Contraction of the ROS scavenging enzyme glutathione *S*-transferase gene family in cetaceans

Ran Tian[*,†], Inge Seim[†,‡], Wenhua Ren[*], Shixia Xu[*], Guang Yang[*]

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

[†] Integrative Biology Laboratory, College of Life Sciences, Nanjing Normal University, Nanjing, Jiangsu, 210046, China

[‡] Comparative and Endocrine Biology Laboratory, Translational Research Institute-Institute of Health and Biomedical Innovation, School of Biomedical Sciences, Queensland University of Technology, Woolloongabba, Queensland, 4102, Australia

Short running title: Cetacean glutathione transferases

Key words: glutathione transferase, GST, gene family, cetaceans, oxidative stress adaptation

Corresponding authors:

Shixia Xu, Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing, Jiangsu, 210046, China. Email: xushixia78@163.com

Guang Yang, Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing, Jiangsu, 210046, China. Email: gyang@njnu.edu.cn

ABSTRACT

Cetaceans are a group of marine mammals whose ancestors were adaptated for life on land. Life in an aquatic environment poses many challenges for air-breathing mammals. Diving marine mammals have adapted to rapid reoxygenation and reactive oxygen species (ROS)-mediated reperfusion injury. Here, we considered the evolution of the glutathione transferase (GST) gene family which has important roles in the detoxification of endogenously-derived ROS and environmental pollutants. We characterized the cytosolic GST gene family in 21 mammalian species; cetaceans, sirenians, pinnipeds, and their terrestrial relatives. All seven GST classes were identified, showing that GSTs are ubiquitous in mammals. Some GST genes are the product of lineage-specific duplications and losses, in line with a birth-and-death evolutionary model. We detected sites with signatures of positive selection that possibly influence GST structure and function, suggesting that adaptive evolution of GST genes is important for defending mammals from various types of noxious environmental compounds. We also found evidence for loss of alpha and mu GST subclass genes in cetacean lineages. Notably, cetaceans have retained a homolog of at least one of the genes *GSTA1*, *GSTA4*, and *GSTM1*; GSTs that are present in both the cytosol and mitochondria. The observed variation in number and selection pressure on GST genes suggest that the gene family structure is dynamic within cetaceans. Taken together, our results indicate that the cytosolic GST family in cetaceans reflects unique evolutionary dynamics related to oxygen-poor aquatic environments.

INTRODUCTION

One of the classical examples of evolution is the return of terrestrial vertebrates to an aquatic environment – manifested by functional (secondary) adaptations in species whose ancestors departed an aquatic environment hundreds of millions of years earlier. The order Cetacea (whales, dolphins, and porpoises) is a model group of air-breathing marine mammals that transitioned to an aquatic lifestyle approximately 55 million years ago (Uhen 2007). The majority of cetaceans make shallow, short dives. This includes the common dolphin (*Delphinus delphis*) which usually dives down to 200 m and remains submerged for 5 minutes (Schreer and Kovacs 1997). Some cetaceans are capable of deep, long dives. For example, the sperm whale (*Physeter catodon*) can dive to a depth of 3,000 m and stay underwater for at least 138 minutes (Schreer and Kovacs 1997). Regardless of their diving abilities, all cetaceans face the tremendous challenge posed by a lack of oxygen during dives, so-called asphyxia (the integration of hypoxia, hypercapnia, and acidosis) (Elsner and Gooden 1983). In response, cetaceans have numerous adaptations of the respiratory system, such as improved oxygen delivery and storage in blood and muscle, as well as increased activity of glycolytic enzymes (Ramirez et al. 2007). Cardiovascular adaptations, including bradycardia (slowed heart rate) and peripheral vasoconstriction, also play a vital role in the conservation of oxygen in cetaceans. They augment or maintain blood flow to the central nervous system and heart but reduce flow in peripheral tissues such as kidney, liver, and skeletal muscle (ischemia). In terrestrial mammals reperfusion injury occurs when blood flow and oxygen delivery is restored to ischemic tissues, resulting in oxidative stress (Panneton 2013; Zenteno-Savín et al. 2011; Cantú-Medellín et al. 2011). Oxidative stress reflects an imbalance between the generation of reactive oxygen species (ROS) and the cell's ability to detoxify reactive

intermediates and repair damage (Birben et al. 2012). Oxidative stress can damage cells by lipid peroxidation and alteration of protein and nucleic acid structures (Apel and Hirt 2004).

The current evidence suggests that diving marine mammals and hibernating terrestrial mammals (e.g. ground-squirrels and some bats) have adapted to rapid reoxygenation and ROS-mediated reperfusion injury (Hermes-Lima et al. 2015). Oxidative damage is limited in cetaceans due to an intrinsic protection against ROS by scavenging enzymes and nonenzymatic antioxidants. Ceteceans have increased blood levels of the reduced form of glutathione (GSH), one of the most important nonenzymatic ROS scavengers (Wilhelm Filho et al. 2002; García-Castañeda et al. 2017). Similarily, the blood levels of vitamin E (α-tocopherol), which acts as a nonenzymatic antioxidant by protecting against peroxidation (Niwa 1999), is elevated in bottlenose dolphins (*Tursiops truncatus*) compared to its terrestrial sister taxa (Kasamatsu et al. 2009). ROS scavenging enzymes include glutathione peroxidase (GPX) which consumes hydrogen peroxide, glutathione reductase (GRS) which recycles glutathione from glutathione disulfide, superoxide dismutase (SOD) which scavenges superoxide radicals, and glutathione-*S*-transferase (GST) which catalyzes the conjugation of glutathione (Birben et al. 2012; Wilhelm Filho et al. 2002; Dröge 2002). Studies have revealed that antioxidant enzyme activies is higher in ceteceans compared to terrestrial mammals (Birben et al. 2012; Wilhelm Filho et al. 2002). Taken together, it is now recognized that cetaceans reduce reperfusion injury in various ways, however the genetic changes associated with these adaptations remain elusive.

