

## Expansion Mechanisms and Functional Divergence of the Glutathione S-Transferase Family in Sorghum and Other Higher Plants

YUNHUA Chi<sup>1,†</sup>, YANSONG Cheng<sup>1,†</sup>, JEEVANANDAM Vanitha<sup>2</sup>, NADIMUTHU Kumar<sup>2,‡</sup>, RENGASAMY Ramamoorthy<sup>3</sup>, SRINIVASAN Ramachandran<sup>1,2,\*</sup>, and SHU-YE JIANG<sup>1,2,\*</sup>

*IOB-TLL Joint R&D Laboratory, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China<sup>1</sup>; Temasek Life Sciences Laboratory, 1 Research Link, the National University of Singapore, Singapore 117604, China<sup>2</sup> and Department of Biological Sciences, the National University of Singapore, Singapore 117543, China<sup>3</sup>*

\*To whom correspondence should be addressed. Tel. +86 65-68727489. Fax. +86 65-68727007.  
E-mail: sri@tll.org.sg (S.R.); shuye@tll.org.sg (S-Y.J.)

Edited by Satoshi Tabata  
(Received 12 July 2010; accepted 21 November 2010)

### Abstract

**Glutathione S-transferases (GSTs) exist in various eukaryotes and function in detoxification of xenobiotics and in response to abiotic and biotic stresses. We have carried out a genome-wide survey of this gene family in 10 plant genomes. Our data show that tandem duplication has been regarded as the major expansion mechanism and both monocot and dicot plants may have practiced different expansion and evolutionary history. Non-synonymous substitutions per site (Ka) and synonymous substitutions per site (Ks) analyses showed that N- and C-terminal functional domains of GSTs (GST\_N and GST\_C) seem to have evolved under a strong purifying selection ( $Ka/Ks < 1$ ) under different selective pressures. Differential evolutionary rates between GST\_N and GST\_C and high degree of expression divergence have been regarded as the major drivers for the retention of duplicated genes and the adaptability to various stresses. Expression profiling also indicated that the gene family plays a role not only in stress-related biological processes but also in the sugar-signalling pathway. Our survey provides additional annotation of the plant GST gene family and advance the understanding of plant GSTs in lineage-specific expansion and species diversification.**

**Key words:** abiotic stress; biotic stress; comparative genomics; functional divergence; glutathione transferase

### 1. Introduction

Glutathione (GSH) is the tripeptide  $\gamma$ -glutamyl-cysteinyl-glycine and plays a central role in the processes of detoxification and redox buffering.<sup>1</sup> Glutathione S-transferases (GSTs, EC. 2.5.1.18) catalyze the conjugation of GSH to an electrophilic substrate.<sup>2</sup> Plant GSTs have been actively investigated during last decades.<sup>2–8</sup> Currently, large numbers of

GST genes have been identified or annotated from at least 17 plant species.<sup>7–9</sup> In *Arabidopsis*, the complete identification in a genome-wide level revealed at least 53 GST genes.<sup>5,10,11</sup> In rice, 59 GST genes have been identified.<sup>12</sup> However, their works were carried out by BLAST searches against the rice expression sequence tag (EST) database; thus, some members may escape from their collection. Recently, 81 GST genes have been genome-widely identified in *Populus trichocarpa* and they exhibit extensive functional diversification.<sup>13</sup> However, no other data have been reported on the genome-wide identification of the GST family, although at least 20 plant genomes

<sup>†</sup> These authors contributed equally to the work.

<sup>‡</sup> Present address: JOil Pte Ltd, 1 Research Link, Singapore 117604.

have been completely sequenced (<http://www.genomesonline.org/gold.cgi>).

Since GST functions have been closely linked to stress responses, expression analyses were carried out under a wide variety of stress conditions. Evidence showed that the transcript of plant GST genes was regulated by various abiotic and biotic stresses as well as hormones including xenobiotic-type stresses, such as herbicide application,<sup>2</sup> chilling,<sup>14</sup> hypoxic stress,<sup>15</sup> dehydration,<sup>16,17</sup> wounding,<sup>18</sup> pathogen attack,<sup>19</sup> ethylene,<sup>20</sup> auxin,<sup>21</sup> 2,4,6-trinitrotoluene,<sup>22</sup> hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) and the defence signal salicylic acid (SA).<sup>23</sup> However, whole family-based expression analyses were carried out only in *Arabidopsis*,<sup>11</sup> *P. trichocarpa*<sup>13</sup> and rice,<sup>24</sup> the transcript profiling of all family members is not available for the other plants.

Although many GST genes have been isolated or annotated, only a small number of them have been functionally characterized. Reports showed that plant GSTs might play important roles in herbicide resistance<sup>2</sup> and detoxification.<sup>25</sup> Overexpression of GSTs in plants improved herbicide and stress tolerance,<sup>26–32</sup> and some of GST genes have been patented.<sup>8</sup> On the other hand, efforts have been put on the understanding of the role of GSTs in endogenous plant developmental processes. For example, studies revealed the role of GSTs in the vacuolar sequestration of anthocyanins in maize, petunia and *Arabidopsis*.<sup>33–36</sup> Evidence also showed that GSTs might function as binding proteins by binding to various hormones including auxin<sup>37–39</sup> and cytokinin<sup>40</sup> as well as porphyrin compounds<sup>41</sup> to regulate their activities.

Generally, although plant GSTs have been discovered for more than 30 years,<sup>5</sup> limited data are available on the genome-wide identification and expression analysis as well as their functional divergence of the GST family. Thus, we do not know the family size in a genome and its member that are involved in biotic and abiotic stress-related biological processes. On the other hand, little is known about how these genes have been evolved or expanded with such functions. In this report, we first genome-wide identified and characterized all GST genes encoded by the sorghum and other nine plant genomes, especially focusing on the Tau and Phi classes, which are the top two subfamilies in the plant kingdom. We then evaluated their expansion and evolutionary mechanisms by investigating their duplication and/or transposition history. Subsequently, we examined their transcript profiling by the full-length cDNA, EST, microarray and massively parallel signature sequencing (MPSS<sup>42</sup>) data sets as well as by RT-PCR and quantitative real-time RT-PCR (qRT-PCR). Finally, we investigated their expression divergence under various stresses to

further annotate their biological functions and retention mechanisms. Our data show that monocot and dicot GST genes exhibit differences in their evolutionary history and they have been involved in lineage-specific expansion and species diversification. Our data further confirmed their functions in biotic and abiotic stress-related developmental processes and also demonstrated that plant GSTs may be involved in the sugar-related signalling pathway.

## 2. Materials and methods

### 2.1. Plant materials and treatments

Grain sorghum (*Sorghum bicolor*) L. Moench cultivar BT × 623 was used for all experiments in this study. Seeds were imbibed in water. After germination, they were planted in greenhouse and were grown under natural light and temperature conditions. The 2-week-old seedlings were used for all treatments. For the drought treatment, seedlings were treated with 30% polyethylene glycol (PEG) and whole plants were collected in various time intervals (0, 0.5 and 2 h), and then frozen in liquid nitrogen for RNA preparation. For salinity and cold treatments, seedlings were subjected to the 250 mM NaCl solution and 4°C conditions, respectively. Samples for RNA extractions were collected in 0, 2 and 8 h time intervals. For glucose and sucrose treatments, seedlings were subjected to 5% of glucose and sucrose solutions, respectively, as suggested by Kojima *et al.*<sup>43</sup> Samples were then collected in 0, 2 and 6 h intervals, respectively.

### 2.2. Expression analysis of sorghum Tau GST subfamily members by RT-PCRs and qRT-PCRs

Expression of sorghum Tau GST members was evaluated using the sorghum EST collections from several databases (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>, <http://www.phytozome.net/sorghum> and [http://compbio.dfci.harvard.edu/cgi-bin/tgi/est\\_ann.pl?gudb=Sorghum](http://compbio.dfci.harvard.edu/cgi-bin/tgi/est_ann.pl?gudb=Sorghum)). Since low percentages of Tau members have been detected with EST expression evidence, expression analysis was carried out by normal and qRT-PCR. All gene-specific primers used for both normal and qRT-PCR analyses in this study were designed by Applied Biosystems Primer Express® software. Supplementary Table S1 lists all the primer sequences used in this study.

QIAGEN RNeasy Mini Kit was used for total RNA isolation. The first-strand cDNA was synthesized using Invitrogen kit. RT-PCRs were performed in 20 µl of reaction mixtures with 20 ng of first-strand cDNA, 200 µM of each dNTPs, 2.5 mM MgCl<sub>2</sub>, 0.5 µM each of primers, 1 U of *Taq*-DNA polymerase in 1 × PCR buffer. These reactions were carried out using PTC-100 thermo-cyclers. The temperature profiling for

PCR is as follows: 94°C for 2 min followed by 35 cycles at 94°C for 10 s, 59°C for 10 s and 72°C for 25 s followed by a 2-min extension step at 72°C. PCR products (10 µl) were visualized in 2% agarose gel, and all pictures were taken in BIORAD UV-Gel documentation system using Quantity one 1-D Analysis software. The qRT-PCR analyses were carried out according to our previous description.<sup>44</sup>

### 2.3. Expression divergence of rice, *Arabidopsis* and soybean Tau GST subfamily members

Rice Tau ESTs were collected from the Michigan State University (MSU) rice genome annotation database (<http://rice.plantbiology.msu.edu/>) and the NCBI EST database. *Arabidopsis* and soybean Tau ESTs were obtained from The *Arabidopsis* Information Resource (TAIR, <http://www.arabidopsis.org/>) and the Soybase (<http://soybase.org/index.php>), respectively. The rice MPSS database (<http://mpss.udel.edu/rice/>)<sup>42</sup> was used to evaluate differentially regulated genes under abiotic (drought, high salinity and cold) and biotic (*Magnaporthe grisea*, *Mg*, and *Xanthomonas oryzae* pv. *oryzae*, *Xoo*) stresses according to our previous description.<sup>45</sup> For *Arabidopsis*, the expression data under abiotic (drought, high salinity and cold) and biotic (*Erysiphe orontii* and *Botrytis cinerea*, *Bcin*) stresses were downloaded, and differentially expressed Tau genes were identified according to the description.<sup>46</sup>

To evaluate the expression divergence of tandemly or segmentally duplicated genes, the Pearson's correlation coefficient ( $r$ ) of their expression was computed for each pair according to the method by Lin *et al.*<sup>47</sup> The cut-off  $r$ -value, below which duplicated genes can be considered divergent in their expression, might be determined by calculating  $r$  between the expression profiles of 10 000 pairs of randomly selected genes.<sup>47,48</sup> Based on the method, we have calculated the cut-off  $r$ -value for the *Arabidopsis* expression data set in this study as 0.53. Thus, we utilized  $r < 0.53$  as an indicator of divergent expression in *Arabidopsis*. We have used  $r$ -value = 0.59 calculated by Lin *et al.*<sup>47</sup> as a cut-off in rice since we employed the same data set (MPSS) for our expression analysis. For sorghum, due to the lack of enough expression data for calculating the cut-off  $r$ -value,  $r < 0.5$  was used as an indicator since  $r$ -values from 0.5 to 0.59 have been used for criteria to determine the diverged expression.<sup>47–49</sup>

### 2.4. Genome-wide identification of plant genes encoding GST proteins

The Pfam database (<http://pfam.sanger.ac.uk/>)<sup>50</sup> was used to locate the GST\_N (PF02798) domains in 53 key GST proteins from a variety of organisms.

