Phi Class of Glutathione S-transferase Gene Superfamily Widely Exists in Nonplant Taxonomic Groups

Jean-Pierre Munyampundu[1], You-Ping Xu[2] and Xin-Zhong Cai[1]

[1]Institute of Biotechnology, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China. [2]Center of Analysis and Measurement, Zhejiang University, Hangzhou, China.

ABSTRACT: Glutathione S-transferases (GSTs) constitute a superfamily of enzymes involved in detoxification of noxious compounds and protection against oxidative damage. GST class Phi (GSTF), one of the important classes of plant GSTs, has long been considered as plant specific but was recently found in basidiomycete fungi. However, the range of nonplant taxonomic groups containing GSTFs remains unknown. In this study, the distribution and phylogenetic relationships of nonplant GSTFs were investigated. We identified GSTFs in ascomycete fungi, myxobacteria, and protists *Naegleria gruberi* and *Aureococcus anophagefferens*. GSTF occurrence in these bacteria and protists correlated with their genome sizes and habitats. While this link was missing across ascomycetes, the distribution and abundance of GSTFs among ascomycete genomes could be associated with their lifestyles to some extent. Sequence comparison, gene structure, and phylogenetic analyses indicated divergence among nonplant GSTFs, suggesting polyphyletic origins during evolution. Furthermore, *in silico* prediction of functional partners suggested functional diversification among nonplant GSTFs.

KEYWORDS: Phi class glutathione S-transferases (GSTFs), fungi, Ascomycota, bacteria, protists

CITATION: Munyampundu et al. Phi Class of Glutathione S-transferase Gene Superfamily Widely Exists in Nonplant Taxonomic Groups. *Evolutionary Bioinformatics* 2016:12 59–71 doi: 10.4137/EBO.S35909.

TYPE: Original Research

RECEIVED: October 02, 2015. **RESUBMITTED:** December 13, 2015. **ACCEPTED FOR PUBLICATION:** December 13, 2015.

ACADEMIC EDITOR: Jike Cui, Associate Editor

PEER REVIEW: Five peer reviewers contributed to the peer review report. Reviewers' reports totaled 1704 words, excluding any confidential comments to the academic editor.

FUNDING: This work was financially supported by grants from the Genetically Modified Organisms Breeding Major Projects (no. 2014ZX0800905B to X-ZC), the Special Fund for Agro-scientific Research in the Public Interest (no. 201103016 to X-ZC), and the SRFDP (no. 20110101110092 to X-ZC). The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: xzhcai@zju.edu.cn

Introduction

Glutathione S-transferases (GSTs, EC.2.5.1.18) are a superfamily of multifunctional enzymes that traditionally catalyze the nucleophilic conjugation of glutathione (GSH: γ Glu–Cys–Gly) to diverse hydrophobic and electrophilic compounds.[1] GSTs are widely distributed in all major kingdoms of living organisms and are distinguished into three subfamilies: (1) cytosolic (or soluble) GSTs, (2) mitochondrial and peroxisomal GSTs (kappa class), and (3) membrane-associated proteins in eicosanoid and glutathione metabolism or microsomal GSTs, depending on their primary sequence, 3D structure, function, and cellular localization. An additional highly divergent family comprises the plasmid-encoded bacterial fosfomycin-resistant GSTs (FosA).[2]

Soluble GSTs constitute a very ancient protein superfamily whose structural fold is believed to have evolved from a thioredoxin/glutaredoxin progenitor in response to the development of oxidative stress. Despite the drastic overall sequence divergence across their classes, all soluble GST protein structures adopt a similar canonical fold. This GST fold is characterized by an N-terminal thioredoxin-fold (GST_N) domain and a C-terminal alpha helical (GST_C) domain, with an active site located in a cleft between the two domains. GST_N domain is highly conserved and provides most of the GSH-binding site (G-site), whereas the less conserved GST_C domain contributes the majority of residues in the adjacent H-site that accommodates the cosubstrate hydrophobic moiety.[3] Based on sequence similarity, immunological cross-reactivity, substrate specificity, and genome organization, soluble GSTs have been further grouped into different species-independent classes, and some of them are specific to kingdoms or phyla. To date, 14 distinct GST classes have been identified in plants. They are Phi (GSTF), Tau (GSTU), Lambda (GSTL), dehydroascorbate reductase (DHAR), Theta (GSTT), Zeta (GSTZ), Elongation factor 1B gamma (EF1BG), tetrachlorohydroquinone dehalogenase (TCHQD), glutathionyl-hydroquinone reductase or Xi (GHR or GSTX), microsomal prostaglandin E synthase type 2 (mPGES2), metaxin, hemerythrin (GSTH), iota (GSTI), and Ure2p, with the last three being limited to nonvascular plants.[4–6]

Plant GSTs exhibit multiple functions apart from their well-documented function in the detoxification of a wide range of xenobiotic and endogenous harmful compounds, thanks to their H-site residue variation within and among classes. Certain plant GSTs protect plants from oxidative damage by catalytically removing endogenous cytotoxic hydroperoxides, function as isomerases, and suppress apoptosis.[7,8] Plant GSTs are involved in antioxidant recycling and primary and

secondary metabolisms. Additionally, plant GSTs display noncatalytic cellular functions such as acting as transporters of small molecules and involving in stress-induced cell signaling. They also contribute to plant growth and development as well as abiotic and biotic stress tolerance.[4,9–11] The first plant GST identified in corn 45 years ago[12] belongs to the Phi class (GSTF). GSTFs are the most studied among all plant GSTs. *Arabidopsis thaliana* genome contains 13 GSTF members, being the second largest GST class in plant after Tau.[4] The current knowledge on the function of plant GSTFs and their genome-wide distribution in terrestrial plants has been recently documented.[13] Even more recent evidence indicate that AtGSTF2 and AtGSTF3 confer protection against oxidative stress in *A. thaliana* as a result of their oxidation reverse by methionine sulfoxide reductases for which they are substrates,[14] and wheat TaGSTF6 is involved in Dn1-mediated resistance to the Russian wheat aphid *Diuraphis noxia*.[15] The solved crystal structure indicated G-site residues of AtGSTF2. These include Lys41, Glu53, Va15, Glu66, Ser67, and Arg68.[16] The distribution of GSTF in various taxonomic groups is still controversial. Along with DHAR, Lambda, and Tau, GSTF has long been regarded as plant specific until it has recently been noticed that basidiomycete fungi and green algae contain GSTFs and DHARs, respectively.[6,17] Moreover, a GSTF from a wood-degrading basidiomycete *Phanerochaete chrysosporium* (*Phchr7971*), which was previously identified as a GTT2-related isoform, has been found to cluster with plant GSTFs in a phylogenetic tree.[18] Given that a *GST* gene, which encodes a protein showing as much as 49% sequence identity to *A. thaliana* GST6 (AtGSTF8), was encountered in the Ascomyceta *Gibberella zeae*,[19] it is likely that the existence of nonplant GSTFs extends beyond basidiomycete fungi. Nevertheless, the range of nonplant taxonomic groups containing GSTFs remains unclear. In the present study, we analyzed the global distribution and phylogenetic relationship of nonplant GSTFs outside basidiomycete fungi. Our results reveal that GSTFs exist widely in ascomycete fungi, bacteria, and even protists. We also predict the functional partners of nonplant GSTFs and provide useful information about the functions of nonplant GSTFs.

Materials and Methods

Database search. Putative nonplant GSTFs were obtained from Joint Genome Institute (JGI; http://genome.jgi.doe.gov/) databases. Putative GSTFs from bacteria and *Aureococcus anophagefferens* CCMP 1984 were obtained by BLASTp search at the Integrated Microbial Genomes (IMG; https://img.jgi.doe.gov/). Fungal sequences were retrieved from fungal genomics Resource database (MycoCosm; http://genome.jgi-psf.org/programs/fungi/index.jsf), while sequences from *Naegleria gruberi* were downloaded from its genome database (http://genome.jgi-psf.org/Naegr1/Naegr1.home.html). IMG and MycoCosm were used as they offer an advantage of searching against specific bacterial or fungal groups. To obtain nonplant GSTF homologs, a BLASTp search was initially performed against the UniProt database (http://www.uniprot.org/) using protein sequences of *A. thaliana* GSTFs[4] as queries. The closest GSTF homologs from bacteria and fungi were subsequently used to screen more sequences in respective databases with the e-value cutoff set to 1e−20. Representatives of other known GST classes from bacteria, fungi, plants, and animals were retrieved following the BLASTp searches against different databases, including the Arabidopsis Information Resource (TAIR; http://www.arabidopsis.org), UniProtKB, Phytozome (http://phytozome.jgi.doe.gov/pz/portal.html), and NCBI's nonredundant (nr) protein database (http://www.ncbi.nlm.nih.gov/) (Supplementary Table 1).

