Genomics 100 (2012) 327–335

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Glutathione S-transferase (GST) genes in the red flour beetle, *Tribolium castaneum*, and comparative analysis with five additional insects

Houxia Shi, Lianghong Pei, Shasha Gu, Shicheng Zhu, Yanyun Wang, Yi Zhang, Bin Li *

Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210046 Jiangsu, PR China

ARTICLE INFO

Article history:
Received 28 March 2012
Accepted 12 July 2012
Available online 21 July 2012

Keywords:
Insect
Glutathione S-transferase
Tribolium castaneum
Insecticide resistance
Comparative analysis
Evolution

ABSTRACT

Glutathione S-transferases are important detoxification enzymes involved in insecticide resistance. Sequencing the *Tribolium castaneum* genome provides an opportunity to investigate the structure, function, and evolution of GSTs on a genome-wide scale. Thirty-six putative cytosolic GSTs and 5 microsomal GSTs have been identified in *T. castaneum*. Furthermore, 40, 35, 13, 23, and 32 GSTs have been discovered the other insects, *Drosophila*, *Anopheles*, *Apis*, *Bombyx*, and *Acyrthosiphon*, respectively. Phylogenetic analyses reveal that insect-specific GSTs, Epsilon and Delta, are the largest species-specific expanded GSTs. In *T. castaneum*, most GSTs are tandemly arranged in three chromosomes. Particularly, Epsilon GSTs have an inverted long-fragment duplication in the genome. Other four widely distributed classes are highly conserved in all species. Given that GSTs specially expanded in *Tribolium castaneum*, these genes might help to resist poisonous chemical environments and produce resistance to kinds of different insecticides.

1. Introduction

Glutathione S-transferases (GSTs, EC 2.5.1.18) are widespread superfamily genes existing in both prokaryotic and eukaryotic organisms, and they are involved in many cellular physiological activities, such as detoxification of endogenous and xenobiotic compounds, intracellular transport, biosynthesis of hormones and protection against oxidative stress [1]. They mainly catalyze the conjugation of tri-peptide glutathione (GSH) and electrophilic compounds, increase the solubility of products and then make them easier to excrete [2].

According to their cellular locations, GSTs are generally divided into the following three major categories: cytosolic, microsomal, and mitochondrial GSTs. To date, only the first two groups have been discovered in insects [3,4]. Cytosolic GSTs are proteins that are typically comprised of 200–250 amino acids and are active as either homodimers or heterodimers. Based on sequence homology in the N-terminus, substrate specificity, immunoreactivity, and sensitivity to different inhibitors, cytosolic GSTs in insects have been generally divided into at least six classes, (Delta, Epsilon, Omega, Sigma, Theta, and Zeta) [5–7]. Among these classes, Delta and Epsilon are unique and exist only in insects, and the Theta class is thought to have given rise to cytosolic GSTs [8]. Microsomal GSTs (MGSTs) usually consist of approximately 150 amino acid residues and are membrane-bound proteins that are active as trimmers. They have recently been designated as MAPEGs (membrane-associated proteins in eicosanoid and glutathione metabolism). Topological structures of GSTs have been analyzed for a few microsomal members, including Human MGST1 [9]. Mitochondrial GSTs, also known as the Kappa class, are dimeric. Its overall topology is similar to that of the disulfide bond isomerase from *E. coli* (DsbA). And Ladner *et al.* proposed that the mitochondrial and cytosolic GST families diverged from a thioredoxin-like ancestor, respectively a DsbA-like ancestor and a glutaredoxin-like ancestor, through parallel mechanisms [10].

Current research of GSTs has focused on the relationship between GSTs and insecticide resistance in insects. Previous studies have demonstrated that cytosolic GSTs (mainly Delta and Epsilon members) are involved in resistance to DDT/organophosphate [1]. Recent studies also showed that GSTe2 and GSTe7 of *Aedes aegypti* are involved in resistance to pyrethroid deltamethrin [11]. Interestingly, as some environmental compounds induce excessive expression of GSTs, certain GSTs have been utilized as biomarkers of environmental pollution [12,13]. Research on human GSTs has focused more on how the GST polymorphisms are associated with susceptibility to cancer [14,15]. GSTs play a role in the protection of cellular structures and DNA structures against toxic or carcinogenic compounds [14,15]. With the recent completion of several insect genome sequencing projects, such as those of *Drosophila melanogaster*, *Anopheles gambiae*, *Apis mellifera*, and *Bombyx mori*, 37, 28, 8, and 23 cytosolic GSTs were estimated to be present in these organisms, respectively [5,6,16–18]. For the microsomal group, only one MGST in *D. melanogaster*, three in *A. gambiae*, and two in *A. mellifera* were discovered [16].

The red flour beetle, *Tribolium castaneum*, is a worldwide agricultural pest that produces significant damage to stored cereals. It is also an important model insect for the study of Coleoptera insects [19]. As pesticide resistance has dramatically increased, research

* Corresponding author. Fax: +86 25 85891763.
E-mail address: libin@njnu.edu.cn (B. Li).

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2012.07.010

into *T. castaneum* GSTs has become an especially important issue [19–21]. Sequencing of the red flour beetle genome enables us to analyze GST structures, functions and evolution, as well as to investigate how GSTs are involved in pesticide resistance and environmental adaptation. In this study, *D. melanogaster* GSTs were used as reference sequences to identify all of the putative GST genes in *T. castaneum*. The genomic organization, intron characteristics and evolutionary relationships were analyzed. Furthermore, GSTs from the newest sequenced hemimetabolic insect *A. pisum* were also annotated and compared to those of genome-sequenced holometabolic insects. This comparative genomic analysis of GST genes through hemimetabolic insect *A. pisum* and other 5 representative holometabolic insects provides new insight into the evolutionary relationships between GST superfamilies of different insect lineages, promotes research about the adaptation of GSTs, and will facilitate the development of novel biological pesticides.

2. Results

2.1. Identification of GST genes

2.1.1. Basic information of GSTs in T. castaneum

A total of 41 genes encoding putative GST proteins, including 36 cytosolic GSTs and 5 microsomal GSTs, were identified in *T. castaneum* (Table 1). They covered all six cytosolic classes and microsomal groups. Among the 41 GSTs, 34 were directly mined from BeetleBase predictions, whereas the other 7 members were mined from the unannotated genome database or NCBI. They were renamed as TcGSTd1, TcGSTd2, etc., according to their relative position on the chromosome.

Interestingly, two insect-specific classes, Delta and Epsilon GSTs, have significantly different profiles in *T. castaneum*. It has fewer Delta GST genes (3 genes), significantly fewer than other species. By contrast, *T. castaneum* has the most Epsilon GSTs genes (19 genes), which were highly duplicated in *T. castaneum*. The Omega, Theta, Zeta and Sigma classes possess three, one, one and seven members respectively. *T. castaneum* possess the most microsomal GSTs (5 genes) among the six insect species (Table 2). There are also two other GSTs, which we could not classify into any known class. A BLAST search of TcGSTu1 in NCBI aligned GST-containing FLYWCH zinc-finger protein and TcGSTu2 across 379 amino acid residues, which were obtained from the query of the longer, atypical GST CG4623 that structurally diverged from other GSTs (Table 1).

2.1.2. GSTs in other insects

Previous studies have reported 38 Drosophila GST genes, including 37 cytosolic GSTs and 1 microsomal GST (CG1742). We compiled all of these data [6,22] and added new GSTs to this list. Three additional genes, 1 cytosolic GST CG4623 and 2 microsomal GSTs CG33177 and CG33178 were included (Supplementary Table S1).

Table 1
Summary of *T. castneum* GSTs, based on BeetleBase v3.0, the whole genome database and NCBI.