Similarly, the GST\_C domains (PF00043) were retrieved from 58 key GST proteins from the same database. These key domain amino acid sequences were aligned by ClustalX 2.0<sup>51</sup> and were then used to generate the hidden Markov model (HMM) profiles for HMM searches with  $E$ -value cut-off of 1.0 against annotated protein databases from 10 plant organisms. These species include four monocots (*Brachypodium distachyon*, *Oryza sativa*, *S. bicolor* and *Zea mays*) and six dicots (*Arabidopsis lyrata*, *Arabidopsis thaliana*, *Carica papaya*, *Glycine max*, *Medicago trunculata* and *P. trichocarpa*). For rice, the *O. sativa* spp. *japonica* cv. Nipponbare release 6.1 pseudomolecules in the MSU Rice Annotation was used for the search. For *A. thaliana*, the latest version (TAIR9) of *Arabidopsis* genome annotation was employed (<http://www.arabidopsis.org>). For the remaining eight species, the latest versions were from the phytozome database (<http://www.phytozome.net/>). The key domain amino acid sequences were also used as queries for BLASTP/TBLASTN searches for possible homologues encoded by the 10 plant organisms with an  $E$ -value cut-off of 0.01 to confirm the HMM searches.

The presence of the domain GST\_N or GST\_C in the putative GST members detected by the HMM or BLAST searches was confirmed by searching the Pfam database with  $E$ -value = 0.01 as the cut-off level. The SMART database (<http://smart.embl-heidelberg.de/>)<sup>52</sup> was also employed to detect conserved domains with default parameters. Proteins confirmed by domain searches were regarded as putative GST\_N or GST\_C domain-containing proteins (referred to GST\_N or GST\_C domain-containing proteins for convenience); otherwise, they were excluded from our data set.

### 2.5. Alignment and phylogenetic analysis

GST\_N/GST\_C domain or full-length amino acid sequences were aligned using ClustalX 2.0 (<http://www.clustal.org/>)<sup>51</sup> and manually edited in Jalview (version 2).<sup>53</sup> The aligned sequences were used for phylogenetic analysis according to the description by Jiang and Ramachandran.<sup>54</sup>

### 2.6. Expansion and evolutionary analysis

Chromosomal distributions of *GST* genes were performed by searching and mapping physical positions of their corresponding locus numbers in their genomes. Tandemly duplicated *GST* genes in 10 plant genomes were identified by three criteria (i)  $\leq 10$  genes apart, (ii) belong to the same *GST* subfamily and (iii) within 100 kb for both *A. lyrata* and *A. thaliana* or 350 kb for the remaining species as suggested by Lehti-Shiu *et al.*<sup>55</sup> Segmentally duplicated



chromosome blocks have been previously identified in a genome-wide level in *Arabidopsis*,<sup>56</sup> rice<sup>57,58</sup> and soybean (<http://www.phytozome.net/soybean.php>). We examined the segmental duplicates by comparing the positions of *GST* genes with known duplicated chromosomal blocks. For the remaining genomes, we identified duplicated blocks using the flanking regions (50 kb upstream and downstream) of *GST* genes according to the method by Kong *et al.*<sup>59</sup>

To determine the contribution of transposable elements to the expansion of the GST family, the flanking genomic sequences of the 50-kb upstream and downstream of *GST* genes in different genomes were used for the identification of major transposon family members including *mutator-like transposable element* (MULE), *hAT*, *CACTA* and *Helitron* families as well as the identification of retrogenes according to the description.<sup>45</sup> Retrotransposon elements were identified by executing the LTR\_Finder program<sup>60</sup> or RetrOryza.<sup>61</sup> Non-synonymous (Ka) and synonymous (Ks) values and their ratios (Ka/Ks) were estimated and were statistically tested according to Jiang *et al.*<sup>45</sup>

### 3. Results

#### 3.1. The sorghum and other plant genomes encode different numbers of the GST family members

To identify and characterize the sorghum GST family members in a genome-wide level, HMM and BLAST searches were carried out against the whole sorghum genome sequences (the 'Materials and methods' section). Our data show that the sorghum genome encodes at least 99 GST\_N or GST\_C domain-containing proteins (Table 1). Supplementary Table S2 lists the details of these members including their locus names, genome positions, coding sequences and amino acid sequences. Among these two domains, GST\_N contains catalytically essential residues and is thought to be a key component of catalysis in GSTs, whereas GST\_C may be responsible for substrate specificity.<sup>62</sup> Therefore, these two domains were investigated either separately or jointly to better understand their evolution or function. Based on the GST\_N domain sequences, the sorghum GST family can be divided into seven different classes by phylogenetic analysis (Supplementary Fig. S1). They are Tau, Phi, Theta, Zeta, Lambda, GSH-dependent dehydroascorbate reductase and tetrachlorohydroquinone dehalogenase, as reported in other studies.<sup>5,7,11,13</sup> Among them, both Tau and Phi are the major classes, consisting of 56 (64%) and 19 (22%) members, respectively (Table 1).

In addition to the sorghum GST family, we have also genome-widely identified GST members from other

**Table 1.** Genome-wide identification of GST encoding genes in the 10 completely sequenced plant genomes

Organism	Total GSTs	Domain			Classes <sup>a</sup>		
		GST_N	GST_C	Both	Tau (%)	Phi (%)	Both (%)
<i>A. lyrata</i>	64	56	58	51	30 (54)	13 (23)	43 (77)
<i>A. thaliana</i>	60	54	51	45	28 (52)	13 (24)	41 (76)
<i>B. distachyon</i>	69	56	64	56	25 (45)	21 (38)	46 (82)
<i>C. papaya</i>	48	40	37	29	23 (58)	6 (15)	29 (73)
<i>G. max</i>	105	87	82	65	56 (64)	11 (13)	67 (77)
<i>M. truncatula</i>	43	40	36	33	31 (78)	6 (15)	37 (93)
<i>O. sativa</i>	84	77	75	68	50 (65)	18 (23)	68 (88)
<i>P. trichocarpa</i>	81	81	81	81	56 (69)	9 (11)	65 (80)
<i>S. bicolor</i>	99	87	87	78	56 (64)	19 (22)	75 (86)
<i>Z. mays</i>	72	67	58	53	37 (55)	23 (34)	60 (90)

<sup>a</sup>Classification was based on the GST\_N domain amino acid sequences.

nine genomes as described in the 'Materials and methods' section. Our data show that the sorghum and other plant genomes encode different numbers of the GST family members, and both Tau and Phi classes of GST members consist of at least 73% of total GSTs as shown in Table 1. Their locus names, genome positions, coding sequences and amino acid sequences were included in Supplementary Table S2. Currently, the GST members from *Arabidopsis*, rice and poplar have been reported<sup>12,13,63</sup> and members from the remaining 7 plant genomes were newly identified. Among the 53 previously identified GSTs in *Arabidopsis* (<http://www.arabidopsis.org/browse/genefamily/gst.jsp>),<sup>63</sup> two of them, *AtT3g55040* and *At5g02780*, were not included in this study since no GST\_N or GST\_C domains could be detected with *E*-value = 0.01 as the cut-off level. However, we have identified nine more GST\_N or GST\_C members (Supplementary Table S3). Soranzo *et al.*<sup>12</sup> have identified 59 rice GST members. However, three members *OsGSTMU357*, *OsGSTU37* and *OsGSTT2* were not included in this study since we cannot detect any GST domains in these members and we have also identified 28 more GST members (Supplementary Table S4). On the other hand, we have identified the same numbers of the poplar GST members as reported by Lan *et al.* (2009)<sup>13</sup> and we have used them for this study.

Besides the full-length GST\_N and/or GST\_C domain-containing genes, we have also identified numbers of partial GST fragments. They encode only partial GST\_N or GST\_C or both of partial domains. These members contain no typical domain structure

and have no expression evidence with the characters of pseudogenes. Due to the low feasibility of phylogenetic analyses by integrating these partial fragments, we removed these members from our analyses although we may underestimate the rate of gene duplication.

