Y., and G.-F.L. provided valuable suggestions for the manuscript. Q.-Y.Z. Y.-J.L., and X.-R.W. wrote the article. Q.-Y.Z. and Y.-J.L. agreed to serve as the authors responsible for contacting and ensuring communication. All authors read, revised, and approved the final manuscript.

Data Availability Statement

Sequencing data can be found in GenBank databases under the accession numbers listed in Table S3. The atomic coordinates and structural factors for PdeGSTU1 have been deposited with the Protein Data Bank under accession code 7Y55.

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J, Keller B, Marczewski T, Mallet J, Martinez-Rodriguez P, Möst M, Mullen S, Nichols R, Nolte AW, Parisod C, Pfennig K, Rice AM, Ritchie MG, Seifert B, Smadja CM, Stelkens R, Szymura JM, Väinölä R, Wolf JBW, Zinner D. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26: 229–246.

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.

Dixon DP, Hawkins T, Hussey PJ, Edwards R. 2009. Enzyme activities and subcellular localization of members of the *Arabidopsis* glutathione transferase superfamily. *Journal of Experimental Botany* 60: 1207–1218.

Edwards R, Dixon DP. 2005. Plant glutathione transferases. *Methods in Enzymology* 401: 169–186.

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

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

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

Gao J, Wang B, Mao J-F, Ingvarsson P, Zeng Q-Y, Wang X-R. 2012. Demography and speciation history of the homoploid hybrid pine *Pinus densata* on the Tibetan Plateau. *Molecular Ecology* 21: 4811–4827.

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

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

Hall TA. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.

Kao H-N, Lan T, Wang X-R, Zeng Q-Y. 2012. Functional divergence of dehydroascorbate reductase genes in *Pinus densata, P. tabulaeformis*, and *P. yunnanensis. Chinese Bulletin of Botany* 47: 1–10.

Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* 20: 1160–1166.

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 Evolutionary Biology* 6: 29.

Labrou NE, Papageorgiou AC, Pavli O, Flemetakis E. 2015. Plant GSTome: Structure and functional role in xenome network and plant stress response. *Current Opinion in Biotechnology* 32: 186–194.

Lan T, Wang X-R, Zeng Q-Y. 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 Z-L, Yang X, Liu Y-J, Wang X-R, Zeng Q-Y. 2009. Extensive functional diversification of the *Populus* glutathione S-transferase supergene family. *Plant Cell* 21: 3749–3766.

Liu Y-J, Han X-M, Ren L-L, Yang H-L, Zeng Q-Y. 2013. Functional divergence of the glutathione S-transferase 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. *Plant Cell* 12: 1939–1950.

Mallet J. 2007. Hybrid speciation. *Nature* 446: 279–283.

Mao J-F, Wang X-R. 2011. Distinct niche divergence characterizes the homoploid hybrid speciation of *Pinus densata* on the Tibetan plateau. *The American Naturalist* 177: 424–439.

Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications* 8: 14363.

Nianiou-Obeidat I, Madesis P, Kissoudis C, Voulgari G, Chronopoulou E, Tsaftaris A, Labrou NE. 2017. Plant glutathione transferase-mediated stress tolerance: Functions and biotechnological applications. *Plant Cell Reports* 36: 791–805.

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

Potterton E, Briggs P, Turkenburg M, Dodson E. 2003. A graphical user interface to the CCP4 program suite. *Acta Crystallographica. Section D, Biological Crystallography* 59: 1131–1137.

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.

Rieseberg LH. 1997. Hybrid origins of plant species. *Annual Review of Ecology, Evolution, and Systematics* 28: 359–389.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301: 1211–1216.

Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, Blazier JC, Sankararaman S, Andolfatto P, Rosenthal GG, Przeworski M. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* 360: 656–660.

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.

Wang B-S, Mao J-F, Gao J, Zhao W, Wang X-R. 2011. Colonization of the Tibetan Plateau by the homoploid hybrid pine *Pinus densata. Molecular Ecology* 20: 3796–3811.

Wang X-R, Szmidt AE. 1994. Hybridization and chloroplast DNA variation in a *Pinus* species complex from Asia. *Evolution* 48: 1020–1031.