Gene family innovation may enable species to adapt to novel or stressful environments (Kondrashov 2012). Here, we consider the ROS scavenging enzyme

gene family glutathione transferase (GST; EC 2.5.1.18). A handful of studies suggest that GST gene gain/loss and positive selection facilitate the adaptation to changing environments (Ding et al. 2017; Low et al. 2007; Monticolo et al. 2017; Khan 2014; Liu et al. 2015). The GST family is abundant and widely distributed in vertebrates, plants, insects, and microbes (Board and Menon 2013). In addition to conjugating GSH with reactive electrophilic compounds, some GSTs can also deteoxify hydroperoxides (Sherratt and Hayes 2002). Mammalian GSTs have been divided into three structurally distinct superfamily classes with separate evolutionary origins (Board and Menon 2013): cytosolic, mitochondrial, and microsomal transferases. Cytosolic GSTs represent the largest class and consists of seven distinct subclasses: alpha (α; encoded by human chr 6), mu (μ; chr 1), theta (θ; chr 22), pi (π; chr 11), zeta (ζ; chr 14), sigma (σ; chr 4), and omega (ω; chr 10). The number of genes in each subclass varies across the phylogenetic tree. For example, human alpha (*GSTA1* to *GSTA5*) and mu (*GSTM1* to *GSTM5*) have five enzymes each, omega (*GSTO1* and *GSTO2*) two members each, and zeta (*GSTZ*) has only one member (Table S1). Members within each GST subclass shares greater than 40% amino acid sequence identity, while members of different subclasses share less than 25% identity (Wu and Dong 2012). Although diverse mammals have retained each cytosolic GSTs subclass, there is variation in the number of genes in each subclass. For example, there is one pi subclass gene (*GSTP*) in humans and two pi subclass genes in mice (Bammler et al. 1994). In this study, we performed a comparative genomics analysis of cytosolic GST genes in 21 mammals, including seven cetaceans, to improve our knowledge on the genetic and evolutionary dynamics of the GST superfamily in mammals.

MATERIALS AND METHODS

Sequence retrieval

We obtained known human GST genes by perusing research articles and recent reviews (Board and Menon 2013; Morel et al. 2002) and by downloading coding sequences (CDS) from GenBank (http://www.ncbi.nlm.nih.gov) (Benson et al. 2018). GenBank accession numbers are listed in Table S1. Employing human protein sequences as queries, BLASTn searches were performed using an in-house Python script (see Supplemental Material, file 2) on local databases constructed from downloaded genomic sequences from 20 species. These included seven cetaceans: bottlenose dolphin (*Tursiops truncatus*), killer whale (*Orcinus orca*), Yangtze river dolphin (*Lipotes vexillifer*), Yangtze finless porpoise (*Neophocaena asiaeorientalis asiaeorientalis*), minke whale (*Balaena acutorostrata*), bowhead whale (*Balaena mysticetus*), and sperm whale (*Physeter macrocephalus*); two pinnipeds: Weddell seal (*Leptonychotes weddellii*), Pacific walrus (*Odobenus rosmarus divergens*); one sirenian: Florida manatee (*Trichechus manatus latirostris*); and ten terrestrial mammals: cow (*Bos taurus*), Tibetan yak (*Bos mutus*), sheep (*Ovis aries*), Tibetan antelope (*Pantholops hodgsonii*), dog (*Canis lupus familiaris*), horse (*Equus caballus*), a bat (little brown bat; *Myotis lucifugus*), mouse (*Mus musculus*), naked mole rat (*Heterocephalus glaber*), and (African) elephant (*Loxodonta africana*). The completeness of the annotated gene set of the 20 specices was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO v3.0) with mammalian-specific single-copy orthologs (mammalia_odb9) (Simão et al. 2015). Genome sequencing and assembly information of each species are listed in Table S2.

To identify GST genes, we set the BLASTn parameter *E*-value cut-off to 10 and hits with the highest score, lowest *E*-value, and a length of $\geq$ 150bp were

retained. We next retrieved multiple non-redundant hits by extending 1,000 bp at both 5' and 3' ends to find exon boundaries following the canonical gt/ag exon/intron junction rule (Cheng et al. 1995). We compared the genomic locations of each coding sequence among all genes to filter out repeated sequences with the same location on the same scaffold. The online resource GENEWISE (http://www.ebi.ac.uk/Tools/psa/genewise) (Birney et al. 2004) was employed to identify open reading frame (ORF) for all obtained sequences. Finally, all the predicted GST sequences were verified by BLAST against its respective species genomes to acquire complete GST gene sets. Mammalian GSTs were named according to the human GST nomenclature (Nebert and Vasiliou 2004). All identified GST genes were categorized into three categories – based on amino acid composition, unique motifs, BLAST and alignment results: 1) intact gene, a complete CDS region with the canonical structure typical of GST families; 2) partial gene, putative functional protein, but missing a start codon and/or stop codon; 3) pseudogene, highly similar to functional orthologs but with (a) inactivating mutation(s) and/or stop codon(s). To achieve a high accuracy in identifying GST genes in mammals, we used Genomicus v93.01 (Nguyen et al. 2017) to identify genes flanking the GST gene clusters in human and searched the mammalian genomes using BLAST to identify orthologous genomic regions. This enabled the identification of the correct arrangement and orientation of the GST genes in each species.