### 3.2. A significantly larger Tau and Phi subfamily of most recent common ancestor in monocots than in dicots

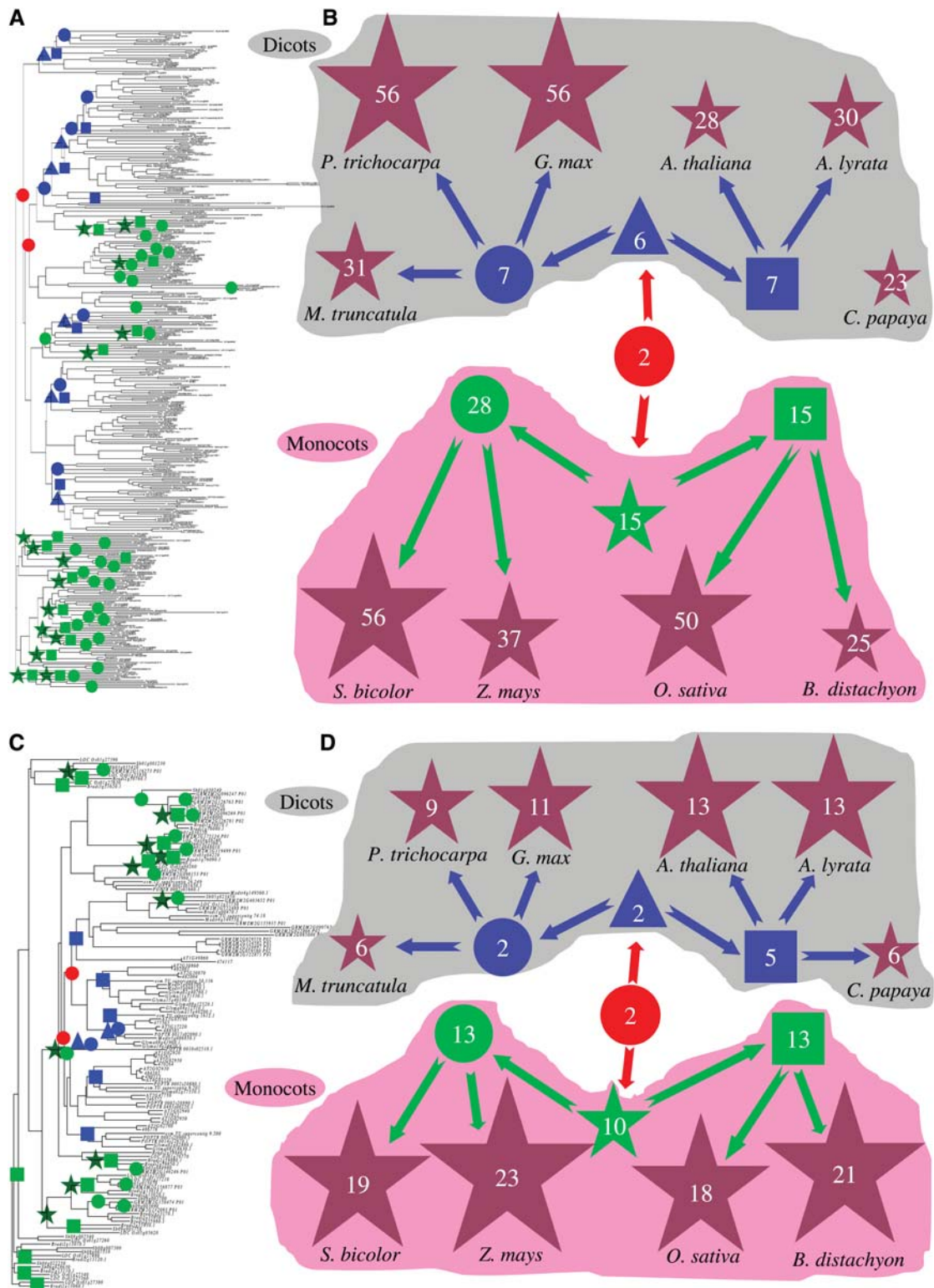
Since both Tau and Phi classes are the largest subfamilies in all analysed 10 genomes, we evaluated the patterns of expansion of these two classes by analysing the phylogenetic relationships of these GST members from different organisms (Fig. 1A–D and Supplementary Fig. S2 and S3). Most of dicot members fell into their own subclasses, separating from monocot plants. Members from each species intended to be clustered together, exhibiting lineage-specific expansion. To determine the degrees of expansion of these two classes among monocot and dicot plant lineages, we broke down the phylogeny into ancestral units according to the method.<sup>64</sup> Due to possible gene loss and pseudogenes, the most recent common ancestor (MRCA) members may be underestimated. Thus, at least 2 Tau or Phi members would have been present in the MRCA among these 10 organisms (red circles in Fig. 1A and B for Tau and C and D for Phi). The MRCA among the 4 monocot plants possessed 15 Tau (green stars in Fig. 1A and B) and 10 Phi (green stars in Fig. 1C and D) members; thus, 7.5-fold and 5-fold expansions have occurred for Tau and Phi, respectively, in this period. Following this era, they underwent relatively lower expansion rates except for rice, resulting in the current Tau or Phi members. These results suggested that the rapid gene expansion for the 4 monocots occurred after the divergence of monocot from dicot plants. However, the MRCA among the 6 dicot plants contained only 6 Tau (blue triangles in Fig. 1A and B) and 2 Phi (blue triangles in Fig. 1C and D) members; thus, only 3-fold or no expansion has occurred for Tau and Phi, respectively, in this period. The large-scale expansion of Tau or Phi members for dicot plants occurred during their divergence from their MRCA (Eurosids I and II) (Fig. 1). These data demonstrate that both monocot and dicot plants exhibit the differences in their Tau and Phi subfamily expansion history.

### 3.3. Significant contributions of tandem duplications to family size

To explore the possible mechanisms of the Tau and Phi GST subfamily expansion, we investigated the

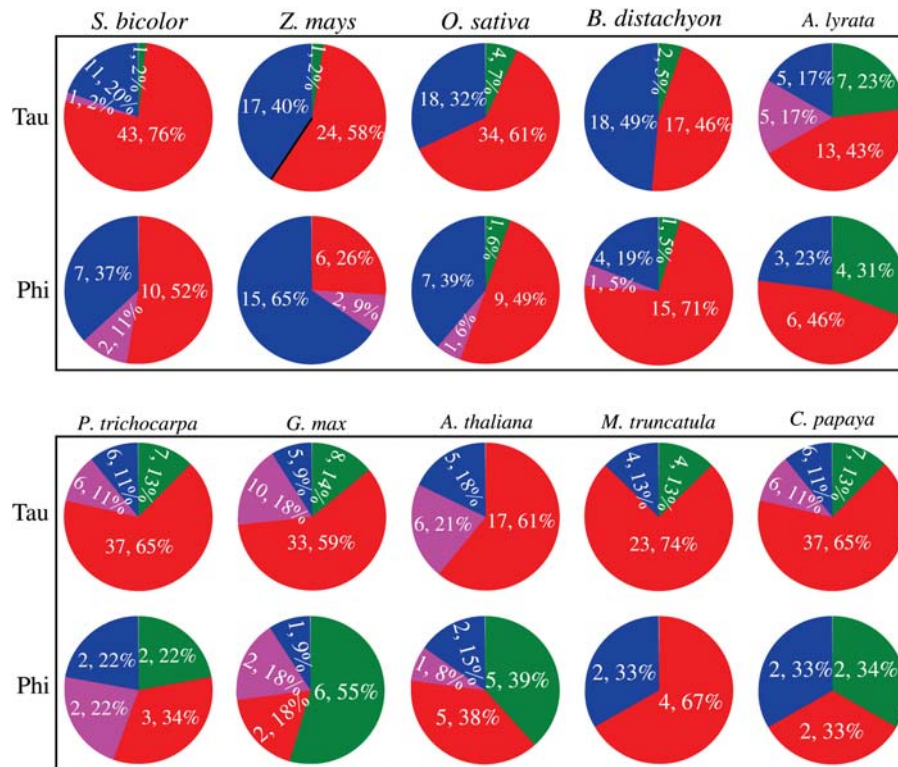
contributions of both tandem and segmental duplications to the expansion of these subfamilies in the sorghum genome. We examined the physical positions of the Tau and Phi GST members on different sorghum chromosomes (Supplementary Fig. S4). The results showed that both Tau/Phi GSTs were located on multiple chromosomes with non-random distributions. For example, chromosome 1 contained the highest density of the Tau or Phi GSTs with 30 or 7 members, respectively. Similar results were observed in rice and *Arabidopsis*.<sup>5,12</sup> However, more even distributions were observed in soybean with fewer members in most of chromosomes (data not shown). Based on the physical positions of the Tau or Phi GSTs, the tandem cluster was defined as described in the Methods. Among the 56 Tau and 19 Phi GSTs in sorghum, 44 Tau (78%) and 12 Phi (63%) GSTs are found in tandem clusters, respectively (red locus name in Supplementary Fig. S4; Fig. 2), indicating tandem duplications as the main mechanism for the Tau and Phi GST expansion. On the other hand, only two members of Tau and Phi GSTs (*Sb01g005990* and *Sb01g013590* for Tau and *Sb01g030240* and *Sb01g047980* for Phi) were involved in segmental duplications (indicated by blue lines in Supplementary Fig. S4). Besides both tandem and segmental duplications, we also investigated the contributions of other expansion mechanisms including transpositions and retrotranspositions to the expansion of these subfamilies (see Methods). However, our data showed that limited contribution by both transposons and retrotransposons to the expansion was detected. Similarly, we have investigated the contributions of tandem and segmental duplication, transpositions and retrotranspositions to the expansion of Tau and Phi subfamilies in the remaining 9 plant species (Fig. 2). The results showed that the expansion of 46–82% of Tau and 33–76% Phi GSTs were found in tandem clusters (Fig. 2), suggesting tandem duplication as the major mechanism for Tau and Phi GST expansion. In addition, 2–23% of Tau and 0–72% Phi GSTs were segmentally duplicated (Fig. 2), suggesting that segmental duplication should also be regarded as the major expansion mechanism for the Phi class in some species such as in soybean and *Arabidopsis*.