Group	Gene name	Identity (%)	Chromosome	Length of putative proteins	Numbers of intron	EST	Accession number
Delta	*TcGSTd1*	61	4	216	2	+	XP_974273[b]
	TcGSTd2	45	4	241	5	−	JN695809[b]
	TcGSTd3	44	7	220	3	−	GLEAN_09482[a]
Epsilon	*TcGSTe1*	42	2	222	2	+	GLEAN_04450[a]
	TcGSTe2	50	2	223	2	−	GLEAN_04449[a]
	TcGSTe3	47	2	215	2	−	GLEAN_04448[a]
	TcGSTe4	51	2	216	2	−	GLEAN_04447[a]
	TcGSTe5	49	2	216	2	+	GLEAN_04940[a]
	TcGSTe6	48	2	224	2	−	GLEAN_04941[a]
	TcGSTe7	51	2	218	1	+	GLEAN_04942[a]
	TcGSTe8	46	2	218	2	−	GLEAN_04446[a]
	TcGSTe9	48	2	216	2	−	GLEAN_04445[a]
	TcGSTe10	51	2	221	2	+	GLEAN_04444[a]
	TcGSTe11	52	2	219	2	−	GLEAN_04443[a]
	TcGSTe12	53	2	219	2	−	GLEAN_04442[a]
	TcGSTe13	55	3	195	2	−	GLEAN_03104[a]
	TcGSTe14	51	3	219	2	−	GLEAN_03345[a]
	TcGSTe15	45	3	214	2	−	XP_971268[b]
	TcGSTe16	47	3	227	2	+	XP_971327[b]
	TcGSTe17	43	3	198	2	−	GLEAN_03347[a]
	TcGSTe18	51	3	217	2	+	GLEAN_03348[a]
	TcGSTe19	40	3	214	2	+	GLEAN_03103[a]
Omega	*TcGSTo1*	37	3	237	3	−	XP_971118[b]
	TcGSTo2	43	3	241	2	−	XP_971184[b]
	TcGSTo3	43	3	239	3	+	GLEAN_00054[a]
Sigma	*TcGSTs1*	39	3	203	3	−	GLEAN_03231[a]
	TcGSTs2	44	3	202	3	+	GLEAN_03232[a]
	TcGSTs3	47	3	204	3	+	GLEAN_03233[a]
	TcGSTs4	49	3	204	3	+	GLEAN_03496[a]
	TcGSTs5	41	3	204	3	−	GLEAN_02878[a]
	TcGSTs6	47	3	204	3	−	GLEAN_00067[a]
	TcGSTs7	45	5	204	0	−	GLEAN_13948[a]
Theta	*TcGSTt1*	48	8	226	4	+	GLEAN_06215[a]
Zeta	*TcGSTz1*	85	7	215	3	−	GLEAN_09842[a]
Microsomal UN	*TcMGST1*	56	7	151	1	+	XP_968617[b]
	TcMGST2	41	7	155	1	−	GLEAN_09129[a]
	TcMGST3	45	7	162	1	−	JN695810[b]
	TcMGST4	45	7	152	2	−	JN695811[b]
	TcMGST5	43	9	153	2	−	GLEAN_11646[a]
	TcGSTu1	39	2	231	4	+	GLEAN_00522[a]
	TcGSTu2	32	3	379	3	+	GLEAN_03336[a]

[a] Accession number of the *T. castaneum* genome database (http://beetlebase.org/cgi-bin/gbrowse/BeetleBase3.gff3/).
[b] GenBank accession number.

Table 2
Comparison of GST gene numbers of six insects lineages.

GST	*T. castaneum*	*A. pisum*	*D. melanogaster*	*A. gambiae*	*A. mellifera*	*B. mori*
Delta	3	16	11	17 (12)[a]	2 (1)[a]	5
Epsilon	19	1	14	8	0	8
Omega	3	2	4 (5)[a]	1	2 (1)[a]	4
Sigma	7	6	1	1	4	2
Theta	1	2	4	2	1	1
Zeta	1	0	2	1	1	2
Microsomal	5	2	3 (1)[a]	3	2	0
UN	2	3	1 (0)[a]	2 (3)[a]	1(0)[a]	1
Total	41	32	40	35	13	23

Numbers without parentheses represent our results which are the same with previous studies.
[a] Numbers within parentheses are cited from the reference documents (*D. melanogaster* GST numbers cited from [6,22], *A. gambiae* numbers cited from [5,6,22], *A. mellifera* numbers cited from [16,17] and *B. mori* numbers cited from [18]).

However, there is a gene, CG6673, which is an isoform that is alternatively spliced to generate CG6673A and CG6673B, which were previously considered to be 2 unique GSTs. We determined that they were only a single GST, thus, a total of 40 GSTs were used in this new GST set to identify GSTs from five additional insect species (Table 2).

Based on the very recently completed hemi-metamorphosis insect genome sequence of *A. pisum*, we identified 30 cytosolic GSTs and 2 microsomal GSTs from *A. pisum* (Supplementary Table S2). Among the 32 GSTs, there are 10 GSTs (7 in Delta, 1 in Epsilon, 1 in Omega and 1 in an unknown class), differing greatly from the usual lengths of 200–250 amino acid residues. Among these 10 GSTs there were 7 with only C- or N- terminal regions. We hypothesized that they might be partial fragments of GST genes and attempted to re-annotate them using EST evidence and genome scaffold information but failed. Because these incomplete genes influenced the alignment of GST sequences, they were excluded in further phylogenetic analysis.

In *A. gambiae*, 32 cytosolic GSTs and 3 microsomal GSTs were found (Supplementary Table S3). The results were largely consistent with earlier reports [5,6,22]. One difference was that AgGSTu2 and AgGSTu3 were classified into the Delta class based on our new analysis as a result of high bootstrap values in the phylogenetic analysis. Another difference was that an unknown member AGAP006132 was added, which was obtained from the longer atypical GST CG4623. In addition, three more GST candidates, AGAP012702, AGAP012838, and AGAP012839, were identified from the floating scaffolds of the *A. gambiae* genome sequence, and their nucleotide sequences were strikingly similar to AgGSTd5 (99.2%), AgGSTd12 (97.9%) and AgGSTd4 (99.0%), respectively. However, their chromosomal locations were unknown, and whether they were newly duplicated GST genes is still unclear.

In *A. mellifera*, 11 cytosolic GSTs and 2 microsomal GSTs were found. In addition to the 10 GSTs that were discovered in previous studies [16,17], three other GSTs, GB19772 (Delta class), GB19678 (Omega class), and GB10031 (unknown class), were newly identified here (Supplementary Table S4). However, GB19772 and GB19678 are likely partial fragments of GST genes, which might be the reason why they were missed in the previous analysis. We combined both genome and EST information and corrected the length of GB19678 from 169 to 211 amino acid residues, but we failed to determine the structure of GB19772.

In *B. mori*, we re-annotated 23 cytosolic GSTs, which were identified by previous research [18] (Supplementary Table S5). The only minor difference is that the previously unclassified BmGSTu2 was sorted into the Epsilon class according to our new phylogenetic analysis as we largely increased the insect GSTs gene set and made it easier to classify them. No MGST members were found in the *B. mori* genome database in SilkDB v2.0 or NCBI.

2.2. Gene structure of GSTs

2.2.1. Genomic organization of T. castaneum GSTs

Thirty eight of 41 TcGSTs were mapped to 4 of the 10 *T. castaneum* chromosomes, whereas the remaining three were assigned to 3 other chromosomes (Fig. 1). Relatively high densities of TcGSTs were discovered on chromosomes 2 and 3, which were arranged into clusterIto V respectively. 19 Epsilon TcGSTs were divided into two clusters, which we named as cluster Iand II, and were located separately on chromosomes 2 and 3. Six members of the Sigma GST class were divided into cluster III and IV and were located at both ends of cluster II of Epsilon. Three Omega GSTs were also located on chromosomes 3 and were designated as cluster V. In addition to the cytosolic GSTs, four MGSTs constituted cluster VI and were located separately on chromosomes 7.

