*Pakistan J. Zool., vol. 49(4), pp 1437-1448, 2017.* **DOI**: http://dx.doi.org/10.17582/journal.pjz/2017.49.4.1437.1448

# Genomic Analysis of Glutathione S-transferases (GST) Family in Common Carp: Identification, Phylogeny and Expression

**Baohua Chen[1,2], Wenzhu Peng[3], Jian Xu[2], Jingyan Feng[1,2], Chuanju Dong[1,2] and Peng Xu[2,3,*]**

[1]*College of Life Science and Technology, Shanghai Ocean University, Shanghai, 201306, China*

[2]*Key Laboratory of Aquatic Genomics, Ministry of Agriculture, CAFS Key Laboratory of Aquatic Genomics and Beijing Key Laboratory of Fishery Biotechnology, Chinese Academy of Fishery Sciences, Beijing 100141, China*

[3]*Fujian Collaborative Innovation Center for Exploitation and Utilization of Marine Biological Resources, College of Ocean and Earth Science, Xiamen University, Xiamen, 361005, China*

**Baohua Chen and Wenzhu Peng contributed equally to this article.**

**Article Information**
**Received 7 March 2016**
**Revised 25 January 2017**
**Accepted 29 January 2017**
**Available online 25 July 2017**

**Authors' Contribution**
**PX conceived the study. BC and WP wrote the manuscript. BC, WP and JX performed the bioinformatics analysis and conducted the phylogenetic analysis. CD and JF helped in data collection.**

**Key words**
**Glutathione S-transferases, Common carp genome, Gene family, Gene duplication.**

## ABSTRACT

Glutathione S-transferases (GSTs) comprise a large and diverse family of enzymes with a wide phylogenetic distribution. They are multifunctional enzymes that play a crucial role in cellular detoxification and oxidative stress tolerance. Comparing with that in mammals, investigation of GSTs is more complicated in teleosts because of the greater pressure they suffer in aquatic environment. In this study, we identified a set of 27 GSTs including 8 classes of members in common carp genome. Both sequences alignment and phylogenetic analysis exhibited that genes derived from the same GST class from different species share more similarity than genes of different classes in the same species. Copy number of GSTs examining showed that five classes of GST genes in common carp have undergone the gene duplications, including MGST1, GSTK, GSTM, GSTA and GSTT. Comparative genomics and syntenic analysis provided new evidences for better understanding on gene fates post whole genome duplication (WGD) of common carp. The expression patterns of all GST genes were established in various tissues, including brain, heart, spleen, kidney, intestine, gill, liver, skin, blood and muscle of common carp. Expression profiles provided us more evidences to understand GST gene functions as well as their functional evolution post duplication. Overall, the whole set of GST genes provide essential genomic resources for future biochemical, toxicological and physiological studies in common carp.

## INTRODUCTION

Glutathione S-transferases (GSTs), also known as the glutathione transferases, comprise a large and diverse family of enzymes with a wide phylogenetic distribution. These enzymes catalyze the conjugation of glutathione (GSH) with a variety of electrophilic compounds and server as intracellular binding and transport proteins (Buetler and Eaton, 1992). On the base of these two characters, GSTs can detoxify electrophilic xenobiotics, such as environmental pollutants, drugs, and carcinogens. For example, some research suggests that polymorphic sites on glutathione-S-transferase P1 (GSTP1) are associated with risk of asthma in people (Hemmingsen *et al*., 2001; Al-Arifa and Jahan, 2016). Besides that, they can also inactivate endogenous quinones, epoxides and hydroperoxides formed as secondary metabolites during oxidative stress (Hayes *et al*., 2005). In addition, the GSTs also activate the biosynthesis of some hormones like prostaglandins and progesterone (Listowsky *et al*., 1988), as well as degradation of tyrosine.

Due to the crucial role GSTs play in detoxication of multiple compounds, especially xenobiotics, and their extensive distribution in almost every species, various investigators have focused on their purification, characterization and expression in plants and mammals. Ever since the first characterization of GST more than fifty years ago, a lot of data has been available on this family of enzymes. So far, 84 GST genes have been identified and grouped into eight classes in barley by sequence alignment and phylogenetic analysis (Rezaei *et*

* Corresponding author: xupeng77@xmu.edu.cn
0030-9923/2017/0004-1437 $ 9.00/0

*al*., 2013). Several classes of GST sequences have been identified and classified from both mammalian and non-mammalian organisms through different techniques, such as immunological methods, amino acid sequencing, molecular cloning and so on (Buetler and Eaton, 1992). In mammalian, this superfamily is composed of three subfamilies, namely cytosolic, mitochondrial, and microsomal GST (Hayes *et al*., 2005). Indeed, several attempts had been taken for the classification and nomenclatures of so much GST enzymes identified by different laboratories through different techniques (Mannervik *et al*., 1988). Eventually a generally accepted nomenclature introduced by Mannervik *et al*. (1988) mainly on human GSTs were published in 1992 (Buetler and Eaton, 1992). So far, the GSTs of mammals have been divided into several classes based on the sequence (Board *et al*., 2001), subunit structure (Ma *et al*., 2009), kinetics, inhibitor specificity (Blanchette *et al*., 2007) and immunological identity (Fan *et al*., 2007). These classes include Alpha, Mu, Pi, Theta, Sigma, Omega, and Zeta of the cytosolic GSTs (Kim *et al*., 2010), Kappa of the mitochondrial GSTs and Mgst1, Mgst2 and Mgst3 of the microsomal GSTs also designated as MAPEG now (Hayes *et al*., 2005).

The studies in teleost is more complicated, because of the aquatic environment and the greater pressure they suffer. Most of the investigations about fish GSTs focus on the expression level and changes when exposed to metal like cadmium rather than their identification. Thus, information is not enough to establish the accurate molecular phylogeny of GSTs in fish.

Common carp, *Cyprinus carpio*, one of the most significant aquaculture fish species, is widespread all over the word especially in Europe and Asia. Great efforts have been made in developing genomic resources in recent years. These genomic resources included a large number of ESTs (Christoffels *et al*., 2006), BAC end sequences (Xu *et al*., 2011), comprehensive transcriptome obtained by RNA-seq (Ji *et al*., 2012), single nucleotide polymorphism (SNPs) (Xu *et al*., 2014a), genetic and physical maps (Zhao *et al*., 2013). The common carp whole genome sequences have recently been published (Xu *et al*., 2014b). It is now known that common carp genome is allotetraploidized genome which had experienced an additional round of whole genome duplication (WGD) compared with many other teleosts. Therefore, the complexity of the tetraploidized genome and gene duplications may cause misidentification in assembly and annotation. Examination of gene families with phylogenic or orthologous analysis would verify the whole genome sequences assembly and annotation (Liu *et al*., 2013). In this study, by utilizing all available common carp genomic resources, we identified 27 GST genes across the genome. Further phylogenetic and syntemic analysis confirmed the annotation. Our study on examining gene families in common carp not only supported the accuracy of the common carp whole genome sequences assembly and annotation, but also provided valuable genomic resources for the future evolutionary, biochemical, toxicological, and physiological studies on common carp and other teleosts.