Sequence alignment and construction of phylogenetic tree. Protein sequences from each database screen were analyzed with NCBI-conserved domain database (CDD; http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to check the presence of GST_N and GST_C domains or GST fold. Following protein multiple sequence alignment, putative GSTF homologs were analyzed to verify the presence of conserved amino acid residues characteristic for GSTFs. Amino acid sequences of 133 full-length nonplant GSTF candidates were aligned along with 45 plant GSTFs, 54 basidiomycete GSTFs, 10 sequences representing uncharacterized GSTs closely related with GSTFs from bacteria, and representatives of 13 different known classes from all kingdoms of life. Sequence alignment was conducted using the MUSCLE program[20] with default parameters and subsequently adjusted manually using GenDoc 2.6 software. The phylogenetic trees were constructed by neighbor-joining (NJ) and maximum likelihood (ML) method using JTT model[21] implemented in MEGA 5.05.[22] Bootstrap of 1000 replicates was performed to evaluate the support of clusters and nodes.

Sequence comparison and gene structure analysis. Nonplant GSTF representatives in the main groups inferred from the phylogenetic tree were selected for sequence similarity and gene structure analyses. To evaluate sequence similarity, amino acid residues of putative GSTFs of interest were compared with known GSTFs from Arabidopsis,[4] moss (*Physcomitrella patens*),[5] Larix (*Larix kaempferi*),[23] and poplar (*Populus trichocarpa*).[13] Protein sequence identity/similarity percentages among sequences were determined using MegAlign™ program of DNASTAR software. Gene structure was analyzed by examining the exon/intron positions and phases using the Gene Structure Display Server (GSDS; http://gsds.cbi.pku.edu.cn/).

Prediction of functional partners of the identified GSTF sequences. To predict the functions of nonplant GSTFs, protein interaction and association analysis was performed with STRING v9.1 database[24] using default settings. Representatives of GSTF protein sequences from different phylogenetic clades were chosen for this analysis. In case that the organism from which a protein sequence to be subjected

to analysis was not represented in the database, a protein sequence of another organism with the highest similarity to the sequence of interest was selected from the list provided by the database. For the predicted proteins whose name was not clear, domain composition was analyzed using CDD or Pfam databases in NCBI.

Results

Putative Phi class GST sequences are widely distributed in nonplant kingdoms. To identify putative Phi class GST (GSTF) sequences in nonplant kingdoms, BLASTp searches at the default setting parameters, except for the number of returned alignments (hits) set to 500, against UniProtKB database were performed using Arabidopsis GSTF protein sequences as queries. These searches retrieved substantial homologs of nonplant origin, including bacteria, Ascomycota fungi, and ameba, generally with an identity of >40% (data not shown), indicating that they may belong to the same class according to the criterion suggested by Hayes et al.[25] The sequences with UniprotKB accession numbers A0A017SXI2, A0A0E1RY08, and D2VAQ3 from the bacterium *Chondromyces apiculatus* DSM 436, the fungus *Coccidioides immitis* RS, and the protozoan *N. gruberi*, respectively, were collected and analyzed with NCBI-CDD. All of them were found to contain GST_N_Phi and/or GST_C_Phi domains in their protein structures (Supplementary Fig. 1). These results demonstrate that these three sequences from the bacterium, fungus, and protozoan are highly likely to be GSTF sequences.

To further explore the degree of putative GSTF occurrence in bacteria, BLASTp search against all bacterial genomes in the IMG database was performed using A0A017SXI2 of *C. apiculatus* DSM 436 as query sequence. A total of 14 with an identity of >50% to A0A017SXI2 were retrieved. Domain analysis of these sequences using CDD program indicated the presence of Phi class specific GST_N domain in these sequences (Supplementary Fig. 1). Accordingly, these sequences were considered as putative GSTFs. They were present in bacteria of Myxococcales order, particularly Myxococcaceae and Polyangiaceae families (Supplementary Table 2). All of these bacteria carried one copy of *GSTF* gene, except *Myxococcus stipitatus* in which two genes were identified. In addition, the same IMG database searches retrieved sequences from other groups of bacteria, including Rhizobiales, Cyanobacteria, Rhodospirillales, sharing identity ~40% with the query sequence A0A017SXI2 (Supplementary Table 2). Domain analyses predicted that they possessed a GST_N domain. However, no known GSTF class specificity was indicated for these sequences. Therefore, whether these sequences belong to GSTF awaits further analysis. Collectively, these results demonstrated that the Phi class of GST is at least widely distributed in bacteria of Myxococcales order.

A similar approach was employed to find Ascomycota GSTF homologs in MycoCosm database with A0A0E1RY08 of *C. immitis* RS serving as query sequence. We checked Ascomycota fungi since existence and distribution of GSTFs in Basidiomycota had been previously reported.[17] Consequently, we identified 110 full-length putative GSTFs in 85 Ascomycota fungal strains among a total of 250 that were examined (Supplementary Table 3). Their distribution was significantly different across and within Ascomycota classes (Supplementary Table 3). The identified GSTF homolog sequences belonged to Xylonomycetes, Sordariomycetes, Leotiomycetes, Lecanoromycetes, Dothideomycetes, Eurotiomycetes, and Orbiliomycetes Ascomycota classes following the evolution order from the most ancient to the most recent. No GSTF homolog was found in Taphrinomycotina, Saccharomycotina, and Pezizomycetes classes. GSTF is most largely expanded in Leotiomycetes, followed by Eurotiomycetes and Sordariomycetes, and the largest gene copy number was identified in *Oidiodendron maius* Zn v1.0 (Leotiomycetes) and *Talaromyces aculeatus* ATCC 10409 v1.0 (Eurotiomycetes), which contained five genes each. Of 13 species of Leotiomycetes under examination, 11 contained at least one GSTF sequence, and the majority of these species carried more than one copy of *GSTF* genes with an average of 2.1, while *Amorphotheca resinae* v1.0 and *Blumeria graminis* f. sp. *hordei* DH14 were exceptions as they were not found to contain any *GSTF* gene. Among 51 species (57 strains) of Eurotiomycetes, 30 were found to have at least one *GSTF* gene. However, unlike Leotiomycetes, most species of Eurotiomycetes contained only one copy of *GSTF* gene. Moreover, a notable disproportion in GSTF distribution among *Aspergillus* and *Penicillium* genera was observed. *Aspergillus brasiliensis* v1.0 and *A. tubingensis* v1.0 comprised three and two GSTFs, respectively. A total of 16 *Aspergillus* fungi carried one *GSTF* gene each, whereas the remaining 6 contained no *GSTF* gene. Similarly, the majority of *Penicillium* fungi did not harbor any *GSTF* gene in their genomes except three species, among which *Penicillium bilaiae* ATCC 20851 v1.0 and *Penicillium glabrum* DAOM 239074 v1.0 each contained one *GSTF* gene, while *Penicillium janthinellum* ATCC 10455 v1.0 comprised two. For Sordariomycetes fungi, 15 out of 54 species possessed at least one *GSTF* gene and contained a total of 26 *GSTF* genes. As for Dothideomycetes, the putative *GSTF* gene copy number ranged from zero to two. Taken together, these results revealed that the Phi class of GST is widely and unequally distributed in at least seven classes of Ascomycota fungi.

Besides bacteria and fungi, nonplant GSTFs were also identified in peculiar species from distinct kingdoms (or phyla) of organisms. Surprisingly, BLASTp search using a protein sequence from *N. gruberi* with UniProtKB accession number D2VAQ3 as query retrieved eight GSTF homolog sequences in the amoeboflagellate *N. gruberi* NEG-M genome database (Supplementary Table 3). Domain analysis demonstrated that all of them possessed both well-defined Phi class specific GST_N and GST_C domains (Supplementary Fig. 1). Finally, the existence of GSTF sequence in 13 heterokonta organisms deposited in the JGI database was examined

using a similar BLASTp approach. Intriguingly, a GSTF sequence was found in a harmful algal bloom (HAB) species, *A. anophagefferens* CCMP 1984 (Supplementary Table 2).

Collectively, our results reveal that the Phi class of GST widely exists in diversified kingdoms of organisms ranging from prokaryotes through protists to eukaryotes.

Putative nonplant GSTFs show phylogenetic and gene structure divergence. To evaluate the phylogenetic position of retrieved putative nonplant Phi GSTs, a large-scale NJ tree containing aligned 131 full-length protein sequences of potential GSTFs from Ascomycota, bacteria, *N. gruberi* NEG-M, and the HAB species *A. anophagefferens* CCMP 1984 was generated along with representatives of plant and Basidiomycota GSTFs and some other well-known GST classes from different kingdoms of life (Supplementary Table 1). In addition, 10 bacterial sequences resulting from IMG BLAST screens with identity ≤40% to query sequence (A0A017SXI2 of *C. apiculatus* DSM 436) were also included in this tree (Supplementary Table 2). The phylogenetic tree indicated that the novel nonplant GSTF sequences identified in this study distinctly clustered only with typical plant and basidiomycete fungal GSTFs but not with any other previously identified GST class. Interestingly, some ascomycete sequences, including the query A0A0E1RY08 of *C. immitis* RS, clustered within the basidiomycete group (Fig. 1). Notwithstanding the lack of strong bootstrap support, bacterial sequences with identity <40% to the query (referred to as the 10 uncharacterized GSTs in this work, Supplementary Table 2) departed from the major Phi group and formed a cluster sharing node with GTT2 class (Fig. 1).