Wang Z, Jiang Y, Bi H, Lu Z, Ma Y, Yang X, Chen N, Tian B, Liu B, Mao X. 2021. Hybrid speciation via inheritance of alternate alleles of parental isolating genes. *Molecular Plant* 14: 208–222.

Wernersson R, Pedersen AG. 2003. RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Research* 31: 3537–3539.

Wersebe MJ, Sherman RE, Jeyasingh PD, Weider LJ. 2022. The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex. *Molecular Ecology* 32: 1478–1496. https://doi.org/10.1111/mec.16383

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

Zhao W, Meng J, Wang B, Zhang L, Xu Y, Zeng Q-Y, Li Y, Mao J-F, Wang X-R. 2014. Weak crossability barrier but strong juvenile selection supports ecological speciation of the hybrid pine *Pinus densata* on the Tibetan plateau. *Evolution* 68: 3120–3133.

Zhao W, Sun Y-Q, Pan J, Sullivan AR, Arnold ML, Mao J-F, Wang X-R. 2020. Effects of landscapes and range expansion on population structure and local adaptation. *The New Phytologist* 228: 330–343.

Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.12953/suppinfo:

Fig. S1. Phylogenetic tree of glutathione *S*-transferases (GSTs) from *Pinus densata* (*Pde*), *Pinus yunnanensis* (*Pyu*), and *Pinus tabuliformis* (*Pta*). GSTs designated as GSTU, F, T, Z, and L correspond to tau, phi, theta, zeta, and lambda class GSTs, respectively. *Escherichia coli* GRX2 was used as an outgroup.

www.jse.ac.cn

J. Syst. Evol. 62(1): 135–148, 2024

Fig. S2. Phylogenetic relationships of the tau and phi glutathione *S*-transferases (GSTs) from *Pinus densata*, *Pinus yunnanensis*, and *Pinus tabuliformis.* The tree was constructed using complementary DNA sequences. Numbers on branches indicated the bootstrap values calculated from 1000 bootstrap replicates. The orthologous groups (OG1–OG19) identified in this study were marked with curly brackets. The tau and phi GSTs were colored in blue and orange, respectively.

Fig. S3. Enzymatic activities of glutathione *S*-transferase (GST) proteins from three *Pinus* species. The phylogenetic trees are from Fig. 1. Values shown are mean $\pm$ SD, as calculated from three replicates. Data cited from Lan et al. (2013) are indicated with asterisks. Each GST name is followed by a code in parentheses describing the associated analysis: A, purified GST protein assayed; I, recombinant protein totally insoluble. To detect whether there were significant differences among the two proteins of each orthologous group for each substrate, the nonparametric Kruskal–Wallis *H* test was performed. ns, not significantly different ($P > 0.05$); •, significantly different ($P < 0.05$).

Fig. S4. Pairwise comparison of the enzymatic activities of the three proteins in each orthologous group. For each comparison, the value with high enzymatic activity was divided by the value with low enzymatic activity.

PdeGST–PyuGST, PdeGST–PtaGST, and PyuGST–PtaGST represent the pairwise comparison between *Pinus densata* and *Pinus yunnanensis*, between *Pinus densata* and *Pinus tabuliformis*, and between *P. yunnanensis* and *P. tabuliformis*, respectively. To detect whether there was significance for each comparison, the Mann–Whitney *U*-test was performed. **Extremely significant difference ($P < 0.001$).

Fig. S5. A cartoon representation of the PdeGSTU1 monomer. The blue helices and yellow arrows represented α-helices and β-strands, respectively. Purple sticks represent glutathione.

Table S1. Data collection and structure refinement statistics of PdeGSTU1.

Table S2. Polymerase chain reaction primers used to clone glutathione *S*-transferase genes of *Pinus densata* and *Pinus yunnanensis*.

Table S3. The CDS and protein sequences of glutathione *S*-transferases from *Pinus densata* and *Pinus yunnanensis*.

Table S4. Polymerase chain reaction primers used to detect the expression patterns of glutathione *S*-transferase genes from *Pinus densata* and *Pinus yunnanensis*.

Table S5. Primers used to construct glutathione *S*-transferase protein expression vectors.

Table S6. Primers used to construct the expression vectors of mutant proteins.