## MATERIALS AND METHODS

### *Identification of GSTs genes and homologs*

To identify the *GSTs* genes, public available databases were searched for GST family homologues in seven species: zebrafish (*Danio rerio*), human (*Homo sapiens*), chicken (*Gallus gallus*), frog (*Xenopus tropicalis*), pufferfish (*Takifugu rubripes*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*). All amino acid sequences of GSTs genes were retrieved by searching the Ensembl genome browser (http:// www.ensembl.org) and GenBank (http://www.ncbi.nlm.nih.gov/genbank/) and used as queries to search against all available common carp genomic resources, including the databases of whole genome sequences, amino acid sequences, transcriptome sequences and cDNAs, by BLAST searches to acquire the candidate genes with an E-value cut off of 1e-10. All the databases used above were sequenced, assembled and annotated by our own laboratory. Methods applied for sequence data production of common carp are described in previous publication (Xu *et al*., 2014b). The resulting alignments were checked manually to identify the best hits as candidate sequences considering score, identity values and alignment position of the query. Then reciprocal BLAST searches were conducted by using the candidate common carp *GST* genes as queries to verify the veracity of candidate genes. Additionally, the coding sequences were confirmed by BLAST searches against NCBI non-redundant protein sequence database (nr). The full-length amino acid sequences as well as the partial sequences coding for the conserved domains were used in the phylogenetic analysis. The GST proteins from other organisms were retrieved from the Ensembl genome database (Release 75) for phylogenetic analysis with exclusion of partial sequences.

### *Nomenclature of GSTs*

The predicted GST genes of common carp were named based on their zebrafish orthologs, corrected by phylogenetic topologies. First, the subfamilies and gene members were determined for each common carp GST orthologs based on the classes of zebrafish (for instance, GSTA, GSTO, *etc*). Then, the closely related zebrafish GST genes were assigned to each common carp GST orthologs, respectively.

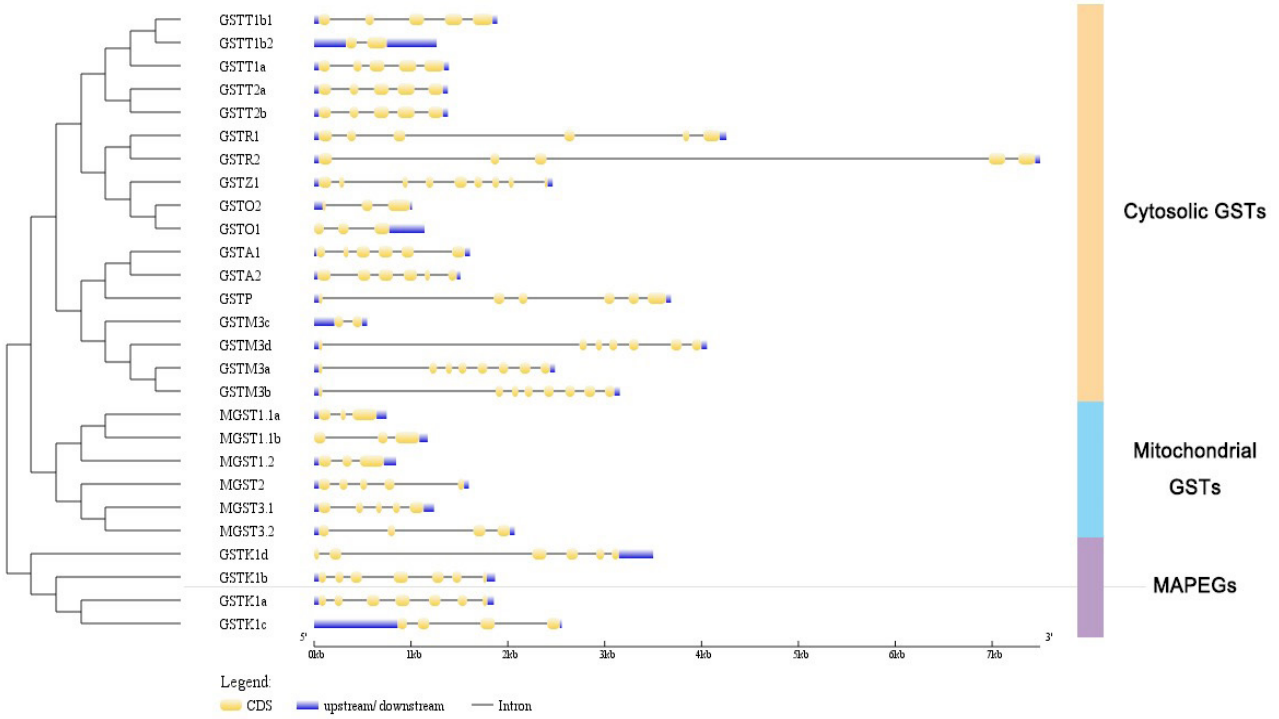


Fig. 1. Sequence alignment and gene structure of common carp *GST* genes. The sequence alignment of carp *GST* genes was conducted by ClustalO. Sketch map of *GST* genes structure was conducted on GSDS.

These GST genes were named after their closely related zebrafish genes by sequences alignment. When more than one copy of common carp *GST* genes was clustered with certain zebrafish GST gene, the alphabetical suffixes were added to each copy (for instance, *GST*K1a, *GST*K1b, *GST*K1c, *GST*K1d, *etc*). After construction of phylogenetic tree, few of these names given for carp *GST* genes were corrected to adjust the phylogenetic topologies. *GST* gene names among different teleost species have not been standardized. To prevent further confusion, we renamed all the *GST* genes which appear in this study based on the rules stated above and the original name of the gene. Names of all *GST* genes in surveyed species and their accession numbers are listed in Table I.

*Gene characterization and sequence alignment*

To characterize the gene structure, we performed exon-intron structure analysis by using Gene Structure Display Server 2.0 online analysis tool (http://gsds.cbi.pku.edu.cn/). The analysis was conducted automatically by providing both CDS and genomic sequence of gene. Gene structure analysis is shown in Figure 1. The incomplete gene sequences are ignored. The predicted common carp GST amino acid sequences together with zebrafish *GST* genes were aligned by MAFFT version 7 (http://mafft.cbrc.jp/alignment/server/) using default parameters and illustrated with GeneDoc. Details about sequences alignment is shown in Supplementary Figure S1 and percent identity matrix between carp and zebrafish is listed in Supplementary Table SI.

*Phylogenetic analysis*

For the sake of annotating the *GST* genes, phylogenetic analysis was conducted with amino acid sequences of GST genes from common carp and other seven vertebrates including four teleost. For nomenclatures of the common carp GSTs, whenever possible we followed those of zebrafish because zebrafish is the most closely related model species to common carp. Finally, a total of 136 protein sequences were aligned by Mega 6 using ClustalW method with default parameters. A maximum likelihood tree of carp and seven representative species GSTs (Fig. 2) was constructed by Mega 6, with LG model, and 1000 bootstrap replicas were utilized to access the strength of the suggested associations.