Phylogenetic analyses

The phylogenetic relationships between putative GST members in each subclass were estimated by maximum likelihood and Bayesian methods, as implemented in RAxML v8.0.26 (Stamatakis 2014) and MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003).

Nucleotide sequences were aligned using MUSCLE v3.6 (Edgar 2004), implemented in SeaView v4.5.4 (Gouy et al. 2009), and manually corrected upon inspection. MrModeltest was used to estimate the best-fit model of nucleotide substitution (SYM+G) and amino acid substitution (JTT+G4) (Nylander 2009). For the RAxML analyses the ML phylogeny was estimated with 1,000 bootstrap replicates. For MrBayes analyses we performed two simultaneous independent runs for 50 million iterations of a Markov Chain, with six simultaneous chains and sampling every 1,000 generations. A consensus tree was obtained after discarding the first 25% trees as burn-in.

Gene family analyses

To identify expanding and contracting gene ortholog groups across the mammalian phylogeny, we estimated the gene numbers on internal branches using a random birth and death process model implemented in the software CAFÉ v3.0, a tool for the statistical analysis of the evolution of the size of gene families (De Bie et al. 2006). We defined gene gain and loss by comparing cluster size differences between ancestors and each terminal branch among the phylogenetic tree. An ultrametric tree, based on the concatenated orthogroups, was estimated with BEAST v1.10 using Markov chain Monte Carlo (MCMC) with fossil calibrations and a Yule tree prior (Suchard et al. 2018). The molecular timescales were obtained from TimeTree (http://www.timetree.org) (Kumar et al. 2017). The analyses ran for 10 million generations, with a sample frequency of 1,000 and a burn-in of 10%.

Adaptive evolution analyses

To evaluate the positive selection of all GST genes during mammalian evolution, codon substitution models implemented in the *codeml* program in PAML v4.4 (Yang 2007) were applied to GST gene alignments. Two pairs of site-specific modes were compared using the likelihood ratio test (LRT): M8 (beta & ω) vs. M8a (beta & ω = 1) (Swanson et al. 2003). M8 estimates the beta-distribution for ω and takes into account positively selected sites (ω > 1), with the neutral model M8a not 'allowing' a site with ω > 1. We next employed branch-site models (test 2) to explicitly assess the rate of evolution on a site along a specific lineage of a tree: branch-site model (Ma) vs. branch-site model with fixed ω1 = 1 (Ma0) (Zhang et al. 2005). The Ma model assumes that sites in the foreground branch (i.e. the branch of interest) are under positive selection. When the LRT was significant ($p < 0.05$) under the M8 and Ma tests, codon sites under positive selection were assessed using the Bayes Empirical Bayes (BEB) method (Yang et al. 2005). The species tree was used as the guide tree in all analyses (Figure S1). Multiple testing for positive selection on genes was corrected by performing a false discovery rate (FDR, Benjamini-Hochberg) test at a cutoff of 0.05 (Benjamini and Hochberg 1995). We only accepted positively selected sites with a posterior probability (PP) > 0.80. A series of models implemented in HyPhy (http://www.datamonkey.org) (Pond and Muse 2005) were also used to estimate ratios of nonsynonymous (d_N) and synonymous (d_S) based on a maximum likelihood (ML) framework (Pond and Frost 2005a). These models tested were Single Likelihood Ancestor Counting (SLAC), Fixed Effect Likelihood (FEL), and Random Effect Likelihood (REL) (Pond and Frost 2005b). As a criteria to identify candidates under selection, we used a Bayes Factor > 50 for REL and *P*-value of 0.10 for SLAC and FEL. The program TreeSAAP (Selection on Amino Acid Properties using

Phylogenetic trees) v3.2 (Steve et al. 2003) was used to evaluate amino acid residue replacement during evolution, taking into account physicochemical properties and the assumption of random replacement under a neutral model of evolution. TreeSAAP complements d_N/d_S analysis, which does not distinguish between different types of non-synonymous substitutions. InterProScan 5 was used to annotate positively selected sites found in functional protein domains (Philip et al. 2014). We also used the PyMOL (http://pymol.sourceforge.net/) to load and manipulate PDB files, highlighting the position of selected residues.

Comparison of mammalian niches

To investigate the potential links between the molecular evolution of GST genes and ecological adaptations, we assigned habitats (aquatic vs. terrestrial) according to data in the literature (Uhen 2007). We used Clade Model C (CmC) to identify the level of divergent selection among clade with different ecological niches and to test what partition of clades best fit the data (Bielawski and Yang 2004). CmC, which allows site variation among a priori defined foreground (cetacean, pinnipeds, sirenians) and background partitions (Figure S1), was compared with the M2a_rel null mode which does not allow variation among divisions. Taking into account the phylogenetic relationships between mammals, a phylogenetic ANOVA (Garland et al. 1993) using the function *phylANOVA* implemented in the R package 'phytools' (Revell 2012) was performed to test for differences in the number of GST genes between marine and terrestrial mammals.

RESULTS

Cytosolic glutathione transferase gene repertoires in mammals

To examine the evolution of cytosolic GST genes in mammals, we interrogated public genomic data of the 21 species, representing all major mammalian taxa. We identified a total of 448 GST genes (333 intact genes, 22 partial genes, and 93 pseudogenes) (Table 1). Amino acid sequence identity between GST paralogous was more than 50%, whereas the identity between all subclasses was less than 30% (Table S3). It has been reported that a low-quality genome can effect gene family analyses by introducing frame shifting errors in coding sequences (Young et al. 2010). In our study we employed genomes with good reported genome assemblies in an effort to minimize this potential bias. In agreement, BUSCO analysis on the protein gene sets showed that most, except for bowhead whale (74.6%) and Weddell seal (87.3%), included >95% complete sequences of mammalian universal single-copy orthologs (*n* = 4,104) (Table S2). This suggests that the genome assembly quality is similar between the marine and the terrestrial species.