To further assess the clustering pattern of these sequences, a separate ML tree of the representative sequences of the identified nonplant GSTFs was constructed along with plant GSTs, using Ure2p from both bacteria and fungi as outgroup with 1000 bootstrap replicates. In this tree, all nonplant GSTFs formed phylogenetic groupings that were separated from those

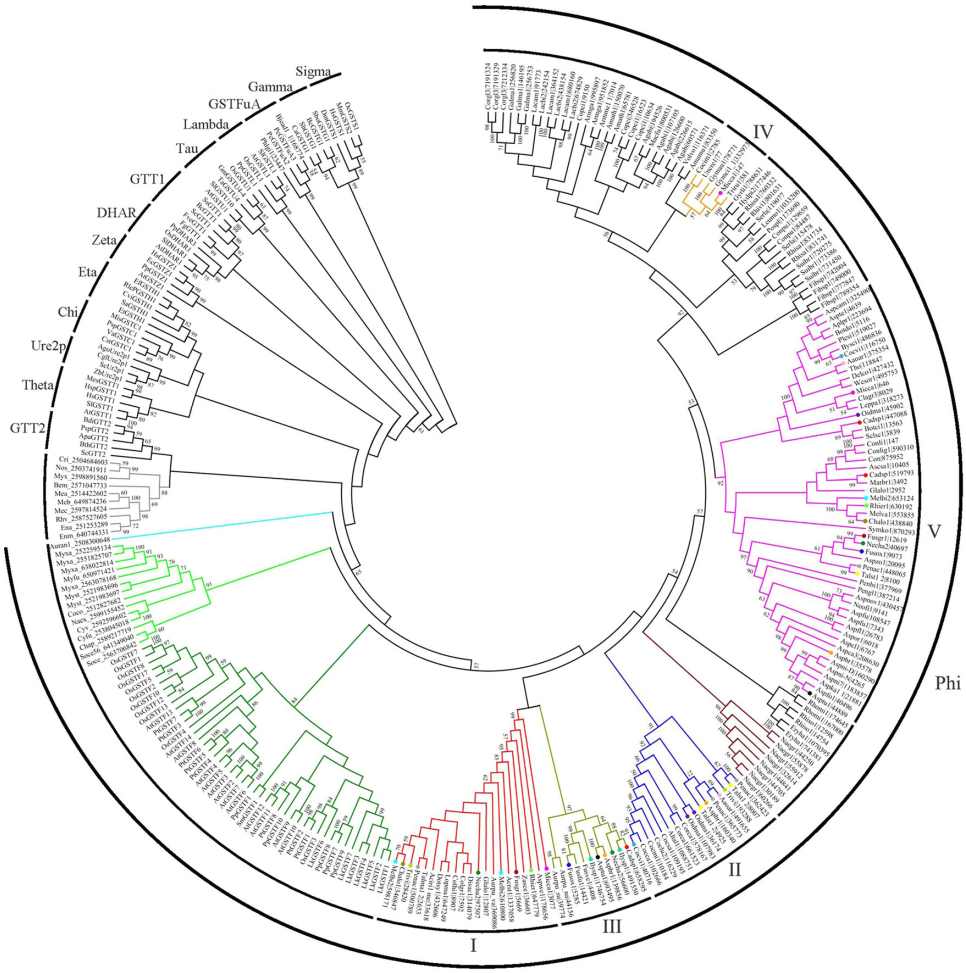

Figure 1. Unrooted phylogenetic tree of Phi class nonplant GSTs. The representatives of mainstream plant GSTFs and some other well-known GST classes from different kingdoms of life were also included. The tree was constructed by the NJ method using MEGA 5.05. Only bootstraps with >50% support in 1000 replications are shown. Different gene copies from the same ascomycete fungi are marked with the same color. Ascomycete GSTF sequences were divided into five clades, which are indicated in red, blue, olive, orange, and magenta for clades I–V, respectively. Branches in purple, green lime, and cyan represent sequences from *N. grubei*, bacteria, and *A. anophagefferens*, respectively. Plant GSTFs are in green branches while basidiomycete GSTFs are in black. Roman numerals I–V show different clades of putative ascomycete GSTFs.

of mainstream plant GSTFs. The GSTF sequences from ascomycetes were grouped into five discrete clades. Clade I constituted sequences mainly from multiple gene copy-bearing fungi of different classes, though single *GSTF* gene-carrying fungi were present as well. In principle, all sequences in clade I displayed an e-value <1e–30 and 45%–37% identity to the query sequence (Fig. 2 and Supplementary Table 3). Unlike the uncharacterized bacterial sequences, which displayed no

Figure 2. Phylogenetic tree and exon/intron structure of nonplant Phi class GST homologs along with plant GSTFs. An ML phylogenetic tree of novel nonplant GSTF representatives from ascomycete fungi, bacteria, *N. gruberi,* and *A. anophagefferens* was generated using MEGA 5.05 with 1000 bootstrap replicates. Bootstrap values for nodes with >50% support are displayed on the branches. Branches in blue, orange, red, cyan, green lime, and green indicate sequences from agaricomycetes, pucciniomycetes, *N. grubei*, *A. anophagefferens*, bacteria, and mainstream plant GSTFs, respectively. Different clades of ascomycete putative GSTF sequences are indicated by roman numerals (I–V). Exons of ascomycete putative GSTF genes are represented by green bars, while those of their homologs from corresponding sources in the phylogenetic tree are denoted by red bars. Introns were represented by a black line, whereas intron phases are designed by numbers 0, 1, and 2. Ure2p protein sequences from *Ashbya gossypii* ATCC 10895 (AgoUre2p1), *Saccharomyces cerevisiae* S288c (ScUre2p1), and *Zygosaccharomyces bailii* ISA1307 (ZbUre2p1) were used as outgroup. The length of each branch is proportional to the average substitutions per site as indicated by the scale.

class-specific GST domain, substantial ascomycete sequences in clade I exhibited GST_N_Phi domain in their structures (Supplementary Fig. 1). Interestingly, exon/intron arrangement and intron phase inquiry disclosed that all putative fungal *GSTF* genes of clade I had two exons and one intron of phase 0 in the coding region (Fig. 2 and Supplementary Table 3). A set of sequences from fungi belonging to different taxonomical classes with multiple gene copies of potential GSTFs formed clade II along with four sequences from *Cochliobolus carbonum* 26-R-13 v1.0 (jgi|Cocca1|102666), *Cochliobolus lunatus* m118 v2.0 (jgi|Coclu2|116229), *Cochliobolus miyabeanus* ATCC 44560 v1.0 (jgi|Cocmi1|10184), and *Alternaria alternata* SRC1lrK2f v1.0 (jgi|Altal1|1088751), which contained only one gene each. All the genes encoding GSTF sequences in Clade II consisted of four exons and three introns apart from one from *Trichoderma virens* Gv29–8 v2.0 (jgi|TriviGv29_8_2|193288), which possessed five exons and four introns. Clade III *GSTF* genes also consisted of four exons and three introns as Clade II. However, their intron phase pattern, 020, was distinguished from that of clade II, 012 (Fig. 2 and Supplementary Table 3). Clade IV was composed of six sequences all originating from Eurotiomycetes members. These included the query sequence A0A0E1RY08 (jgi|Cocim1|2785) of *C. immitis* RS, jgi|Uncre1|77 of *Uncinocarpus reesii* 1704, jgi|Micca1|147 of *Microsporum canis* CBS 113480, jgi|Triru1|581 from *Trichophyton rubrum* CBS 118892, jgi|Gymau1|78771| of *Gymnascella aurantiaca* v1.0, and jgi|Gymci1_1|332973 from *Gymnascella citrina* v1.1. This clade clustered amid sequences from basidiomycete GSTFs in the tree (Figs. 1 and 2 and Supplementary Table 3). All the genes of clade IV possessed five exons and four introns with a phase pattern 0202, as did their basidiomycete homologs (Fig. 2 and Supplementary Tables 1 and 3). Finally, clade V was the largest one, comprising 53 GSTFs from various classes of ascomycetes. All clade V *GSTF* genes carried three exons and two introns but with diversified phase patterns (Fig. 2 and Supplementary Table 3). Regarding bacterial GSTFs, as depicted in Figure 1, the supposed true bacterial GSTF sequences from Myxococcales showing GST_N_Phi domain formed a discrete clade departing from the one grouping representatives of other bacterial sequences sharing low amino acid sequence identity and lacking Phi-specific domain but recognized as GST members by Prosite database (Supplementary Table 2). On the other hand, the sequence from *A. anophagefferens* made its own clade, while the sequences from *N. gruberi* NEG-M curiously established a special clade that intercepted the fungal GSTF candidate clusters in the tree. It is important to note that the *GSTF* genes identified from *N. gruberi* NEG-M, *A. anophagefferens*, and bacteria were all intron free in their structures (Fig. 2).

Taken together, the phylogenetic and gene structural analyses revealed that the gene structure was the driving force of clustering pattern of the novel nonplant GSTFs with exception of those sequences from *N. gruberi* NEG-M. This was substantiated by the finding from exon/intron organization and intron phases, which were highly conserved among genes encoding protein sequences clustering together in the phylogenetic trees (Fig. 2).