*Syntenic analysis*

Syntenic analysis of the evolution relationship about GST superfamily genes were performed among seven species by identifying the common genes both up- and

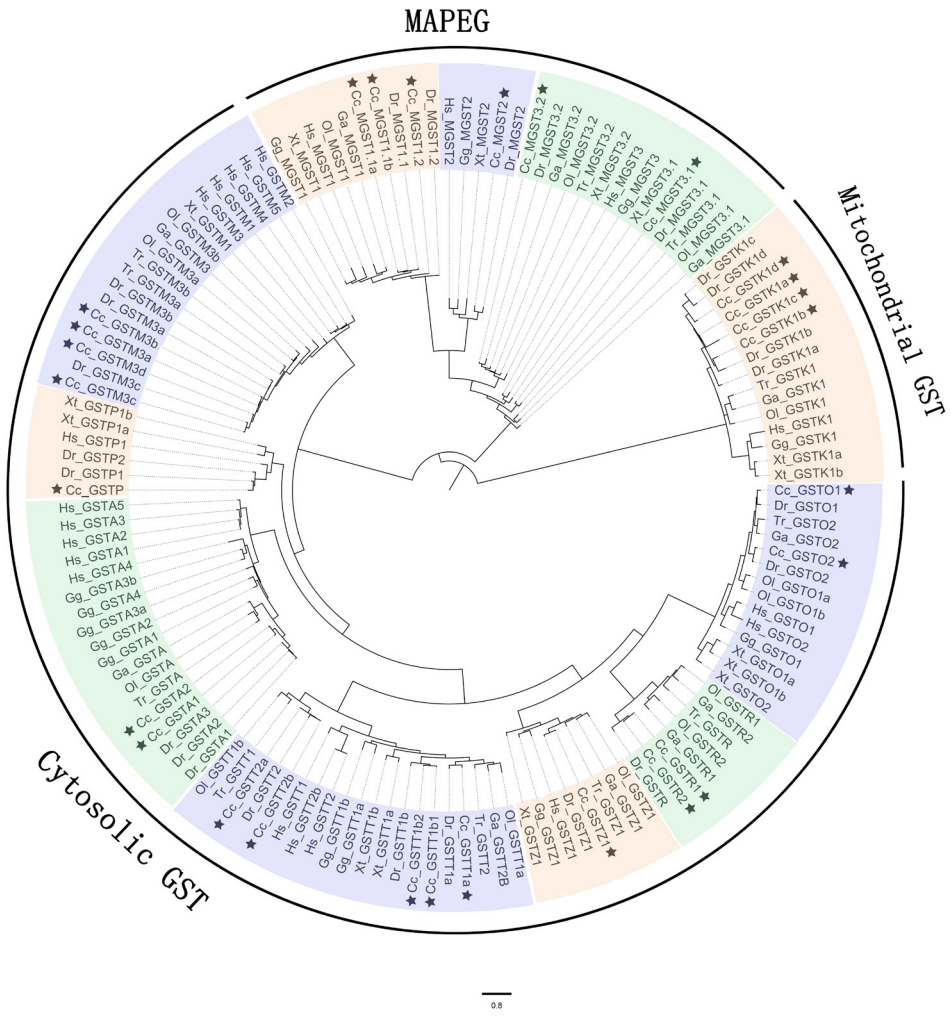


Fig. 2. Phylogenetic tree of *GST* gene family. Neighbor-joining-based phylogenetic tree of GST protein sequences. Including several kinds of typical vertebrates: human (Hs), chicken (Gg), frog (Xt), zebrafish (Dr), pufferfish (Tr), medaka (Ol), stickleback (Ga) and common carp (Cc). GST superfamily is labeled by three subfamilies: cytosolic GST, mitochondrial GST and MAPEG.

downstream of the focal genes in zebrafish and common carp. Annotation information of genes distribution along chromosomes of common carp is available on CarpBase database (http://www.carpbase.org/). The distribution information of genomic regions in other species was downloaded from Ensembl (http://www.ensembl.org/). Then we confirmed the conservative regions between zebrafish and common carp by comparing annotation information of genes with the help of Perl program. Syntenic maps were constructed mainly based on the information regarding the location of genes and draw manually.

*Tissue expression profiling of GST genes*

Total RNA from various adult common carp tissues (brain, heart, spleen, kidney, intestine, gill, liver, skin, blood, muscle) was extracted using Trizol reagent (Life Technologies, NY, USA), and the cDNA was synthesized by the RT-PCR using the SuperScript III Synthesis System (Life technologies, NY, USA). ß-actin gene was used as an internal positive control, with forward primer (5'-TGCAAAGCCGGATTCGCTGG-3') and reverse primer (5'-AGTTGGTGACAATACCGTGC-3'). The PCR thermal cycle comprised an initial denaturation step of 2 min at 94°C followed by 35 cycles of denaturation (30 sec at 94°C), annealing (30 sec at 62°C or 64°C), and extension (20 sec at 72°C), and a final elongation step of 2 min at 72°C. The PCR products were separated by gel electrophoresis (1.0% agarose gel at 140 V) in the presence of ethidium bromide and visualized under ultraviolet light.

**Table I.-** **GST gene names corresponding with their accession number in eight species.**

| *Cyprinus carpio* | | *Danio rerio* | | *Homo sapiens* | | *Gallus gallus* | |
|---|---|---|---|---|---|---|---|
| **Gene name** | **Accession No.** | **Gene name** | **Accession NO.** | **Gene name** | **Accession No.** | **Gene name** | **Accession No.** |
| Cc_MGST1.1a | LC071490 | Dr_MGST1.1 | ENSDARG00000022165 | Hs_MGST1 | ENSG00000008394 | Gg_MGST1 | ENSGALG00000013098 |
| Cc_MGST1.1b | LC071491 | | | | | | |
| Cc_MGST1.2 | LC071492 | Dr_MGST1.2 | ENSDARG00000032618 | | | | |
| Cc_MGST2 | LC071493 | Dr_MGST2 | ENSDARG00000071345 | Hs_MGST2 | ENSG00000085871 | Gg_MGST2 | ENSGALG00000009803 |
| Cc_MGST3.1 | LC071494 | Dr_MGST3.1 | ENSDARG00000102744 | Hs_MGST3 | ENSG00000143198 | Gg_MGST3 | ENSGALG00000003445 |
| Cc_MGST3.2 | LC071495 | Dr_MGST3.2 | ENSDARG00000033364 | | | | |
| Cc_GSTK1a | LC071486 | Dr_GSTK1a | ENSDARG00000056510 | Hs_GSTK1 | ENSG00000197448 | Gg_GSTK1 | ENSGALG00000014708 |
| Cc_GSTK1b | LC071487 | Dr_GSTK1b | ENSDARG00000019585 | | | | |
| Cc_GSTK1c | LC071488 | Dr_GSTK1c | ENSDARG00000093119 | | | | |
| Cc_GSTK1d | LC071489 | Dr_GSTK1d | ENSDARG00000092052 | | | | |
| Cc_GSTM3a | LC071496 | Dr_GSTM3a | ENSDARG00000042533 | Hs_GSTM1 | ENSG00000134184 | | |
| Cc_GSTM3b | LC071497 | Dr_GSTM3b | ENSDARG00000029473 | Hs_GSTM2 | ENSG00000213366 | | |
| Cc_GSTM3c | LC071499 | Dr_GSTM3c | ENSDARG00000088116 | Hs_GSTM3 | ENSG00000134202 | | |
| Cc_GSTM3d | LC071498 | | | Hs_GSTM4 | ENSG00000168765 | | |
| | | | | Hs_GSTM5 | ENSG00000134201 | | |
| Cc_GSTA1 | LC071500 | Dr_GSTA1 | ENSDARG00000039832 | Hs_GSTA1 | ENSG00000243955 | Gg_GSTA1 | ENSGALG00000016328 |
| Cc_GSTA2 | LC071501 | Dr_GSTA2 | ENSDARG00000039832 | Hs_GSTA2 | ENSG00000244067 | Gg_GSTA2 | ENSGALG00000016322 |
| | | Dr_GSTA3 | ENSDARG00000090228 | Hs_GSTA3 | ENSG00000174156 | Gg_GSTA3a | ENSGALG00000016324 |
| | | | | Hs_GSTA4 | ENSG00000170899 | Gg_GSTA3b | ENSGALG00000016325 |
| | | | | Hs_GSTA5 | ENSG00000182793 | Gg_GSTA4 | ENSGALG00000028551 |
| Cc_GSTT1a | LC071504 | Dr_GSTT1a | ENSDARG00000042428 | Hs_GSTT1 | ENSG00000184674 | Gg_GSTT1a | ENSGALG00000006344 |
| Cc_GSTT1b1 | LC071505 | Dr_GSTT1b | ENSDARG00000017388 | | | Gg_GSTT1b | ENSGALG00000005204 |
| Cc_GSTT1b2 | LC071506 | | | | | | |
| Cc_GSTT2a | LC071502 | Dr_GSTT2 | ENSDARG00000095464 | Hs_GSTT2 | ENSG00000099984 | | |
| Cc_GSTT2b | LC071503 | | | Hs_GSTT2b | ENSG00000133433 | | |
| Cc_GSTP | LC071507 | Dr_GSTP1 | ENSDARG00000104068 | Hs_GSTP1 | ENSG00000084207 | | |
| | | Dr_GSTP2 | ENSDARG00000103019 | | | | |
| Cc_GSTO1 | LC071509 | Dr_GSTO1 | ENSDARG00000022183 | Hs_GSTO1 | ENSG00000148834 | Gg_GSTO1 | ENSGALG00000008409 |
| Cc_GSTO2 | LC071508 | Dr_GSTO2 | ENSDARG00000033285 | Hs_GSTO2 | ENSG00000065621 | | |
| Cc_GSTZ1 | LC071510 | Dr_GSTZ1 | ENSDARG00000033285 | Hs_GSTZ1 | ENSG00000100577 | Gg_GSTZ1 | ENSGALG00000010432 |