Since tandem duplication is regarded as the major expansion mechanism, we re-constructed their expansion history using the largest tandem cluster with 23 members in the sorghum chromosome 1. The 23 tandem genes are in two clades (Supplementary Fig. S5A), suggesting that this cluster is the results of two ancestral units, which may be evolved from ancient tandem duplication events. One of them contains only 3 genes whereas the secondary clade contains 20



**Figure 1.** Phylogenetic analysis and evolutionary dynamics of the Tau and Phi classes of the GST superfamily. (A and C) Phylogenetic analyses of the Tau and Phi subfamily members, respectively, in four monocot and six dicot plants. GST\_N domain amino acid sequences were employed to construct phylogenetic trees using the bootstrap method with a heuristic search of the PAUP 4.0b8 program. The results were confirmed by the Bayesian analyses. Ancestral units were defined according to Shiu *et al.*<sup>64</sup> Their enlarged phylogenetic trees and their analyses are shown in Supplementary Figs S2 and S3, respectively. (B and D) Evolutionary history of the Tau and Phi subfamily members in 10 organisms, respectively. Red circles represent the MRCA Tau/Phi units among all 10 organisms, blue triangles indicate the MRCA Tau/Phi units among dicot plants and green stars show the MRCA Tau/Phi units among monocot plants. Blue circles and squares represent the MRCA Tau/Phi units in Eurosid I (*M. truncatula*, *P. trichocarpa* and *G. max*) and Eurosid II (*A. thaliana* and *A. lyrata*), respectively. Green circles and squares show the MRCA Tau/Phi units between *S. bicolor* and *Z. mays* as well as between *O. sativa* and *B. distachyon*, respectively. Brown stars indicate the expanded Tau/Phi members in all 10 organisms. Grey and pink shadows in (B) and (D) indicate dicot and monocot plant species and their MRCA, respectively.



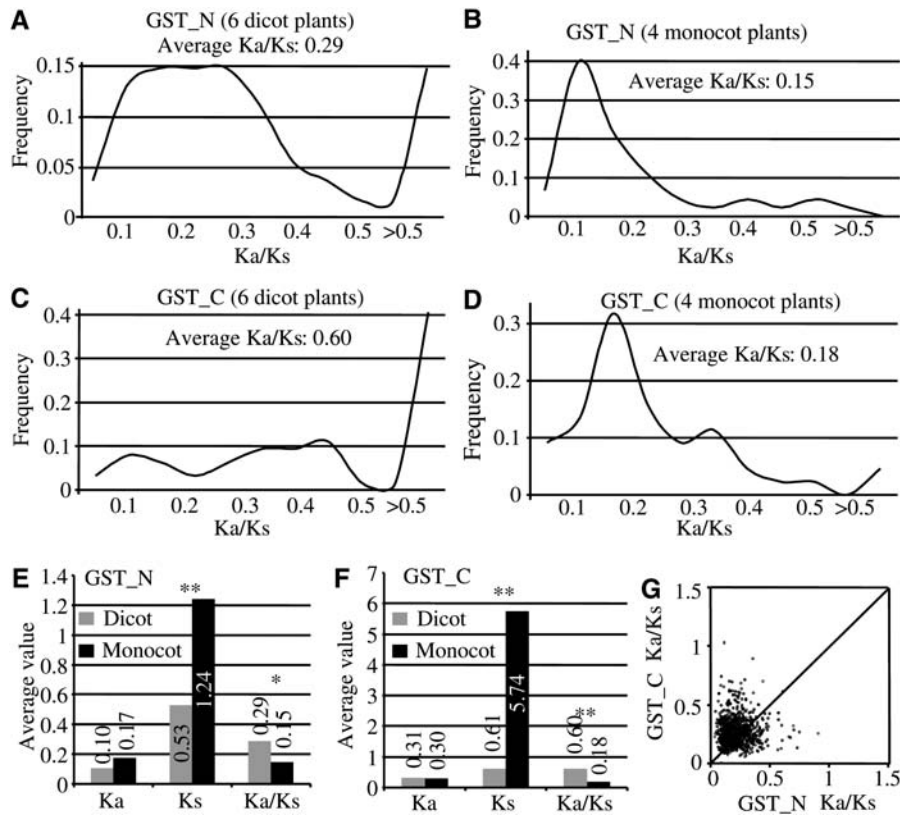


**Figure 2.** Pie diagrams showing expansion mechanisms of the Tau and Phi classes of the GST superfamilies in 10 plants. The figure indicates the contributions of tandem (red), segmental (green) duplications and both of them (pink) as well as other mechanisms (blue) to the expansion of the Tau and Phi class members in 10 genomes. In each pie diagram, the number prior to the comma indicates the total expanded members by different expansion mechanisms and its percentage is given following the comma.

members. The data suggested that different ancestral genes duplicated themselves by differential expansion rates. On the basis of the phylogenetic tree, we deduced the hypothetical origins of 23 genes from two ancestral units by tandem duplication (Supplementary Fig. S5B). For the ancestral unit with 3 descendents, only two rounds of tandem duplications were required to generate these genes. However, for the secondary unit with 20 descendents, at least 9 rounds of tandem duplications were expected to produce such a population. After expansion, these genes were then not always inserted into the loci according to their physical orders. For example, the putative tandem pair of the *Sb01g030810* gene is *Sb01g030990* but not its physical neighbour *Sb01g030820*. On the other hand, we found that most of tandem duplications occurred by a one-gene mode, i.e. only one gene was duplicated in a one tandem duplication event. This case was also observed in our previous report on tandem duplication of the lectin gene superfamily.<sup>65</sup> Taken together, our data showed that parental genes were not always physically linked to their descendant genes and different expansion rates were observed for different parental genes.

### 3.4. Different selection forces between monocot and dicot plants and between GST<sub>N</sub> and GST<sub>C</sub> domains in a species

Since both monocot and dicot plants exhibited the distinct difference in their expansion history, we investigated whether they were under different selection forces. We first identified reciprocal best matches for all Tau or Phi GST members either from 4 monocot or 6 dicot plants. These identified matches were then used to calculate Ka/Ks ratios for their GST<sub>N</sub> domain (Fig. 3A and B). Evolutionarily, the ratio of Ka to Ks can be used as an indicator of selective pressure acting on a protein-coding gene. A Ka/Ks ratio < 1 indicates functional constraint with purifying selection on a gene, and a Ka/Ks ratio > 1 indicates accelerated evolution with positive selection and a Ka/Ks ratio = 1 indicates neutral selection. Although the period with large scale of expansion of Tau or Phi GSTs from 6 dicot plants was later than 4 monocot plants, the average Ka/Ks for 6 dicots (0.29) is significantly higher than that for 4 monocots (0.15). Furthermore, the frequency distribution of the Ka/Ks ratios is also significantly different between dicot and monocot plants. For 6 dicots, most of the mass were centred near Ka/Ks = 0.2 or > 0.5



**Figure 3.** Ka/Ks ratio analysis. Frequency distributions of Ka/Ks ratios were analysed using the best matched pairs among six dicot plants (A and C) or four monocot plants (B and D). (A) and (B) show the data analysed from the GST\_N regions. (C) and (D) show the analysis from the GST\_C regions. The average Ka, Ks and their ratios in monocot and dicot pairs were indicated in (E) for GST\_N and in (F) for GST\_C. Asterisks indicate significant differences between monocot and dicot plants at  $P < 0.05$  (\*) and  $P < 0.01$  (\*\*) by  $t$ -test, respectively. The Ka/Ks plots for the GST\_C versus GST\_N domains of the sorghum Tau members are shown in (G).

whereas most of the mass for 4 monocots were centred near  $Ka/Ks = 0.12$ . These data suggested that GST\_N in dicot plants evolved faster than that in monocot plants. Similar results were observed when GST\_C domain regions were used for such analyses (Fig. 3C and D). To further analyse the reason why both dicot and monocot plants exhibited the difference in their evolutionary rates, we compared their Ka and Ks values separately. We found that their Ka values between dicot and monocot plants were similar (Fig. 3E and F). However, both dicot and monocot plants exhibited significant differences in their Ks values (Fig. 3E and F). As a result, they showed significant differences in their Ka/Ks ratios.

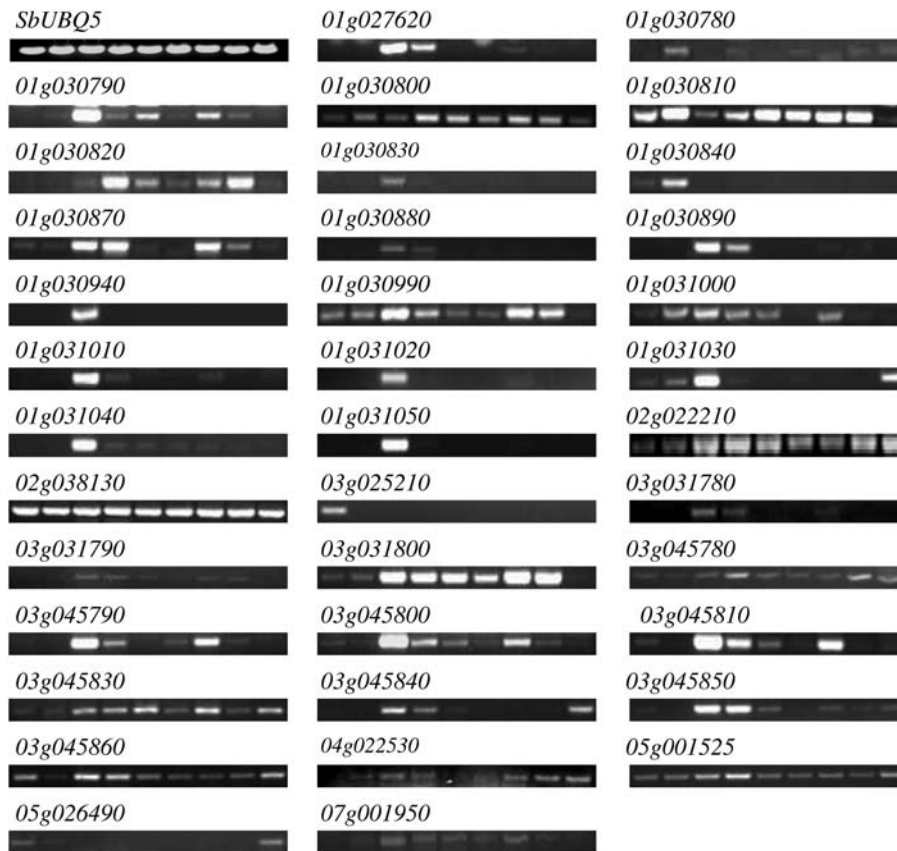
Since most of plant GST proteins consist of two functional domains including GST\_N and GST\_C, we were wondering if these two domains exhibited different evolutionary rates. We analysed the Ka/Ks ratios of these two domains of the sorghum Tau members separately. The ratios of GST\_C plotted against that of GST\_N of same proteins were shown in Fig. 3G. The results suggested that both domains were under differential selective pressures and the GST\_C domain might have been subjected to more relaxed

functional constraints. Similar results were observed in the remaining 9 organisms (data not shown). The recent analysis on the *Populus* GST superfamily also showed similar results.<sup>13</sup>