Tandem duplications are a general feature of the distribution of GST genes. This was found in all genome-sequenced insects, except that it was unclear in *A. pisum*, as its raw genome sequence was difficult to assign chromosomal locations. This feature is more remarkable in the Delta and Epsilon classes than that in other classes, especially in *D. melanogaster* and *T. castaneum*. In *D. melanogaster*, all Delta members are tandem, as are 10 members of Epsilon. In *T. castaneum*, two Delta members are tandem, and the Epsilon class is made up of two tandem clusters.

2.2.2. Intron-exon structure of T. castaneum GSTs

Of the 41 *T. castaneum* GSTs, only one intronless gene (TcGSTs7) has been found (Fig. 2 and Table 1). This is similar to the number of intronless genes observed in *A. mellifer* and *B. mori* (one intronless GST) [16,18] and *A. pisum* and *A. gambiae* (three intronless GSTs) [5] but far less than the number observed in *D. melanogaster*, which has 20 intronless GSTs (10 Delta and 10 Epsilon GSTs) [23]. In total, at least 94 introns have been identified from the 41 *T. castaneum* GST genes (Table 1). Each class of *T. castaneum* GSTs had similar intron numbers. Furthermore, most cytosolic TcGSTs possessed 2–3 introns, and they shared two conserved intron positions at approximately the 45th codon and 115th/100th codon from the 5′ end of the gene. Few of the TcGSTs possess a different number of introns. For example, TcGSTd2 has the largest number of introns (5 introns), and TcGSTs7 is intronless. Moreover, most microsomal GSTs possess fewer introns (1–2 introns) in insects (Supplementary Tables S1–5).

Of the 94 introns, their sizes ranged from 37 to 2978 bp among the *T. castaneum* GSTs, for an average of 248 bp. There are 74 short introns (<100 bp), 13 longer introns (100–1000 bp) and 7 especially long introns (>1000 bp). Of these, 64, 24, and 6 belong to phase 0, phase 1, and phase 2 introns, respectively. Phase 0 introns show significant dominance compared with the other two groups. The insect-specific classes (the Delta and Epsilon classes) only contain Phase 0 introns except that TcGSTd2 contains 2 Phase 2 introns. Phase 1 introns mainly existed in non-insect specific classes, which include Omega, Sigma, Theta, and Zeta GSTs. 4 additional Phase 2 introns were present in two TcMGSTs, one Theta and one unknown GSTs.

2.3. Evolution of GSTs

2.3.1. Hypothetical expansion history of T. castneum GSTs

A phylogenetic analysis of each sub-class of GST was performed. By combining these results and the distributions of GST genes in the

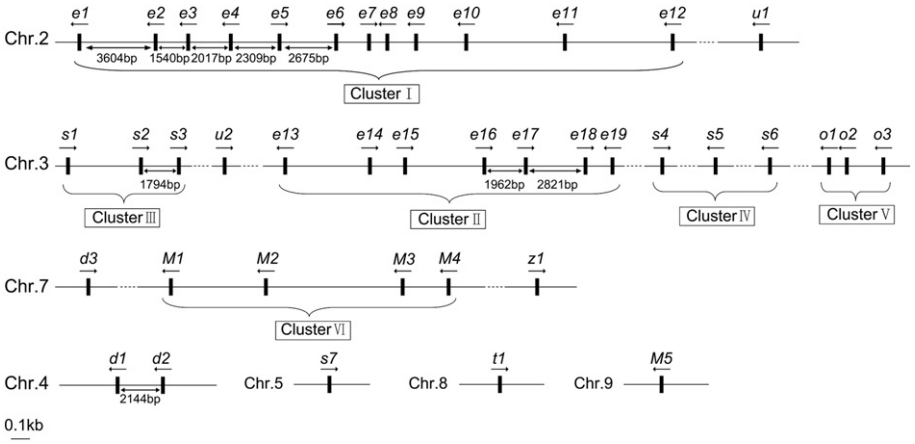

Fig. 1. The location of GST genes in the *T. castaneum* genome. Each vertical bar represents a GST gene. The arrows indicate gene orientation. Dashed lines represent an especially long chromosome distance. The regions with more than three tandem GST genes from the same class are considered to be a cluster.

genome, the hypothetical expansion history of *T. castneum GSTs* was reconstructed (Fig. 3, Supplementary Fig. S1). Analysis of the GST families of *A. gambiae, D. melanogaster* and *A. aegypti* shows that the Epsilon class is organized as a single gene cluster [24]. However, Epsilon GSTs in *T. castaneum* consisted of two clusters, which were named clusterIand II, on the second and third chromosomes, respectively. There are four pairs of corresponding genes within clusterIand II in the tree, TcGSTe13 and TcGSTe10 to TcGSTe12, TcGSTe14 and TcGSTe3 to TcGSTe5, TcGSTe15 to TcGSTe18 and TcGSTe2, and TcGSTe19 and TcGSTe1. This shows that the original Epsilon GST gene was duplicated and formed four copies on Chromosome 2, and then had a one-time long-fragment duplication generating another four copies that were inserted into Chromosome 3 (Figs. 1 and 3). The arrangement of the genes in cluster Iis in inverted order with those of cluster II. This indicated that this might represent an inverted long-fragment duplication with these four GST genes. After this process, these genes seemed to have experienced a series of additional tandem duplications and rearrangements and finally formed 19 Epsilon GSTs. Clusters III and IV, mostly on the same chromosome except TcGSTs7, seemed to have experienced six rounds of tandem duplication and a rearrangement. In the last round, a newly generated gene, TcGSTs7, translocated to another chromosome. Cluster V has only three members and possibly only experienced 2 duplications. The expansion of cluster VI is similar to that of clusters III and IV, with a jumping and a rearrangement, and three rounds of tandem duplication.