*Continued on next page*

| *Cyprinus carpio* | | *Danio rerio* | | *Homo sapiens* | | *Gallus gallus* | |
|---|---|---|---|---|---|---|---|
| **Gene name** | **Accession No.** | **Gene name** | **Accession NO.** | **Gene name** | **Accession No.** | **Gene name** | **Accession NO.** |
| Cc_GSTR1 | LC071511 | Dr_GSTR | ENSDARG00000042620 | | | | |
| Cc_GSTR2 | LC071512 | | | | | | |
| Xt_MGST1 | ENSXETG00000007664 | | | Ol_MGST1 | ENSORLG00000010999 | Ga_MGST1 | ENSGACG00000019267 |
| Xt_MGST2 | ENSXETG00000030456 | | | | | | |
| Xt_MGST3.1 | ENSXETG00000004756 | Tr_MGST3.1 | ENSTRUG00000000749 | Ol_MGST3.1 | ENSORLG00000010182 | Ga_MGST3.1 | ENSGACG00000016252 |
| Xt_MGST3.2 | ENSXETG00000003689 | Tr_MGST3.2 | ENSTRUG00000001586 | Ol_MGST3.2 | ENSORLG00000017532 | Ga_MGST3.2 | ENSGACG00000007329 |
| Xt_GSTK1a | ENSXETG00000016238 | Tr_GSTK1 | ENSTRUG00000014778 | Ol_GSTK1 | ENSORLG00000012057 | Ga_GSTK1 | ENSGACG00000010408 |
| Xt_GSTK1b | ENSXETG00000032788 | | | | | | |
| Xt_GSTM1 | ENSXETG00000002017 | Tr_GSTM3a | ENSTRUG00000006620 | Ol_GSTM3a | ENSORLG00000005927 | Ga_GSTM3 | ENSGACG00000007655 |
| | | Tr_GSTM3b | ENSTRUG00000001751 | Ol_GSTM3b | ENSORLG00000005961 | | |
| | | Tr_GSTA | ENSTRUG00000014918 | Ol_GSTA | ENSORLG00000009674 | Ga_GSTA | ENSGACG00000006489 |
| Xt_GSTT1a | ENSXETG00000022358 | Tr_GSTT1 | ENSTRUG00000013937 | Ol_GSTT1a | ENSORLG00000018586 | Ga_GSTT2b | ENSGACG00000014744 |
| Xt_GSTT1b | ENSXETG00000024924 | | | Ol_GSTT1b | ENSORLG00000020134 | | |
| | | Tr_GSTT2 | ENSTRUG00000009188 | | | | |
| Xt_GSTP1a | ENSXETG00000021016 | | | | | | |
| Xt_GSTP1b | ENSXETG00000008727 | | | | | | |
| Xt_GSTO1a | ENSXETG00000026601 | | | Ol_GSTO1a | ENSORLG00000006192 | | |
| Xt_GSTO1b | ENSXETG00000016907 | | | Ol_GSTO1b | ENSORLG00000006201 | | |
| Xt_GSTO2 | ENSXETG00000026602 | Tr_GSTO2 | ENSTRUG00000009691 | | | Ga_GSTO2 | ENSGACG00000009287 |
| Xt_GSTZ1 | ENSXETG00000011079 | Tr_GSTZ1 | ENSTRUG00000018231 | Ol_GSTZ1 | ENSORLG00000016318 | Ga_GSTZ1 | ENSGACG00000007752 |
| | | Tr_GSTR | ENSTRUG00000014824 | Ol_GSTR1 | ENSORLG00000013712 | Ga_GSTR1 | ENSGACG00000007518 |
| | | | | Ol_GSTR2 | ENSORLG00000019461 | Ga_GSTR2 | ENSGACG00000007537 |

## RESULTS AND DISCUSSION

*Identification and nomenclature of GSTs*

Previous reports have described 9 classes of carp *GST* genes from three GST subfamilies: MAPEGs (mGST1, mGST2 and mGST3), Kappa from the mitochondrial GSTs, and the cytosolic GSTs (Mu, Alpha, Theta, Pi and the Rho which is special in fish and shares no molecular homologue in mammals) (Konishi *et al*., 2005; Fu and Xie, 2006). In this study, the Blast searching of zebrafish GSTs against all available genomic resources of common carp revealed a total of 27 *GST* genes, including 2 new classes of cytosolic GSTs that have not been identified previously in common carp: Omega and Zeta. All coding sequences of *GST* genes were deposited to DDBJ database with continuous accession number of LC071486 to LC071512 (Table I). We have also downloaded different amino acid sequences of GSTs in several organism genomes available in Ensemble genome browser and GenBank. However, during the process of data analysis, we found it using hard to classify these sequences using abbreviations of gene names due to different nomenclatures of these species. We thus renamed all these genes based on nomenclature of zebrafish GSTs according to the amino acid sequences alignment. Gene names and their corresponding accession numbers are shown in Table I.

*Sequences analysis and alignment of common carp GSTs*

Among all the *GST*s genes discovered in common carp, 4 of them (GSTK1c, GSTK1d, GSTM3c, GSTM3d) are fragments due to the absence of complete coding sequence. Detailed information of their genomic sequences, coding sequences and location are summarized in Table II. Most GST proteins include 200 to 250 amino acids except for MAPEGs (MGST1.1a, MGST1.1b, MGST1.2, MGST2, MGST3.1, MGST3.2) which are much shorter. The exon numbers of the MAPEGs are also less than other GSTs which implies their special role.