Protein sequence analyses further demonstrate that the identified nonplant GSTs belong to Phi class. Full-length protein sequence pairwise comparison of nonplant GSTF candidates, shown in Figure 2, indicated that the overall percentage identity among Ascomycota members ranged from 34% to 96% with the exception of sequences in clade III, especially one from *Cadophora* sp. DSE1049 v1.0 with a JGI number 658295, which was markedly less similar (28%–39% identity with sequences in either clade). The 56 Ascomycota GSTF candidates were 29%–52%, 32%–45%, 31%–57%, 39%–66%, and 28%–58% similar to their homologs from bacterial clade II, *A. anophagefferens* CCMP 1984, *N. gruberi* NEG-M, basidomycetes, and mainstream plant GSTF sequences, respectively, except OsGSTF6, OsGSTF14, and AtGSTF14, which showed great unlikeness to all homolog sequences with which they were compared. Likewise, bacterial GSTF homologs in clade I shared sequence similarity of 53%–100% among themselves, 29%–42% with those in clade II, 35%–40% with *A. anophagefferens* sequence, 32%–53% with those from *N. gruberi* NEG-M, and 45%–61% with rice GSTFs, excluding OsGSTF6 and OsGSTF14, whereas *N. gruberi* NEG-M class Phi GST homologs were 56%–90% identical among them, 62%–71% similar to clade V Ascomycota homologs, and 38%–62% identical to plant GSTFs apart from OsGSTF6 and OsGSTF14. In brief, pairwise protein sequence comparisons revealed high sequence similarities between most of rice GSTFs and their nonplant homologs, up to 61% with bacterial sequences, 58% with Ascomycota sequences, and 60% with *N. gruberi* NEG-M sequences, though the highest similarities with the latter was found to be GSTFs from the deciduous conifer *L. kaempferi* (up to 62% with LkGSTF1). The exception was the sequence from *A. anophagefferens*, which showed only 45% as the highest identity to the ascomycete *Symbiotaphrina kochii* v1.0 homolog (jgi|Symko1|870293), while its plant closest homolog was the moss *P. patens* PpGSTF9 (39% identity). Globally, GSTF homolog sequences from *N. gruberi* NEG-M showed astonishingly high similarity with all plant GSTFs analyzed (from 46% to 62%), except OsGSTF6 and OsGSTF14 as mentioned above, 61% identity to AtGSTF10 and PtGSTF8 in *A. thaliana* and *P. trichocarpa*. Pairwise protein identities of putative nonplant GSTFs as compared with typical plant GSTFs are shown in Figure 3. The evidence from phylogenetic clustering and sequence similarities between the newly identified nonplant GSTFs and the typical plant GSTFs clearly shows that they belong to the same class. The divergence in pairwise identity among newly identified GSTFs and their plant homologs predicts functional and evolutionary diversification along with their hosting genomes.

Furthermore, side-by-side comparison of nonplant GSTFs aligned with AtGSTF2 and PtGSTF1, for which

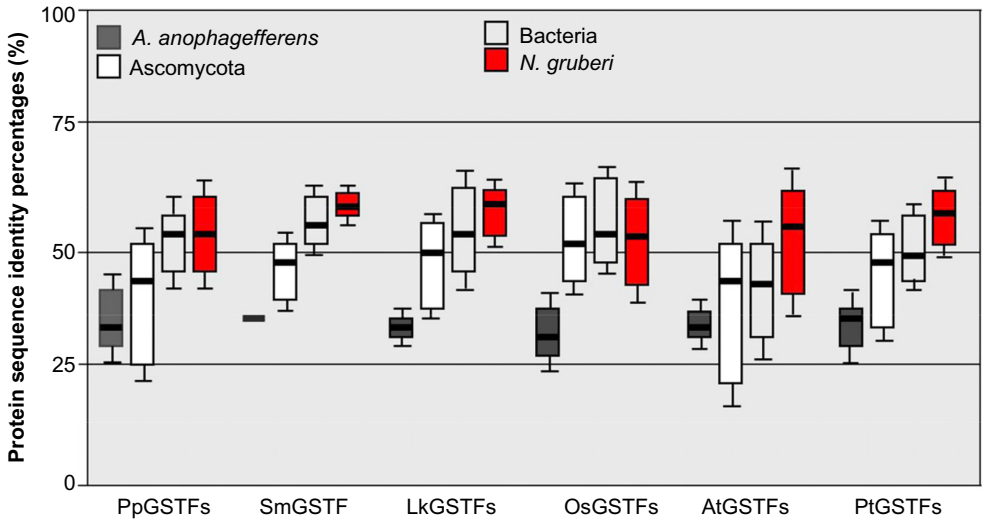

Figure 3. Protein pairwise comparison of Phi class GSTs from various taxonomic groups. Full-length protein pairwise comparison was performed among putative nonplant Phi class GSTs from a bloom alga *A. anophagefferens*, ascomycete fungi, bacteria, a protozoan *N. gruberi,* and mainstream plant Phi class GSTs from moss (PpGSTFs), spikemoss (SmGSTFs), Larix (LkGSTFs), rice (OsGSTFs), Arabidopsis (AtGSTFs), and poplar (PtGSTFs). The tubes delimit the highest and lowest pairwise similarity percentages, and heavy black lines indicate average similarity percentages.

crystal structures are solved,[13,16] pointed out that amino acid residues reported to be involved in catalysis and GSH binding (G-site) in either of the two structures were highly conserved in the novel nonplant GSTFs, except those uncharacterized GSTs (Fig. 4). Close examination of signature motif in the active site characteristic for thioredoxin superfamily revealed 11 different types of active site motif in nonplant GSTFs, namely, STNT, ST/SAT/S, STCT, STCV/A, SXCT, STS/A, STRT, ATCT, ST/IYT, TTCS, and SVNA (Fig. 4 and Supplementary Fig. 2). In contrast to the first three, which are common in plants,[13] the other eight were not found in plant GSTFs and were scattered in different phylogenetic clusters, except SXCT and STS/A, which were exclusively associated with GSTFs of ascomycete clade II and clade V with conserved intron phase of 22 (Fig. 2). This tremendous divergence of active site motif of these GSTFs may explain the low bootstrap support of ancestral nodes of the identified GSTFs (Figs. 1 and 2). The G-site-binding residues Lys41, Val/Ile54, Glu66, Ser67, and Arg68 of AtGSTF2 and their corresponding residues in PtGSTF1 are conserved thoroughly in identified nonplant GSTFs, except for Lys41 in a few sequences. His40/Gln42 in AtGSTF2 and PtGSTF1, respectively, are also conserved at the same positions of the nonplant GSTFs. The remarkable difference between identified nonplant GSTF sequences and these two plant GSTFs was the substitution of Gln53/54 in AtGSTF2/PtGSTF1, which is dominant in plant GSTFs, by lysine (K) in the bloom algae and ascomycete clades I, II, V, and part of clade III, and by valine (V) in GSTFs of myxobacteria, *N. gruberi*, and some remaining ascomycete sequences that did not contain glutamine as plants (Fig. 4). Strikingly, the AtGSTF2 residue Phe123, which is functionally important in binding bacterial and plant natural compounds,[26] is highly conserved in nonplant GSTFs, covering 88% of total identified sequences. It is found in all bacterial GSTFs, all *N. gruberi* GSTFs, and ~43% of ascomycetes GSTFs, especially those in clades II, III, and V. More importantly, GSTF catalytic residue (Ser11 for AtGSTF2) was conserved throughout nonplant GSTFs with a few exceptions (residues in red background in Fig. 4). In addition, other important residues such as those participating in substrate interaction (H-site) and monomer dimerization were conserved as well. The variation in active site signature in combination with the substitution of binding residues in G-site and the difference in substrate binding residues may not only account for the diversity of nonplant GSTF ligands but also reflect functional closeness among the majority of these proteins.