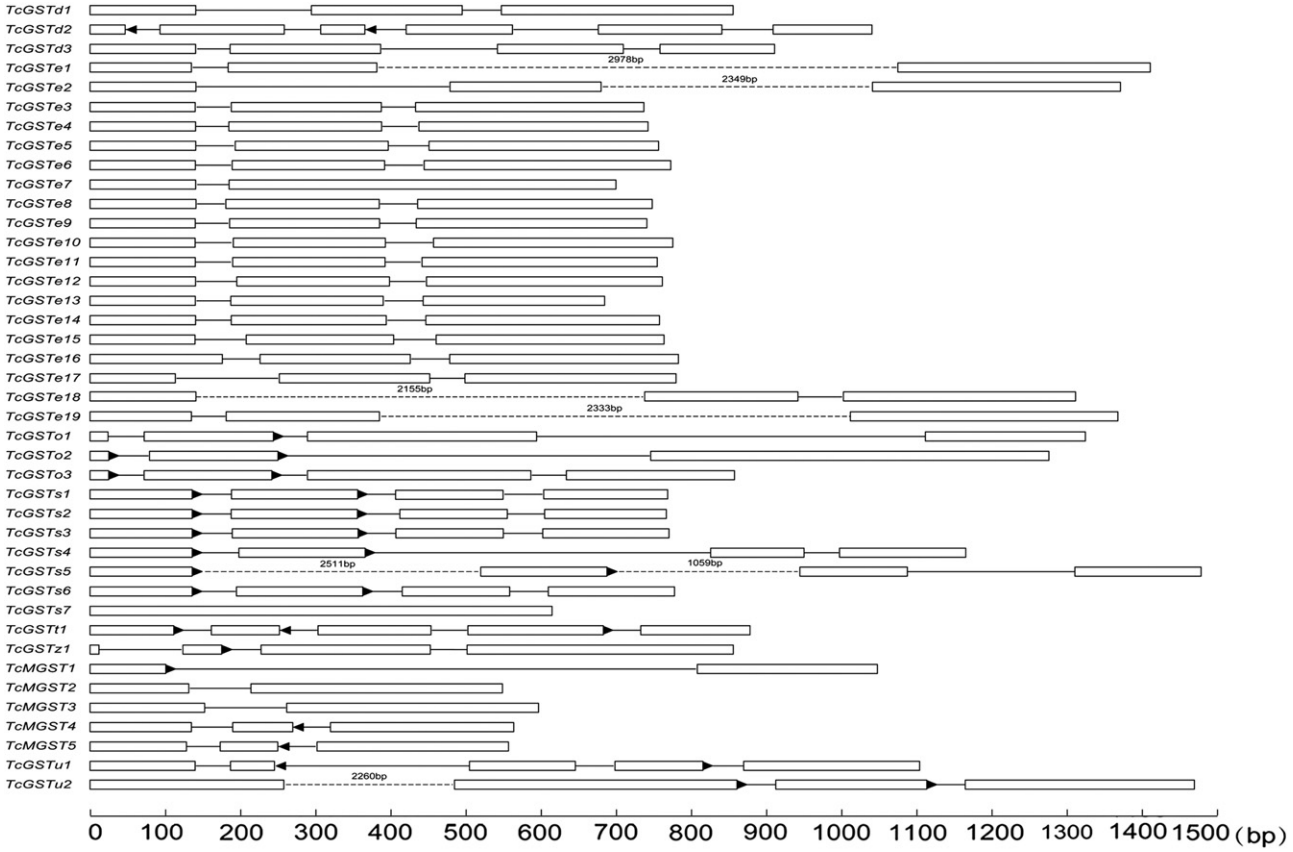

Fig. 2. Exon–intron structure of each *T. castaneum* GST gene. Boxes and horizontal lines represent exons and introns. Dashed lines indicate introns with a length longer than 900 bp. Exons and short-length introns were drawn to scale. Phase 0, -1, -2 are marked by no, right and left arrowheads, respectively.

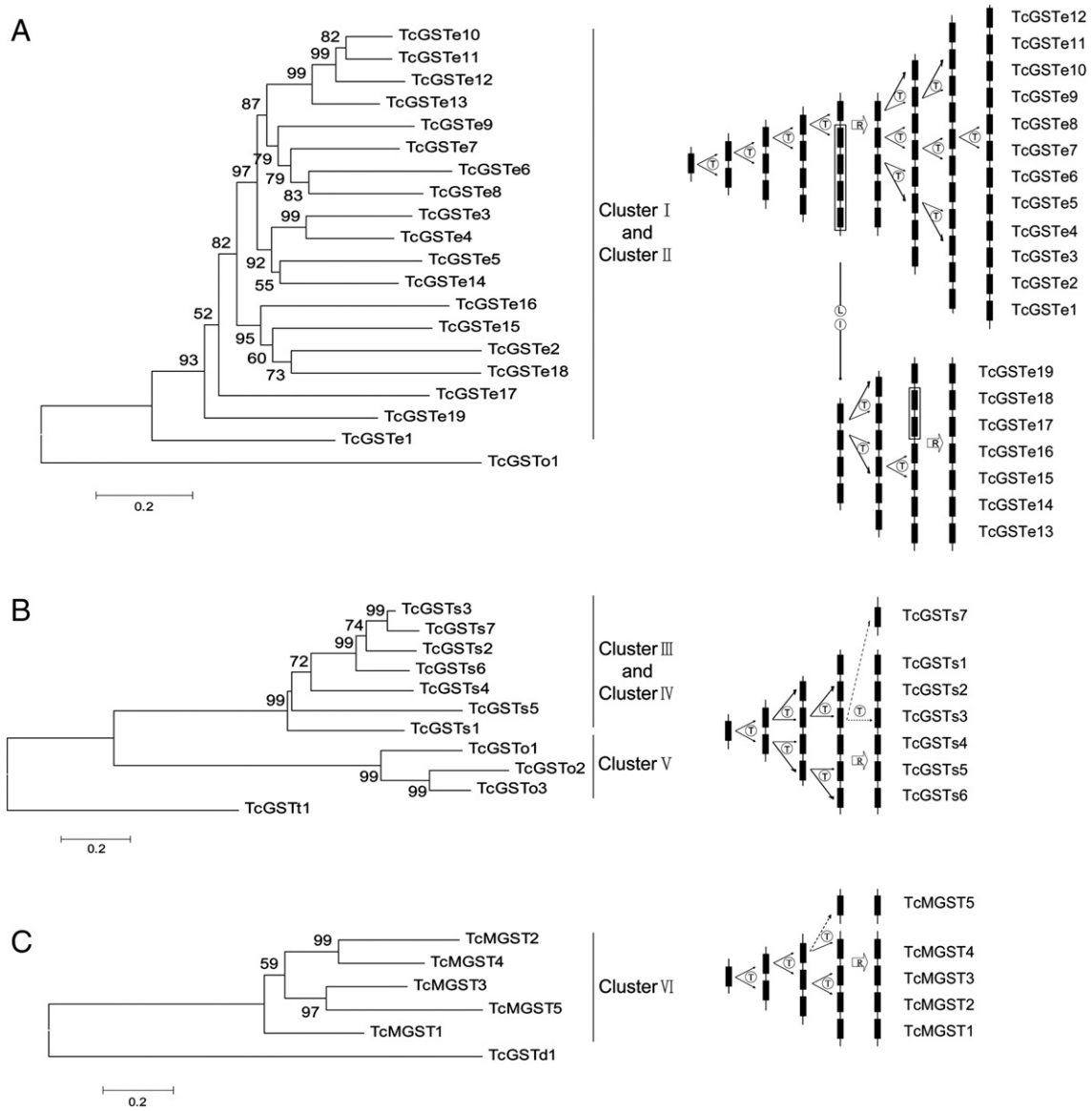

Fig. 3. Hypothetical expansion histories of *T. castneum* GSTs in each cluster. The most parsimonious scenario of expansion was assumed by combining both the gene tree and distribution of genes within clusters. Neighbor-joining trees are shown and ML trees are available in Supplemental data Fig. 2. The letters T, L, R, I indicate putative tandem duplications, long-fragment duplications, rearrangements and inversions, respectively.

2.3.2. Species specificity and evolutionary conservation

A phylogenetic analysis of all six species GSTs was performed using DmMGST1 as an outgroup. These experiments revealed that there are two types of evolutionary patterns on the phylogenetic tree. One type is the genes that had experienced species specific expansion (Fig. 4 and Supplementary Fig. S2) and another type is those genes that had near orthologous clusters with few duplications of non-insect-specific GSTs, which were highly conserved.

Epsilon GSTs are insect-specific GSTs, and they were highly expanded in *D. melanogaster*, *A. gambiae, B. mori* and *T. castneum* (Fig. 4A), suggesting that this class should have larger variation between different species and higher conservation within the species. This indicated that this family was highly expanded after insect speciation. However, this insect class was absent in *A. mellifera*. Delta, another insect-specific class also showed a similar species-specific expansion pattern (Fig. 4B).

The widely distributed, non-insect-specific GSTs had near orthologous clusters with few duplications. Zeta GST is the most highly conserved class, shows 37.9–86.0% (average 64.2%) identity within the class and likely formed as one orthologous cluster (Supplementary Figs. S1 and S4). Each insect has one copy of a Zeta GST, except *D. melanogaster* and *B. mori*, which have two copies and *A. pisum*, which lacks the genes entirely. The second conserved class is Theta, which shows 27.3–65.4 (average 41.5%) identity within the class. Similar to the Zeta class, every insect has one Theta GST, except *A. pisum, D. melanogaster* and *A. gambiae* have 2, 4 and 2 copies, respectively (Supplementary Figs. S1 and S4). The other two non-insect-specific GSTs, Omega and Sigma, were also less highly duplicated in insects, but they have slightly higher expansion than those of Zeta and Theta. The Omega class has expanded in *B. mori* and *T. castneum,* and the Sigma class is also larger in *A. pisum* and *T. castneum,* which shows a certain degree of species-specificity (Supplementary Fig. S5).