To better understand gene structure and their differences, we have aligned those GST sequences of common carp and their orthologs of zebrafish. Figure 1 and Supplementary Figure S1 show gene structure and the amino acid sequences which display significant diversities among different subfamilies and even different classes in the same subfamily. For example, GSTK exhibited a closer evolutionary relationship with members of MAPEGs than other classes of *GST*s, whereas sequence structures showed great differences. This is corresponding to the fact that different GSTs have different subunits and are involved in different reactions (Hayes *et al*., 2005). However, sequences under the same class, like genes in

*GST*Ts, *GST*Ms, *GST*As or *GST*Ks unusually retain similar gene structures (Fig. 1). Protein sequence alignment revealed that carp *GST*s share much more identity with zebrafish *GST*s under the same class, like GSTA, than other members of this superfamily (Supplementary Table SI). So this leads to the conclusion that sequences of the same class of GSTs are highly conserved.

*Phylogenetic analysis of GSTs*

Phylogenetic tree of GST proteins from the predicted genes in common carp and the other seven vertebrates including four teleost were constructed using Maximum Likelihood method performed by MEGA6 (Fig. 2). Based on the resultant tree, it is inferred that major functional diversification within the GST family predated the divergence of vertebrates, and most classes of the teleost GSTs are present in all the species involved in this tree. As shown in Figure 2, all GSTs fall into three main branches that are comprised of eleven sub-branches, with GSTK solely in a clade, MAPEGs in a clade, and GSTO, GSTZ, GSTR, GSTA, GSTP, GSTM and GSTT in a clade, respectively. This result is well matched with the classification relationship of the subfamilies. Classes of the cytosolic GST subfamily clustered together with a step-by-step evolutionary relationship and GSTM seemed to be the most primitive one. The phylogeny of the eleven classes are consistent with sequence similarity analysis between carp and zebrafish (Supplementary Table SI), which showed that genes derived from the same GST class from different species share more similarity than genes of different classes in the same species. The carp GSTs in the phylogenetic tree are usually clustered with their zebrafish orthologs and then to other three teleosts, which agree with their evolutionary relationships.

**Table II.- Summary of GST gene family in common carp genome.**

| Gene name | Nucleotide size (bp) | Predicted cDNA size (bp) | Predicted peptide size (amino acids) | CDS status | No. of exons | Location |
|---|---|---|---|---|---|---|
| MGST1.1a | 595 | 417 | 138 | complete | 3 | LG8 |
| MGST1.1b | 1087 | 462 | 153 | complete | 3 | LG8 |
| MGST1.2 | 670 | 468 | 155 | complete | 3 | LG5 |
| MGST2 | 1498 | 426 | 141 | complete | 5 | LG28 |
| MGST3.1 | 1081 | 465 | 154 | complete | 5 | scaffold28912 |
| MGST3.2 | 1971 | 423 | 140 | complete | 4 | LG25 |
| GSTK1a | 1739 | 690 | 229 | complete | 7 | scaffold5619 |
| GSTK1b | 1732 | 663 | 220 | complete | 7 | scaffold5619 |
| GSTK1c | - | 624 | 207 | partial | - | scaffold3734 |
| GSTK1d | - | 678 | 225 | partial | - | scaffold3734 |
| GSTM3a | 2386 | 660 | 219 | complete | 8 | LG9 |
| GSTM3b | 3050 | 660 | 219 | complete | 8 | LG9 |
| GSTM3c | - | 186 | 61 | partial | - | LG3 |
| GSTM3d | - | 564 | 187 | partial | - | LG15 |
| GSTA1 | 1531 | 672 | 223 | complete | 6 | LG11 |
| GSTA2 | 1435 | 672 | 223 | complete | 6 | LG11 |
| GSTT1a | 1290 | 729 | 242 | complete | 5 | LG35 |
| GSTT1b1 | 1792 | 729 | 242 | complete | 5 | LG42 |
| GSTT1b2 | - | 729 | 242 | complete | - | scaffold2140 |
| GSTT2a | 1279 | 684 | 227 | complete | 5 | LG48 |
| GSTT2b | 1279 | 684 | 227 | complete | 5 | LG46 |
| GSTP | 3584 | 627 | 208 | complete | 6 | LG10 |
| GSTO1 | - | 723 | 240 | complete | - | LG25 |
| GSTO2 | - | 723 | 240 | complete | - | LG12 |
| GSTZ1 | 2360 | 663 | 220 | complete | 9 | scaffold2542 |
| GSTR1 | 4138 | 681 | 226 | complete | 6 | LG30 |
| GSTR2 | 7383 | 681 | 226 | complete | 5 | scaffold3096 |

The GST family consists of three subfamilies: the cytosolic, mitochondrial, and microsomal proteins, which are shown in the phylogenetic tree (Fig. 2). The GSTK members, which cluster in the most ancient branch in the phylogenetic tree, are distinct from other GSTs in sequence similarity and protein structure and shows similarity to prokaryotic 2-hydroxychromene-2-carboxylate isomerases (Robinson *et al*., 2004). Based on Figure 2, we can deduce that the original GST gene differentiated into two distinct subsets, the mitochondrial GSTs and the common ancestor of microsomal GSTs and cytosolic GSTs. Subsequently, the common ancestor is subdivided into many more different classes. As we previously mentioned, the cytosolic GST subfamily contains seven classes in mammals. Teleost cytosolic GSTs also own the same number of classes. Sigma are absent in teleost genome, and are replaced by a new class, Rho.

*Gene duplications and losses of GSTS in common carp*

Bridges (1936) reported one of the earliest observations of doubling of a chromosomal band of the fruit fly *Drosophila melanogaster*, which exhibited extreme reduction in eye size. Since then, importance of gene duplication in supplying raw genetic material to biological evolution has been recognized and several studies on comparative analysis have been conducted. Ohno (1970) suggested that two rounds of whole-genome duplication (WGD) occurred in the early phase of the vertebrate evolution; whereas, Meyer and Schartl (1999) showed third round duplication in the ray-finned fish lineage. Furthermore, on some cyprinids such as common carp, an additional WGD (the 4R WGD) has been hypothesized to have occurred during the evolution (Wang *et al*., 2012; Zhang *et al*., 2013). Comprehensive estimation based on whole genome dataset suggested that the latest WGD (4R) event has occurred around 8.2 MYA (Xu *et al*., 2014b). As a result of genome duplication, common carp ought to own more gene copies than most other teleosts. However, in fact, additional gene copies derived from WGD event usually accumulate mutations because of relaxed selection, and many of them became pseudogenes due to detrimental substitutions. Only a few duplicates can survive from acquisition of new function or shared different function of the original gene with its sister duplicates (Postlethwait *et al*., 2004). Common carp genome resources provide us the good genome models to look into the gene fates after the latest round of WGDs. We used GST gene family as an instance for exploration (Supplementary Table SII).