### 3.5. Expression abundance of the Tau subfamily members among 9 sorghum tissues

Tau class is the largest GST subfamily in plants and is a good candidate for surveying the expression divergence of duplicated members. Therefore, we further investigated the expression profiling of this class. We have detected 33 out of 56 sorghum Tau members with EST expression evidence. To examine if the remaining genes are expressed in sorghum and to explore their expression divergence, we have investigated the expression of 50 annotated Tau genes in 9 different tissues including young and mature leaves, panicles, seeds and roots as well as stems by RT-PCR analysis. The results showed that not all annotated Tau genes were expressed under normal growth conditions. Among 50 tested Tau members, 37 genes (74%) were expressed in at least one of 9 tissues (Fig. 4) and the remaining 13 genes showed no expression in all tested





**Figure 4.** Transcript profiles of the sorghum Tau class of GST superfamily members in nine different sorghum tissues shown by RT–PCR analyses. Nine bands from left to right for each gene represent amplified products from young leaves, mature leaves, young roots, mature roots, young panicles, mature panicles, young seeds, mature seeds and stems, respectively. The amplification of the gene *SbUBQ5* with the locus name *Sb09g004630* was used as control in this study as this gene showed the similar expression level among different tissues or under various stresses in this study.

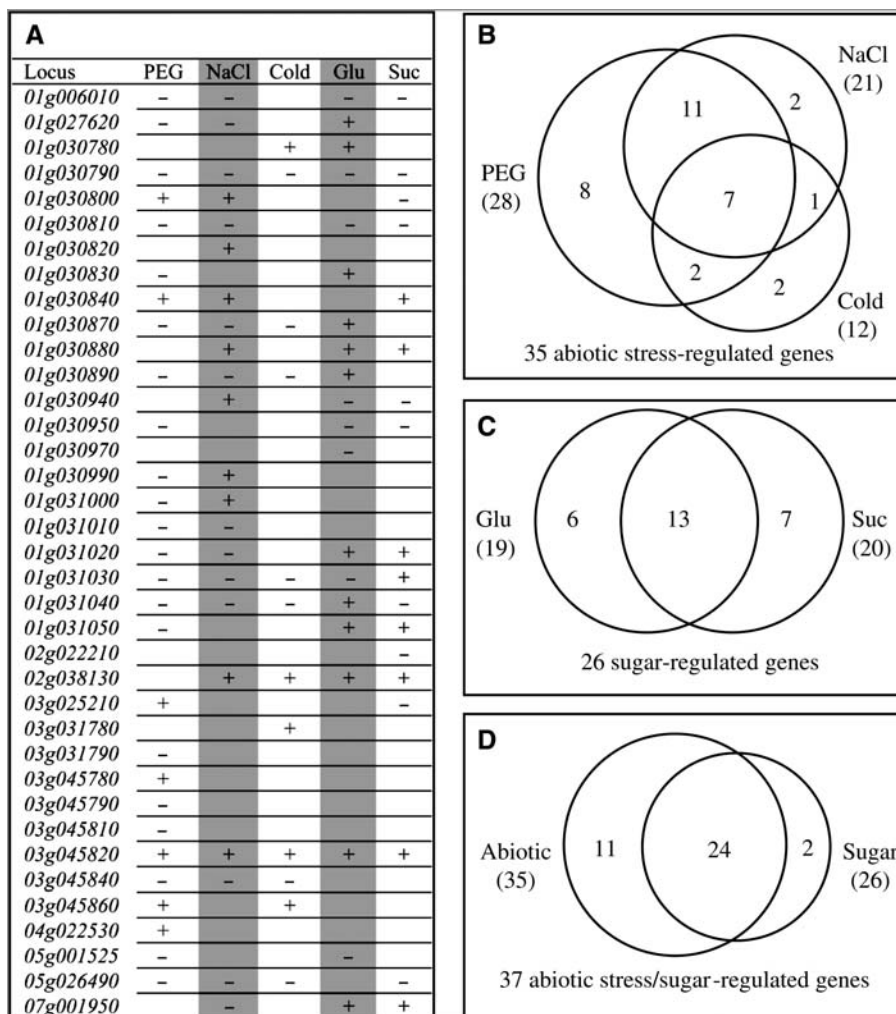
tissues. All these results have been confirmed by qRT–PCR (Supplementary Fig. S6). Among the 37 expressed genes, nine of them were expressed in all tested tissues. These genes were listed as follows: *01g030800*, *01g030810*, *01g030990*, *02g022210*, *02g038130*, *03g045780*, *03g045830*, *03g045860* and *05g001525*. Interestingly, 10 genes showed root-specific expression including *01g027620*, *01g030830*, *01g030880*, *01g030890*, *01g030940*, *01g031010*, *01g031020*, *01g031040*, *01g031050* and *03g031780*. They were mainly expressed in either young or mature roots. On the other hand, we have detected two genes with leaf-specific expression patterns including *01g030840* and *03g025210*.

### 3.6. Regulated expression of the Tau subfamily members in sorghum under various abiotic stresses and sugar treatments

To investigate if some members are expressed only under stressed conditions and to explore their roles and functional divergence, RT–PCRs were carried out to detect their expression abundance under various abiotic stress conditions (cold, PEG or high

salinity) and sugar (glucose and sucrose) treatments. The results indicated that 37 Tau members showed expression under one or more stressed conditions (Supplementary Fig. S7A). All of them were also expressed under normal growth conditions (Fig. 4). The results were confirmed by qRT–PCR (data not shown). Furthermore, we have detected additional 6 genes with significant expression signals under various stresses (Supplementary Fig. S7B). Most of them showed low expression level, which may explain why their signals could not be detected by normal RT–PCR. Thus, we have detected total of 43 GST Tau members with detectable expression signals under normal or stressed conditions.

By comparing the RT–PCR with qRT–PCR results and by statistic analysis, 37 genes were detected with significant differences in their expression level under one or more stress conditions or treatments. These genes and their responses to stresses and treatments were shown in Fig. 5A. Total of 35 Tau genes were significantly regulated by abiotic stresses including cold, PEG and high salinity (Fig. 5B). Among them, 12 genes were regulated by only one of these stresses



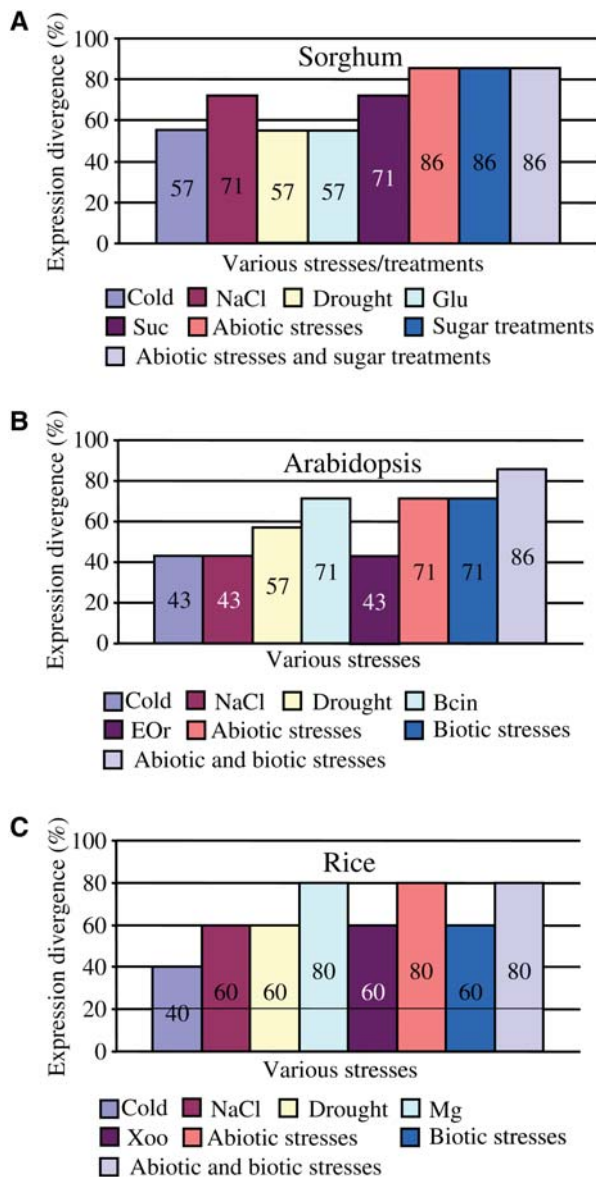
**Figure 5.** A summary of the expression analyses of the sorghum Tau class of GST superfamily members under various treatments. (A) A list of expression patterns of 37 Tau members under five different treatments. The results were from RT-PCR and two biologically replicated qRT-PCR analyses. The '+' indicates that the corresponding gene was significantly up-regulated under a stress treatment by *t*-test. The '-' indicates down-regulation by *t*-test. (B) to (D) Venn diagrams showing the classification of genes regulated by various abiotic stresses and sugar treatments based on the result from (A).

(8 by PEG and 2 by salinity or cold). Another set of 14 genes were regulated by two of these three stresses and the remaining 7 genes were regulated by all three stresses. Besides abiotic stresses, sugar treatment was also shown to play important roles in regulating Tau GST gene expression. Totally, 26 Tau GST members were detected with significant differences in their expression level under both glucose and sucrose treatments (Fig. 5C). Among them, 6 genes were regulated by only glucose treatment and 7 genes were by sucrose. The remaining 13 genes were regulated by both treatments.