The phylogenetic relationships of cytosolic GST genes

We next wished to classify the 448 cytosolic GST genes into their respective subclass. Phylogenetic relationships were reconstructed using maximum likelihood and Bayesian methods. Both topologies yielded similar branch patterns, indicating a reliable tree structure (Figure S2). Each GST subclass was clustered into a monophyletic group with high node bootstrap values (94-100% of bootstrapping) (Figure S2). However, subtrees of species within each subclass were not clearly resolved (< 50% node bootstrap support) (Figure S2) – especially for the alpha and mu subclasses where duplication events were common. According to the phylogenic

tree, the cytosolic GST gene family is highly conserved in mammals (Figure S2). We detected 105 alpha subclass (81 genes and 24 pseudogenes), 133 mu subclass (90 genes and 43 pseudogenes), 47 theta subclass (42 genes and 5 pseudogenes), 74 pi (54 genes and 20 pseudogenes), 21 zeta subclass (21 genes and 0 pseudogenes), 47 omega subclass (46 genes and 1 pseudogenes), and 21 sigma subclass (21 genes and 0 pseudogenes) genes in the 21 species examined.

Notably, ML and Bayesian phylogenies based on the 448 GST genes did not arrange the seven subclasses into well-supported clades (< 50% node bootstrap support). This could result from the 93 pseudogenes and 22 partial GST genes that are likely no longer under natural selection pressure. Therefore, only the 333 intact (complete CDS) sequences were used to infer the phylogenetic tree and determine the evolutionary relationship of subclasses. The tree generated using an amino acid substitution model (Figure S3) was consistent with the tree structure obtained using nucleotide sequences (Figure 1). According to the new phylogeny, the mu subclass clustered with the pi subclass but the bootstrap support value and posterior probability were low (BS = 18%; PP = 0.34, Figure 1). The alpha subclass grouped with sigma with high bootstrapping and posterior probability (BS = 91%; PP = 0.58, Figure 1). The mu, pi, alpha, and sigma subclasses were much more closely related to each other, with 98% bootstrap support and 0.50 posterior probability (Figure 1). The theta subclass was placed as sister to a clade containing mu, pi, alpha, and sigma genes with high support (BS = 85%; PP = 1.00, Figure 1). Moreover, these five subclasses formed a monophyletic sister clade to the zeta subclass. The phylogenetic tree recovered the omega subclass as the most diverged lineage within GSTs (Figure 1).

Further strengthening our GST gene predictions, all seven cytosolic GST subclasses showed conserved synteny and similar arrangement across mammalian

genomes (Figure 2). Duplicated genes within each subclass were arranged in a tandem cluster, with two or more gene copies in tandem in the subclasses alpha, mu, theta, pi, and omega. The zeta and sigma classes had a single copy per species. The subclasses were flanked by the same pair of genes in all species (Figure 2), with the exception of the dog where pi subclass genes were present on separate scaffolds and possibly reflecting a sequencing or genome assembly artifact.

Lineage-specific gene duplications and deletions

We next contrasted the number of GST gene copy number in mammalian lineages. Of the 21 species studied, mouse (25 intact GSTs), naked mole rat (20 intact GSTs), Tibetan yak (21 intact GSTs), Tibetan antelope (20 intact GSTs), sheep (22 intact GSTs), Pacific walrus (21 intact GSTs), and horse (21 intact GSTs) have the largest GST repertoires (Table 1, Figure 2). On the opposite spectrum, cetaceans appear to have the smallest GST repertoire – about ten functional GSTs per species (Table 1, Figure 2). In agreement, the fraction of GST pseudogenes is the highest in cetaceans (mean, 36%, Table 1), which is three times higher than terrestrial artiodactyls (mean, 11%, Table 1). Considering the alpha subclass, we found that the cetacean alpha subclass consists of two functional GSTs, but their relatives (i.e., artiodactyls) have at least four intact alpha GSTs. Sheep (6 intact GSTAs), horse (9 intact GSTAs), and bat (6 intact GSTAs) have a relatively large number of alpha GSTs. Similarly, only two functional mu GSTs were identified in cetaceans. In contrast, the number of functional GSTM genes in artiodactyls (6) is almost six times that of cetaceans. Notably, a large group of species, including all rodents and two afrotherians (West Indian manatee and African elephant), harbor six to ten functional mu GSTs (Table 1, Figure 2). Additionally, bat (4 intact GSTTs) has the largest gene number of theta