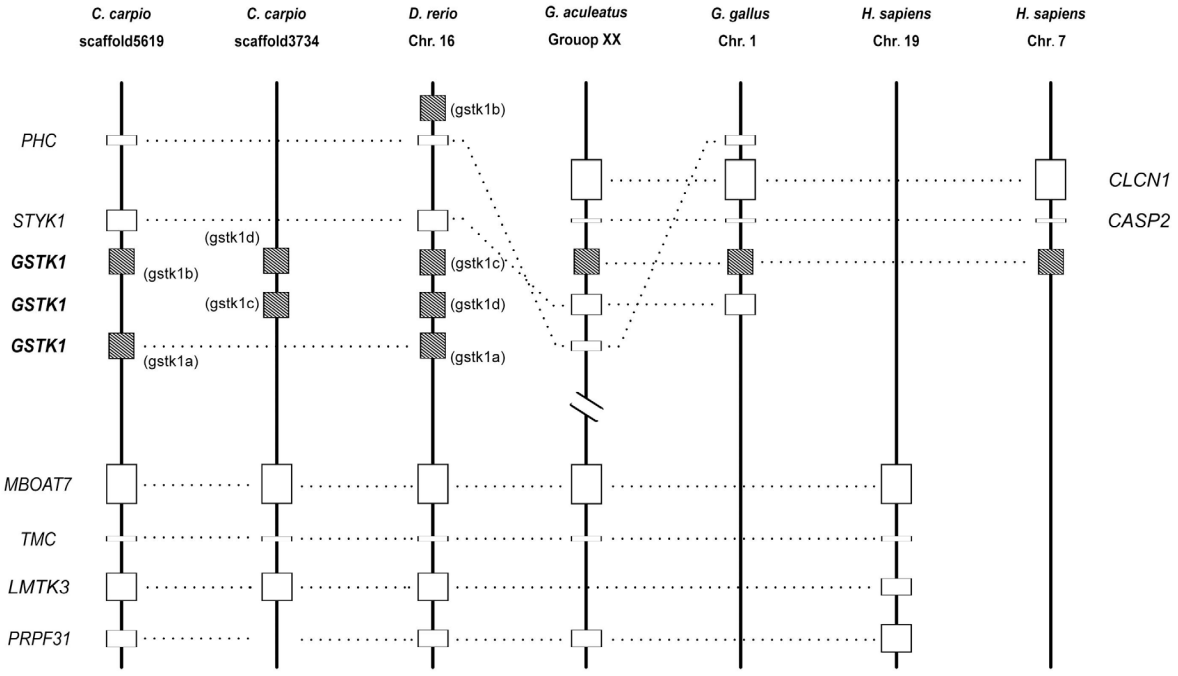


Fig. 3. Analysis of conserved synteny blocks harboring *GST*K genes in several vertebrates. Horizontal lines denote orthologous relationships. Abbreviations: PHC: Polyhomeotic homolog; STYK1: Serine/threonine/tyrosine kinase 1; GSTK1: Glutathione S-transferase kappa 1; MBOAT7: Membrane bound O-acyltransferase domain containing 7; TMC: Transmembrane channel-like; LMTK3: Lemur tyrosine kinase 3; PRPF31: Pre-mRNA processing factor 31 homolog; CLCN1: Chloride channel, voltage-sensitive 1; CASP2: Caspase 2, apoptosis-related cysteine peptidase.

As we can see in Figure 2 and Supplementary Table SII, there are at least 17 *GST* genes in common carp which have undergone gene duplication, including classes of MGST1, *GST*K, *GST*M, *GST*A and *GST*T. The gene duplication in common carp may lead to the speculation that these duplicates are highly likely derived from the 4R WGD. However, we also observed significant segmental gene duplications in several *GST* genes, suggesting the complexity of *GST* gene evolution in common carp genome. To better understand the complexity, we selected GSTK subfamily as the typical instance. We performed comparative genomic analysis to identify potential syntenic regions in common carp and related vertebrate genomes (Fig. 3). As shown in Figure 3, four GSTK genes are located into two distinct scaffolds of common carp genome, which suggested that GSTKa/GSTKb and GSTKc/GSTKd may have been derived from the latest round of WGD. GSTKa and GSTKb are located on the same genome region, suggesting the segmental duplication or tandem duplication origin. GSTKc and GSTKd have the similar inference of their segmental duplication origin. The phylogenetic topology demonstrated that CcGSTKa and DrGSTKa have higher similarity than that of CcGSTKa and CcGSTKb, which suggested that GSTKa and GSTKb may diverge earlier than the divergence time of zebrafish and common carp. However, CcGSTKc and CcGSTKd obviously diverged post zebrafish and common carp divergence. Surprisingly, we observed all zebrafish *GST*K genes are tandemly located on chromosome 16. Multiple rounds of gene losses and segmental duplications/relocations may be involved in zebrafish.

On the contrary of gene gains from WGD and segmental duplication, gene loss is the most typical fate post WGD events during evolution. Although the latest common carp specific WGD just occurred around 8.2 MYA, we already observed gene losses in *GST* gene superfamily in carp genome. Some GST classes retain only one copy, such as MGST2, GSTP and GSTZ, which suggest potential gene loss after WGD (Supplementary Table SII). To demonstrate gene loss, we constructed syntenic block across common carp and other five vertebrate genomes (Fig. 4). A single copy of MGST2 can be identified in higher vertebrates such as human, chicken and frog. In teleost, we identified either single copy of MGST2 gene (such as common carp, zebrafish and platyfish, etc) or absent (such as stickleback, medaka and pufferfish, etc), which suggested that *MGST*2 gene was lost in some teleost genomes completely, but still retained one copy in Cyprinids such as carp and zebrafish.

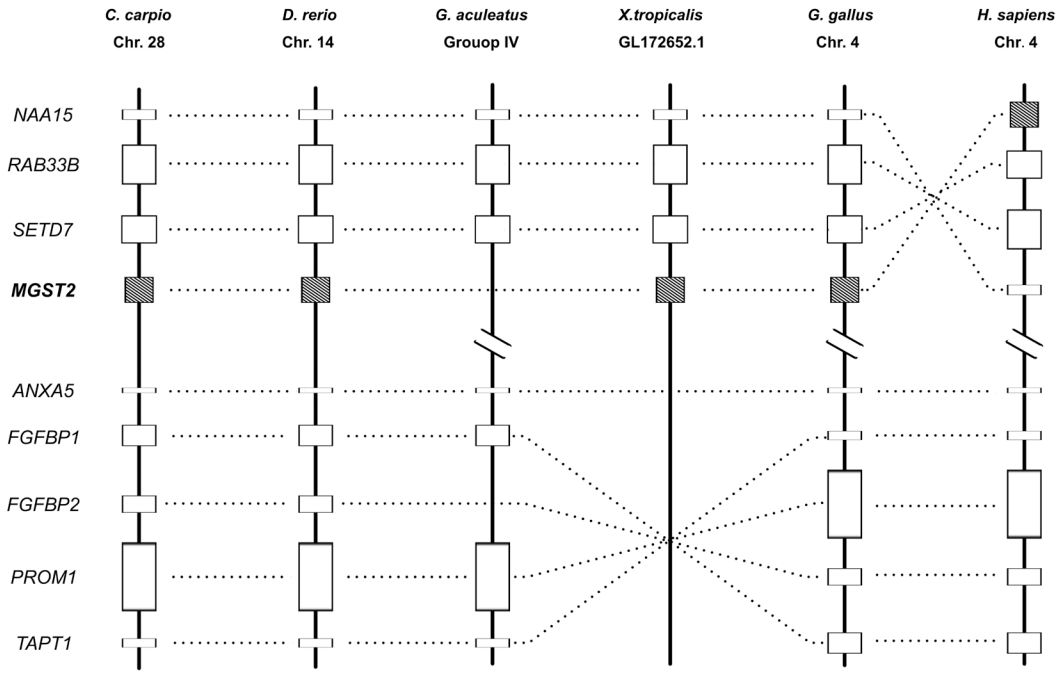


Fig. 4. Analysis of conserved synteny blocks harboring MGST2 genes in several vertebrates. Horizontal lines denote orthologous relationships. Abbreviations: NAA15: N(alpha)-acetyltransferase 15, NatA auxiliary subunit; RAB33B: Member RAS oncogene family; SETD7: SET domain containing (lysine methyltransferase) 7; MGST2: Microsomal glutathione S-transferase 2; ANXA5: Annexin A5; FGFBP1: Fibroblast growth factor binding protein 1; FGFBP2: Fibroblast growth factor binding protein 2; PROM1: Prominin 1; TAPT1: Transmembrane anterior posterior transformation 1.