Protein partner prediction reveals functional divergence of nonplant GSTFs. Functions of nonplant GSTFs were predicted based on their predicted association partners provided by STRING V9.1 database that amalgamates information from diverse data such as gene neighborhood, gene fusion, gene cooccurrence, gene coexpression, and protein–protein interactions. A set of proteins were selected as the representatives of bacteria, harmful bloom alga, protozoan, and different clades of ascomycete fungi. They were *Sorangium cellulosum* GSTF (IMG Id 641349040) for all Myxobacteria bacterial GSTFs and *A. anophagefferens* GSTF, *G. zeae* GSTF (UniProtKB ID I1RLP9) for ascomycete clade I, *Laccaria bicolor* GSTF (jgi|Penac1|365773) and *P. chrysosporium* GSTF (jgi|Phchr2|2971755) for clade II symbiont and saprophyte fungi, respectively, *A. niger* CBS 513.88 GSTF (jgi| Aspni_DSM_1|160290) and *Neosartorya fischeri* (UniProtKB ID A1DKY4) for clade III, and *Coccidioides posadasii* GSTF (UniProtKB ID C5PF64) for clade IV, and *A. terreus* GSTF (jgi|Aspte1|4639), *Fusarium graminearum*

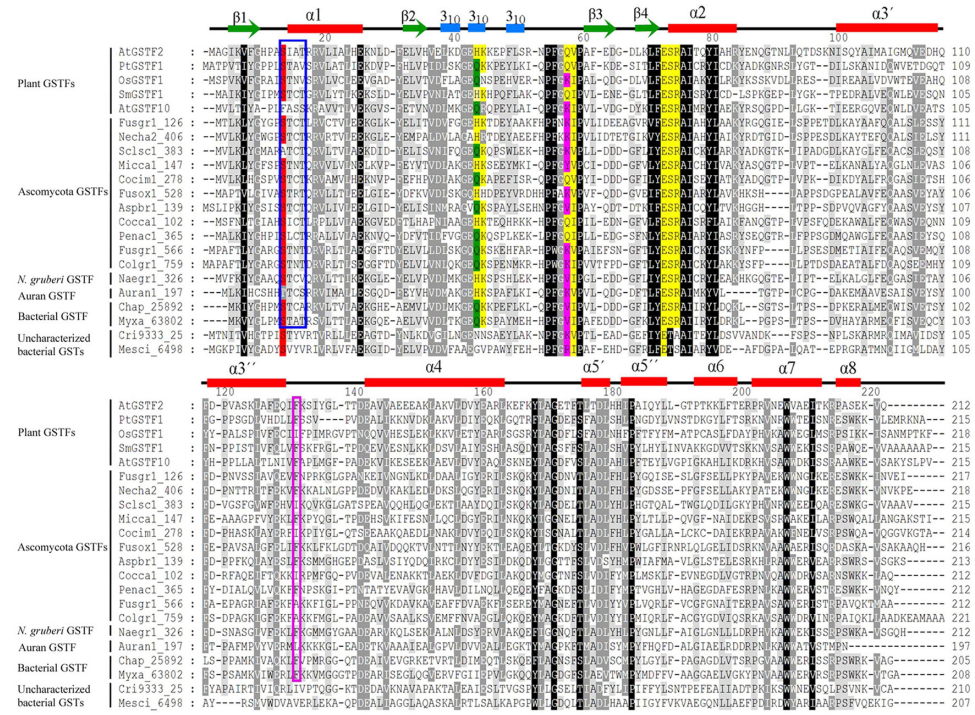

Figure 4. Multiple sequence alignment of representatives of plant GSTFs with their nonplant homologs. Five plant GSTFs, AtGSTF2, PtGSTF1, OsGSTF1, SmGSTF1, and AtGSTF10, were subjected to this alignment analysis.

Notes: AtGSTF2 secondary structure is shown above the alignment. α-Helices, 3_{10}-helices, and β-strands are represented as red cylinders, blue cylinders, and green arrows, respectively. Conserved residues are shaded gray and black, while catalytic residues are marked in red. Conserved GSH-binding residues in AtGSTF2 and PtGSTF1 are shown in yellow and green background. Highlighted in pink are conserved residues in GSH-binding position of plant GSTFs whose crystal structures are not yet solved, whereas the blue and purple boxes indicate active site motifs and the ligand-binding residue corresponding to Phe123 of AtGSTF2. The "Auran GSTF" indicates a GSTF sequence from *A. anophagefferens*.

GSTF (jgi|Fusgr1|12619), and *Sclerotinia sclerotiorum* GSTF (Sclsc1|3839) for clade V ascomycete GSTFs with conserved intron phases of 22, 10, and 02, respectively. *S. sclerotiorum* GSTF (Sclsc1|3839) was also used as the representative of the protozoan *N. gruberi* GSTF as well (Supplementary Table 4). Prediction results showed that glutathione reductases and glutathione peroxidases were ubiquitously associated with all GSTF sequences analyzed at high confidence score, except the clade V GSTFs with conserved intron phase of 22 whose partners were only members of GST classes EF1B and Ure2p (Fig. 5 and Supplementary Table 4). Gamma-glutamyl transferase was often retrieved by more than one copy with high confidence as a partner of all analyzed GSTFs, except clade I ascomycete fungal GSTFs. Glutathione synthetase was also widely present in the list of partners of all analyzed GSTFs, except clade I and clade II mycorrhizal group Ascomycete fungal GSTFs. This result implies that nonplant GSTFs may have a role in secondary metabolism during perturbation of redox homeostasis. In addition, unlike bacterial GSTFs, ascomycete fungal GSTFs generally associated with other classes of GSTs, dominated by EF1B and GTT1. This indicates that Ascomycete fungal GSTFs function along with other classes of GSTs such as EF1B and GTT1. The other predicted partners included transporter proteins for almost all ascomycete fungal GSTFs, plant cell wall-degrading enzymes for bacterial GSTFs, and a set of clade V GSTFs with conserved intron phase of 02 and nucleotide-modifying enzymes Uracil-DNA glycosylase and endonuclease III for bacterial GSTFs, suggesting that these GSTFs may be involved in detoxification and plant infection, respectively (Supplementary Table 4). From these data, it is discernable that nonplant GSTFs may have divergent functions and are apparently responsive to various stresses.

Discussion

GSTFs have long been considered as plant specific until they have recently been identified in basidiomycete fungi.[16] Nonetheless, how broadly the GSTFs are distributed in the nonplant taxonomic groups beyond basidiomycete fungi remains unclear. This has been addressed in the present study. We systemically identified 131 full-length GSTF homologs in ascomycete fungi, bacteria, the protist *N. gruberi*, and the harmful bloom alga *A. anophagefferens*. The majority of these sequences contain Phi class-specific GST_N domain, carry the G-site and H-site, and cluster with the canonical plant GSTFs in the phylogenetic tree. Our results reveal that GSTFs exist widely in nonplant taxonomic groups.

From 250 ascomycete fungi in the MycoCosm database, 85 were found to contain GSTF homologs in their genomes. The Ascomyceta species possessing GSTFs exhibit diverse

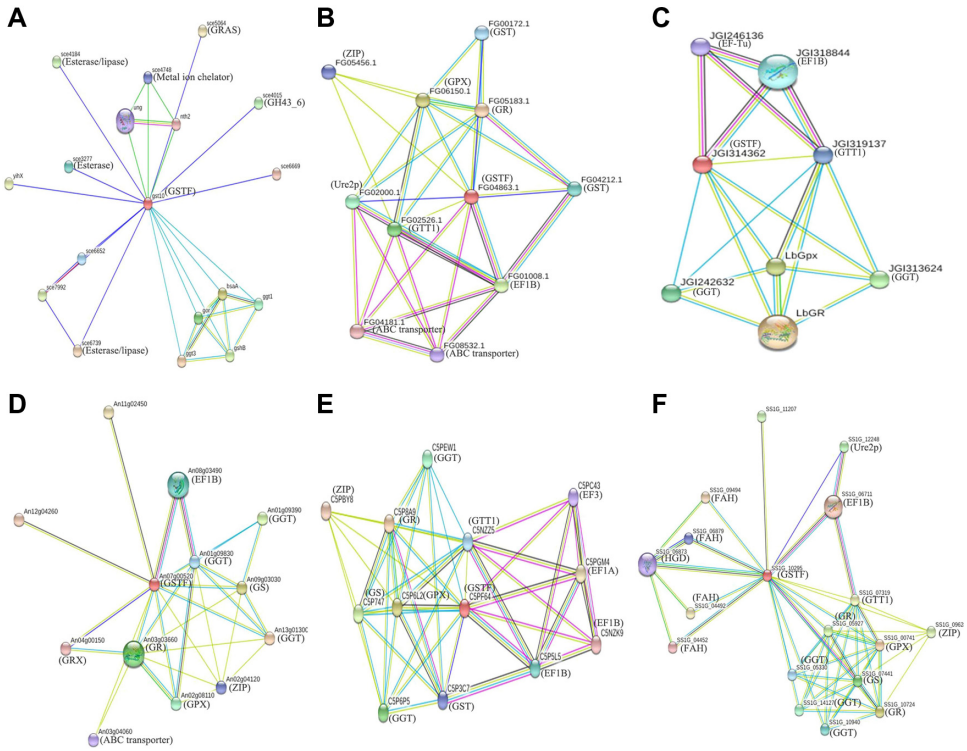

Figure 5. Predicted functional partners of nonplant GSTF representatives from different phylogenetic groups. Partners of GSTFs from the bacterium *Sorangium cellulosum* (**A**) and the ascomycete fungi *Fusarium graminearum* (**B**), *Laccaria bicolor* (**C**), *Aspergillus niger* (**D**), *Coccidioides posadasi* (**E**), and *Sclerotinia sclerotiorum* (**F**) are shown. (**B**–**F**) represent ascomycete GSTFs in clades I–V.

