Structural and Functional Evolution of Positively Selected Sites in Pine Glutathione S-Transferase Enzyme Family*[S]

Ting Lan[‡], Xiao-Ru Wang[§], and Qing-Yin Zeng[‡1]

From the [‡]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China and [§]Department of Ecology and Environmental Science, Umeå University, SE-901 87 Umeå, Sweden

Background: The functional significance of positive selection in enzyme families is largely unknown.
Results: The structural and biochemical properties of positively selected sites in pine GST enzyme family were characterized.
Conclusion: The positively selected sites significantly affect enzyme activity and specificity.
Significance: This study sheds light on the selective aspects of the functional and structural divergence of plant enzyme family.

Phylogenetic analyses have identified positive selection as an important driver of protein evolution, both structural and functional. However, the lack of appropriate combined functional and structural assays has generally hindered attempts to elucidate patterns of positively selected sites and their effects on enzyme activity and substrate specificity. In this study we investigated the evolutionary divergence of the glutathione S-transferase (GST) family in *Pinus tabuliformis*, a pine that is widely distributed from northern to central China, including cold temperate and drought-stressed regions. GSTs play important roles in plant stress tolerance and detoxification. We cloned 44 GST genes from *P. tabuliformis* and found that 26 of the 44 belong to the largest (Tau) class of GSTs and are differentially expressed across tissues and developmental stages. Substitution models identified five positively selected sites in the Tau GSTs. To examine the functional significance of these positively selected sites, we applied protein structural modeling and site-directed mutagenesis. We found that four of the five positively selected sites significantly affect the enzyme activity and specificity; thus their variation broadens the GST family substrate spectrum. In addition, positive selection has mainly acted on secondary substrate binding sites or sites close to (but not directly at) the primary substrate binding site; thus their variation enables the acquisition of new catalytic functions without compromising the protein primary biochemical properties. Our study sheds light on selective aspects of the functional and structural divergence of the GST family in pine and other organisms.

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. The enzyme family is a group of enzymes with similar amino acid sequence and protein folds encoded by a group of genes created by gene duplication. Numerous studies (1–3) have shown that different members of enzyme families have distinct substrate specificities or catalytic activities, which are shaped by selective pressures during the course of evolution (2, 4, 5). Phylogeny-based analyses of codon substitution have shown that many protein families have been subject to positive selection (6–8), and positive selection is now recognized as an important driver of protein evolution (9–11). However, the lack of appropriate combined functional and structural assays has generally hindered attempts to elucidate patterns of positively selected sites and their effects on enzyme activity and substrate specificity.

Glutathione S-transferases (EC 2.5.1.18) comprise a large, ubiquitous gene family of multifunctional proteins that mainly catalyze the conjugation of intracellular glutathione (GSH) to a wide variety of electrophilic, cytotoxic, and genotoxic molecules of endogenous or exogenous origin. Detoxification of xenobiotics is considered to be the main function of plant GSTs, but other functions include protecting cells from diverse biotic and abiotic stressors, such as pathogens, heavy metal toxins, oxidative agents, and UV radiation (12–15). Because of the diversity of potential xenobiotics and stressors, functional divergence of this enzyme family has major adaptive significance. Accordingly, the supergene GST family displays extensive functional diversification in gene expression, enzymatic activities, and substrate specificities (2, 3). However, although much is known about the biochemistry of GST, little attention has been paid to the patterns of selective pressures that have directed its evolutionary divergence and acquisition of novel activities.

The molecular mechanisms that create structural modifications of protein copies and the subsequent evolutionary genetic processes that fix mutations, alter functions, and improve adaptation should differ among plant lineages due to differences in selective pressures associated with differences in their life history traits. Gymnosperms are a large group of plants with a long evolutionary history. Many are important coniferous components of northern ecosystems and valuable forestry trees. Unlike annual plants, the perennial conifers have long generation times, large effective population sizes, and wide distributions over heterogeneous mountainous regions. These life history traits could theoretically favor mutations that facilitate adaptation to changing environments. However, due to the lack of genome information, knowledge of the functional, structural, and selective aspects of natural enzyme family evolution in conifers is limited. In this study we used a combined evolu-

tionary and experimental approach to elucidate evolution and selection in the GST gene family in *Pinus tabuliformis*, a pine that is widely distributed from northern to central China including cold temperate and drought regions. It grows from 0 to 2700 m above sea level and forms extensive forests (16). Through analyses of substitution and gene expression patterns, enzyme assays, site-directed mutagenesis, and structural comparisons, we examined 1) the selective context and functional divergence in the GST family in the pine and 2) the structural and functional importance of positively selected sites.

EXPERIMENTAL PROCEDURES

Molecular Cloning and Nomenclature—Primers, based on putative *Pinus taeda* GST gene sequences identified from the GenBank™ EST database and listed in supplemental Table S1 were designed to amplify mRNA and genomic DNA of *P. tabuliformis*. Total RNA was isolated from fresh needles of *P. tabuliformis* using an Aurum Total RNA kit (Bio-Rad), and first-strand cDNA was synthesized using a TaKaRa RNA PCR kit (avian myeloblastosis virus) Version 3.0 (TaKaRa). PCR was performed in a 25-μl reaction mixture containing 1–2 μg of template DNA, 0.1 units of TaqDNA polymerase (Invitrogen), 200 μM each of dNTP (Invitrogen), 1.5 mM MgCl$_2$, and 10 pmol of each primer. Optimized PCR conditions consisted of 3-min initial denaturation at 94 °C, then 35 cycles of 30 s at 94 °C, 40 s at 55 °C, and 60 s at 72 °C followed by a 3-min final extension step at 72 °C. PCR products were separated and recovered from a 1% agarose gel using a GFX PCR DNA and Gel Band Purification kit (Amersham Biosciences), then cloned into the pGEM-T Easy Vector (Promega) and sequenced in both directions to verify the gene sequences and structures. The sequences were further analyzed by the NCBI conserved domain database (www.ncbi.nlm.nih.gov) to confirm the presence of typical GST N- and C-terminal domains in the putative encoded proteins.

Following Edwards *et al.* (17), we assigned a univocal name to each *P. tabuliformis* GST gene (supplemental Table S2), consisting of *Pta* followed by a letter denoting the subfamily class (*e.g.* GSTU, -F, -T, -Z, and -L corresponding to Tau, Phi, Theta, Zeta and Lambda, respectively) then a progressive number for each gene of that class (*e.g. PtaGSTU1*).

Expression of GST Genes in P. tabuliformis Tissues—To investigate expression patterns of GST gene family members during growth under normal conditions, we sampled five mature *P. tabuliformis* trees in the Beijing Botanical Garden and five 3-month-old *P. tabuliformis* seedlings grown in a greenhouse. We isolated total RNA from top bud, needle, root tip, and phloem (stem and root) tissues of each mature *P. tabuliformis* tree and young needle, stem, and root tissues of each seedling. We then amplified GST sequences in the samples using PCR, as described above, and primers listed in supplemental Table S3.

Phylogenetic and Molecular Evolution Analyses—The GST protein sequences were aligned using the MUSCLE program (18), with subsequent manual adjustment using BioEdit (19). A phylogenetic tree of *Pinus* GSTs was then reconstructed using a maximum likelihood (ML) procedure by PHYML (20) with the JTT (Jones, Taylor, and Thornton) amino acid substitution model. For analysis of Tau-class GSTs, members of the sister class Lambda of GSTs were used as an outgroup. One thousand

bootstrap replicates were performed in each analysis to calculate the support for identified clades and nodes.

We identified two major clades of *P. tabuliformis* Tau GSTs (Fig. 1*A*). To evaluate variation in selective pressure between them, we used CODEML branch models in PAML (21) to estimate the ratio of nonsynonymous *versus* synonymous substitutions ($\omega = d_N/d_S$) under two *a priori* assumptions: a one-ratio model in which one ω value was assumed for the entire tree and a two-ratio model in which ω values were allowed to vary between the two major clades (Fig. 1*A*). To verify which of the models best fitted the data, likelihood ratio tests were applied by comparing twice the difference in log-likelihood values between pairs of the models using a χ^2 distribution, with the degrees of freedom equal to the differences in the number of parameters between the models (22). The branch model tests indicated that selective pressures significantly differed between the two major clades of *P. tabuliformis* Tau GSTs and that one clade (*clade A* in Fig. 1*A*) has undergone relaxed negative selection. Thus, we used two branch-site models (A0 and A) to test whether positive selection has occurred at some amino acid sites in clade A. Clades A and B were assigned as foreground and background branches, respectively, and Model A allowed ω to vary among sites in foreground branches, whereas Model A0 did not allow for positive selection; instead ω was fixed at a value of 1. The positive selection (A) and neutral (A0) branch-site models were evaluated using likelihood ratio tests, then the Bayes Empirical Bayes method (23) was applied to the results from Model A to identify candidate positively selected sites.