Among 37 abiotic stress or sugar-responsive Tau members, 11 genes were regulated by only abiotic and 2 genes were by sugar; the remaining 24 genes were regulated by both abiotic and sugar treatments (Fig. 5D). These data suggested the possible interaction between abiotic stresses and sugar signalling.

### 3.7. Expression divergence among tandemly and fragmentally duplicated genes in sorghum, rice and Arabidopsis

To understand the functional divergence of duplicated Tau genes, we compared the expression profiles of these genes under various stresses or different sugar treatments in sorghum. Totally, 7 tandem arrays and 1 pair of segmentally duplicated genes were submitted for such analyses. If  $r < 0.5$ , 0.59, and 0.53 (see Methods) between a tandem array or a segmental pair in sorghum, rice and *Arabidopsis*, respectively, the tandem array or segmental pair was regarded as a divergent array or pair in their expression. Based on the criteria, no expression divergence has been observed between segmentally duplicated gene pair 01g005990 and 01g013590. Among 7 tandem arrays, 57–71% of them exhibited significant divergence in their expression patterns in



**Figure 6.** Expression divergence among tandemly and fragmentally duplicated Tau genes in sorghum, rice and *Arabidopsis*. (A) The effect of tandem and segmental duplications on gene expression divergence of the Tau subfamily members under abiotic stresses and sugar treatments. (B and C) Gene expression divergence of tandemly/fragmentally duplicated Tau members under abiotic and biotic stresses in rice and *Arabidopsis*, respectively.

response to cold, NaCl, PEG, glucose or sucrose treatments (Fig. 6A). Statistically, 86% of arrays showed significant divergence under abiotic stresses, sugar treatments or both of them (Fig. 6A). More detail analysis by comparing the expression data with phylogenetic relationship showed that expression divergence was observed even in closely related genes.

Similarly, we have also investigated the expression divergence of duplicated Tau genes in *Arabidopsis* and rice, respectively (Fig. 6B and C). For *Arabidopsis*, we have analysed total of 3 pairs of segmentally

duplicated pairs (*At1g10360* and *At1g59670*, *At1g17190* and *At1g78340*, *At1g27130* and *At1g69920*), two of them (67%) showed expression divergence in response to either abiotic or biotic stresses. Among 7 tandem arrays, 43–71% of them were differentially expressed under cold, NaCl, PEG, *Bcin* or *EQr* treatments (Fig. 6B). As a result, 71% of them showed significant divergence under abiotic or biotic stresses and 86% of them exhibited expression divergence under both abiotic and biotic stresses (Fig. 6B). For rice, two pairs of segmentally duplicated Tau genes (*LOC\_Os03g57200* and *LOC\_Os07g05800*, *LOC\_Os11g03210* and *LOC\_Os12g02960*) have been detected. The former pair exhibited divergence in their transcript abundance among various tissues and no expression divergence has been detected in the latter pair. On the other hand, among 5 tandemly duplicated Tau gene arrays, 40–80% of them were divergent in their expression under cold, NaCl, PEG, *Mg* or *Xoo* treatments (Fig. 6C). Thus, 60–80% of them exhibited expression divergence under either abiotic/biotic or both of them (Fig. 6C).

## 4. Discussion

### 4.1. Evolutionary origins of GST domains and their combinations

In this study, we have genome-widely identified more than 700 GST\_N or GST\_C domain-containing GSTs from 10 higher plants. Besides these, our searches and reports show that GSTs are ubiquitous not only in higher plants but also in other eukaryotes and prokaryotes.<sup>62,66</sup> Thus, GSTs are evolutionarily ancient proteins and one may be interested in the origin of this family. Reports suggested Theta, Zeta and Omega GSTs as the most ancestral classes.<sup>66,67</sup> Our phylogenetic analysis also supported both Theta and Zeta as the ancestral classes in plants (Supplementary Fig. S1). On the other hand, GSTs were thought to have evolved from a thioredoxin-like ancestor in response to the development of oxidative stress<sup>68,69</sup> and glutaredoxins are the suggested ancestors of the GST\_N domain.<sup>70</sup> In this study, we indeed detected several GST\_N domains, for example, in the sorghum locus *Sb02g003090*, with very high sequence similarity to the glutaredoxin domain, supporting the putative origin of GSTs from glutaredoxins.

Among these identified hundreds of GSTs, both GST\_N and GST\_C domains are usually encoded in a single gene, indicating that most of GSTs are homodimers in plants. The fact also suggests the co-evolution of both GST\_N and GST\_C domains. However, we have also identified some GST genes encoding only GST\_N



or GST\_C domain (Table 1). The fact may imply the loss of GST\_N or GST\_C domain during long evolution. One of such examples was shown in Supplementary Fig. S8A. The figure shows a tandem cluster consisting of 4 rice GSTs and their domain organizations. After duplication, only the GST\_C domain was detected in one of duplicates *LOC\_Os10g38150* whereas both domains could be detected in the remaining tandemly duplicated genes.

Besides the detection of putative domain loss events, we are also interested in the detection of domain combinations, which may generate new proteins and thereby functional divergences.<sup>71,72</sup> Our data showed that most of GSTs contained only GST\_N/GST\_C or both of them, suggesting the limited domain combination in this family. However, we have detected some exceptions. For example, both members of the Theta class *Sb04g007760* and *Sb10g022570* contained the Elongation factor 1 gamma domain (EF1G) except for both GST\_N and GST\_C domains (Supplementary Fig. S8B). These cases were also detected in *Arabidopsis* (Supplementary Fig. S8B) and other plants (data not shown). In *Arabidopsis*, we have also detected the integration of other domains such as EF-1 guanine nucleotide exchange domain (EF1\_GNE) and Myb transcription factor domain (Myb\_DNA-binding) (Supplementary Fig. S8B). The integration of the transcription factor domain resulted in the localization of this protein to the nucleus,<sup>73</sup> suggesting the functional divergence by domain combination.

#### 4.2. Differential evolutionary history of the Tau and Phi subfamilies in monocot and dicot plants

In eukaryotes, most of genes with structural and regulatory functions are members of gene families. They are descendants from gene duplication, which plays a major role in plant evolution.<sup>74</sup> These duplicates may be lost or be survival by either retaining their original functions, subfunctionalization or neofunctionalization.<sup>75</sup> However, limited information is available on the patterns of functional diversification governing the evolution of most classes of gene families in the plant kingdom.<sup>13</sup>

Sorghum has been regarded as a biofuel crop of growing importance for ethanol production. Now both methylation filtration-based genome sequencing and 8-time-assembly draft sequencing have been finished and their data have been freely released.<sup>76,77</sup> Thus, it is now becoming more feasible and imperative to further characterize its genes and their families for better understanding their evolutionary mechanisms. However, comparing with other plants, limited data has been reported on the genome-wide analysis of a gene family in sorghum.

We have constructed a hypothetical evolutionary history of the Tau and Phi classes of the GST family and found numbers of pseudogenes with partial domain structures, indicating that some duplicated genes failed to persist similar to other gene families.<sup>75</sup> On the other hand, we found that the era for a large-scale expansion was different between dicot and monocot plants. Our data from 6 dicot and 4 monocot plant genomes showed that different dicot plants exhibited similar evolutionary patterns, so did monocot plants. The MRCA of all analysed dicot plants had a small family of the Tau and Phi GST subfamilies and the MRCA of all monocot plants had evolved into bigger sizes of these subfamilies (Fig. 1). The data suggested that the MRCA of monocot plants duplicated Tau and Phi GST members faster than the MRCA of dicot plants. However, during the period from the MRCA of monocot or dicot plants to current species, dicot plants evolved faster than monocot plants with more rapid expansion and higher Ka/Ks ratios (Figs. 1 and 3). Our data also show that sorghum plants exhibit high percentages of regulated expression patterns among different tissues and under various abiotic stresses (Figs. 4 and 5) and both monocot and dicot plants have also been detected with high expression divergence under these stresses (Fig. 6). Therefore, it is reasonable to suggest that differential evolutionary history between monocot and dicot plants may be due to the changed adaptability to various environmental conditions during the divergence of monocot plants from dicot plants.

#### 4.3. Lineage-specific expansion and species diversification

We have detected two large-scale expansion events of the Tau or Phi subfamily by investigating the expansion history of these classes in 6 dicot and 4 monocot plants. The MRCA of monocot plants experienced the first large-scale expansion with 7.5-fold increase for Tau and 5-fold for Phi while dicot plants required only 3-fold more members for Tau or no expansion for Phi GSTs (Fig. 1). The second large-scale expansion occurred mainly in dicot plants with the maximum 8-fold expansion during the species diversification of some dicot plants from their common ancestors (Fig. 1). These results suggested that the large-scale expansions might be required for species diversification although more GST members from such expansion might be also for the environmental adaptability or biological processes. On the other hand, we have also detected some species-specific Tau or Phi members as reported by Lan *et al.*<sup>13</sup> In fact, Tau or Phi classes of GSTs are plant-specific and species-specific GST members have been detected not only

in plants but also in non-plant organisms.<sup>5,7,8</sup> Thus, both evolutionary history and lineage-specific expansion suggested that GST members may have contributed to species diversification.

#### 4.4. Expansion patterns and mechanisms of duplications

We have demonstrated that dicot and monocot plants exhibited different expansion patterns. We have also shown that tandem duplication represented the major mechanism for the subfamily expansion. At least eight classes of the GST subfamily have been identified<sup>7</sup> and they exhibited different rates of expansion, resulting in different sizes of classes. Further investigation showed that the sizes of subfamilies were proportional to the tandem duplication rates (data not shown). The results suggested that tandem duplication could be regarded as the major driver for the expansion of not only the Tau/Phi GST subfamily but also the other subfamilies.