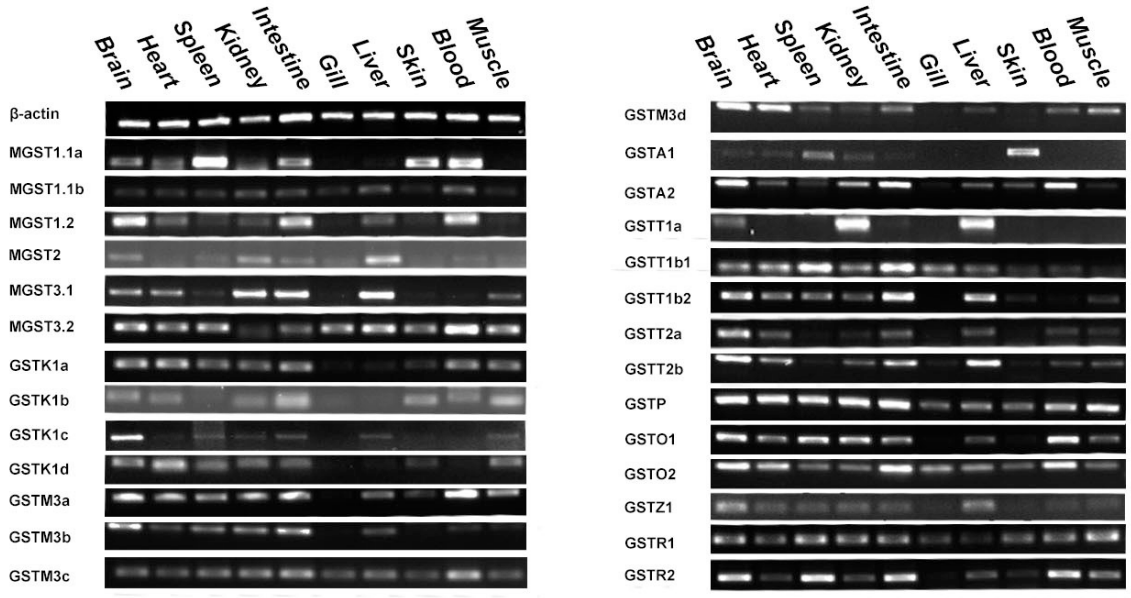


Fig. 5. RT-PCR based expression analysis of common carp glutathione S-transferase genes. The amplification of β-actin was used as an internal control.

The observation implies that *MGST*2 gene function would be redundant along with other MGST members in teleost. Therefore, MGST2 duplicates derived from the multiple rounds of WGD were lost quickly. The gene loss of MGST2 may not affect their survival, and even benefit their adaptation in aquatic environment. More surveys and investigations are required to confirm the inference.

*Tissue expression profiles of GST genes of common carp*

Functional inferences of genes in teleost fish, especially those that have undergone duplications or losses, would be very interesting because they are potentially underlying the adaptations to aquatic environments. Due to the important role of *GST* genes on cellular detoxification and the expansion in common carp, it was necessary to examine how many of these genes are expressed. It was also important to confirm the expression pattern of these genes for identification of functional differentiation post duplication. Thus, we conducted RT-PCR using gene-specific primers to examine the expression pattern of all members of GST superfamily in 10 tissues of common carp. The expression profiles are shown in Figure 5. Overall, *GST* genes are widely expressed in all tissues with relatively higher expression in brain, heart, spleen, kidney, intestine and liver. All classes of MAPEG were mainly expressed in brain, heart, spleen, kidney, intestine and liver. We observed significant expression differences among a number of duplicated GST genes. For instance, MGST1.1b was universally expressed in all tissues, while its duplicate copy, MGST1.1a, was absent in gill, liver and muscle. MGST3.2 was widely expressed in all tissues, while MGST3.1 was not expressed in gill, skin and blood. Similar expression differences were also identified in GSTM3a/GSTM3b/GSTM3c, of which GSTM3a was not expressed in gill, and GSTM3b was not expressed in gill and skin. In the expression profiles of GSTT1b1/GSTT1b2, GSTT1b2 was absent in gill and blood. Overall, we observed similar expression profiles in all *GST* genes across their duplicate copies, suggesting that they are still retain similar gene functions after duplications. However, significant differences on expression profiles of some specific pairs of duplicated *GST* genes implied that substantial subfunctionalization did occur after the gene duplications and potentially evolved new functions.

## CONCLUSION

A total of 27 *GST* genes were identified from common carp genome. Sequences analysis and alignment exhibited that genes under the same class are highly conserved while members of different classes shows great difference. Phylogenetic analysis, which provided the basis for accurate nomenclature and annotation of these genes, indicated that GSTK ought to be the most ancient branch and MAPES may have a common ancestor with cytosolic GSTs. Besides that, phylogenetic results based on sequence alignment show that same GST class members from different species share more identity than different classes

of the same species. Comparative genomics and syntenic analysis provided new evidences for better understanding of gene fates after WGD of common carp. Some of the glutathione S-transferases genes were ubiquitously expressed in common carp and their high expression in tissues like kidney, intestine and liver, indicated the critical roles of this gene family in detoxication. However, detailed functions of each gene need further studies. The complete set of GST genes provided the essential genomic resources for future biochemical, toxicological and physiological studies in common carp.

## ACKNOWLEDGMENTS

We acknowledge grant support from the National Natural Science Foundation of China (No. 31422057 and No. 31502151), National High-Technology Research and Development Program of China (863 program; 2011AA100401), Fundamental Research Funds for the Central Universities, Xiamen University (20720160110), Funding Program for Outstanding Dissertations of Shanghai (A1-0209-14-0902-6).

*Supplementary material*

There is supplementary material associated with this article. Access the material online at: http://dx.doi.org/10.17582/journal.pjz/2017.49.4.1437.1448

*Statement of conflict of interest*

The authors declare that they have no conflict of interest.

## REFERENCES

Al-Arifa, N. and Jahan, N., 2016. Association of glutathione-S- transferase P1 (GSTP1) and group-specific component (GC) polymorphism with the risk of asthma in Pakistani Population. *Pakistan J. Zool.*, **48**: 937-942.

Blanchette, B., Feng, X. and Singh, B.R., 2007. Marine glutathione S-transferases. Mar. *Biotechnol.* (NY), **9**: 513-542. https://doi.org/10.1007/s10126-007-9034-0

Board, P., Chelvanayagam, G., Jermiin, L., Tetlow, N., Tzeng, H.F., Anders, M. and Blackburn, A., 2001. Identification of novel glutathione transferases and polymorphic variants by expressed sequence tag database analysis. *Drug Metab. Dispos.*, **29**: 544-547.