Expression and Purification of GST Proteins—To investigate the enzymatic functions of *P. tabuliformis* Tau GST proteins, the *P. tabuliformis* Tau GST genes were subcloned into pET30a expression vectors (Novagen) to obtain an N-terminal His$_6$ tag. The primers used to construct the GST expression vectors are listed in supplemental Table S4. Colonies containing appropriate inserts were identified by sequencing.

The likelihood ratio test and Bayes Empirical Bayes analysis identified five positive-selection sites with posterior probabilities >0.99. To investigate whether these sites are important for catalysis, we selected protein PtaGSTU17 as a model, subcloned *PtaGSTU17* cDNA into protein vector pET30a (Novagen), then performed site-directed mutagenesis using the pET30a/PtGSTU17 plasmid as a template, the mutagenesis primers listed in supplemental Table S4, and methods described by Zeng and Wang (24). The resultant plasmids were used to transform *Escherichia coli* BL21 and verified by sequencing. Expression and purification of the wild-type and mutant PtaGSTU17 proteins followed the procedures presented by Zeng *et al.* (25).

Enzyme Activity and (1-Anilino-8-naphthalene-sulfonate) (ANS)[2] Binding Assays—To assess the specificity of the wild-type and mutant GST enzymes, we used the following six substrates: 1-chloro-2,4-dinitrobenzene (CDNB), 1,2-dichloro-4-nitrobenzene (DCNB) and 4-nitrobenzyl chloride (NBC), as described by Habig *et al.* (26); 7-chloro-4-nitrobenzo-2-oxa-

---

[2] The abbreviations used are: ANS, 1-anilino-8-naphthalene-sulfonate; CDNB, 1-chloro-2,4-dinitrobenzene; DCNB, 1,2-dichloro-4-nitrobenzene; NBC, 4-nitrobenzyl chloride; NBD-Cl, 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole; Cum-OOH, cumene hydroperoxide; MRPP, multiple response permutation procedure.

1,3-diazole (NBD-Cl), as described by Ricci *et al.* (27); and the diphenyl ethers (Fluorodifen) and cumene hydroperoxide (Cum-OOH), as described by Edwards and Dixon (28). These substrates were frequently used to determine the enzymatic characteristics of plant GSTs (28). All assays were carried out at 25 °C. Protein concentrations in the enzyme preparations were determined by measuring their absorbance at 280 nm.

The fluorescent dye ANS is a sensitive probe for detecting and analyzing structural changes in proteins (29). To detect protein-folding differences between the mutant and wild-type enzymes, we obtained fluorescence measurements using a HITACHI F-4500 FL spectrophotometer after adding 100 μl of 2 mM ANS to a final concentration of 0.05 mg/ml enzyme (sodium phosphate buffer, pH 7.4) in 1-ml reaction mixtures. A total of three scans each for blank and sample were collected and averaged for each enzyme. Reported spectra are the means from at least three independent experiments.

Homology Modeling—Three-dimensional structures of *P. tabuliformis* Tau GSTs were constructed using the InsightII software package (Accelrys) and x-ray structure of a soybean Tau GST (Protein Data Bank accession number 2VO4) as template. The *P. tabuliformis* Tau GSTs showed 34–50% of protein sequence identity to the template protein. Initially, the sequences were aligned using the Align 2D program of the InsightII package Homology Module, structures were then automatically built using the Modeler Module, and the resulting models were evaluated by the profile-3D program. Substrate binding sites are detected as cavities present on the surface of the protein. In this study the binding sites were identified by grid analysis using the Binding-site Analysis module in InsightII.

RESULTS

Pinus GST Gene Family Contains Many Diverse Members—Based on the *P. taeda* EST database, we cloned 44 full-length genes encoding putative GST proteins from *P. tabuliformis* (supplemental Table S2). This list is most likely incomplete due to gaps in the currently available genome information. The predicted proteins encoded by these 44 genes were initially classified based on the NCBI conserved domain analysis, which divided them into eight classes: Tau, Phi, Theta, Zeta, DHAR, Lambda, TCHQD and EF1Bγ. The Tau GSTs were most numerous (26 members) followed by the Phi and Lambda classes (seven and three, respectively), then Zeta, DHAR, and EF1Bγ classes (two members each), and finally Theta and TCHQD classes (just one member each). Phylogenetic analysis of these *P. tabuliformis* GSTs with the GST family from *Populus trichocarpa* showed that GSTs of each subfamily class from both genomes grouped as a single clade with high bootstrap support (data not shown), which further supports the subfamily designations among the 44 *P. tabuliformis* GSTs.

Expression Patterns of GST Gene Family in Pine—The expression patterns of 44 *P. tabuliformis* GST genes were examined in the radicle, seedlings, and mature tree in normal conditions. Seven Phi GSTs and all minor classes GSTs (*PtaGSTL1*, -2, and -3, *PtaGSTZ1* and -2, *PtaGSTT1*, *PtaDHAR1* and -2, *PtaEF1Bγ1* and -2, and *PtaTCHQD1*) were expressed in all tissues or organs examined (Fig. 1*B*). But expression divergences were observed among the Tau GSTs. The phylogenetic

tree showed the 26 Tau GSTs were divided into distinct two clades (*clades A* and *B*, Fig. 1*A*). The six genes (*PtaGSTU1*, -14, -25, -24, -17, and -26) in clade B were expressed in all of the tissues examined. In contrast, of the 20 Tau GSTs in clade A, only nine were expressed in all tissues. Another two (*PtaGSTU11* and -19) were not detectably expressed in any tissue under normal growth and development conditions, possibly because they are expressed at sub-detectable levels. Alternatively, they may only be induced in response to treatments and/or in tissues not examined in our study, or they may be pseudogenes. The other nine Tau GSTs in clade A were selectively expressed in specific tissues (Fig. 1*B*). These results indicate that functional divergence is more pronounced among the Tau GST genes in clade A than in clade B.

Molecular Evolution of Pinus Tau GSTs—Tau class GSTs were the most numerous in plant GST gene family. Our previous study showed the moss *Physcomitrella patens* did not contain any Tau GST, whereas the lycophyte *Selaginella moellendorffii* had 47 Tau GSTs (30). The GST gene families in *Oryza sativa*, *Arabidopsis thaliana*, and *P. trichocarpa* have been previously characterized (2, 31, 32), and they reportedly contain 52, 28, and 58 Tau members, respectively. Our search of the NCBI EST database identified 35 Tau GSTs from four additional gymnosperm species: *P. taeda*, *Gnetum gnemon*, *Ginkgo biloba*, and *Cycas rumphii* (supplemental Table S5). Genome-wide phylogenetic analysis of Tau GST genes in the above nine land plant species revealed two major clades (*A* and *B*, Fig. 2). Clade A contained Tau GSTs of eight seed plant species but not lycophyte, indicating that this group of GSTs might have been lost in lycophyte (Figs. 2 and 3). The 20 *P. tabuliformis* Tau GSTs of clade A in Fig. 1*A* grouped with other gymnosperms and formed a separate group (*A2b*, Fig. 2) that was diverged from angiosperm GSTs. The clade B can be divided into two major groups, B1 and B2. B1 contained GSTs from all species, whereas B2 contained only lycophyte and gymnosperm GSTs (Fig. 2). This phylogeny suggests that the B2 GSTs might have been lost in angiosperms (Fig. 3). All the six *P. tabuliformis* Tau GSTs of clade B in Fig. 1*A* were placed in clade B of Fig. 2.