2.4. Microsomal GSTs

Microsomal GSTs, which are now grouped into MAPEG, are also ubiquitous. Multiple sequence alignments and calculations of evolutionary trees revealed that MAPEG consisted of MGST1, MGST2, MGST3, and three additional subfamilies, leukotriene C_4 synthase (LTC_4S), 5-lipoxigenase activating protein (FLAP), and prostaglandin

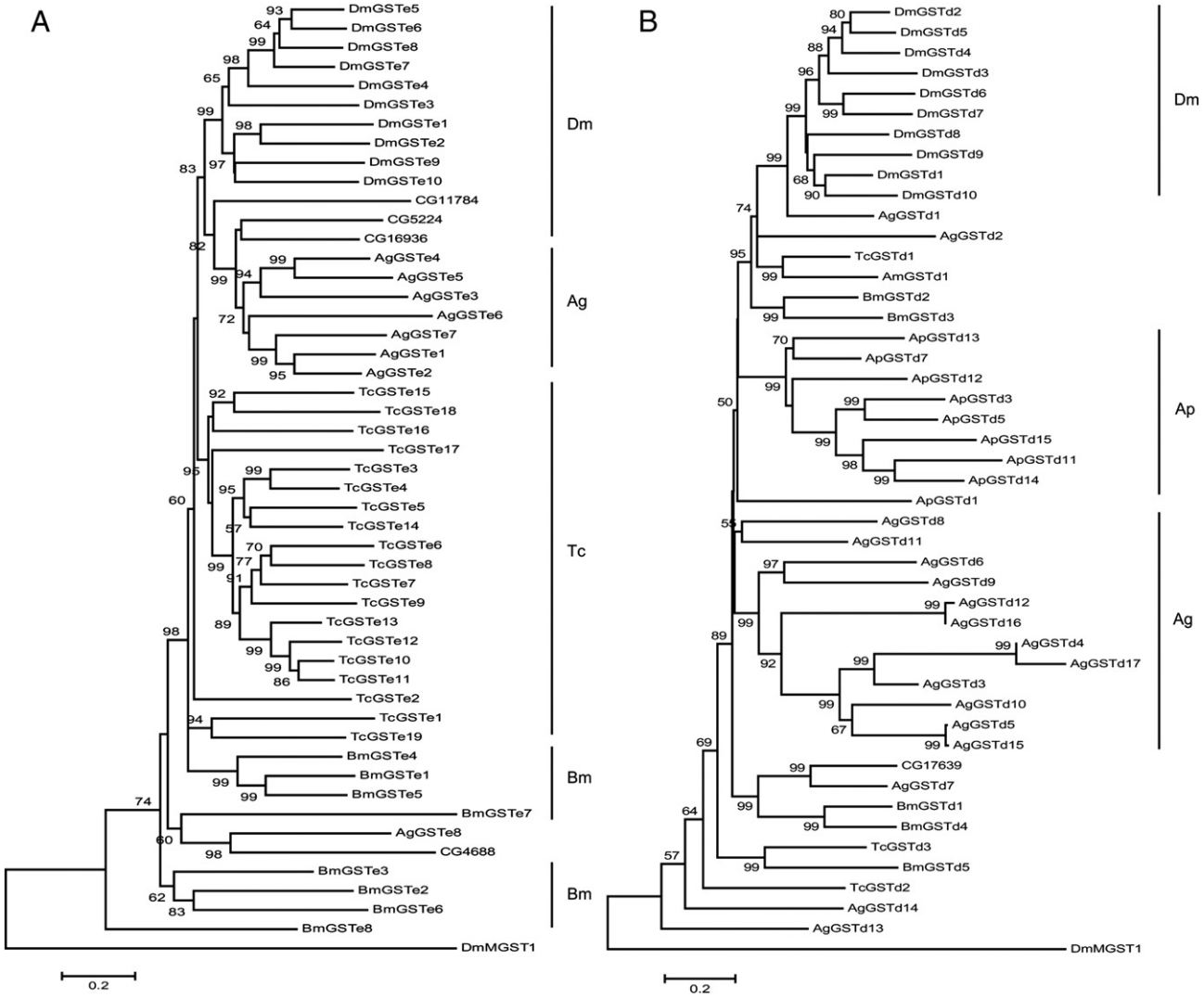

Fig. 4. Neighbor-joining consensus tree of Epsilon and Delta GSTs. A. Phylogenetic tree of insect Epsilon GSTs. B. Phylogenetic tree of insect Delta GSTs. Dm, Ag, Tc, Bm and Ap are abbreviations of *Drosophila melanogaster*, *Anopheles gambiae*, *Tribolium castaneum*, *Bombyx mori* and *Acyrthosiphon pisum*. The ML tree is shown in Supplemental Fig. S2.

E synthase (PGES) [9,10]. The insect MAPEG members are most similar to MGST1 and PGES [25]. Evolutionarily, it is thought to be formed by domain insertion from GRX-1, whereas cytosolic GSTs formed by domain addition [10].

The number of microsomal GSTs is significantly less than the number of cytosolic GSTs. All 15 MGSTs of the five insect species were aligned (Fig. 5), and there was a highly conservative motif, which has 16 amino acids (VERVRRAHRNDLENIL, marked by a box in Fig. 5). This motif also exists in human and rat MGSTs [26], but it is different from cytosolic GSTs, which have the consensus sequence SNAIL/TRAIL in their N-termini [27,28].

3. Discussion

According to GST cellular location, GSTs are generally divided into three major categories, cytosolic, microsomal, and mitochondrial. Previous studies usually focused on cytosolic GSTs, while research of microsomal GSTs usually concentrated on human diseases. Considering microsomal GSTs may also play a role in xenobiotic detoxification [29], both of these groups were analyzed in the six genome sequenced insect species. 36 putative cytosolic GSTs and 5 microsomal GSTs were identified in *T. castaneum*, and 2 of them could not be classified into any known classes. A total of 41 GSTs were identified, and this is comparable to the number found in Dipteran insects (40 in *D. melanogaster*, and 35 in *A. gambiae)*. It is far more than that of in the other insect species. *T. castaneum* has largest number of Epsilon GSTs (19 genes), however it has fewer Delta GSTs.

For multiple reasons, the GST numbers are variable among the six insect lineages. Delta and Epsilon are especially expanded in insects, and they occupy over 50% of the entire cytosolic subgroup except in *A. mellifera*, which is consistent with previous studies from the dipteran insects [6]. Gene duplication is a fundamental process in evolution and is believed to play leading roles for the creation of novel gene functions [30]. Gene family expansion is mainly through gene duplication. The hypothetical expansion history for Epsilon class contains a number of tandemly arrayed Genes (TAGs), which account for about one third of the duplicated genes in eukaryotes [31]. But long-fragment duplications are relatively rare. It is now estimated that ~5% of human genetic material is composed of segmental duplications [32]. To date, the formation mechanism of long-fragment duplications remains obscure [33].