On the other hand, the rice genome has been reported to undergo large-scale duplication 40–50 million years ago<sup>78</sup> and rice is likely an ancient aneuploid rather than a polyploidy<sup>57</sup>; however, we do not see evidence for the significant contribution of large-scale duplications within a relative short period to the expansion of the rice GST family. On the other hand, we also investigated the contributions of other mechanisms to the expansion of this family since different organisms have evolved into different sizes of GST families. We have identified all possible transposons (*MULE*, *hAT*, *CACTA* and *Helitron*) and retrotransposons (LTR and non-LTR) using flanking genomic sequences of the 50-kb upstream and downstream of GST members in four monocot and six dicot plants. We found very limited contribution of transposons/retrotransposons to the expansion of this family in multiple organisms. For example, only one of the rice Tau members *LOC\_Os10g38710* was located in a *PACK-MULE* member T10007202, which was identified by Juretic *et al.*<sup>79</sup> However, we cannot detect its parental gene.

#### 4.5. Expression divergence and biological functions

At least 26 sorghum Tau members were regulated by glucose or sucrose treatments (Fig. 5C). In *Arabidopsis*, we have also detected five Tau genes with up-regulated expression under sucrose treatment based on the microarray analysis.<sup>80</sup> These data suggested that the Tau subfamily members in plants might play a role in sugar signalling.

The complexity of the gene family functions was mainly reflected by the very high expression

divergence under various stresses.<sup>81</sup> Our data revealed that the Tau subfamily has evolved into a highly divergent group in their expression under various abiotic stresses and sugar treatments. These phenomena have been observed in all tested organisms including sorghum, rice and *Arabidopsis* (Fig. 6). Furthermore, the expression divergence was also observed among different individuals within a tandem array. For example, among 23 tandem duplicates (Supplementary Fig. S5), 20 GST genes show detectable expression and none of them exhibit the same expression pattern when their expression patterns were examined among different tissues (Fig. 4) or under different stresses/treatments (Fig. 5 and Supplementary Fig. S7). In addition, expression divergence of the Tau members was observed not only under these stresses but also under various hormones. For example, in sorghum, we have detected at least 40% of tandemly duplicated Tau genes with diverse expression patterns under various hormones including SA, methyl jasmonate or the ethylene precursor aminocyclopropane based on the microarray data.<sup>82</sup> All these data suggested that expression divergence should be regarded as one of the major drivers to facilitate the retention of the tandemly duplicated genes in this subfamily. Additionally, these data also suggested that the Tau subfamily members should play important roles in stress regulation as well as sugar and hormone signalling.

On the other hand, functions of plant Tau and Phi members may be retained with a relatively low divergence as shown by Ka/Ks analysis (Fig. 3). We also showed that both GST\_N and GST\_C domains were under different selective pressures (Fig. 3). GSTs have been thought to function in detoxification of xenobiotics and in response to biotic and abiotic stresses. Thus, both more relaxed functional constraints in the GST\_C domain and high degree of expression divergence should be regarded as the major mechanisms to facilitate the retention of duplicated genes and the adaptability to the diversity of potential xenobiotics and stressors.

**Supplementary data:** Supplementary data are available at [www.dnaresearch.oxfordjournals.org](http://www.dnaresearch.oxfordjournals.org).

**Acknowledgements:** We thank the Joint Genome Institute, the MSU Buell Lab and the *Arabidopsis* Information Resource for providing annotations and sequences. We also thank the National Center for Biotechnology Information, the rice MPSS database, the RIKEN Plant Science Center, the DFCI *Sorghum bicolor* Gene Index, the Soybase and Texas A&M University for making the EST and/or stress expression data sets available.

## Funding

This work was supported by Strategic Research Program from Temasek Life Sciences Laboratory in Singapore.

## References

- Noctor, G. and Foyer, C.H. 1998, Ascorbate and glutathione: keeping active oxygen under control, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **49**, 249–79.
- 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**, 193–8.
- Marrs, K.A. 1996, The functions and regulation of glutathione S-transferases in plants, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **47**, 127–58.
- Droog, F. 1997, Plant glutathione S-transferases, a tale of theta and Tau, *J. Plant Growth Regul.*, **16**, 95–107.
- Dixon, D.P., Laphorn, A. and Edwards, R. 2002, Plant glutathione transferases, *Genome Biol.*, **3**, 3004.1–10.
- Moons, A. 2005, Regulatory and functional interactions of plant growth regulators and plant glutathione S-transferases (GSTs), *Vitam. Horm.*, **72**, 155–202.
- Basantani, M. and Srivastava, A. 2007, Plant glutathione transferases: a decade falls short, *Can. J. Bot.*, **85**, 443–56.
- Chronopoulou, E.G. and Labrou, N.E. 2009, Glutathione transferases: emerging multidisciplinary tools in red and green biotechnology, *Recent Pat. Biotechnol.*, **3**, 211–23.
- Conn, S., Curtin, C., Bézier, A., Franco, C. and Zhang, W. 2008, Purification, molecular cloning, and characterization of glutathione S-transferases (GSTs) from pigmented *Vitis vinifera* L. cell suspension cultures as putative anthocyanin transport proteins, *J. Exp. Bot.*, **59**, 3621–34.
- Sappl, P.G., Heazlewood, J. and Millar, A.H. 2004, Untangling multi-gene families in plants by integrating proteomics into functional genomics, *Phytochemistry*, **65**, 1517–30.
- Sappl, P.G., Carroll, A.J., Clifton, R., Lister, R., Whelan, J., Harvey Millar, A. and Singh, K.B. 2009, The *Arabidopsis* glutathione transferase gene family displays complex stress regulation and co-silencing multiple genes results in altered metabolic sensitivity to oxidative stress, *Plant J.*, **58**, 53–68.
- 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**, 511–21.
- 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**, 3749–66.
- Seppanen, M.M., Cardi, T., Hyokki, M.B. and Pehu, E. 2000, Characterisation and expression of cold-induced glutathione S-transferase in freezing tolerant *Solanum commersonii*, sensitive *S. tuberosum* and their interspecific somatic hybrids, *Plant Sci.*, **153**, 125–33.
- Moons, A. 2003, *Osgstu3* and *osgtu4*, encoding tau class glutathione S-transferases, are heavy metal- and hypoxic stress-induced and differentially salt stress-responsive in rice roots, *FEBS Lett.*, **553**, 427–32.
- Kiyosue, T., Yamaguchi-Shinozaki, K. and Shinozaki, K. 1993, Characterization of two cDNAs (ERD11 and ERD13) for dehydration-inducible genes that encode putative glutathione S-transferases in *Arabidopsis thaliana* L., *FEBS Lett.*, **335**, 189–92.
- Bianchi, M.W., Roux, C. and Vartanian, N. 2002, Drought regulation of GST8, encoding the *Arabidopsis* homologue of ParC/Nt107 glutathione transferase/peroxidase, *Physiol. Plant*, **116**, 96–105.
- Vollenweider, S., Weber, H., Stolz, S., Chetelat, A. and Farmer, E.E. 2000, Fatty acid ketodienes and fatty acid ketotrienes: Michael addition acceptors that accumulate in wounded and diseased *Arabidopsis* leaves, *Plant J.*, **24**, 467–76.
- Mauch, F. and Dudler, R. 1993, Differential induction of distinct glutathione-S-transferases of wheat by xenobiotics and by pathogen attack, *Plant Physiol.*, **102**, 1193–201.
- Zhou, J. and Goldsbrough, P.B. 1993, An *Arabidopsis* gene with homology to glutathione S-transferases is regulated by ethylene, *Plant Mol. Biol.*, **22**, 517–23.
- Chen, W. and Singh, K.B. 1999, The auxin, hydrogen peroxide and salicylic acid induced expression of the *Arabidopsis* GST6 promoter is mediated in part by an ocs element, *Plant J.*, **19**, 667–77.
- Brentner, L.B., Mukherji, S.T., Merchie, K.M., Yoon, J.M., Schnoor, J.L. and Van Aken, B. 2008, Expression of glutathione S-transferases in poplar trees (*Populus trichocarpa*) exposed to 2,4,6-trinitrotoluene (TNT), *Chemosphere*, **73**, 657–62.
- Chen, W., Chao, G. and Singh, K.B. 1996, The promoter of a H<sub>2</sub>O<sub>2</sub>-inducible, *Arabidopsis* glutathione S-transferase gene contains closely linked OBF- and OBP1-binding sites, *Plant J.*, **10**, 955–66.
- Jain, M., Ghanashyam, C. and Bhattacharjee, A. 2010, Comprehensive expression analysis suggests overlapping and specific roles of rice glutathione S-transferase genes during development and stress responses, *BMC Genomics*, **11**, 73.
- Coleman, J., Blake-Kalff, M. and Davies, E. 1997, Detoxification of xenobiotics by plants: chemical modification and vacuolar compartmentation, *Trends Plant Sci.*, **2**, 144–51.
- Roxas, V.P., Smith, R.K., Allen, E.R. and Allen, R.D. 1997, Overexpression of glutathione S-transferase/glutathione peroxidase enhances the growth of transgenic tobacco seedlings during stress, *Nat. Biotechnol.*, **15**, 988–91.
- Roxas, V.P., Lodhi, S.A., Garrett, D.K., Mahan, J.R. and Allen, R.D. 2000, Stress tolerance in transgenic tobacco seedlings that overexpress glutathione S-transferase/glutathione peroxidase, *Plant Cell Physiol.*, **41**, 1229–34.
- Milligan, A.S., Daly, A., Parry, M.A.J., Lazzeri, P.A. and Jepson, I. 2001, The expression of a maize glutathione S-transferase in transgenic wheat confers herbicide tolerance, both in planta and *in vivo*, *Mol. Breed.*, **7**, 301–5.