To test for significant differences in selective pressures between clades A and B in Fig. 1*A*, we performed a two-ratio branch model test using PAML, in which different ω ($= d_N/d_S$) values were assigned to the two clades (Table 1). The log likelihood values under the one-ratio and two-ratio models were ln $L = -9394.796577$ and ln $L = -9378.129047$, respectively. A likelihood ratio test indicated that the one-ratio model should be rejected ($p < 0.001$); hence, selective pressures differed between the two clades. The mean ω values for clades A and B were 0.43 and 0.20, respectively, indicating that clade B has been under stronger purifying selection, whereas some amino acid changes in clade A might have been preserved by positive selection. To test this hypothesis, we applied a branch-site test (33) to identify target sites that are potentially under positive selection in clade A (Table 1). For this, we assigned clades A and B as foreground and background branches, respectively, and obtained log likelihood values under the positive selection model A and the null model A0 of ln $L = -9186.073757$ and ln $L = -9188.425166$, respectively. The likelihood ratio test indicated that the null model should be rejected ($p < 0.03$), corrob-
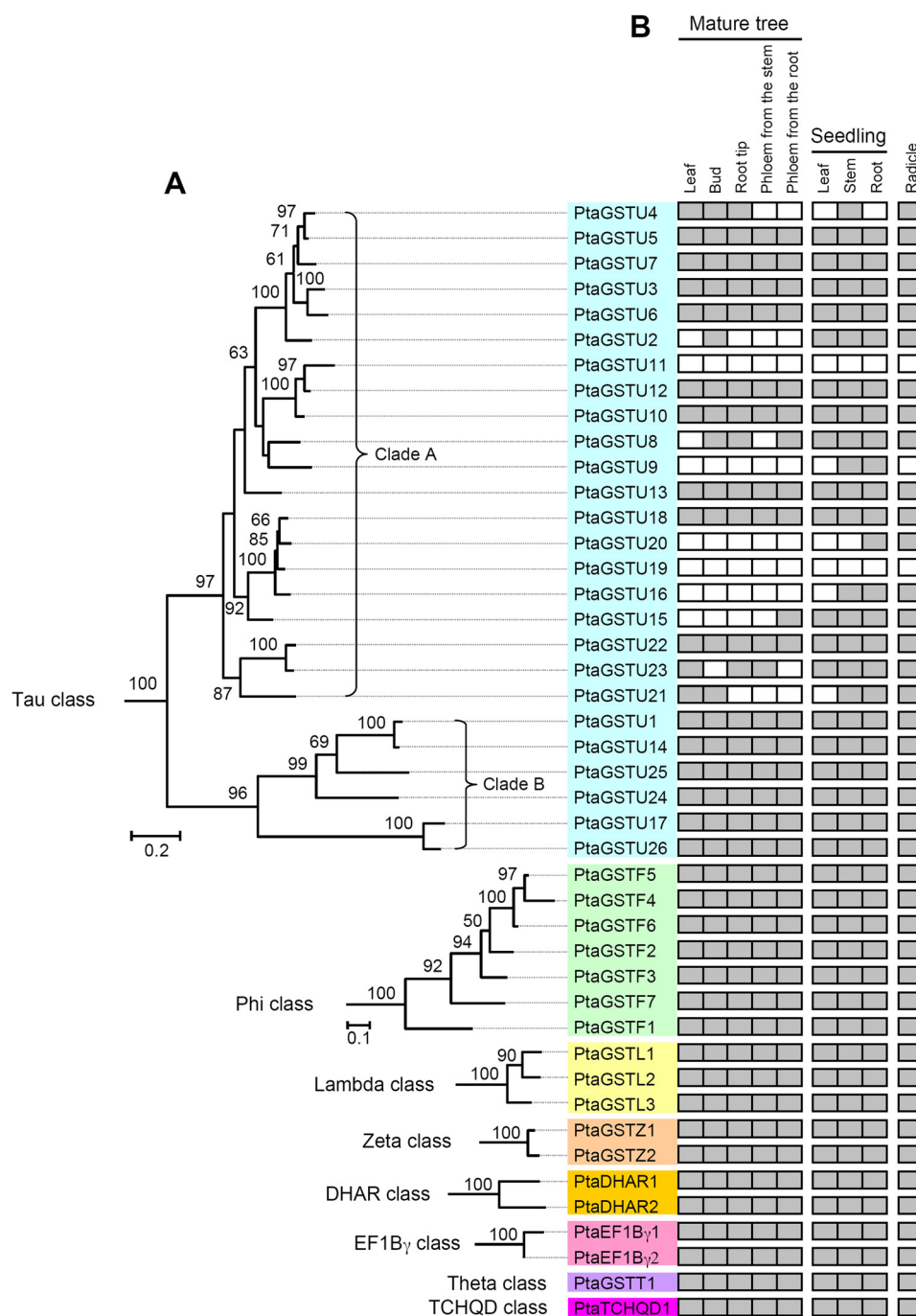
FIGURE 1. **Phylogenetic relationships of *P. tabuliformis* GSTs (*A*) and their expression patterns (*B*).** *Numbers at each node* in the phylogenetic tree signify bootstrap values, and only bootstrap values >50% are shown. GST genes designated GSTU, -F, -T, -Z, and -L correspond to Tau, Phi, Theta, Zeta, and Lambda class GSTs, respectively. The *gray box* indicates positive detection of gene expression in the corresponding tissue under normal growth conditions.

orating the hypothesis that some amino acid sites in clade A have been under positive selection. Further analysis using a Bayes empirical Bayes procedure identified five sites (alignment positions 17, 43, 87, 129, and 183; Fig. 4) that are under positive selection with posterior probabilities >0.99. The two clades differed markedly in amino acid composition at these five sites, and at two of the five sites they shared no common residues (Fig. 4). Further PAML evaluation by including Tau GSTs of *P. taeda* also supported the positive selection on these five sites (data not shown).

Gene conversion or recombination can contribute to sequence divergence in some gene families (34). In this study we used GENECONV (35) and the Recombination Detection Program (RDP) (36) to detect gene conversion and recombination events among *P. tabuliformis* Tau GST genes. Only two gene conversion events and two recombination events were detected among all pairwise GST gene comparisons (data not shown). This result suggests that gene conversion and recombination do not seem to have played significant roles in GST evolution in pine. Similar results are also observed in other gene families, *e.g. Arabidopsis* MADS-box and vertebrates arylamine *N*-acetyltransferase gene families (37, 38).
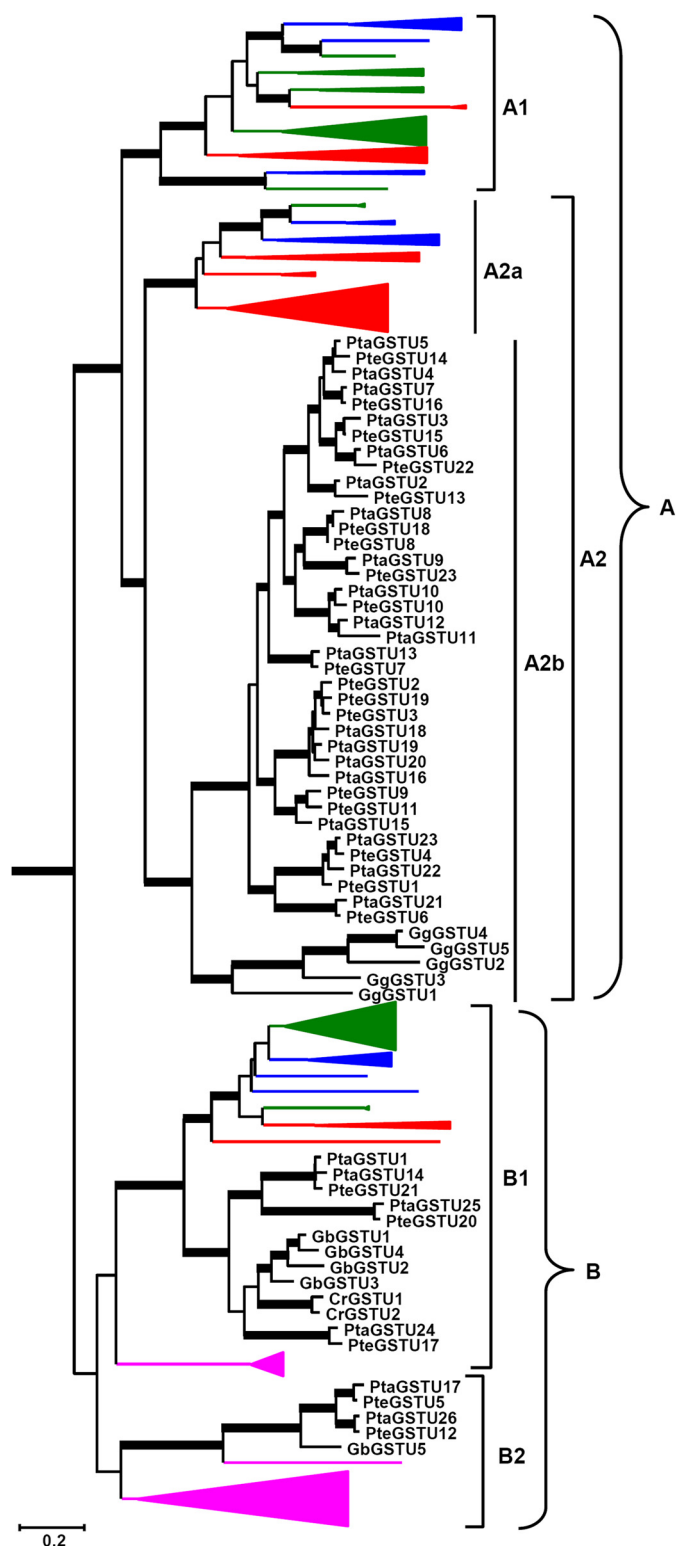
FIGURE 2. **Phylogenetic tree of 244 Tau GSTs from *S. moellendorffii* (*purple triangle*), *O. sativa* (*red triangle*), *A. thaliana* (*blue triangle*), *P. trichocarpa* (*green triangle*), *P. tabuliformis* (*Pta*), *P. taeda* (*Pte*), *G. gnemon* (*Gg*), *G. biloba* (*Gb*), and *C. rumphii* (*Cr*).** Thick branches indicate clades with strong statistical support (>50%). The protein sequences used for the tree reconstruction are provided in supplemental Table S5.
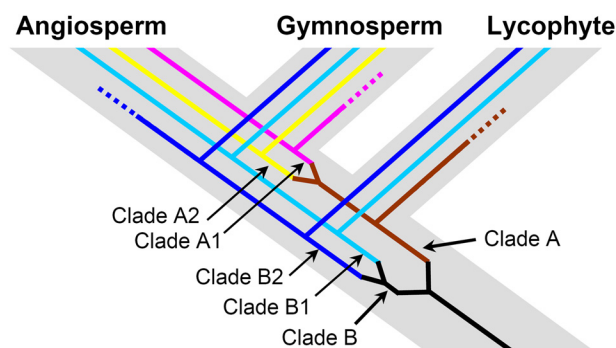
FIGURE 3. **Schematic representation of the duplication history of Tau GSTs in land plants.** Gene loss events are indicated by *dashed lines*. The designation of each clade corresponds to the Fig. 2.