Different functions are suggested by changes in the size of GST protein families, which can indicate adaptation [22]. *T. castaneum* is a seriously harmful insect to granaries, and it is noteworthy that *T. castaneum* has a strikingly larger Epsilon class than the others. Previous studies have demonstrated that Epsilon GSTs correlate with detoxification. For example, Ortelli *et al.* illustrated that GSTe2-2 in *A. gambiae* was able to metabolize DDT [34]. GSTe2 in *A. aegypti*

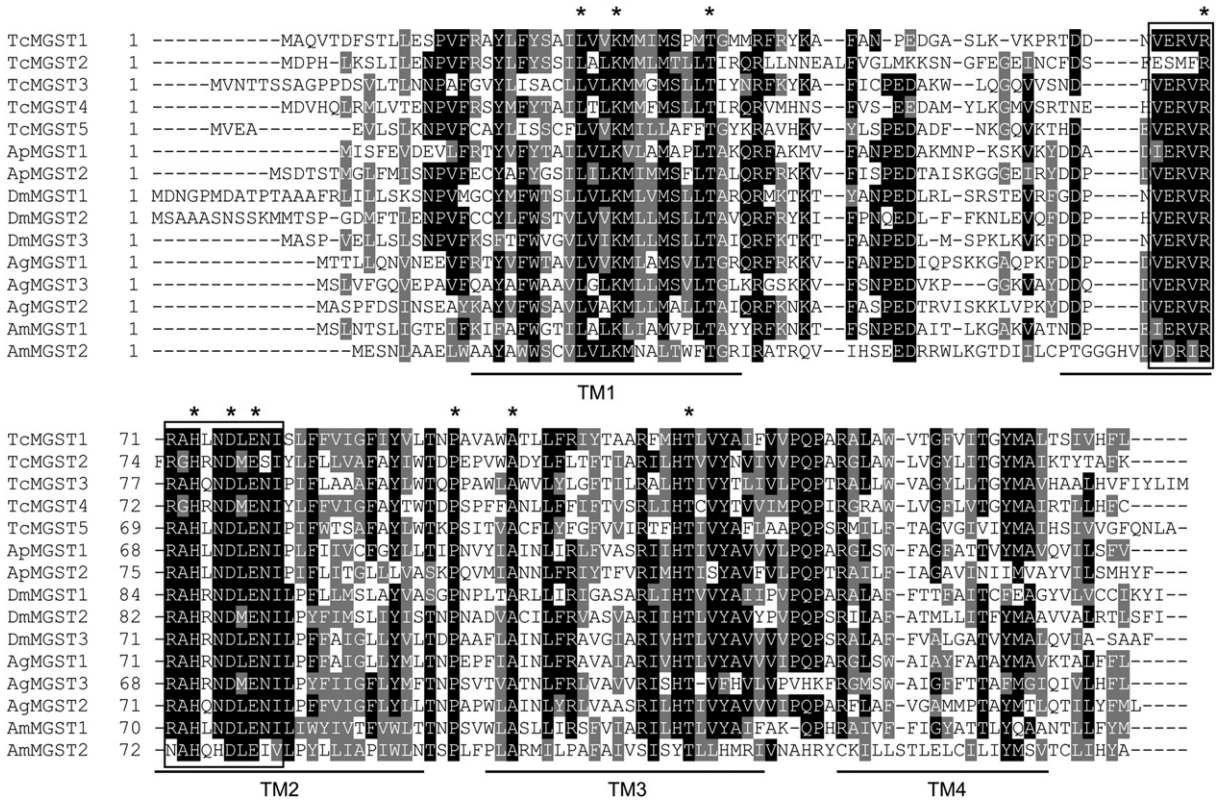

Fig. 5. Multiple alignments of microsomal GSTs. TM indicates transmembrane regions, using *Rattus norvegicus* MGST1 (GenBank accession number: NP_599176) as a template sequence from the website (http://swissmodel.expasy.org/). Conserved residues are marked with an asterisk.

confers resistance to DDT as well [11]. The association of amplification of GST gene and resistance has also been proved [35–38]. Rason and Claudianos thought that the multiple independent radiation of Delta and Epsilon class in *D.melanogaster* and *A. gambiae* suggested their important roles in the adaptation to environmental selection pressures [6,16]. Ayres et al. conjectured that Epsilon clusters might be important for adaptation to different habitats and calculated the likelihood ratio test of positive selection at sites of Epsilon class. Their results demonstrated that at least one gene GSTe5 was under positive selection [39]. These inferred that Epsilon clusters could help to survival under the selection pressures. Associated to the detoxification function of Epsilon class, it suggested that *T. castaneum* has higher insecticides resistance and adaptability may be due to the expansion of this class. And then, it could help the beetle survive in the presence of poisonous or rugged environments.

To date, Delta and Epsilon GSTs are the only two classes that have been postulated to affect pesticide sensitivity. It is not clear that how many GST genes are necessary to provide adequate detoxification capacity. Papadopoulos *et al.* found two main GST isoenzymes were present in *A. mellifera*, playing essential roles in the survival of insects exposed to endogenous or exogenous xenobiotics and these two GST isoenzymes were controlled independently [40]. Later it was characterized that *A. mellifera* has only two Delta GSTs and no Epsilons. In the opinion of Claudianos *et al.* this might be the reason that the honeybee was one of the main non-target species in toxicological studies of new insecticides [16]. A plausible explanation is that *A. mellifera* does not have significant exposure to insecticides, and these two Delta GSTs are capable of combating the environmental selection pressures, which might involve behavioral adaptation [16,40].

MGSTs were found in the insects mentioned above except *B. mori*. It was assumed this was a consequence of food habit. MGSTs are relevant for the biosynthesis of arachidonic acid, which is the prosoma of prostaglandin. Arachidonic acid and prostaglandin both have intense inhibitory action toward gastric juice secretion [41–43]. Compared with the other insects analyzed in this manuscript, *B. mori* needs more gastric acid to digest the abundant mulberry leaf and it may maintain the digestive function by degrading the microsomal subfamily. However, this is not consistent with the fact that there are microsomal members in other herbivorous insects, such as *Heliothis virescens* (Lepidoptera).

Variations in gene (exon-intron) organization are often used to understand evolutionary relationships [44]. Given that cytosolic and microsomal GSTs had significant differences in both sequence and gene structure similarity, only intron numbers in the cytosolic GSTs group were analyzed and compared among 6 insect species. The average number was 3.3 introns per gene in *A. pisum* (Homoptera), 2.8 in *A. mellifera* (Hymenoptera), 2.4 in *T. castaneum* (Coleoptera), 3.1 in *B. mori* (Lepidoptera), 1.5 in *A. gambiae* (Diptera) and 1.0 in *D. melanogaster* (Diptera). This analysis shows that the intron numbers of GSTs were reduced as their evolutional status increased from the basal Homoptera to advanced Diptera insects, with the exception of *B. mori* (Supplementary Fig. S6) [45]. This might be because genes with fewer introns have more rapid response to endogenous and xenobiotic compounds than those with more introns [46]. Higher insects need to be able to generate fast responses to multiple endogenous and xenobiotic compounds, and thus they have possessed more rapid GST transcription and expression regulation by reducing the number of redundant introns. So far the whole genome of about 40 insects have been sequenced. These insects belong to seven orders, respectively Homoptera, Hymenoptera, Coleoptera, Lepidoptera, Diptera, Hemiptera and Phthiraptera. The representative insects of the first five orders have been chosen and analyzed here, of course, datas of other insects such as *Rhodnius prolixus* (Hemiptera) and *Pediculus humanus corporis* (Phthiraptera) will enrich the credibility in the future study.

In addition to the best-characterized function in insecticide resistance, the Alpha, Pi and Mu classes in humans are mediators of signaling pathways that are involved in cell proliferation and cell death by interacting with important signaling proteins in a non-enzymatic way [47]. This suggests that GSTs have far more functions than was

previously believed. Here we presented an overview of the genomic organization and evolution of six insect GSTs. However, little is known about how these genes are related to the degradation of insecticides and other xenobiotics, and what the mechanisms are that produce insecticide resistance for each individual GST. Thus, further functional research of TcGSTs is needed to identify their potential functions and mechanisms. Moreover, limited to the bulkiness of the GSTs superfamily, high throughput techniques such as proteomics and transcriptomics are expected to facilitate the resolution of this issue.