GSTs, while the lowest gene copy number was found in cetaceans (just one gene in six cetaceans; two in killer whale). Almost all mammalian species have two omega subclass genes, with the exception of artiodactyls (3 or 4 genes) (Table 1, Figure 2). The zeta and sigma subclass gene number appear to be more conserved in mammals, with one copy identified, respectively. Furthermore, the gene gain and loss of GSTs at each ancestral node was estimated by the software CAFÉ (De Bie et al. 2006). We found that the terrestrial groups have similar GST gene repertoires – carnivorans (mean 24.8), rodents (mean 25), artiodactyls (mean 22.29) (Welch's *t*-test, $p > 0.05$). In contrast, the number of intact GST genes in cetaceans (mean 16.92) was significantly lower than that of the terrestrial groups (Welch's *t*-test , $p < 0.001$) (Figure 2). To compare gene numbers and account for statistical non-independence of closely-related species, we performed phylogenetic ANOVA (phylANOVA). After accounting for phylogeny, habitat (aquatic vs. terrestrial) was a signficant predictor of the gene copy number of all cytosolic GSTs combined (phylANOVA; $F = 23.135$, $p = 0.009$) (Figure 3A) and alpha-class GSTs (GSTA) alone (phylANOVA; $F = 21.599$, $p = 0.007$), with a smaller number of genes apparent in cetaceans (Figure 3B). In agreement, when we compared each marine order (Cetacea, Pinnipedia, Sirenia) to terrestrial species separately, cytosolic gene copy numbers was only significantly different between cetacean and terrestrial species (Figure 3C). This included all cytosolic GST combined (phylANOVA; $F = 142.458$, $p = 0.001$), GSTA alone (phylANOVA; $F = 24.300$, $p = 0.026$), and GSTM alone (phylANOVA; $F = 27.557$, $p = 0.011$).

Evolutionary model of the cytosolic GST gene family in mammals

To investigate the possible role of natural selection on the evolutionary history of the cytosolic GST gene family, a series of site-specific and branch-specific evolutionary models were evaluated. Site-specific selection tests implemented in PAML (Yang 2007) were performed to assess the selective pressure acting on mammalian GSTs. Site-specific positive selection with posterior probabilities > 0.80 were detected for *GSTA1* (16 sites), *GSTM1* (5 sites), *GSTO1* (14 sites), *GSTO2* (3 sites), *GSTP1* (2 sites), *GSTP2* (6 sites), *GSTT2* (1 sites), and *GSTZ1* (3 sites) (Table S4); suggesting that GSTs have evolved under diversifying selection in mammals. Similarly, positively selected codons in eight genes were identified using SLAC (23 codons), FEL (28), and REL (27) models implemented in Datamonkey (Pond and Muse 2005). In total, 38 codons from eight genes (*GSTA1*: 10; *GSTM1*: 1; *GSTO1*: 9; *GSTO2*: 5; *GSTP1*: 2; *GSTP2*: 5; *GSTT2*: 3; and *GSTZ1*: 3) with evidence of positive selection were detected with at least two of the ML methods (Table 2). Among these codons, 31 sites were also detected by amino-acid level selection analyses using TreeSAAP (Steve et al. 2003). Intriguingly, site enrichment analysis revealed that the sites under positive selection (28/38, 73%) are located or close to a GSH binding site, a substrate binding pocket, a C-terminal domain interface, or a N-terminal domain interface (Table S5, Figure S4).

We also employed the PAML branch-site model (Yang 2007) to identify episodic adaptations that affect amino acids along specific lineages. Few internal branches and several terminal branches showed evidence of positive selection after FDR correction (Table 3, Figure S5). In cetaceans the lineage leading to *GSTO1* in bottlenose dolphin and *GSTP2* in sperm whale were under selection. In pinnipeds the branches leading to *GSTM1* in Pacific walrus and its ancestor, *GSTP2* in the ancestor,

as well as *GSTA1* in Weddell seal were under positive selection. *GSTA1* and *GSTM1* were under positive selection only in pinnipeds (Pacific walrus and Weddell seal). The lineage leading to human *GSTA4*, elephant and bat *GSTT1*, cow, horse and manatee *GSTT2*, sheep and antelope *GSTP2*, naked mole rat *GSTO1*, as well as cetartiodactyla (includes whales and dolphins, and even-toed ungulates) *HPGDS* were also under positive selection.

A more thorough investigation of the evolutionary history of mammalian GSTs was conducted by extending the analysis to comparing marine and terrestrial species. Clade model C allows for more than two clades to be defined as separate partitions, estimating ω separately for each partition. A model which assumes divergent selection along the cetacean, pinniped, sirenian, and marine mammal branches fitted the data better than the null model in the case of the GST genes *GSTA1*, *GSTA4*, *GSTM1*, *GSTM3*, and *GSTT1* ($p < 0.05$, Table S6), suggesting there is a divergent selection pressure between marine and terrestrial mammals and within marine mammal groups.

DISCUSSION

Molecular evolution of GST genes

Several previous surveys have documented the distribution of GST genes. Nebert and Vasiliou (2004) provided a comprehensive assessment of GST superfamily genes. Other studies examined the phylogeny (Pearson 2005) and evolution (da Fonseca et al. 2010) of the superfamily across the tree of life. In the present study, we expanded on the previous surveys by performing a comprehensive search for cytosolic GST genes in 21 mammals representative of all major mammalian taxa. Of these, 15 species had no previous information on their cytosolic GST gene repertoire (Figure

2). We provide evidence for positive selection acting on several GST genes in divergent taxonomic groups (*GSTA1*, *GSTM1*, *GSTO1*, *GSTO2*, *GSTP1*, *GSTP2*, *GSTT2*, and *GSTZ1*), indicative of pervasive adaptive evolution. Considering that cytosolic GST genes play critical roles in the detoxification and metabolic activation of xenobiotics (Board and Menon 2013), these changes are likely to reflect previous and ongoing adaptions to diverse environments by mammals. In addition, 31 positively selected sites with radical amino acid changes were identified by gene- and protein-level selection analyses. Interestingly, 73% of the total amino acids under positive selection were found to be located in or close to functional domains. These results indicate that positively selected amino acid changes might play an important role in modulating the specificity or potency of detoxification and antioxidant defenses during mammalian evolution. For example, residue 45 of *GSTA1* is known to be involved in GSH binding (Balogh et al. 2009), while *GSTA1* residues 212 and 215 are close to a residue (216) important for thiolester substrate hydrolysis (Hederos et al. 2004). Residue 107 (a histidine) of *GSTM1* is a second substrate-binding site and has five radical amino acid changes in mammals. Site-directed mutatagenesis of this residue to an aspargine caused a 50% reduction in catalytic activity (Patskovsky et al. 2006).