Tau GST (PDB accession number 2VO4) as reference. The simulated structures of the 26 Tau GSTs were superimposed to evaluate the goodness of fit of the overall topologies (Fig. 5*A*). This examination showed that all members shared the same conformation of the structural elements of α-helices and β-sheets, but structural modifications are present in loop regions. Among the five putative positively selected sites, four are located in loop regions and one in an α-helix of the C-terminal domain (Figs. 5*A* and 4).

GST proteins typically have two catalytically active sites: the prime GSH substrate binding site (G-site) and a secondary hydrophobic substrate binding site (H-site). The G-site within each GST class is formed by a group of highly conserved amino acid residues in the N-terminal domain of the protein, and the H-site is formed by residues in the C-terminal domain that are generally less conserved (32). Using the x-ray structures of soybean Tau GST as references, we identified five sites (alignment positions 19, 48, 62, 74, and 75 in Fig. 4; *pink arrows* in Fig. 5*A*) in *P. tabuliformis* Tau GSTs that correspond to G-site residues. To identify H-site residues, we used PtaGSTU17 as a model to predict possible ligand binding cavities in the protein by the Binding Site Analysis program in InsightII. One large cavity located between the N- and C-terminal domains was identified as the substrate binding pocket (*blue dots*, Fig. 5*B*). Residues Lys-119 and Trp-171 of PtaGSTU17 are located at the entrance and on the wall of this hydrophobic substrate-binding pocket, respectively (Fig. 5*B*), indicating that these two sites are likely H-site residues.

Among the five putative positively selected sites, two (alignment positions 17 and 43, corresponding to Arg-12 and Asn-37 of PtaGSTU17; Figs. 4 and 5*A*) are adjacent to the G-site in the N-terminal domain. Two others (alignment positions 129 and 183, corresponding to Lys-119 and Trp-171 of PtaGSTU17; Figs. 4 and 5*A*) have predicted H-site locations, and one site (alignment position 87, corresponding to Pro-80 of PtaGSTU17; Figs. 4 and 5*A*) is in the linker region between the two domains and distant from both active centers.

Substrate Specificities and Activities of Pinus Tau GSTs— Among 44 *P. tabuliformis* GSTs, the Tau GSTs were the most numerous. Seventeen Tau GSTs with differing evolutionary relationships were selected for protein expression and purification. All of these GSTs were expressed as soluble proteins in *E. coli* (Table 2), except for four Tau GSTs (PtaGSTU18, -22,

*Structural Distribution of the Putative Positively Selected Sites—*We modeled the three-dimensional structures of the 26 *P. tabuliformis* Tau GSTs using x-ray structure of a soybean

TABLE 1
Results of CODEML analyses of selective pattern for *P. tabuliformis* Tau GSTs

Model	Estimates of parameters[a]	ln L	χ^2	P	Positively selected sites[b]
Branch model					
One ratio	$\omega = 0.35157$ for all branches	−9394.796577			
Two ratios	$\omega_0 = 0.20429$ for clade B	−9378.129047	33.33	<0.001	
	$\omega_1 = 0.43029$ for clade A				
Branch site model					
Mode A0	Clade B (background)	−9188.425166			Not allowed
	$\omega_0 = 0.16225$ ($p_0 = 0.80589$)				
	$\omega_1 = 1.00000$ ($p_1 = 0.19411$)				
	Clade A (foreground)				
	$\omega_0 = 0.16225$ ($p_0 = 0.63453$)				
	$\omega_1 = 1.00000$ ($p_1 = 0.36547$)				
Mode A	Clade B (background)	−9186.073757	4.70	<0.03	17[c],43[c],87[c],129[c],183[c]
	$\omega_0 = 0.18336$ ($p_0 = 0.77377$)				
	$\omega_1 = 1.00000$ ($p_1 = 0.22623$)				
	Clade A (foreground)				
	$\omega_0 = 0.18336$ ($p_0 = 0.66553$)				
	$\omega_1 = 1.00000$ ($p_1 = 0.19459$)				
	$\omega_2 = 1.57576$ ($p_2 = 0.13988$)				