lifestyles such as saprotrophs, mycorrhizae, and symbionts of a wide range of plants, mycangial associates of insects, as well as pathogens and parasites of virtually all groups of organisms, including bacteria, plants, other fungi, and animals.[27] All these lifestyles are associated with extremely large gene copy numbers encoding oxidoreductases and carbohydrate-active enzymes related to plant cell wall degradation, including class II peroxidases.[28] These enzymes are known to strategically catalyze reactions that lead to inevitable production of reactive oxygen species (ROS).[29] To prevent oxidative damage, aerobic organisms have evolved antioxidant systems comprising high efficient enzymes for dismutation of ROS, such as heme catalases, manganese catalases, and catalase/peroxidases, as well as electron ROS-reducing peroxidases conjugated to electron donors, including heme peroxidases and nonheme peroxidases such as ascorbate peroxidases, various types of peroxiredoxins, glutathione/thioredoxin peroxidases (GPX), and GSTs.[30,31] Therefore, it is not surprising that ascomycete fungi and other organisms mentioned above might possess GSTFs, given their crucial role in ROS-induced defense responses during oxidative stresses. GSTs are involved in phase II detoxification reactions and GSTFs have been reported to have a role in protection against oxidative damage.[7] This is in accordance with increasing evidence that some plant GSTFs function as GPX to prevent oxidative injury through direct reduction of reactive organic hydroperoxides such as those resulting from lipid peroxidation. Moreover, it is has been noticed that most GSTFs tested displayed GPX activity regardless of their source.[13,32]

We found that GSTFs were widespread in Leotiomycetes (Supplementary Table 3). This class encompasses fungi that colonize a large variety of habitats, including parasites, saprobes, endophytes, and symbionts of various plants, and even marine inhabitants.[33] More than 77% (21 out of 27) of GSTF sequences identified in this fungal class are from mycorrhizal fungi, among which 57.1% (12 out of 21) are contributed by the so-called *Rhizoscyphus ericae* aggregate, which are composed of *R. ericae* UAMH 7357 v1.0, *Meliniomyces variabilis* F v1.0, *M. bicolor* E v2.0, and *Cadophora sp.* DSE1049 v1.0.[34] Moreover, it is known that mycorrhizal symbiosis establishment is empowered by the fungal ability to respond to plant defensive deployment of flavonoids and strigolactones during the initial stage of fungal colonization.[35,36] It is important to mention that besides ericoid mycorrhizal ascomycetes, *GSTF* genes were also found to occur, often in more than one copies, among plant endosymbionts such as *Ascocoryne sarcoides* NRRL50072, *Ilyonectria sp, Oidiodendron maius* Zn, and *Trichoderma harzianum* CBS, while they have been lost in all ectomycorrhizal fungi analyzed, especially those in Pezizomycetes class. This suggests that *GSTF* gene distribution and abundance among these species are somehow correlated with the fungal lifestyles. Furthermore, given that some plant GSTFs are recognized in acting as carrier proteins of defense-related compounds, including flavonoids and

indole-derived phytoalexin among others,[26,37,38] it is tempting to assume that the prevalence of *GSTF* genes in mycorrhizal fungi may be suggestive of their important role in success of the fungal lifestyle. Interestingly, GSTFs were identified in a significant number of ascomycetes from different classes such as pathogenic species in *Aspergillus*, *Fusarium, Alternaria*, *Penicillium*, and *Cochliobolus* genera, to name a few, known to produce a plethora of bioactive secondary metabolites such as antibiotics, hormones, and toxins, which are of significant human health, veterinary, and agricultural importance. Even more strikingly, we found that GSTFs exist in animal pathogenic fungi, including dermatophytes *T. rubrum* and *M. canis*, as well as the Valley fever causal agent *C. immitis* and its close nonpathogenic relative *U. reesii*. The genomes of these dermatophytes and *Coccidioides* were revealed to be rich in genes encoding regulators of secondary metabolism as well.[39,40]

GSTFs are present in ascomycete fungi associated with stressful environments and those with capacity to degrade noxious compounds as well. For example, two genes encoding full-length GSTF proteins exist in the genome of *T. virens* Gv29–8 v2.0 (Supplementary Table 3), which is known to be able to degrade exogenous hazardous compounds, including pesticides, polyphenols, and polyaromatic hydrocarbons, and to sequester heavy metals.[41] Similarly, *GSTF* genes are present in species of *Aureobasidium* genus that can grow in such highly radioactive environments in addition to their ability to tolerate other various stresses such as extremely low temperature, high temperature, and high salinity.[42] Curiously, a *GSTF* gene was found in an endosymbiont of beetle fungi *S. kochii*, which is beneficial to the insect, thanks to its enzymes that degrade diverse hazardous compounds, including flavonoids, tannic acid, phenol, plant meal toxins, certain mycotoxins and insecticides, and herbicides.[43] *Nectria haematococca* genome, a member of *Fusarium solani* complex, was found to consist of three *GSTF* genes. Fungi in *F. solani* complex are adaptive to various habitats, including extreme environments such as highly radioactive inner parts of the damaged nuclear reactor at Chernobyl, and tolerant to many compounds shown to be toxic to other fungi such as antibiotics, heavy metals, and metabolic poisons.[44] Therefore, occurrence of *GSTF* genes in these fungi is in line with the degree of stress exposure in their living environments. Given their distribution in ascomycetes with diversified habitats and lifestyles, fungal GSTFs obviously respond to a wide range of stresses and likely have various functions. In plants, GSTFs play different roles, including detoxification of noxious compounds, involvement in biosynthesis and transport of secondary metabolites, and acting as peroxidases, as well.[13]

As for bacteria, putative true GSTFs were only identified in myxobacteria. This is in accord with the lifestyle of these bacteria. Living mainly in soil, myxobacteria are found everywhere, in all climate zones, preferentially in places rich in microbial life and organic matter such as rotting plant material, dung of various animals, and on the bark of living and dead trees.[45] Myxobacteria are notable producers of secondary metabolites with antibiotic or cytotoxic activities and are considered as natural pharmaceutical factories.[46] It is noteworthy to mention that GSTFs have been proved to bind bacterial natural product as well.[26] In a recent study, a sequence from *Rhizobium meliloti* (UnprotKB accession number Q92Q06) was shown to belong to Phi-like GST group (main.5 of subgroup level 2 GSTs) and have activity toward isothiocyanate.[47] However, none of the sequences from Rhizobiaceae, including *Sinorhizobium/Ensifer* group, clustered with the mainstream GSTFs; they rather formed their own group of unknown GST class (Fig. 1). The position of these sequences in the phylogenetic tree is further supported by protein sequence comparison (Fig. 4). Unlike to seemingly true myxobacterial GSTFs, the bacterial uncharacterized GSTs used in this study apparently do not belong to Phi class, but seem to be members of closely related unknown class, though further studies such as crystal structure determination are required. Moreover, cyanobacterial GST class Chi also showed high catalytic activities toward naturally occurring isothiocyanates.[48] Besides fungi and bacteria, GSTFs are rare in protists, and the species found to have them are associated with peculiar lifestyle compared to their closely related species. *A. anophagefferens* blooms when levels of light and inorganic nutrients are low while organic matter and metal concentrations are elevated,[49] whereas *N. gruberi* is a bacterial predator.[50]

Altogether, the difference in distribution and gene copies of GSTF was obviously significant among nonplant organisms analyzed (Supplementary Tables 2 and 3). GSTF existence in myxobacteria and protists *A. anophagefferens* CCMP 1984 and *N. gruberi* EG-M is correlated with their habitats and genome sizes, which is in accordance with the reported distribution and abundance pattern in basidiomycetes.[16] Myxobacteral genomes are among the largest bacterial genome sequenced to date.[51] *A. anophagefferens* genome was also reported to be larger than those of other competing algae,[49] and the eight *GSTF* genes in *N. gruberi* may correspond with the reported distinctive features of its genome.[50] Unlike basidiomycetes, bacteria, and protists, the distribution and gene abundance of GSTFs among ascomycete fungi appear to be cryptic since they do not match with the fungal lifestyle and/or genome size. This is most remarkable in the order of Dothideomycetes, which contains the most important plant pathogens worldwide. For example, the analyzed species of *Cochliobolus* genus possess *GSTF* genes except the two *Cochliobolus heterostrophus* strains (C4 and C5), while *C. heterostrophorus* C5 possesses even larger genome size than the remaining six fungi. *Leptosphaeria maculans* and species of *Mycosphaerella* genus are among the plant pathogens with large genomes, but they do not have GSTF as well. Similarly, Pezizomycetes, lacking *GSTF* genes as noted earlier, are among the largest ascomycete genomes (Supplementary Table 3 and JGI database for genome size details). However, the correlation could be spotted with lifestyle such

as GSTF expansion in ericoid mycorrhizal and endosymbiotic. Additionally, GSTFs were distributed in a considerable number of plant pathogenic species, which are known to produce a range of secondary metabolites.