Mammalian cytosolic GSTs phylogeny and birth-and-death evolution

The phylogenetic relationships among 333 cytosolic GSTs were reconstructed by two independent methods which gave a similar estimation of mammalian phylogeny. The branch pattern indicated that the omega, theta, and zeta subclasses are ancient in mammals, while the alpha, mu, pi, and sigma subclasses evolved later (Figure 1). This is in agreement with previous studies based on structural and functional data

(Armstrong 1997; da Fonseca et al. 2010; Frova 2006). The omega subclass has a cysteine residue at the active site, while the theta and zeta subclasses employ catalytic serine hydroxyl to activate GSH. They are thus predicted to be the progenitors of GSTs (Frova 2006). According to our phylogenetic tree, the omega subclass harbors the most ancient genes (Figure 1). Omega GSTs have strong homology to glutaredoxins, the predicted ancestors of the N-terminal topology of GST (Frova 2006; Oakley 2005). Taken together, it is thus reasonable to assume that the omega subclass evolved earlier. A switch from serine to tyrosine in the alpha, mu, pi, and sigma subclasses is another evolutionary scenario of GSTs which would explain why these four subclasses cluster together in the tree with high bootstrapping (Figure 1). It would appear that the sigma subclass diverged before the mammalian alpha, mu, and pi group due to its presence in invertebrates and vertebrates (Frova 2006). In our study, however, sigma appeared as the sister group of the alpha subclass. This scenario is in line with a previous study (da Fonseca et al. 2010). The mammalian sigma subclass, known as prostaglandin synthases, has a hydrophilic interface with a lock-and-key motif – similar to the alpha, mu, and pi subclasses (Sheehan et al. 2001). Therefore, the observed clustering might be related to their specialized structure and function in mammals. On the other hand, our result also suggests that subclass mu proteins arose most recently. The theta subclass was placed as sister to a clade containing mu, pi, alpha, and sigma subclass with high support – supporting the prediction that alpha, mu, pi, and sigma subclasses arose from the duplication of theta subclass (Armstrong 1997).

The GST gene family has been previously described in various prokaryotes and eukaryotes (Pearson 2005). In this study, we examined cytosolic GST gene repertories in 21 species representing all major mammalian taxa. Our results show an

unequal copy number of GST genes across the mammalian phylogeny, as has previously been suggested (da Fonseca et al. 2010; Pearson 2005). For instance, the largest gene expansion was observed in the mouse (30 copies), whereas only 16 GSTs were identified in bowhead whale. This could be due to a lineage or species-specific duplication or deletions of this gene family in mammals. In support of this possibility, a diverse pseudogene proportion was found in mammalian GSTs; ranging from 41% in Yangtze river dolphin and Yangtze finless porpoise to 0% in the African elephant and Florida manatee (Table 1). Studies of gene duplicates have shown that new genes are usually created by gene duplication (Lynch and Conery 2000; Nei and Rooney 2005). Some duplicated genes are maintained in the genome for a long time, while others are nonfunctional or deleted from the genome. This phenomenon is termed the birth-and-death evolution model (Lynch and Conery 2000; Nei and Rooney 2005). We argue that this model is supported by the gene duplication events observed in our data of mammalian GSTs. It is also important to note that extensive duplication (12 paralogous) of pi GSTs was found in the dog, with half being pseudogenes. It has been reported that most duplicated genes tend to experience a brief period of relaxed selection early in their history, often resulting in non-functionalization or pseudogenes (Lynch and Conery 2000). Therefore, the high fraction of pi GSTs pseudogenes in dog further supports a birth-and-death model of GSTs evolution. Notably, seven pi subclass paralogous identified in mouse are apparently intact and functional, as is the case for the naked mole rat mu GSTs (10 copies). This could be the outcome of adaptations to environmental toxins (mouse) and very low oxygen levels in underground burrows (naked mole rat) via diversification of duplicated copies. Moreover, our results reveal extensive positive selection in *GSTP2* along five mammalian lineages (Figure S5), suggestive of episodic selection pressure – possibly

in response to changes in xenobiotic exposure. In contrast, there is no evidence of positive selection in *GSTP1* along specific lineages, indicating a functional conservation which is consistent with a critical role of *GSTP1* in ethacrynic acid metabolism in the liver (Henderson et al. 1998). These results reveal that divergent selective regimes occurred in paralogs within a cytosolic GST class.

Divergent selection of mammalian cytosolic GSTs

Numerous studies have reported that gene family evolution is closely tied with environments. This includes opsin genes in cichlid fish (Henderson et al. 1998), hemoglobins in vertebrates (Nery et al. 2013), keratin-associated proteins in vertebrates (Khan 2014; Sun et al. 2017), and olfactory receptor genes in mammals (Hughes et al. 2018; Niimura et al. 2018). An interesting result of our analysis is the correlation between distinct ecological milieus and cytosolic GST gene copy number (Table 1). For example, in Carnivora 29 genes were identified in dog, while 22 and 25 genes were found in the Weddell seal and Pacific walrus, respectively. In Cetartiodactyla, 21 to 25 genes were found in Artiodactyla, while 16 to 17 genes were found in Cetacea. We, therefore, hypothesize that the dynamic evolution of the cytosolic GST gene family reflects ecological adaptations in aquatic and terrestrial species. We also show that hypoxia-tolerant species with different ecological niches show evidence of divergent selection (Table S6), further suggesting that habitat plays a role in GST gene family evolution.