[a] The proportion of sites (p_0, p_1, . . .) estimated to have ω_0, ω_1, ...
[b] The residues were identified by Bayes Empirical Bayes analysis (23); numbers correspond to their alignment positions in Fig. 4.
[c] Posterior probability >99% of having $\omega > 1$. Clade A and clade B refer to the two clades of Tau class GSTs shown in Fig. 1*A*.
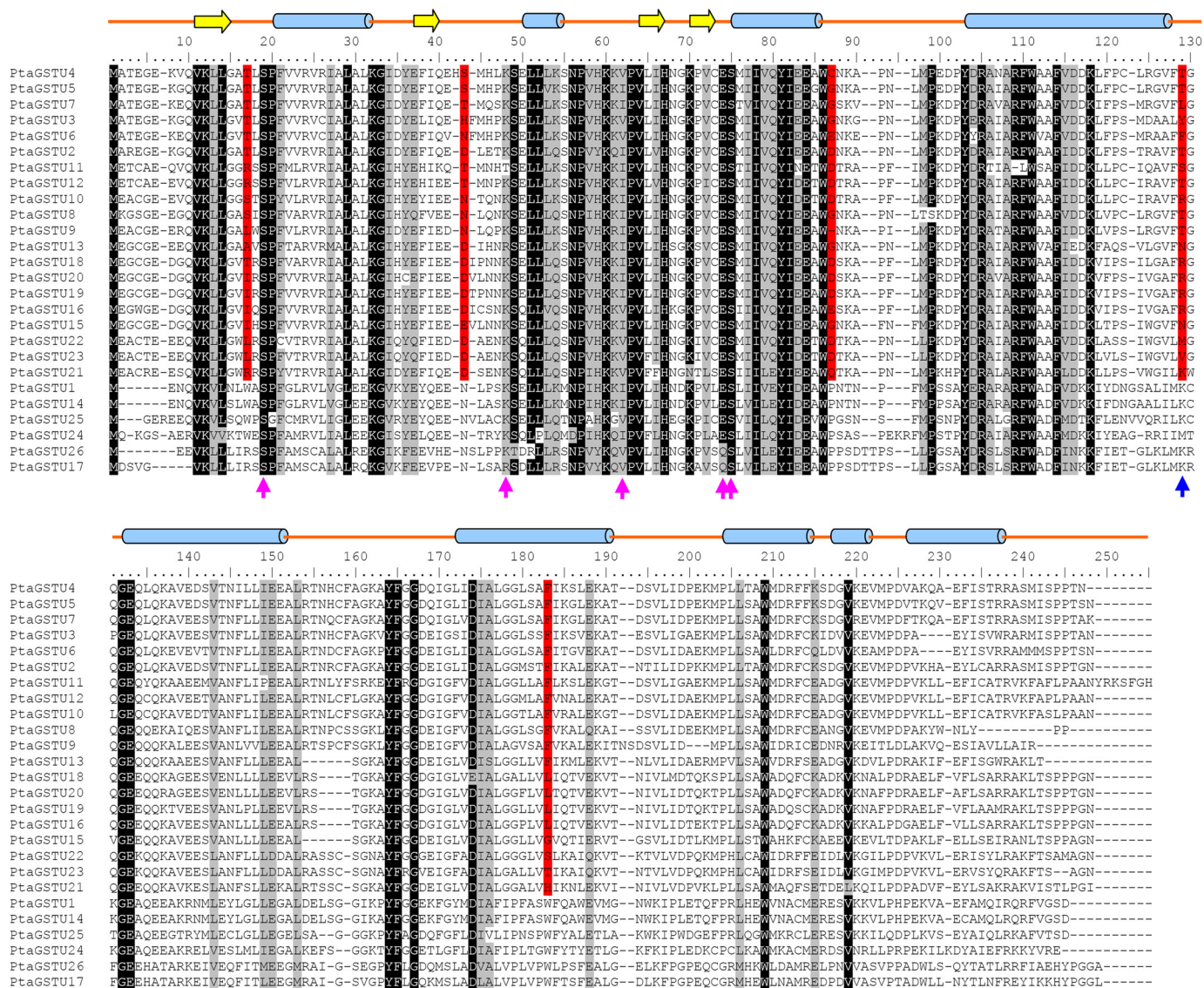
FIGURE 4. **Sequence alignment of *P. tabuliformis* Tau GSTs and predicted secondary structure elements.** α-Helices and β strands are represented as *blue cylinders* and *yellow arrows*, respectively. Residues conserved in all *P. tabuliformis* Tau GSTs are shaded *gray* and *black*. The five positive selection sites predicted by the branch-site model test are colored *red*. Predicted G-sites and H-sites are marked with *pink* and *blue arrows*, respectively.
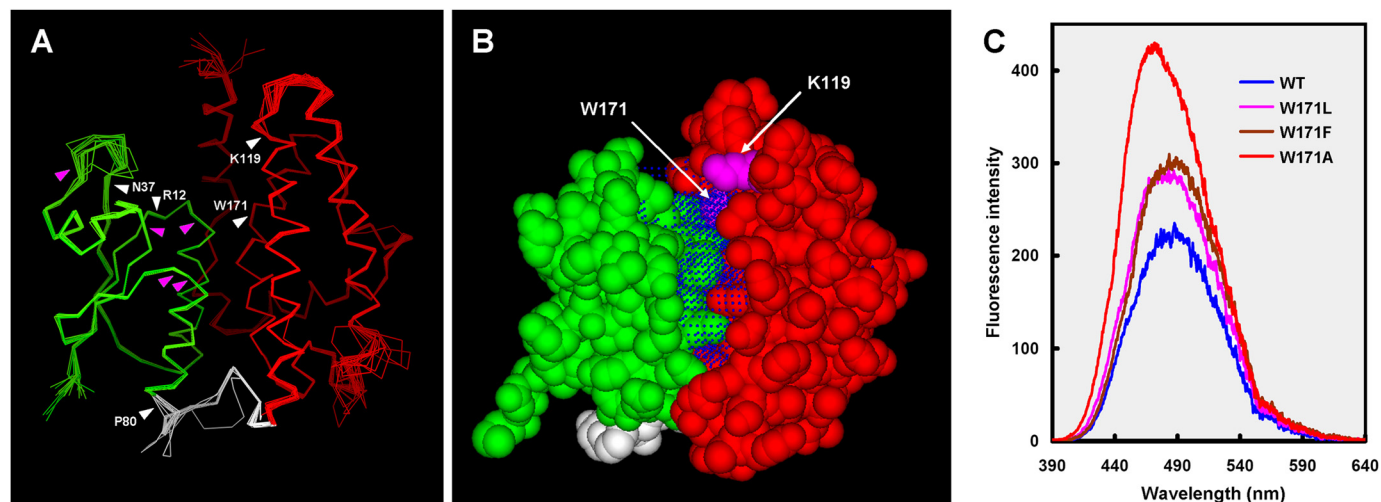
FIGURE 5. **Structural comparison of the *P. tabuliformis* Tau GSTs.** *A*, shown is structural superposition of the 26 Tau GSTs. *B*, shown are predicted ligand binding cavities (*blue dots*) in the PtaGSTU17 model by Binding Site Analysis. *C*, shown are fluorescence emission spectra of ANS binding to the W171L, W171F, and W171A mutants and wild-type PtaGSTU17. In *A* and *B*, the N- and C-terminal domains are illustrated in *green* and *red*, respectively, and the linker between the two domains is shown in *white*. In *A*, positions of the positively selected sites listed in Table 1 are indicated by *white arrows*; the GSH binding sites are indicated by *purple arrows*. In *B*, positions of Lys-119 and Trp-171 residues of PtaGSTU17 are shown in *purple*.

TABLE 2

Specific activities of the *P. tabuliformis Tau* GSTs and mutants towards six substrates (mean ± S.D. obtained from at least three independent determinations)

ND, no activity detected.

GSTs	Specific activity					
	CDNB	NBD-Cl	DCNB	NBC	Fluorodifen	Cum-OOH
			μmol/min per mg			
PtaGSTU3[a]	0.168 ± 0.023	ND	ND	ND	0.147 ± 0.061	ND
PtaGSTU4[a]	2.561 ± 0.173	1.321 ± 0.371	0.059 ± 0.013	0.321 ± 0.006	2.387 ± 0.373	0.072 ± 0.001
PtaGSTU5[a]	7.505 ± 0.112	12.681 ± 0.019	0.030 ± 0.003	0.911 ± 0.042	1.437 ± 0.007	0.566 ± 0.004
PtaGSTU6[a]	2.835 ± 0.062	1.882 ± 0.009	0.023 ± 0.004	ND	4.118 ± 3.426	1.210 ± 0.005
PtaGSTU10[a]	ND	0.042 ± 0.004	ND	0.156 ± 0.024	0.563 ± 0.003	0.020 ± 0.006
PtaGSTU12[a]	2.297 ± 0.011	2.199 ± 0.021	ND	0.754 ± 0.172	7.079 ± 1.233	0.445 ± 0.011
PtaGSTU13[a]	0.513 ± 0.063	0.078 ± 0.004	0.013 ± 0.005	ND	0.727 ± 0.038	ND
PtaGSTU15[a]	0.123 ± 0.010	0.044 ± 0.013	ND	0.106 ± 0.039	0.406 ± 0.035	ND
PtaGSTU7[a]	12.168 ± 0.012	10.657 ± 0.096	0.067 ± 0.005	0.552 ± 0.088	2.377 ± 0.998	0.265 ± 0.002
T16R (PtaGSTU7[a])	10.462 ± 1.670[b]	9.909 ± 3.140	0.059 ± 0.002[b]	0.443 ± 0.015[b]	2.197 ± 0.280	0.234 ± 0.004[b]
T16W (PtaGSTU7[a])	6.077 ± 0.615[b]	3.557 ± 0.356[b]	0.037 ± 0.001[b]	0.505 ± 0.028	0.172 ± 0.043[b]	0.220 ± 0.012[b]
T41N (PtaGSTU7[a])	11.590 ± 0.513[b]	23.890 ± 0.474[b]	0.069 ± 0.010	0.413 ± 0.058	3.197 ± 0.106	0.269 ± 0.024
G84P (PtaGSTU7[a])	14.769 ± 2.462[b]	21.012 ± 1.861[b]	0.059 ± 0.004	0.451 ± 0.023	2.274 ± 0.095	0.279 ± 0.007[b]
L121K (PtaGSTU7[a])	7.290 ± 1.988[b]	5.472 ± 3.009[b]	0.037 ± 0.008[b]	0.453 ± 0.085	1.797 ± 0.038	0.267 ± 0.035
L121 M (PtaGSTU7[a])	5.379 ± 0.182[b]	7.328 ± 0.886[b]	0.044 ± 0.003[b]	0.456 ± 0.066	2.091 ± 0.022	0.176 ± 0.002[b]
F175W (PtaGSTU7[a])	4.431 ± 0.070[b]	11.721 ± 0.569[b]	0.007 ± 0.001[b]	0.951 ± 0.113[b]	1.218 ± 0.065[b]	0.099 ± 0.005[b]
PtaGSTU14[c]	16.956 ± 0.266	1.914 ± 0,074	0.230 ± 0.015	0.419 ± 0.064	75.069 ± 2.763	0.046 ± 0.002
PtaGSTU24[c]	ND	ND	ND	ND	0.064 ± 0.041	ND
PtaGSTU26[c]	0.348 ± 0.007	0.489 ± 0.006	ND	ND	0.451 ± 0.071	0.012 ± 0.002
PtaGSTU17[c]	6.170 ± 0.039	1.575 ± 0.003	0.033 ± 0.001	0.506 ± 0.048	0.345 ± 0.013	0.092 ± 0.004
R12A (PtaGSTU17[c])	6.103 ± 0.087	0.494 ± 0.022[b]	0.077 ± 0.008[b]	1.114 ± 0.103[b]	ND[b]	0.101 ± 0.004[b]
R12I (PtaGSTU17[c])	5.166 ± 0.136[b]	0.686 ± 0.015[b]	0.004 ± 0.002[b]	0.637 ± 0.057[b]	0.541 ± 0.072[b]	0.076 ± 0.002[b]
R12L (PtaGSTU17[c])	5.361 ± 0.077[b]	0.501 ± 0.032[b]	0.082 ± 0.011[b]	0.895 ± 0.191[b]	ND[b]	0.11 ± 0.008[b]
R12T (PtaGSTU17[c])	5.705 ± 0.223[b]	1.502 ± 0.020[b]	0.042 ± 0.002[b]	0.254 ± 0.033[b]	ND[b]	0.073 ± 0.001[b]
N37A (PtaGSTU17[c])	5.009 ± 0.134[b]	1.434 ± 0.049[b]	0.039 ± 0.001[b]	0.934 ± 0.046[b]	ND[b]	0.081 ± 0.005[b]
N37D (PtaGSTU17[c])	29.346 ± 1.024[b]	1.869 ± 0.034[b]	0.052 ± 0.003[b]	1.050 ± 0.170[b]	ND[b]	0.27 ± 0.007[b]
N37T (PtaGSTU17[c])	9.376 ± 0.325[b]	1.127 ± 0.023[b]	0.049 ± 0.003[b]	0.646 ± 0.036[b]	ND[b]	0.105 ± 0.005[b]
P80A (PtaGSTU17[c])	6.644 ± 0.244[b]	1.153 ± 0.013[b]	0.017 ± 0.001[b]	0.324 ± 0.027[b]	0.398 ± 0.015[b]	0.068 ± 0.002[b]
P80G (PtaGSTU17[c])	5.052 ± 0.208[b]	1.310 ± 0.007[b]	0.031 ± 0.003	0.342 ± 0.005[b]	ND[b]	0.08 ± 0.002[b]
P80D (PtaGSTU17[c])	7.073 ± 0.229[b]	1.463 ± 0.015[b]	0.036 ± 0.001	0.426 ± 0.036[b]	ND[b]	0.078 ± 0.001[b]
K119A (PtaGSTU17[c])	0.066 ± 0.003[b]	0.163 ± 0.008[b]	0.008 ± 0.002[b]	0.121 ± 0.041[b]	0.452 ± 0.113	ND[b]
K119T (PtaGSTU17[c])	1.382 ± 0.033[b]	0.433 ± 0.020[b]	0.006 ± 0.003[b]	0.254 ± 0.009[b]	0.033 ± 0.004[b]	0.013 ± 0.001[b]
K119R (PtaGSTU17[c])	1.577 ± 0.033[b]	0.713 ± 0.006[b]	0.024 ± 0.001[b]	0.336 ± 0.036[b]	0.056 ± 0.006[b]	0.032 ± 0.002[b]
W171A (PtaGSTU17[c])	0.160 ± 0.006[b]	0.385 ± 0.022[b]	ND[b]	0.061 ± 0.022[b]	1.510 ± 1.105[b]	0.008 ± 0.001[b]
W171L (PtaGSTU17[c])	0.397 ± 0.010[b]	0.975 ± 0.009[b]	ND[b]	0.229 ± 0.024[b]	0.715 ± 0.075[b]	0.056 ± 0.001[b]
W171F (PtaGSTU17[c])	0.989 ± 0.019[b]	1.878 ± 0.028[b]	0.015 ± 0.001[b]	0.268 ± 0.045[b]	0.413 ± 0.042[b]	0.072 ± 0.002[b]