Phylogenetic analysis of nonplant GSTFs together with their plant and basidiomycete counterparts showed different clustering patterns, which depicts a different evolution history (Figs. 1 and 2) from a single gene inherited from the last common ancestor bacterial genome. In contrast to GSTFs from bacteria and *N. gruberi* that made a single clade each, different clades in ascomycete GSTFs, each containing sequences from unrelated species, are most likely indicative of independent evolution from different lineages. Estimated evolutionary distances (Supplementary Table 5) revealed that while plant GSTF evolution preceded that of their most nonplant homologs, ascomycete *GSTF* genes in clade III might have appeared earlier and/or evolved concurrently with plant GSTFs (Figs. 1 and 2) since it groups sequences from different fungal classes with various habitats and divergent lifestyles. On the other hand, clades IV and V are likely the most recently evolved considering their position in the phylogenetic tree and the source of sequences it clusters, as well as their gene configuration (Figs. 1 and 2 and Supplementary Table 3), which is typically similar to that of plant GSTFs.[11] This is consistent with the previous studies that have illustrated the repeated independent evolution of fungal lifestyles.[36,52] The conservation of exon/intron configuration and intron phase patterns in each clade (Fig. 2) supports the common origin of fungal GSTFs, which is in line with the belief that conserved intron positions and phases were gained only once in evolution.[53] Phylogenetic clustering patterns of proteins can also provide an insight into the functional relationships. Generally, phylogenetic divergence of nonplant GSTFs may imply functional diversification. As evidence, GSTFs from the same ascomycete fungi clustered in separate clades. This pattern was in contrast to that of basidiomycete GSTFs in which sequences from the same species clustered together.[17] In addition, there are obvious similarities and residue conservation within GSTF protein sequences in the same clade. Interestingly, bacterial GSTFs share with plants the STCT and STAT active site motifs, signatures characteristic for thioredoxin superfamily.[13] For example, all bacterial GSTFs from Myxococcaceae contain STAT motif signature (Fig. 4), which is also found in Arabidopsis AtGSTF6, AtGSTF7, AtGSTF8, and poplar PtGSTF1. Similarly, STCT, which is found in the remaining bacterial sequences and a set of ascomycete GSTFs clade V, along with STNT in clade I, is also extensively represented in plants.[13] Moreover, the amino acid at position 53 (Gln54 in this study) is conserved in the alignment with Arabidopsis AtGSTF2 (Fig. 4), while the dominant Gln54 in plant GSTFs is replaced by lysine (K) in most sequences or valine in nonplant GSTFs. More importantly, the residue revealed to be responsible for ligand binding in plant GSTFs is extremely conserved in nonplant GSTFs as well. Mutagenesis studies in Arabidopsis AtGSTF2 revealed that Phe123 residue was important in binding both plant and bacterial heterocyclic compounds.[26] This residue is found at the same position in all bacterial GSTFs, all *N. gruberi* GSTFs, and ~43% of total GSTFs identified in ascomycetes, especially those of clades II, III, and V (purple box in Fig. 4). It is possible that plant and nonplant GSTFs with the same characteristic sequence features may have similar functions. Therefore, their role in secondary metabolism and binding and transport of natural products is compelling, but this remains to be determined. It is noteworthy that a group of GSTFs in clade V seem to have lost one residue in their catalytic site motif (Supplementary Fig. 2). This feature is associated only with ascomycete GSTFs, which have apparently recently diverged from their clade counterparts, as reflected in their gene structures with intron phase of 22 (Fig. 2). This recent gene rearrangement of GSTFs in these fungi may indicate either acquisition of new function or loss of function.

Information from STRING database provided further insights into the functional diversification of nonplant GSTFs. Partner prediction results revealed that nonplant GSTFs belonging to distinct phylogenetic clusters contain both conserved and cluster-specific functional partners. The most conserved partners are glutathione reductases, glutathione peroxidases, glutathione synthetases, and gamma-glutamyltranspeptidase (Supplementary Table 4). All GSTFs under analysis from bacteria, protists, and ascomycete fungi are predicted to associate with these enzymes, except ascomycete fungal GSTFs of clade I, the mycorrhizal group of clade II, and the group with an intron phase of 22 of clade V, which lack one or two of the enzymes, (Supplementary Table 4). These enzymes are involved in glutathione modification and cycling, cellular redox regulation, and secondary metabolism in both fungi and plants,[54–56] suggesting that nonplant GSTFs generally play an important role following redox balance shift during oxidative stresses and secondary metabolism. Except for these highly conserved partners for almost all nonplant GSTFs, there are some taxonomy- or even clade-specific partners. For example, non-Phi classes of GSTs, dominated by EF1B, GTT1, and some unclassified GSTs, are generally predicted to be partners of ascomycete fungal GSTFs. However, no non-Phi class of GST seems to be a partner of bacterial GSTFs. Similarly, ascomycete fungal GSTFs are predicted to partner with various transporters, whereas bacterial GSTFs are not. In contrast, they associate with DNA-modifying enzymes, which are not associated with ascomycete fungal GSTFs (Supplementary Table 4). These data indicate the functional difference of GSTFs in ascomycete fungi and bacteria. Additionally, a group of GSTFs of clade V ascomycetes, which contain an intron phase profile of 22, are unique in that they have only three partner proteins belonging to two other classes of GSTs (EF1BG and Ure2p) (Supplementary Table 4). This occurs for all members of this group since we

obtained identical prediction results when all other members of this group were input in STRING database. To explain this, we pairwise compared GSTFs of this group and other nonplant and plant GSTFs and found that the fourth amino acid of the active site motif is deleted in this group of GSTFs (Supplementary Fig. 2). This result indicates that GSTFs in this group, which seem to be the most recently diverged of all ascomycete GSTFs, may have unique functions that are distinct from other GSTFs.

Conclusion

It is apparent that the *GSTF* gene exists widely in nonplant kingdoms or phyla, including bacteria, protists, and ascomycete fungi, besides basidiomycetes. GSTFs in Myxobacteria, *N. gruberi*, and the harmful bloom alga *A. anophagefferens* are likely to have evolved along with genome expansions during evolutionary adaptation to variable stressful habitats. Nonetheless, the distribution of these genes in ascomycete fungi seems not to correlate with their genome sizes and habitats, suggesting the independent evolution among different fungi lineages. However, this complexity is consistent with the polyphyletic origins over the evolutionary history of fungi, and it is possible that GSTFs play important roles in some fungal lifestyles, as exemplified by their expansion in plant endosymbionts and mycorrhizal fungi, while a complete gene loss is noticed in all ectomycorrhizal asccomycetes examined. From the analyses in this study, we predicted that nonplant GSTFs may have different functions, including detoxification of various exogenous and endogenous deleterious compounds, second metabolite biosynthesis and/or transport, and serving as peroxidases to remove products of oxidative stress and lipid peroxidation caused by different stresses. This study significantly extends the scope of taxonomic groups that contain Phi class GSTs and provides insights in new research directions toward understanding the roles of GST in abiotic and biotic stress responses.

Author Contributions

Conceived and designed the experiments: J-PM, X-ZC. Analyzed the data: J-PM, Y-PX. Wrote the first draft of the manuscript: J-PM. Contributed to the writing of the manuscript: J-PM, X-ZC. Agreed with manuscript results and conclusions: J-PM, YP-X, X-ZC. Jointly developed the structure and arguments for the paper: J-PM, Y-PX, X-ZC. Made critical revisions and approved the final version: J-PM, X-ZC. All the authors reviewed and approved the final manuscript.

Supplementary Material

Supplementary Table 1. List of plant and basidiomycete GSTF sequences and the representatives of other known GST classes used in this study.

Supplementary Table 2. List of putative GSTF sequences of bacteria and Heterokonta retrieved during Blast searches.

Supplementary Table 3. List of putative GSTF sequences of the Ascomycota fungi and the amoeboflagellate *Naegleria gruberi* retrieved during Blast searches.

Supplementary Table 4. Functional partners of representatives of nonplant GSTF clades predicted by STRING database.

Supplementary Table 5. Estimation of evolutionary divergence between GSTF proteins.

Supplementary Figure 1. Domain composition of all full-length putative GSTF protein sequences identified in bacteria, protists, and Ascomycota fungi.

Supplementary Figure 2. Eleven different signature motifs in the catalytic site of nonplant GSTFs.