4. Materials and methods

4.1. Identification and annotation of GST genes in T. castaneum and other insects

To search for putative GST genes from *T. castaneum* and other insect genomes, *D. melanogaster* GST genes were utilized as reference sequences to perform TBLASTN searches. Combing the results from Rason and Low [6,48], 38 *D. melanogaster* GSTs were collected. Thus this primitive query set was used to validate the search power on *D. melanogaster* genome database itself. As results, not only those query GSTs were validated, but also three more putative GST genes were raised out. Finally a set of 40 DmGSTs was characterized, including 37 cytosolic GSTs and 3 microsomal GSTs as references. With a few changes, our set has the four following differences:

a. Two Delta members CG4371 and CG4381 were kept, and they are classified as pseudogenes by previous research. However, later Sawicki, R *et al.* demonstrated that they were expressed [23].
b. CG6673A and CG6673B were considered to be one gene, and previous research considered them as two independent genes.
c. CG4623 was included, which was classified as ganglioside-induced differentiation-associated protein-1 (GDAP1) by Antonio Marco [49]. However, it contained the feature domains of GSTs in its N-terminal and C-terminal regions. Previous research excluded it because it was too long and too divergent from other GSTs.
d. As for microsomal GSTs, three genes CG1742, CG33177, and CG33178 were adopted [17], and in the anterior set only CG1742 was adopted.

To further certify the integrity of the reference set, GSTs of relatively deeply researched insect *A. gambiae* were first analyzed. The results completely covered the earlier reports besides our new adopted referenced genes. Thus, these GST genes were utilized as queries to perform tblastn searches against the whole genome database of *T. castaneum* (http://beetlebase.org/) and the NCBI database (National Center for Biotechnology Information). Similarly, *A. pisum* GSTs were searched from AphidBase (http://www.aphidbase.com/aphidbase/) and NCBI. Although GSTs of *A. gambiae, A. mellifera* and *B. mori* have been analyzed previously, as their genomic data has been re-assembled or renewed recently, we re-analyzed those data independently to validate the results from the genomic databases, VectorBase (http://www.vectorbase.org/index.php), BeeBase (http://genomes.arc.georgetown.edu/drupal/beebase/) and SilkBase (http://silkworm.swu.edu.cn/silkdb/).

4.2. Multiple sequence alignments and phylogenetic analyses

Multiple alignments of the amino acid sequences were performed using ClustalW2 at the website (http://www.ebi.ac.uk/Tools/clustalw2/index.html). The phylogenetic trees were constructed using MEGA 5.0 software using the Neighbor-Joining (NJ) method [16,50]. Poisson models and pairwise deletions were selected. 1000 bootstrap tests were performed and values lower than 50% were not shown. The maximum-likelihood (ML) trees were generated using PHYML3.0 (http://www.atgc-montpellier.fr/phyml/) [51] by the LG amino acid substitution model.

4.3. Genomic location and hypothetical expansion history

The chromosome locations and orientations were analyzed for each GST gene of *T. castaneum*. Some tandem GSTs were designated as clusters. According to both the gene tree and the positions of genes within clusters, the hypothetical expansion history of each cluster was reconstructed. Gene duplication, loss, and rearrangement were shown with the most parsimonious scenario.

4.4. EST evidence of T. castaneum GSTs

The EST database of *T. castaneum* was downloaded from BeetleBase and NCBI, which was utilized to search for EST evidence by local Blast. Putative GST coding sequences were used as queries, and a 95% or greater identity and minimum cut-off E-value (e^{-20} or smaller) were employed to discriminate between duplicated genes [18].

Acknowledgments

This project was supported by the Key Project of Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 10KJA180023), the National Natural Science Foundation of China (No. 31172146), the Natural Science Foundation of Jiangsu Province, China (BK2011785), and PAPD of Jiangsu Higher Education Institutions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2012.07.010.

References

[1] A.A. Enayati, H. Ranson, J. Hemingway, Insect glutathione transferases and insecticide resistance, Insect Mol. Biol. 14 (2005) 3–8.
[2] R.N. Armstrong, Structure, catalytic mechanism, and evolution of the glutathione transferases, Chem. Res. Toxicol. 10 (1997) 2–18.
[3] P.J. Jakobsson, J.A. Mancini, A.W. Ford-Hutchinson, Identification and characterization of a novel human microsomal glutathione S-transferase with leukotriene C4 synthase activity and significant sequence identity to 5-lipoxygenase-activating protein and leukotriene C4 synthase, J. Biol. Chem. 271 (1996) 22203–22210.
[4] J.D. Hayes, J.U. Flanagan, I.R. Jowsey, Glutathione transferases, Annu. Rev. Pharmacol. Toxicol. 45 (2005) 51–88.
[5] Y. Ding, F. Ortelli, L.C. Rossiter, J. Hemingway, H. Ranson, The *Anopheles gambiae* glutathione transferase supergene family: annotation, phylogeny and expression profiles, BMC Genomics 4 (2003) 35.
[6] H. Ranson, C. Claudianos, F. Ortelli, C. Abgrall, J. Hemingway, M.V. Sharakhova, et al., Evolution of supergene families associated with insecticide resistance, Science 298 (2002) 179–181.
[7] H. Ranson, L. Rossiter, F. Ortelli, B. Jensen, X. Wang, C.W. Roth, et al., Identification of a novel class of insect glutathione S-transferases involved in resistance to DDT in the malaria vector *Anopheles gambiae*, Biochem. J. 359 (2001) 295–304.
[8] S.E. Pemble, J.B. Taylor, An evolutionary perspective on glutathione transferases inferred from class-theta glutathione transferase cDNA sequences, Biochem. J. 287 (Pt 3) (1992) 957–963.
[9] C. Frova, Glutathione transferases in the genomics era: new insights and perspectives, Biomol. Eng. 23 (2006) 149–169.
[10] A.J. Oakley, Glutathione transferases: new functions, Curr. Opin. Struct. Biol. 15 (2005) 716–723.
[11] N. Lumjuan, S. Rajatileka, D. Changsom, J. Wicheer, P. Leelapat, L.A. Prapanthadara, et al., The role of the *Aedes aegypti* Epsilon glutathione transferases in conferring resistance to DDT and pyrethroid insecticides, Insect Biochem. Mol. Biol. 41 (2011) 203–209.
[12] M. Perez-Lopez, M.C. Novoa-Valinas, M.J. Melgar-Riol, Glutathione S-transferase cytosolic isoforms as biomarkers of polychlorinated biphenyl (Arochlor-1254) experimental contamination in rainbow trout, Toxicol. Lett. 136 (2002) 97–106.
[13] J.A. Ayoola, J. Garcia-Alonso, J.D. Hardege, Glutathione-S-transferase in *Nereis succinea* (Polychaeta) and its induction by xeno-estrogen, Environ. Toxicol. (2010).
[14] D. Vrana, H. Pikhart, B. Mohelnikova-Duchonova, I. Holcatova, R. Strnad, A. Slamova, et al., The association between glutathione S-transferase gene polymorphisms and pancreatic cancer in a central European Slavonic population, Mutat. Res. 680 (2009) 78–81.
[15] B. Chen, Y. Zhou, P. Yang, X.T. Wu, Glutathione S-transferase M1 gene polymorphism and gastric cancer risk: an updated analysis, Arch. Med. Res. 41 (2010) 558–566.