Marine mammals (cetaceans, pinnipeds, and sirenians) encompass phenotypic convergences that accommodate the challenges of aquatic life. They present a similar respiratory and cardiovascular solutions, such as improved oxygen storage, to low oxygen levels (Ramirez et al. 2007). Our data suggests that their shared adaptations

are not reflected in the evolution of the GST gene family. A total of four, three, and one positively selected genes were identified along cetaceans, pinnipeds, and sirenians, respectively (Figure S5). This suggests that there is a difference in the evolutionary history of the GST family in marine mammals. There are several non-mutually exclusive scenarios that could explain this pattern: 1) Antioxidant status is directly related to diving capacity. Marine mammals that perform shallow/short and deep/long divers might experience different oxidative stress challenges, suggesting distinct mechanisms to maintain redox balance (Cantú-Medellín et al. 2011); 2) Pinnipeds and sirenians possess enhanced enzymatic antioxidant capacities, whilst non-enzymatic antioxidant (e.g., levels of glutathione) seems to play an important role in cetaceans, indicating a different strategy for antioxidant defenses adaptation (Ninfali and Aluigi 1998; Wilhelm Filho et al. 2002); 3) We do not rule out the potential impact of different number of species and the length of branches used in this study. As an ever-increasing amount of high-quality genome assemblies are generated, future studies is likely to resolve these issues.

Oxidative stress adaptation in cetaceans

Cetaceans are faced by chronic oxidant stress stemming from chemical pollutants in aquatic environment and reoxygenation following hypoxia (diving) (Valavanidis et al. 2006; Li and Jackson 2002). Therefore, it might be expected that cetaceans have a large GST repertoire. However, we found that the number of cytosolic GST genes in cetaceans was significantly smaller than terrestrials (Figure 3C). The contraction of alpha GSTs is striking, with only two functional *GSTAs* (*GSTA1* and *GSTA4*) in cetaceans. Similarly, five mu GSTs were identified in cetaceans, but only one gene (*GSTM1*) appears intact compared with 5-10 GSTM genes in other mammals (Figure

2). We observed that ancestral branches of cetaceans also showed reduced GST repertoires, which suggested that contraction of cetacean GSTs could be related to aquatic adaptations after the divergence of cetacea from artiodactyla approximately 55 Mya (Thewissen et al. 2007). The presence of four *GSTM* pseudogenes provides further evidence for the contraction of the cetacean locus from a large GST family in the ancestral cetacean genome. Moreover, these gene losses in cetaceans probably a consequence of relaxed selection after adaptations of aquatic environment (Table S7). This raises the probability of an alternative gene family responsible for enhanced oxidative stress resistance or perhaps selection on particular genes responsible for detecting toxicants and activating oxidative defenses, rather than a large gene repertoire. Regarding the former explanation, it has been reported that the peroxiredoxin (*PRDX*) and glutathione peroxidase (*GPX*) gene families have expanded in whale lineages (Yim et al. 2014; Zhou et al. 2018). On the other hand, it is also possible that the retained GSTs in cetaceans have improved or essential antioxidative properties. For example, single nucleotide polymorphisms that cause amino acid substitutions in *GSTA1* alters its activity towards xenobiotics (Coles and Kadlubar 2005). *GSTA4* protects against oxidative stress by clearing toxic lipid peroxidation by-products (Hayes et al. 2005). *GSTM1* is critical for the detoxification of various oxidants, as well as carcinogens and toxins (Mcilwain and Townsend DMTew 2006). Supporting an essential role of these cytosolic GST genes (*GSTA1*, *GSTA4*, and *GSTM1*), human and rodent studies have reported that these GSTs are also present in mitochondria where they likely play a role in protecting against mitochondrial injury during oxidative stress (Gallagher et al. 2006; Raza et al. 2002) We speculate that the cetacean homologs, in particular *GSTA1* which was intact in all seven cetaceans, play an essential role in protecting against oxidative stress by

localizing to both the cytosol and mitochondria. That is, complete loss of these genes are not tolerated. We also show evidence of positive selection on the retained GSTs in cetaceans. These observations lead us to speculate that widely dispersed xenobiotics in aquatic ecosystems and high oxidative stress drive the adaptive evolution of retained functional and essential GST genes in cetaceans.

Previous studies demonstrated that gene loss in cetaceans could play an important role for natural phenotypic adaptations. For example, loss of genes with hair- and epidermis-related functions contribute to their unique skin morphology, a thicker epidermis and hairlessness (Sharma et al. 2018). Some investigators have reported that gene loss may carry detrimental fitness consequences in modern environments. An intriguing recent example includes the loss of paraoxonase 1 (*PON1*) in marine mammals which likely eliminates their main defense against neurotoxicity from man-made organophosphorus compounds (Meyer et al. 2018). Loss of GST genes in cetaceans may also have negative consequences. For instance, *Gsta3* knockout mice are not only sensitive to acute cytotoxic and genotoxic effect of aflatoxin B1 (Zoran et al. 2010), but also have increased oxidative stress marker levels (Crawford et al. 2017), suggesting that *GSTA3* loss in cetaceans might also weaken their oxidative damage defenses. Nevertheless, more research is necessary to validate whether a cetacean gene loss event is deleterious.