29. Dixon, D.P., McEwen, A.G., Laphorn, A.J. and Edwards, R. 2003, Forced evolution of a herbicide detoxifying glutathione transferase, *J. Biol. Chem.*, **278**, 23930–5.
30. Yu, T., Li, Y.S., Chen, X.F., Hu, J., Chang, X. and Zhu, Y.G. 2003, Transgenic tobacco plants overexpressing cotton glutathione S-transferase (GST) show enhanced resistance to methyl viologen, *J. Plant Physiol.*, **160**, 1305–11.
31. Light, G.G., Mahan, J.R., Roxas, V.P. and Allen, R.D. 2005, Transgenic cotton (*Gossypium hirsutum* L.) seedlings expressing a tobacco glutathione S-transferase fail to provide improved stress tolerance, *Planta*, **222**, 346–54.
32. Karavangeli, M., Labrou, N.E., Clonis, Y.D. and Tsaftaris, A. 2005, Development of transgenic tobacco plants overexpressing maize glutathione S-transferase I for chloroacetanilide herbicides phytoremediation, *Biomol. Eng.*, **22**, 121–8.
33. Marrs, K.A., Alfenito, M.R., Lloyd, A.M. and Walbot, V. 1995, A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2, *Nature*, **375**, 397–400.
34. Alfenito, M.R., Souer, E., Goodman, C.D., Buell, R., Mol, J., Koes, R. and Walbot, V. 1998, Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases, *Plant Cell*, **10**, 1135–49.
35. 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**, 1561–70.
36. Kitamura, S., Shikazono, N. and Tanaka, A. 2004, TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*, *Plant J.*, **37**, 104–14.
37. Zettl, R., Schell, J. and Palme, K. 1994, Photoaffinity labeling of *Arabidopsis thaliana* plasma membrane vesicles by 5-azido-[7-3H]indole-3-acetic acid: identification of a glutathione S-transferase, *Proc. Natl Acad. Sci. USA*, **91**, 689–93.
38. Bilang, J. and Sturm, A. 1995, Cloning and characterization of a glutathione S-transferase that can be photo-labeled with 5-azidoindole-3-acetic acid, *Plant Physiol.*, **109**, 253–60.
39. Smith, A.P., Nourizadeh, S.D., Peer, W.A., Xu, J., Bandyopadhyay, A., Murphy, A.S. and Goldsbrough, P.B. 2003, *Arabidopsis* At-GSTF2 is regulated by ethylene and auxin, and encodes a glutathione S-transferase that interacts with flavonoids, *Plant J.*, **36**, 433–42.
40. Gonneau, M., Pagant, S., Brun, F. and Laloue, M. 2001, Photoaffinity labelling with the cytokinin agonist azido-CPPU of a 34 kDa peptide of the intracellular pathogenesis-related protein family in the moss *Physcomitrella patens*, *Plant Mol. Biol.*, **46**, 539–48.
41. Lederer, B. and Boger, P. 2003, Binding and protection of porphyrins by glutathione S-transferases of *Zea mays* L., *Biochim. Biophys. Acta*, **1621**, 226–33.
42. Nobuta, K., Venu, R.C., Lu, C., et al. 2007, An expression atlas of rice mRNAs and small RNAs, *Nat. Biotechnol.*, **25**, 473–7.
43. Kojima, H., Suzuki, T., Kato, T., et al. 2007, Sugar-inducible expression of the nucleolin-1 gene of *Arabidopsis thaliana* and its role in ribosome synthesis, growth and development, *Plant J.*, **49**, 1053–63.
44. Jiang, S.Y., Bachmann, D., La, H., Ma, Z., Venkatesh, P.N., Ramamoorthy, R. and Ramachandran, S. 2007, *Ds* insertion mutagenesis as an efficient tool to produce diverse variations for rice breeding, *Plant Mol. Biol.*, **65**, 385–402.
45. Jiang, S.Y., Christoffels, A., Ramamoorthy, R. and Ramachandran, S. 2009, Expansion mechanisms and functional annotations of hypothetical genes in the rice genome, *Plant Physiol.*, **150**, 1997–2008.
46. Matsui, A., Ishida, J., Morosawa, T., et al. 2008, *Arabidopsis* transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array, *Plant Cell Physiol.*, **49**, 1135–49.
47. Lin, H., Ouyang, S., Egan, A., et al. 2008, Characterization of paralogous protein families in rice, *BMC Plant Biol.*, **8**, 18.
48. Blanc, G. and Wolfe, K.H. 2004, Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution, *Plant Cell*, **16**, 1679–91.
49. Gu, Z., Nicolae, D., Lu, H.H. and Li, W.H. 2002, Rapid divergence in expression between duplicate genes inferred from microarray data, *Trends Genet.*, **18**, 609–13.
50. Finn, R.D., Mistry, J., Schuster-Böckler, B., et al. 2006, Pfam: clans, web tools and services, *Nucleic Acids Res.*, **34**, D247–51.
51. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997, The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res.*, **25**, 4876–82.
52. Letunic, I., Doerks, T. and Bork, P. 2009, SMART 6: recent updates and new developments, *Nucleic Acids Res.*, **37**(Database issue), D229–32.
53. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. 2009, Jalview version 2—a multiple sequence alignment editor and analysis workbench, *Bioinformatics*, **25**, 1189–91.
54. Jiang, S.Y. and Ramachandran, S. 2006, Comparative and evolutionary analysis of genes encoding small GTPases and their activating proteins in eukaryotic genomes, *Physiol. Genomics*, **24**, 235–51.
55. Lehti-Shiu, M.D., Zou, C., Hanada, K. and Shiu, S.H. 2009, Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes, *Plant Physiol.*, **150**, 12–26.
56. Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M. and van de Peer, Y. 2002, The hidden duplication past of *Arabidopsis thaliana*, *Proc. Natl Acad. Sci. USA*, **99**, 13627–32.
57. Vandepoele, K., Simillion, C. and Van de Peer, Y. 2003, Evidence that rice and other cereals are ancient aneuploids, *Plant Cell*, **15**, 2192–202.
58. Lin, H., Zhu, W., Silva, J.C., Gu, X. and Buell, C.R. 2006, Intron gain and loss in segmentally duplicated genes in rice, *Genome Biol.*, **7**, R41.

59. Kong, H., Landherr, L.L., Frohlich, M.W., Leebens-Mack, J., Ma, H. and de Pamphilis, C.W. 2007, Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth, *Plant J.*, **50**, 873–85.
60. Xu, Z. and Wang, H. 2007, LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons, *Nucleic Acids Res.*, **35**, D265–8.
61. Chaparro, C., Guyot, R., Zuccolo, A., Piégu, B. and Panaud, O. 2007, RetrOryza: a database of the rice LTR-retrotransposons, *Nucleic Acids Res.*, **35**, D66–70.
62. Sheehan, D., Meade, G., Foley, V.M. and Dowd, C.A. 2001, Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily, *Biochem. J.*, **360**, 1–16.
63. Wagner, U., Edwards, R., Dixon, D.P. and Mauch, F. 2002, Probing the diversity of the *Arabidopsis* glutathione S-transferase gene family, *Plant Mol. Biol.*, **49**, 515–32.
64. Shiu, S.H., Karlowski, W.M., Pan, R., Tzeng, Y.H., Mayer, K.F. and Li, W.H. 2004, Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice, *Plant Cell*, **16**, 1220–34.
65. Jiang, S.Y., Ma, Z. and Ramachandran, S. 2010, Evolutionary history and stress regulation of the lectin superfamily in higher plants, *BMC Evol. Biol.*, **10**, 79.
66. da Fonseca, R.R., Johnson, W.E., O'Brien, S.J., Vasconcelos, V. and Antunes, A. 2010, Molecular evolution and the role of oxidative stress in the expansion and functional diversification of cytosolic glutathione transferases, *BMC Evol. Biol.*, **10**, 281.
67. Frova, C. 2006, Glutathione transferases in the genomics era: New insights and perspectives, *Biomol. Eng.*, **23**, 149–69.
68. Koonin, E.V., Mushegian, A.R., Tatusov, R.L., Altschul, S.F., Bryant, S.H., Bork, P. and Valencia, A. 1994, Eukaryotic translation elongation factor 1c contains a glutathione transferase domain - Study of a diverse, ancient protein superfamily using motif search and structural modeling, *Protein Sci.*, **3**, 2045–54.
69. Martin, J.L. 1995, Thioredoxin—a fold for all reasons, *Structure*, **3**, 245–50.
70. Oakley, A.J. 2005, Glutathione transferases: new functions, *Curr. Opin. Struct. Biol.*, **15**, 716–23.
71. Apic, G., Gough, J. and Teichmann, S.A. 2001, Domain combinations in archaeal, eubacterial and eukaryotic proteomes, *J. Mol. Biol.*, **310**, 311–25.
72. Chothia, C., Gough, J., Vogel, C. and Teichmann, S.A. 2003, Evolution of the protein repertoire, *Science*, **300**, 1701–3.
73. 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**, 1207–18.
74. Fligel, L.E. and Wendel, J.F. 2009, Gene duplication and evolutionary novelty in plants, *New Phytol.*, **183**, 557–64.
75. Demuth, J.P. and Hahn, M.W. 2009, The life and death of gene families, *Bioessays*, **31**, 29–39.
76. Paterson, A.H., Bowers, J.E., Bruggmann, R., et al. 2009, The *Sorghum bicolor* genome and the diversification of grasses, *Nature*, **457**, 551–6.
77. Bedell, J.A., Budiman, M.A., Nunberg, A., et al. 2005, Sorghum genome sequencing by methylation filtration, *PLoS Biol.*, **3**, e13.
78. Goff, S.A., Ricke, D., Lan, T.H., et al. 2002, A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*), *Science*, **296**, 92–100.
79. Juretic, N., Hoen, D.R., Huynh, M.L., Harrison, P.M. and Bureau, T.E. 2005, The evolutionary fate of MULE-mediated duplications of host gene fragments in rice, *Genome Res.*, **15**, 1292–7.
80. Gonzali, S., Loreti, E., Solfanelli, C., Novi, G., Alpi, A. and Perata, P. 2006, Identification of sugar-modulated genes and evidence for in vivo sugar sensing in *Arabidopsis*, *J. Plant Res.*, **119**, 115–23.
81. Jiang, S.Y., Ramamoorthy, R. and Ramachandran, S. 2008, Comparative transcriptional profiling and evolutionary analysis of the GRAM domain family in eukaryotes, *Dev. Biol.*, **314**, 418–32.
82. Salzman, R.A., Brady, J.A., Finlayson, S.A., et al. 2005, Transcriptional profiling of sorghum induced by methyl jasmonate, salicylic acid, and aminocyclopropane carboxylic acid reveals cooperative regulation and novel gene responses, *Plant Physiol.*, **138**, 352–68.