[16] C. Claudianos, H. Ranson, R.M. Johnson, S. Biswas, M.A. Schuler, M.R. Berenbaum, et al., A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee, Insect Mol. Biol. 15 (2006) 615–636.
[17] M. Corona, G.E. Robinson, Genes of the antioxidant system of the honey bee: annotation and phylogeny, Insect Mol. Biol. 15 (2006) 687–701.
[18] Q. Yu, C. Lu, B. Li, S. Fang, W. Zuo, F. Dai, et al., Identification, genomic organization and expression pattern of glutathione S-transferase in the silkworm, *Bombyx mori*, Insect Biochem. Mol. Biol. 38 (2008) 1158–1164.
[19] S. Richards, R.A. Gibbs, G.M. Weinstock, S.J. Brown, R. Denell, R.W. Beeman, et al., The genome of the model beetle and pest *Tribolium castaneum*, Nature 452 (2008) 949–955.
[20] D. Andreev, M. Kreitman, T.W. Phillips, R.W. Beeman, R.H. ffrench-Constant, Multiple origins of cyclodiene insecticide resistance in *Tribolium castaneum* (Coleoptera: Tenebrionidae), J. Mol. Evol. 48 (1999) 615–624.
[21] D. Wool, S. Noiman, O. Manheim, E. Cohen, Malathion resistance in Tribolium strains and their hybrids: inheritance patterns and possible enzymatic mechanisms (Coleoptera, Tenebrionidae), Biochem. Genet. 20 (1982) 621–636.
[22] E.M. Zdobnov, C. von Mering, I. Letunic, D. Torrents, M. Suyama, R.R. Copley, et al., Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*, Science 298 (2002) 149–159.
[23] R. Sawicki, S.P. Singh, A.K. Mondal, H. Benes, P. Zimniak, Cloning, expression and biochemical characterization of one Epsilon-class (GST-3) and ten Delta-class (GST-1) glutathione S-transferases from *Drosophila melanogaster*, and identification of additional nine members of the Epsilon class, Biochem. J. 370 (2003) 661–669.
[24] B.P. Niranjan Reddy, G.B. Prasad, K. Raghavendra, In silico characterization and comparative genomic analysis of the *Culex quinquefasciatus* glutathione S-transferase (GST) supergene family, Parasitol. Res. 109 (2011) 1165–1177.
[25] A. Bresell, R. Weinander, G. Lundqvist, H. Raza, M. Shimoji, T.H. Sun, et al., Bioinformatic and enzymatic characterization of the MAPEG superfamily, FEBS J. 272 (2005) 1688–1703.
[26] P.J. Holm, P. Bhakat, C. Jegerschold, N. Gyobu, K. Mitsuoka, Y. Fujiyoshi, et al., Structural basis for detoxification and oxidative stress protection in membranes, J. Mol. Biol. 360 (2006) 934–945.
[27] A. Torres-Rivera, A. Landa, Glutathione transferases from parasites: a biochemical view, Acta Trop. 105 (2008) 99–112.
[28] E.V. Koonin, A.R. Mushegian, R.L. Tatusov, S.F. Altschul, S.H. Bryant, P. Bork, et al., Eukaryotic translation elongation factor 1 gamma contains a glutathione transferase domain — study of a diverse, ancient protein superfamily using motif search and structural modeling, Protein Sci. 3 (1994) 2045–2054.
[29] J.H. Kim, S. Raisuddin, J.S. Rhee, Y.M. Lee, K.N. Han, J.S. Lee, Molecular cloning, phylogenetic analysis and expression of a MAPEG superfamily gene from the pufferfish *Takifugu obscurus*, Comp. Biochem. Physiol. C Toxicol. Pharmacol. 149 (2009) 358–362.
[30] S. Ohno, Evolution by Gene Duplication, Springer, Berlin, 1970.
[31] L. Zhou, B. Huang, X. Meng, G. Wang, F. Wang, Z. Xu, et al., The amplification and evolution of orthologous 22-kDa alpha-prolamin tandemly arrayed genes in coix, sorghum and maize genomes, Plant Mol. Biol. 74 (2010) 631–643.
[32] R.V. Samonte, E.E. Eichler, Segmental duplications and the evolution of the primate genome, Nat. Rev. Genet. 3 (2002) 65–72.
[33] A.S. Fiston-Lavier, D. Anxolabehere, H. Quesneville, A model of segmental duplication formation in *Drosophila melanogaster*, Genome Res. 17 (2007) 1458–1470.
[34] F. Ortelli, L.C. Rossiter, J. Vontas, H. Ranson, J. Hemingway, Heterologous expression of four glutathione transferase genes genetically linked to a major insecticide-resistance locus from the malaria vector *Anopheles gambiae*, Biochem. J. 373 (2003) 957–963.
[35] Z.H. Zhou, M. Syvanen, A complex glutathione transferase gene family in the housefly *Musca domestica*, Mol. Gen. Genet. 256 (1997) 187–194.
[36] M. Syvanen, Z. Zhou, J. Wharton, C. Goldsbury, A. Clark, Heterogeneity of the glutathione transferase genes encoding enzymes responsible for insecticide degradation in the housefly, J. Mol. Evol. 43 (1996) 236–240.
[37] T.A. Gaines, W. Zhang, D. Wang, B. Bukun, S.T. Chisholm, D.L. Shaner, et al., Gene amplification confers glyphosate resistance in *Amaranthus palmeri*, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 1029–1034.
[38] M.G. Paton, S.H. Karunaratne, E. Giakoumaki, N. Roberts, J. Hemingway, Quantitative analysis of gene amplification in insecticide-resistant Culex mosquitoes, Biochem. J. 346 (Pt 1) (2000) 17–24.
[39] C.F. Ayres, P. Muller, N. Dyer, C.S. Wilding, D.J. Rigden, M.J. Donnelly, Comparative genomics of the anopheline glutathione S-transferase epsilon cluster, PLoS One 6 (2011) e29237.
[40] A.I. Papadopoulos, I. Polemitou, P. Laifi, A. Yiangou, C. Tananaki, Glutathione S-transferase in the insect *Apis mellifera macedonica* kinetic characteristics and effect of stress on the expression of GST isoenzymes in the adult worker bee, Comp. Biochem. Physiol. C Toxicol. Pharmacol. 139 (2004) 93–97.
[41] P. Tao, W. Scruggs, D.E. Wilson, The effects of a prostaglandin endoperoxide analogue on canine gastric acid and mucus secretion, Dig. Dis. Sci. 24 (1979) 449–454.
[42] S.J. Konturek, E. Mikos, W. Pawlik, K. Walus, Direct inhibition of gastric secretion and mucosal blood flow by arachidonic acid, J. Physiol. 286 (1979) 15–28.
[43] D.M. Chang, T.H. Young, C.T. Hsu, S.Y. Kuo, T.C. Hsieh, Endoscopic comparison of the gastroduodenal safety and the effects on arachidonic acid products between meloxicam and piroxicam in the treatment of osteoarthritis, Clin. Rheumatol. 20 (2001) 104–113.
[44] D. Sanchez, M.D. Ganfornina, G. Gutierrez, A. Marin, Exon–intron structure and evolution of the Lipocalin gene family, Mol. Biol. Evol. 20 (2003) 775–783.
[45] J. Savard, D. Tautz, S. Richards, G.M. Weinstock, R.A. Gibbs, J.H. Werren, et al., Phylogenomic analysis reveals bees and wasps (*Hymenoptera*) at the base of the radiation of Holometabolous insects, Genome Res. 16 (2006) 1334–1338.
[46] D.C. Jeffares, C.J. Penkett, J. Bahler, Rapidly regulated genes are intron poor, Trends Genet. 24 (2008) 375–378.
[47] E. Laborde, Glutathione transferases as mediators of signaling pathways involved in cell proliferation and cell death, Cell Death Differ. 17 (2010) 1373–1380.
[48] W.Y. Low, H.L. Ng, C.J. Morton, M.W. Parker, P. Batterham, C. Robin, Molecular evolution of glutathione S-transferases in the genus *Drosophila*, Genetics 177 (2007) 1363–1375.
[49] A. Marco, A. Cuesta, L. Pedrola, F. Palau, I. Marin, Evolutionary and structural analyses of GDAP1, involved in Charcot-Marie-Tooth disease, characterize a novel class of glutathione transferase-related genes, Mol. Biol. Evol. 21 (2004) 176–187.
[50] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol. Biol. Evol. 28 (2011) 2731–2739.
[51] S. Guindon, F. Lethiec, P. Duroux, O. Gascuel, PHYML Online — a web server for fast maximum likelihood-based phylogenetic inference, Nucleic Acids Res. 33 (2005) W557–W559.