[a] The proteins in clade A Fig. 1*A*.
[b] Significant difference ($p < 0.05$) in enzymatic activities between the wild-type and its mutants, respectively.
[c] Protein in clade B in Fig. 1*A*.

-23, and -25), which were expressed as inclusion bodies. In this study only soluble proteins were used to investigate their biochemical functions.

Thirteen purified Tau GSTs listed in Table 2 showed specific activities toward substrate fluorodifen. Of 13 purified Tau GSTs, 11 showed specific activity toward CDNB and NBD-Cl, 9
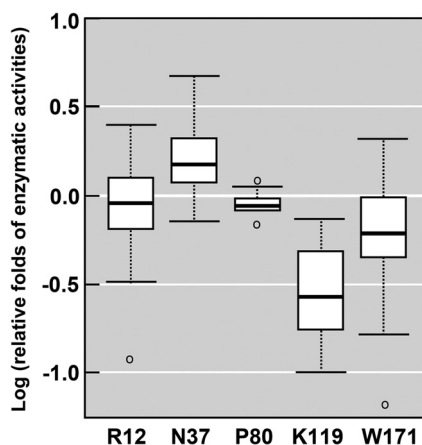
FIGURE 6. **Enzymatic activity changes, relative to wild-type PtaGSTU17 protein, induced by mutations at the five selected sites.** The boxplot shows the median (*black line*), interquartile range (*box*), and maximum and minimum scores (*whiskers*) for each data set. Outliers are shown as *circles outside the whiskers*.

toward Cum-OOH, 8 toward NBC, and 7 toward DCNB. Among the 13 *P. tabuliformis* Tau GSTs, eight GSTs (PtaGSTU3, -6, -10, -12, -13, -15, -24, and -26) shared overlapping but different substrate spectrum, and the other five (PtaGSTU4, -5, -7, -14, and -17) shared similar substrate spectrum. Although these five GSTs showed similar substrate spectrums, their enzymatic activities toward each substrate were very different, *e.g.* PtaGSTU4 and -5, which are closely related paralogs (Fig. 1*A*), showed 1.6–9.6-fold differences in their enzymatic activities toward each substrate. These results indicated *P. tabuliformis* Tau GSTs had divergence in substrate specificities and activities.

Functional Significance of the Putative Positively Selected Sites—To investigate the functional significance of the five putative positively selected sites, we constructed three sets of mutants for biochemical assays. First, we selected PtaGSTU17 of clade B (Fig. 1*A*; background branch in the branch-site model test) and mutated the corresponding sites in PtaGSTU17 to the amino acid residues present in clade A (foreground branch in the branch-site model test). There were 6–10 variants at each of these five sites in the 20 GSTs of clade A (Fig. 4). If a residue variant was represented by >15% of the sequences in clade A, we then replaced the corresponding amino acid residue in PtaGSTU17 with this common residue in clade A. Based on this criterion, we constructed and purified 11 PtaGSTU17 mutants (R12I, R12L, R12T, N37D, N37T, P80G, P80D, K119T, K119R, W171L, and W171F) to represent the major variants at the five positively selected sites. Second, we arbitrarily mutated each of the five sites in PtaGSTU17 to an amino acid not present in clade A GSTs, in this case to Ala. The enzymatic activities of the purified recombinant proteins were then explored using six GST substrates: CDNB, DCNB, NBC, Cum-OOH, NBD-Cl, and fluorodifen. Significant differences in enzyme activity ($p < 0.05$, Mann-Whitney U test) were found between the mutants and wild type (WT) toward all six substrates, except for mutants P80D and P80G toward substrate DCNB (Table 2). The mutants differing at the same residue site also differed markedly in activities toward the substrates, except for the two Pro-80 mutants (Fig. 6); thus the variants generally displayed a broad spectrum of activity ranges.

Third, we selected a GST, PtaGSTU7, of clade A (Fig. 1*A*, foreground branch in the branch-site model test) and replaced its amino acid at the five selected sites to the corresponding residues in clade B (background branch in the branch-site model test). Altogether, seven PtaGSTU7 mutants were constructed, and they were measured for enzymatic activities toward the six GST substrates (Table 2). To examine the activity changes between these mutants and the natural GSTs in clade A and clade B in Fig. 1*A*, we conducted a multiple response permutation procedure (MRPP) test. The MRPP is a nonparametric test that is flexible over unequal sample sizes and violation of normality assumptions. The MRPP test showed a significant difference between the PtaGSTU7 mutants and GSTs in clade A ($p < 0.02$) but not between the mutants and clade B ($p > 0.05$). This result lends further support to the functional significance of the putative positively selected sites and their contribution to the divergence of pine GST family.

The first two positively selected sites (corresponding to Arg-12 and Asn-37 of PtaGSTU17) were located very close to the G-site in the spatial conformation of the protein. Thus, substitution of the residues at these selected sites in clade A could have caused structural changes in the GSH binding pocket and hence pronounced diversification in biochemical function (Table 2 and Fig. 6). The third positively selected site is located in the linker region between the N- and C-terminal domains of the protein. Mutants at this position (P80G and P80D) were less functionally divergent both from each other and from WT. They showed less dispersed enzyme activities than mutants with variants at the other four positive selection sites, demonstrating that changes in this region have less dramatic effects on the enzyme function (Table 2 and Fig. 6).

The last two positively selected sites, corresponding to Lys-119 and Trp-171 of PtaGSTU17, are located in the C-terminal domain and are predicted H-site residues. When Lys-119 of PtaGSTU17 was mutated to Thr, Arg, and Ala, except for K119A toward fluorodifen, the resulting mutants K119T, K119R, and K119A showed significantly lower activities than WT toward all six substrates ($p < 0.05$, Mann-Whitney U test) (Table 2). When Trp-171 was replaced by Leu, Phe, or Ala, the mutants W171L, W171F, and W171A showed decreased activity toward four substrates (CDNB, DCNB, NBC, and Cum-OOH) but significantly increased (1.2–4.4-fold) activity toward fluorodifen (Table 2). In PtaGSTU17, the side chain of Lys-171 formed part of the wall of the hydrophobic substrate binding pocket (Fig. 5*B*). The Trp residue had a large indole (benzopyrrole) side chain. When Trp-171 was replaced by Leu, Phe, or Ala, which have smaller side chain residues, the hydrophobic substrate binding pocket of the three mutants could become larger than that of WT, thereby increasing accessibility for large substrates such as fluorodifen. In ANS fluorescence probe binding analysis, all three mutants showed marked increases in fluorescence intensity compared with the wild-type enzyme (Fig. 5*C*), indicating that more hydrophobic surface was exposed in them. Positive selection at this amino acid site may have driven adaptation of the enzyme to large molecular substrates by increasing the size of the hydrophobic substrate binding pocket. This finding suggests that the substrate specificity of

individual GSTs is influenced by the architecture of their access tunnels.