Bridges, C.B., 1936. The Bar "gene" a duplication. *Science (New York, NY),* **83**: 210-211. https://doi.org/10.1126/science.83.2148.210

Buetler, T.M. and Eaton, D.L., 1992. Glutathione S-transferases: Amino acid sequence comparison, classification and phylogenetic relationship. *J. environ Sci. Hth.* Part C: *Environ. Carcinog. Ecotoxicol. Rev*. **10**: 181-203.

Christoffels, A., Bartfai, R., Srinivasan, H., Komen, H. and Orban, L., 2006. Comparative genomics in cyprinids: common carp ESTs help the annotation of the zebrafish genome. *BMC Bioinform.*, **7**: S2. https://doi.org/10.1186/1471-2105-7-S5-S2

Fan, C., Zhang, S., Liu, Z., Li, L., Luan, J. and Saren, G., 2007. Identification and expression of a novel class of glutathione-S-transferase from amphioxus *Branchiostoma belcheri* with implications to the origin of vertebrate liver. *Int. J. Biochem. Cell Biol.*, **39**: 450-461. https://doi.org/10.1016/j.biocel.2006.09.013

Fu, J. and Xie, P., 2006. The acute effects of microcystin LR on the transcription of nine glutathione S-transferase genes in common carp *Cyprinus carpio* L. *Aquat. Toxicol.*, **80**: 261-266. https://doi.org/10.1016/j.aquatox.2006.09.003

Hayes, J.D., Flanagan, J.U. and Jowsey, I.R., 2005. Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.*, **45**: 51-88. https://doi.org/10.1146/annurev.pharmtox.45.120403.095857

Hemmingsen, A., Fryer, A. A., Hepple, M., Strange, R.C. and Spiteri, M. A., 2001. Simultaneous identification of GSTP1 Ile105→Val105 and Ala114→Val114 substitutions using an amplification refractory mutation system polymerase chain reaction assay: studies in patients with asthma. *Respir. Res.*, **2**: 255–260. https://doi.org/10.1186/rr64

Jakoby, W.B., 1978. The glutathione S-transferases: a group of multifunctional detoxification proteins. *Adv. Enzymol. Rel. Areas Mol. Biol.*, **46**: 383-414. https://doi.org/10.1002/9780470122914.ch6

Ji, P., Liu, G., Xu, J., Wang, X., Li, J., Zhao, Z., Zhang, X., Zhang, Y., Xu, P. and Sun, X., 2012. Characterization of common carp transcriptome: sequencing, de novo assembly, annotation and comparative genomics. *PLoS One*, **7**: e35152. https://doi.org/10.1371/journal.pone.0035152

Kim, J.H., Dahms, H.U., Rhee, J.S., Lee, Y.M., Lee, J., Han, K.N., Lee. J.S., 2010. Expression profiles of seven glutathione S-transferase (GST) genes in cadmium-exposed river pufferfish (Takifugu obscurus). *Comp. Biochem. Physiol. Part C: Toxicol. Pharmacol.,* **151**: 99-106. https://doi.org/10.1016/j.cbpc.2009.09.001

Konishi, T., Kato, K., Araki, T., Shiraki, K., Takagi, M. and Tamaru, Y., 2005. A new class of glutathione

S-transferase from the hepatopancreas of the red sea bream *Pagrus major*. *Biochem. J.*, **388**: 299-307. https://doi.org/10.1042/BJ20041578

Listowsky, I., Abramovitz, M., Homma, H. and Niitsu, Y., 1988. Intracellular binding and transport of hormones and xenobiotics by glutathione-S-transferases. *Drug Metab. Rev.*, **19**: 305-318. https://doi.org/10.3109/03602538808994138

Liu, S., Li, Q. and Liu, Z., 2013. Genome-wide identification, characterization and phylogenetic analysis of 50 catfish atp-binding cassette (ABC) transporter genes. *PLoS One,* **8:** 1-17. https://doi.org/10.1371/journal.pone.0063895

Ma, X.X., Jiang, Y.L., He, Y.X., Bao, R., Chen, Y. and Zhou, C.Z., 2009. Structures of yeast glutathione-S-transferase Gtt2 reveal a new catalytic type of GST family. *EMBO Rep.*, **10**: 1320-1326. https://doi.org/10.1038/embor.2009.216

Mannervik, B., Danielson, U.H. and Ketterer, B., 1988. Glutathione transferases—structure and catalytic activity *Crit. Rev. Biochem. Mol. Biol.*, **23**: 283-337. https://doi.org/10.3109/10409238809088226

Meyer, A. and Schartl, M., 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.*, **11**: 699-704. https://doi.org/10.1016/S0955-0674(99)00039-3

Ohno, S., 1970. *Evolution by gene duplication.* Springverlag. New York. https://doi.org/10.1007/978-3-642-86659-3

Postlethwait, J., Amores, A., Cresko, W., Singer, A. and Yan, Y.L., 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.*, **20**: 481-490. https://doi.org/10.1016/j.tig.2004.08.001

Rezaei, M.K., Shobbar, Z.S., Shahbazi, M., Abedini, R. and Zare, S., 2013. Glutathione S-transferase (GST) family in barley: identification of members, enzyme activity, and gene expression pattern. *J. Pl. Physiol.*, **170**: 1277-1284. https://doi.org/10.1016/j.jplph.2013.04.005

Robinson, A., Huttley, G.A., Booth, H.S. and Board, P.G,. 2004. Modelling and bioinformatics studies of the human Kappa-class glutathione transferase predict a novel third glutathione transferase family with similarity to prokaryotic 2-hydroxychromene-2-carboxylate isomerases. *Biochem. J.*, **379**: 541-552. https://doi.org/10.1042/bj20031656

Tan, K, and Board, P., 1996. Purification and characterization of a recombinant human Theta-class glutathione transferase (GSTT2-2). *Biochem. J.*, **315**: 727-732. https://doi.org/10.1042/bj3150727

Wang, J.T., Li, J.T., Zhang, X.F. and Sun, X.W., 2012. Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (*Cyprinus carpio*). *BMC Genom.*, **13**: 96. https://doi.org/10.1186/1471-2164-13-96

Xu, J., Zhao, Z., Zhang, X., Zheng, X., Li, J., Jiang, Y., Kuang, Y., Zhang, Y., Feng, J. and Li. C., 2014a. Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). *BMC Genom.*, **15**: 307. https://doi.org/10.1186/1471-2164-15-307

Xu, P, Li, J., Li, Y., Cui, R., Wang, J., Wang, J., Zhang, Y., Zhao, Z. and Sun, X., 2011. Genomic insight into the common carp (*Cyprinus carpio*) genome by sequencing analysis of BAC-end sequences. *BMC Genom.*, **12**: 188. https://doi.org/10.1186/1471-2164-12-188

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L. and Wang. G., 2014b. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio. Nature Genet.,* **46**: 1212-1219. https://doi.org/10.1038/ng.3098

Zhang, X., Zhang, Y., Zheng, X., Kuang, Y., Zhao, X., Zhao, L., Li, C., Jiang, L., Cao, D. and Lu. C., 2013. A consensus linkage map provides insights on genome character and evolution in common carp (*Cyprinus carpio* L.). *Mar. Biotechnol.*, **15**: 275-312. https://doi.org/10.1007/s10126-012-9485-9

Zhao, L., Zhang, Y., Ji, P., Zhang, X., Zhao, Z., Hou, G., Huo, L., Liu, G., Li, C., Xu, P. and Sun, X., 2013. A dense genetic linkage map for common carp and its integration with a BAC-based physical map. *PLoS One,* **8**: e63928. https://doi.org/10.1371/journal.pone.0063928