REFERENCES

1. Pickett CB, Lu AYH. Glutathione S-transferases: gene structure, regulation, and biological function. *Annu Rev Biochem*. 1989;58:743–64.
2. Jacquot J, Dietz KJ, Rouhier N, et al. Redox regulation in plants: glutathione and "redoxin"-related families. In: Jakob U, Reichmann D, eds. *Oxidative Stress and Redox Regulation*. Dordrecht: Springer Science and Business Media; 2013:213–31.
3. Sheehan D, Meade G, Foley VM, Dowd CA. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem J*. 2001;360:1–16.
4. Dixon DP, Edwards R. Glutathione transferases. *Arabidopsis Book*. 2010;8:e0131.
5. Liu YJ, Han XM, Ren LL, Yang HL, Zeng QY. Functional divergence of the glutathione S-transferase supergene family in *Physcomitrella patens* reveals complex patterns of large gene family evolution in land plants. *Plant Physiol*. 2013;161:773–86.
6. Lallement PA, Brouwer B, Keech O, Hecker A, Rouhier N. The still mysterious roles of cysteine-containing glutathione transferases in plants. *Front Pharmacol*. 2014;5:192.
7. Marrs KA. The functions and regulation of glutathione S-transferases in plants. *Annu Rev Plant Physiol Plant Mol Biol*. 1996;47:127–58.
8. Dixon DP, Davis BG, Edwards R. Functional divergence in the glutathione transferase superfamily in plants. Identification of two classes with putative functions in redox homeostasis in *Arabidopsis thaliana*. *J Biol Chem*. 2002;277:30859–69.
9. Dixon DP, Steel PG, Edwards R. Roles for glutathione transferases in antioxidant recycling. *Plant Signal Behav*. 2011;6:1223–7.
10. Kumar S, Asif MH, Chakrabarty D, Tripathi RD, Dubey RS, Trivedi PK. Expression of a rice lambda class of glutathione S-transferase, OsGSTL2, in *Arabidopsis* provides tolerance to heavy metal and other abiotic stresses. *J Hazard Mater*. 2013;24(8–249):228–37.
11. Labrou NE, Papageorgiou AC, Pavli O, Flemetakis E. Plant GSTome: structure and functional role in xenome network and plant stress response. *Curr Opin Biotechnol*. 2015;32:186–94.
12. Frear DS, Swanson HR. Biosynthesis of S-(4-ethylamino-6-isopropylamino-2-s-triazino) glutathione: partial purification and properties of a glutathione S-transferase from corn. *Phytochemistry*. 1970;9:2123–32.
13. Pégeot H, Koh CS, Petre B, et al. The poplar phi class glutathione transferase: expression, activity and structure of GSTF1. *Front Plant Sci*. 2014;5:712.
14. Lee SH, Li CW, Koh KW, et al. MSRB7 reverses oxidation of GSTF2/3 to confer tolerance of *Arabidopsis thaliana* to oxidative stress. *J Exp Bot*. 2014;65:5049–62.
15. Schultz T, van Eck L, Botha AM. Phi-class glutathione-S-transferase is involved in Dn1-mediated resistance. *Physiol Plant*. 2015;154:1–12.
16. Reinemer P, Prade L, Hof P, et al. Three-dimensional structure of glutathione S-transferase from *Arabidopsis thaliana* at 2.2 A resolution: structural characterization of herbicide-conjugating plant glutathione S-transferases and a novel active site architecture. *J Mol Biol*. 1996;255:289–309.
17. Morel M, Meux E, Mathieu Y, et al. Xenomic networks variability and adaptation traits in wood decaying fungi. *Microb Biotechnol*. 2013;6:248–63.
18. Morel M, Ngadin AA, Droux M, Jacquot JP, Gelhaye E. The fungal glutathione S-transferase system. Evidence of new classes in the wood-degrading basidiomycete *Phanerochaete chrysosporium*. *Cell Mol Life Sci*. 2009;66:3711–25.
19. McGoldrick S, O'Sullivan SM, Sheehan D. Glutathione transferase-like proteins encoded in genomes of yeasts and fungi: insights into evolution of a multifunctional protein superfamily. *FEMS Microbiol Lett*. 2005;242:1–12.
20. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.

21. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*. 1992;8:275–82.
22. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–9.
23. Yang Q, Liu YJ, Zeng QY. Biochemical functions of the glutathione transferase supergene family of *Larix kaempferi*. *Plant Physiol Biochem*. 2014;77:99–107.
24. Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013;41(Database issue):D808–15.
25. Hayes JD, Flanagan JU, Jowsey IR. Glutathione transferases. *Annu Rev Pharmacol Toxicol*. 2005;45:51–88.
26. Dixon DP, Sellars JD, Edwards R. The *Arabidopsis* phi class glutathione transferase AtGSTF2: binding and regulation by biologically active heterocyclic ligands. *Biochem J*. 2011;438:63–70.
27. Lutzoni F, Kauff F, Cox CJ, et al. Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. *Am J Bot*. 2004;91:1446–80.
28. Alfaro M, Oguiza JA, Ramirez L, Pisabarro AG. Comparative analysis of secretomes in basidiomycete fungi. *J Proteom*. 2014;102:28–43.
29. Ruiz-Dueñas FJ, Lundell T, Floudas D, et al. Lignin-degrading peroxidases in polyporales: an evolutionary survey based on 10 sequenced genomes. *Mycologia*. 2013;105:1428–44.
30. Fawal N, Li Q, Savelli B, et al. PeroxiBase: a database for large-scale evolutionary analysis of peroxidases. *Nucleic Acids Res*. 2013;41(Database issue):D441–4.
31. Shalaby S, Horwitz BA. Plant phenolic compounds and oxidative stress: integrated signals in fungal-plant interactions. *Curr Genet*. 2014;61:347–57.
32. Cummins I, Cole DJ, Edwards R. A role for glutathione transferases functioning as glutathione peroxidases in resistance to multiple herbicides in black-grass. *Plant J*. 1999;18:285–92.
33. Gnavi G, Ercole E, Panno L, Vizzini A, Varese GC. Dothideomycetes and leotiomycetes sterile mycelia isolated from the Italian seagrass *Posidonia oceanica* based on rDNA data. *Springerplus*. 2014;3:508.
34. Vrålstad T, Fossheim T, Schumacher T. *Piceirhiza bicolorata*-the ectomycorrhizal expression of the *Hymenoscyphus ericae* aggregate? *New Phytol*. 2000; 145:549–63.
35. Kloppholz S, Kuhn H, Requena N. A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Curr Biol*. 2011;21:1204–9.
36. Jung SC, Martinez-Medina A, Lopez-Raez JA, Pozo MJ. Mycorrhiza-induced resistance and priming of plant defenses. *J Chem Ecol*. 2012;38:651–64.
37. Mueller LA, Goodman CD, Silady RA, Walbot V. AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. *Plant Physiol*. 2000;123:1561–70.
38. Sun Y, Li H, Huang JR. *Arabidopsis* TT19 functions as a carrier to transport anthocyanin from the cytosol to tonoplasts. *Mol Plant*. 2012;5:387–400.
39. Sharpton TJ, Stajich JE, Rounsley SD, et al. Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. *Genome Res*. 2009;19:1722–31.
40. Martinez DA, Oliver BG, Graser Y, et al. Comparative genome analysis of *Trichophyton rubrum* and related dermatophytes reveals candidate genes involved in infection. *MBio*. 2012;3:e259–12.
41. Harman GE, Howell CR, Viterbo A, Chet I, Lorito M. Trichoderma species–opportunistic, avirulent plant symbionts. *Nat Rev Microbiol*. 2004;2:43–56.
42. Gostinčar C, Ohm RA, Kogej T, et al. Genome sequencing of four *Aureobasidium pullulans* varieties: biotechnological potential, stress tolerance, and description of new species. *BMC Genom*. 2014;15:549.
43. Shen SK, Dowd PF. Detoxification spectrum of the cigarette beetle symbiont *Symbiotaphrina kochii* in culture. *Entomol Exp Appl*. 1991;60:51–9.
44. Coleman JJ, Rounsley SD, Rodriguez-Carres M, et al. The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet*. 2009;5:e1000618.
45. Dawid W. Biology and global distribution of myxobacteria in soils. *FEMS Microbiol Rev*. 2000;24:403–27.
46. Diez J, Martinez JP, Mestres J, Sasse F, Frank R, Meyerhans A. Myxobacteria: natural pharmaceutical factories. *Microb Cell Fact*. 2012;11:52.
47. Mashiyama ST, Malabanan MM, Akiva E, et al. Large-scale determination of sequence, structure, and function relationships in cytosolic glutathione transferases across the biosphere. *PLoS Biol*. 2014;12:e1001843.
48. Wiktelius E, Stenberg G. Novel class of glutathione transferases from cyanobacteria exhibit high catalytic activities towards naturally occurring isothiocyanates. *Biochem J*. 2007;406:115–23.
49. Gobler CJ, Berry DL, Dyhrman ST, et al. Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc Natl Acad Sci U S A*. 2011;108:4352–7.
50. Opperdoes FR, De Jonckheere JF, Tielens AGM. *Naegleria gruberi* metabolism. *Int J Parasitol*. 2011;41:915–24.
51. Barberán A, Ramirez KS, Leff JW, Bradford MA, Wall DH, Fierer N. Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecol Lett*. 2014;17:794–802.
52. Ohm RA, Feau N, Henrissat B, et al. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathog*. 2012;8:e1003037.
53. Putnam NH, Srivastava M, Hellsten U, et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science*. 2007; 317:86–94.
54. Su T, Xu J, Li Y, et al. Glutathione-indole-3-acetonitrile is required for camalexin biosynthesis in *Arabidopsis thaliana*. *Plant Cell*. 2011;23:364–80.
55. Li MY, Liang XF, Rollins JA. *Sclerotinia sclerotiorum* gamma-glutamyl transpeptidase (Ss-Ggt1) is required for regulating glutathione accumulation and development of sclerotia and compound appressoria. *Mol Plant Microbe Interact*. 2012;25:412–20.
56. Spitzmüller Z, Kwon NJ, Szilagyi M, et al. gamma-Glutamyl transpeptidase (GgtA) of *Aspergillus nidulans* is not necessary for bulk degradation of glutathione. *Arch Microbiol*. 2015;197:285–97.