DISCUSSION

Mechanisms that expand protein functional diversity are clearly key evolutionary issues. Large enzyme families with promiscuous functions often acquire specific activities through divergent evolution involving structure- and function-altering amino acid substitutions (39). In nature, advantageous amino acid substitutions may be promoted by positive selection (5, 40), but the general structural distributions of positively selected mutations and their effects on protein function remain poorly examined.

Conventional theory emphasizes changes in function produced by amino acid replacements in active sites (41). GSTs have two characteristic catalytic active centers: the primary GSH substrate binding site (G-site) and the secondary hydrophobic substrate binding site (H-site). The G-site defines the primary biochemical function of all GST proteins and thus is highly conserved within each GST class across divergent groups of plants (42, 43). The H-site is less highly conserved, enabling the enzymes to interact with diverse substrates and facilitating functional diversification of the GST enzyme family. In this study we detected five positively selected sites in *P. tabuliformis* Tau GSTs, three of which are not located in active sites, two are located in the H-site; none is located in the G-site of the protein.

In plant GSTs, highly conserved G-site residues are critical for GSH binding and catalytic activity. In particular, one almost absolutely conserved Ser residue (alignment position 19 in Fig. 4) in all plant Tau GSTs is responsible for stabilization of the thiolate anion of enzyme-bound GSH (42, 44). We have previously shown that replacement of this Ser residue with Ala abolishes the enzyme catalytic activity toward various hydrophobic substrates (25, 45), demonstrating that mutations in the critical active G-site disrupt the essential function of the protein. This strong functional constraint is reflected in the absence of G-site residues among the five putative positively selected sites. In contrast, two of the five positively selected sites are H-site residues. H-sites are located in the C terminus of the protein and are highly variable both in sequence and topology within the Tau GST class (43). In the present study, substitution of the two H-site residues resulted in significant changes in enzymatic activities toward all six tested substrates (Table 2). These functional alterations could be interpreted in terms of the specific structural requirements for enclosing different substrate molecules. Variation at the H-site could provide an effective way for the GST family to acquire new functions toward environmental stimulants while maintaining the primary GSH binding property through conservation of the G-site. In accordance with this hypothesis, an H-site mutation in a human mu class GST, M2-2, elicited a 1000-fold increase in specific activity toward the substrate *trans*-stilbene oxide (46).

Our study further showed that mutations outside the active centers may also have marked effects on the protein function, although mutations close to the active sites seem to be more effective than distant ones. In pine Tau GSTs, three of the five positively selected sites are non-active sites, of which two are adja-

cent to the G-site and one (Pro-80 in PtaGSTU17) is distant from both active centers. The mutations at Pro-80 in PtaGSTU17 had milder effects on the enzyme function than mutations at the other four positively selected sites (Fig. 6). Studies on *Candida antarctica* lipase, aspartate aminotransferase, and biphenyl dioxygenase (47) suggest that mutations at sites <10 Å from a protein active site could alter substrate selectivity or catalytic activity very strongly. This is probably because mutations close to the active site are more likely to induce subtle conformational changes in the substrate binding pocket than distant mutations. Such a pattern supports the notion that strong evolutionary constraints in some regions of proteins can be potentially altered by mutations elsewhere in the structure, *i.e.* changing the amino acid at one position may modify the "landscape" of evolutionary constraints at neighboring positions (48). Thus, positive selection acting on residues adjacent to, rather than directly at, a critical active site is likely a general mechanism for functional diversification of enzyme families. Accordingly, positive selection for amino acid replacements outside the active site of the *Drosophila* JGW protein produced a novel dehydrogenase with altered substrate specificity compared with the ancestral protein (40).

Protein structural and functional properties are defined by rigid and flexible regions. Rigidity is an important factor for the integrity of a protein native folded structure, whereas conformational flexibility is important for binding and catalysis of diverse substrates (49). Protein loop regions are relatively flexible compared with the α-helices and β-strands they connect. Four of the five putative positively selected sites in pine Tau GSTs identified in this study are located in loop regions, indicating that loop regions of the GST protein can tolerate moderate structural modifications. Furthermore, large numbers of x-ray structures of plant and animal enzyme proteins have revealed that active site residues are often in loop regions (50). This suggests that adaptive functional diversification of enzyme families is achieved through modifications in flexible regions more often than modifications of rigid structural elements, as they allow the acquisition of new functions without disrupting native folding structures.

Plant Tau GSTs play important roles in defense responses against both biotic and abiotic stresses by detoxifying xenobiotics and combating oxidative stress (51). Due to their sessile nature, perennial plants are constantly exposed to a multitude of environmental stresses. Hence, possessing GSTs with diverse activities and high evolutionary flexibility is likely to have high adaptive value, enabling perennial plants to respond to the diverse environmental challenges they are likely to encounter. Secondary metabolism plays key roles in a multitude of plant ecological interactions, so the evolution of efficient, flexible enzyme systems should greatly enhance their fitness. In natural systems, adaptive mutations that generate advantageous new traits (*e.g.* beneficial shifts in substrate specificity) are particularly difficult to identify, because they may be obscured by the accumulation of subsequent neutral mutations. In addition, as pointed out by Camps *et al.* (11), whereas adaptive mutations drive positive selection, biological adaptation involves numerous, concomitant compensatory mutations. These compensatory mutations may also be positively selected to suppress del-

eterious pleiotropic effects of adaptive mutations in highly constrained regions (11). Distinguishing between adaptive and compensatory mutations remains a challenging task. Phylogenetic approaches alone are insufficient to elucidate the mechanisms that govern the adaptive molecular evolution of enzyme families. Relationships between the structures and functions of proteins are often unclear even when x-ray structures are available. In this study we combined molecular evolutionary analyses, protein structural modeling, and site-directed mutagenesis to define the structural and functional properties of the positively selected sites in Tau GSTs in the *P. tabuliformis* genome. The use of site-directed mutagenesis to complement molecular evolutionary analyses is a powerful approach that can provide insights into the evolution of enzyme functional divergence. The finding that most of the identified positively selected sites (four of five) in pine Tau GSTs have significant effects on substrate specificity suggests that adaptive evolution of this gene family has been partly driven by selective pressures that have broadened its substrate spectrum and activities. We propose that amino acid replacements accumulated by GST have widened its functions as a consequence of adaptive evolution. Our results provide new insights into the adaptive molecular evolution of the GST enzyme family in perennial plants. They have interesting implications for our understanding of how enzyme families evolve and contribute to our understanding of the selective context leading to biological divergence.

REFERENCES

1. Falara, V., Akhtar, T. A., Nguyen, T. T., Spyropoulou, E. A., Bleeker, P. M., Schauvinhold, I., Matsuba, Y., Bonini, M. E., Schilmiller, A. L., Last, R. L., Schuurink, R. C., and Pichersky, E. (2011) The tomato terpene synthase gene family. *Plant Physiol.* **157,** 770–789
2. Lan, T., Yang, Z. L., Yang, X., Liu, Y. J., Wang, X. R., and Zeng, Q. Y. (2009) Extensive functional diversification of the *Populus* glutathione *S*-transferase supergene family. *Plant Cell* **21,** 3749–3766
3. Dixon, D. P., Hawkins, T., Hussey, P. J., and Edwards, R. (2009) Enzyme activities and subcellular localization of members of the *Arabidopsis* glutathione transferase superfamily. *J. Exp. Bot.* **60,** 1207–1518
4. Varadarajan, N., Gam, J., Olsen, M. J., Georgiou, G., and Iverson, B. L. (2005) Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 6855–6860
5. Barkman, T. J., Martins, T. R., Sutton, E., and Stout, J. T. (2007) Positive selection for single amino acid change promotes substrate discrimination of a plant volatile-producing enzyme. *Mol. Biol. Evol.* **24,** 1320–1329
6. Weinberger, H., Moran, Y., Gordon, D., Turkov, M., Kahn, R., and Gurevitz, M. (2010) Positions under positive selection– key for selectivity and potency of scorpion α-toxins. *Mol. Biol. Evol.* **27,** 1025–1034
7. Gingerich, D. J., Hanada, K., Shiu, S. H., and Vierstra, R. D. (2007) Large-scale, lineage-specific expansion of a bric-a-brac/tramtrack/broad complex ubiquitin-ligase gene family in rice. *Plant Cell* **19,** 2329–2348
8. Smith, N. G., and Eyre-Walker, A. (2002) Adaptive protein evolution in *Drosophila*. *Nature* **415,** 1022–1024
9. Bishop, J. G., Dean, A. M., and Mitchell-Olds, T. (2000) Rapid evolution in plant chitinases. Molecular targets of selection in plant-pathogen coevolution. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 5322–5327
10. Yang, Z., and Bielawski, J. P. (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15,** 496–503
11. Camps, M., Herman, A., Loh, E., and Loeb, L. A. (2007) Genetic constraints on protein evolution. *Crit. Rev. Biochem. Mol. Biol.* **42,** 313–326
12. Loyall, L., Uchida, K., Braun, S., Furuya, M., and 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,**

13. Kampranis, S. C., Damianova, R., Atallah, M., Toby, G., Kondi, G., Tsichlis, P. N., and Makris, A. M. (2000) A novel plant glutathione *S*-transferase/peroxidase suppresses Bax lethality in yeast. *J. Biol. Chem.* **275,** 29207–29216
14. Mueller, L. A., Goodman, C. D., Silady, R. A., and Walbot, V. (2000) AN9, a petunia glutathione *S*-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. *Plant Physiol.* **123,** 1561–1570
15. Agrawal, G. K., Jwa, N. S., and Rakwal, R. (2002) A pathogen-induced novel rice (*Oryza sativa* L.) gene encodes a putative protein homologous to type II glutathione *S*-transferases. *Plant Sci.* **163,** 1153–1160
16. Mao, J. F., and Wang, X. R. (2011) Distinct niche divergence characterizes the homoploid hybrid speciation of *Pinus densata* on the Tibetan plateau. *Am. Nat.* **177,** 424–439
17. Edwards, R., Dixon, D. P., and Walbot, V. (2000) Plant glutathione *S*-transferases. Enzymes with multiple functions in sickness and in health. *Trends Plant Sci.* **5,** 193–198
18. Edgar, R. C. (2004) MUSCLE. Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797
19. Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41,** 95–98
20. Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704
21. Yang, Z. (2007) PAML 4. Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24,** 1586–1591
22. Yang, Z., and Nielsen, R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17,** 32–43
23. Yang, Z., Wong, W. S., and Nielsen, R. (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22,** 1107–1118
24. Zeng, Q. Y., and Wang, X. R. (2005) Catalytic properties of glutathione-binding residues in a Tau class glutathione transferase (PtGSTU1) from *Pinus tabulaeformis*. *FEBS Lett.* **579,** 2657–2662
25. Zeng, Q. Y., Lu, H., and Wang, X. R. (2005) Molecular characterization of a glutathione transferase from *Pinus tabulaeformis* (Pinaceae). *Biochimie* **87,** 445–455
26. Habig, W. H., Pabst, M. J., and Jakoby, W. B. (1974) glutathione *S*-transferases. The first enzymatic step in mercapturic acid formation. *J. Biol. Chem.* **249,** 7130–7139
27. Ricci, G., Caccuri, A. M., Lo Bello, M., Pastore, A., Piemonte, F., and Federici, G. (1994) Colorimetric and fluorometric assays of glutathione transferase based on 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole. *Anal. Biochem.* **218,** 463–465
28. Edwards, R., and Dixon, D. P. (2005) Plant glutathione transferases. *Methods Enzymol.* **401,** 169–186
29. Schonbrunn, E., Eschenburg, S., Luger, K., Kabsch, W., and Amrhein, N. (2000) Structural basis for the interaction of the fluorescence probe 8-anilino-1-naphthalene sulfonate (ANS) with the antibiotic target MurA. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 6345–6349
30. Liu, Y. J., Han, X. M., Ren, L. L., Yang, H. L., and Zeng, Q. Y. (2013) Functional divergence of the GST supergene family in *Physcomitrella patens* reveals complex patterns of large gene family evolution in land plants. *Plant Physiol.* **161,** 773–786
31. Soranzo, N., Sari Gorla, M., Mizzi, L., De Toma, G., and Frova, C. (2004) Organisation and structural evolution of the rice glutathione *S*-transferase gene family. *Mol. Genet. Genomics* **271,** 511–521
32. Dixon, D. P., Lapthorn, A., and Edwards, R. (2002) Plant glutathione transferases. *Genome Biol.* **3,** REVIEWS3004
33. Zhang, J., Nielsen, R., and Yang, Z. (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22,** 2472–2479
34. Mondragon-Palomino, M., and Gaut, B. S. (2005) Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **22,** 2444–2456
35. Sawyer, S. (1989) Statistical tests for detecting gene conversion. *Mol. Biol.*

1939–1950

Evol. **6,** 526–538

36. Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D., and Lefeuvre, P. (2010) RDP3. A flexible and fast computer program for analyzing recombination. *Bioinformatics* **26,** 2462–2463

37. Martinez-Castilla, L. P., and Alvarez-Buylla, E. R. (2003) Adaptive evolution in the *Arabidopsis* MADS-box gene family inferred from its complete resolved phylogeny. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 13407–13412

38. Sabbagh, A., Marin, J., Veyssière, C., Lecompte, E., Boukouvala, S., Poloni, E. S., Darlu, P., and Crouau-Roy, B. (2013) Rapid birth-and-death evolution of the xenobiotic metabolizing *NAT* gene family in vertebrates with evidence of adaptive selection. *BMC Evol. Biol.* **13,** 62

39. Yoshikuni, Y., Ferrin, T. E., and Keasling, J. D. (2006) Designed divergent evolution of enzyme function. *Nature* **440,** 1078–1082

40. Zhang, J., Dean, A. M., Brunet, F., and Long, M. (2004) Evolving protein functional diversity in new genes of *Drosophila. Proc. Natl. Acad. Sci. U.S.A.* **101,** 16246–16250

41. Chen, R., Greer, A., and Dean, A. M. (1995) A highly active decarboxylating dehydrogenase with rationally inverted coenzyme specificity. *Proc. Natl. Acad. Sci. U.S.A.* **92,** 11666–11670

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

43. Thom, R., Cummins, I., Dixon, D. P., Edwards, R., Cole, D. J., and Lapthorn, A. J. (2002) Structure of a Tau class glutathione *S*-transferase from wheat active in herbicide detoxification. *Biochemistry* **41,** 7008–7020

44. Axarli, I., Dhavala, P., Papageorgiou, A. C., and Labrou, N. E. (2009) Crystallographic and functional characterization of the fluorodifen-inducible glutathione transferase from *Glycine max* reveals an active site topography suited for diphenylether herbicides and a novel L-site. *J. Mol. Biol.* **385,** 984–1002

45. Yang, X., Sun, W., Liu, J. P., Liu, Y. J., and Zeng, Q. Y. (2009) Biochemical and physiological characterization of a Tau class glutathione transferase from rice (*Oryza sativa*). *Plant Physiol. Biochem.* **47,** 1061–1068

46. Ivarsson, Y., Mackey, A. J., Edalat, M., Pearson, W. R., and Mannervik, B. (2003) Identification of residues in glutathione transferase capable of driving functional diversification in evolution - A novel approach to protein redesign. *J. Biol. Chem.* **278,** 8733–8738

47. Morley, K. L., and Kazlauskas, R. J. (2005) Improving enzyme properties. When are closer mutations better? *Trends Biotechnol.* **23,** 231–237

48. Williams, S. G., and Lovell, S. C. (2009) The effect of sequence evolution on protein structural divergence. *Mol. Biol. Evol.* **26,** 1055–1065

49. Huber, R. (1987) Flexibility and rigidity, requirements for the function of proteins and protein pigment complexes. Eleventh Keilin memorial lecture. *Biochem. Soc. Trans.* **15,** 1009–1020

50. Padhi, S. K., Fujii, R., Legatt, G. A., Fossum, S. L., Berchtold, R., and Kazlauskas, R. J. (2010) Switching from an esterase to a hydroxynitrile lyase mechanism requires only two amino acid substitutions. *Chem. Biol.* **17,** 863–871

51. Wisser, R. J., Kolkman, J. M., Patzoldt, M. E., Holland, J. B., Yu, J., Krakowsky, M., Nelson, R. J., and Balint-Kurti, P. J. (2011) Multivariate analysis of maize disease resistances suggests a pleiotropic genetic basis and implicates a *GST* gene. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 7339–7344

Access the most updated version of this article at doi: 10.1074/jbc.M113.456863

Find articles, minireviews, Reflections and Classics on similar topics on the JBC Affinity Sites.

Alerts:

- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:

http://www.jbc.org/content/suppl/2013/07/11/M113.456863.DC1.html

This article cites 50 references, 29 of which can be accessed free at
http://www.jbc.org/content/288/34/24441.full.html#ref-list-1