In conclusion, we here characterized the cytosolic glutathione-*S*-transferase (GST) gene family in 21 mammalian species. In particular, our study shows that the gene family has contracted in cetaceans despite the important role of GST genes in the protection against various stressors. Our findings add another piece to the puzzle of understanding how ceteaceans adapt to an oxygen-poor aquatic environments. An

ever-increasing amount of genomic research data and associated tools is likely to address this important research question.

ACKNOWLEDGEMENTS

This work was financially supported by the National Key Program of Research and Development, Ministry of Science and Technology of China (grant no. 2016YFC0503200 to G.Y. and S.X.), the Key Project of the National Natural Science Foundation of China (NSFC) (Grant no. 31630071 to G.Y.), the National Natural Science Foundation of China (NSFC) (grant nos. 31570379, 31772448 to S.X., grant no. 31872219 to W.R.), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), and the China Postdoctoral Science Foundation (grant no. 2018M642278 to R.T.).

REFERENCES

Apel, K., and H. Hirt, 2004 Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.* 55:373-399.

Armstrong, R.N., 1997 Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chemical research in toxicology* 10 (1):2-18.

Balogh, L.M., I. Le Trong, K.A. Kripps, K. Tars, R.E. Stenkamp *et al.*, 2009 Structural analysis of a glutathione transferase A1-1 mutant tailored for high catalytic efficiency with toxic alkenals. *Biochemistry* 48 (32):7698-7704.

Bammler, T.K., C.A. Smith, and C.R. Wolf, 1994 Isolation and characterization of two mouse Pi-class glutathione S-transferase genes. *Biochemical journal* 298 (2):385-390.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* 57 (1):289-300.

Benson, D.A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, J. Ostell *et al.*, 2018 GenBank. *Nucleic Acids Res* 46 (D1):D41-D47.

Bielawski, J.P., and Z. Yang, 2004 A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *Journal of molecular evolution* 59 (1):121-132.

Birben, E., U.M. Sahiner, C. Sackesen, S. Erzurum, and O. Kalayci, 2012 Oxidative stress and antioxidant defense. *World Allergy Organization Journal* 5 (1):9.

Birney, E., M. Clamp, and R. Durbin, 2004 GeneWise and genomewise. *Genome research* 14 (5):988-995.

Board, P.G., and D. Menon, 2013 Glutathione transferases, regulators of cellular metabolism and physiology. *Biochimica et Biophysica Acta (BBA)-General Subjects* 1830 (5):3267-3288.

Cantú-Medellín, N., B. Byrd, A. Hohn, J.P. Vázquez-Medina, and T. Zenteno-Savín, 2011 Differential antioxidant protection in tissues from marine mammals with distinct diving capacities. Shallow/short vs. deep/long divers. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 158 (4):438-443.

Cheng, J., C. Liu, W.J. Koopman, and J.D. Mountz, 1995 Characterization of human Fas gene. Exon/intron organization and promoter region. *The Journal of Immunology* 154 (3):1239-1245.

Coles, B.F., and F.F. Kadlubar, 2005 Human alpha class glutathione S-transferases: genetic polymorphism, expression, and susceptibility to disease. *Methods in enzymology* 401:9-42.

Crawford, D.R., Z. Ilic, I. Guest, G.L. Milne, J.D. Hayes *et al.*, 2017 Characterization of liver injury, oval cell proliferation and cholangiocarcinogenesis in glutathione S-transferase A3 knockout mice. *Carcinogenesis* 38 (7).

da Fonseca, R.R., W.E. Johnson, S.J. O'Brien, V. Vasconcelos, and A. Antunes, 2010 Molecular evolution and the role of oxidative stress in the expansion and functional diversification of cytosolic glutathione transferases. *BMC evolutionary biology* 10 (1):281.

De Bie, T., N. Cristianini, J.P. Demuth, and M.W. Hahn, 2006 CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22 (10):1269-1271.

Ding, N., A. Wang, X. Zhang, Y. Wu, R. Wang *et al.*, 2017 Identification and analysis of glutathione S-transferase gene family in sweet potato reveal

divergent GST-mediated networks in aboveground and underground tissues in response to abiotic stresses. *BMC plant biology* 17 (1):225.

Dröge, W., 2002 Free radicals in the physiological control of cell function. *Physiological Reviews*.

Edgar, R.C., 2004 MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* 5 (1):113.

Elsner, R., and B. Gooden, 1983 *Diving and asphyxia: a comparative study of animals and man*: Cambridge University Press.

Frova, C., 2006 Glutathione transferases in the genomics era: new insights and perspectives. *Biomolecular engineering* 23 (4):149-169.

Gallagher, E.P., J.L. Gardner, and D.S. Barber, 2006 Several glutathione -transferase isozymes that protect against oxidative injury are expressed in human liver mitochondria. *Biochemical Pharmacology* 71 (11):1619-1628.

García-Castañeda, O., R. Gaxiola-Robles, S. Kanatous, and T. Zenteno-Savín, 2017 Circulating glutathione concentrations in marine, semiaquatic, and terrestrial mammals. *Marine Mammal Science* 33.

Garland, T., A.W. Dickerman, C.M. Janis, and J.A. Jones, 1993 Phylogenetic Analysis of Covariance by Computer Simulation. *Systematic Biology* 42 (3):265-292.

Gouy, M., S. Guindon, and O. Gascuel, 2009 SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution* 27 (2):221-224.

Hayes, J.D., J.U. Flanagan, and I.R. Jowsey, 2005 Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.* 45:51-88.

Hederos, S., K.S. Broo, E. Jakobsson, G.J. Kleywegt, B. Mannervik *et al.*, 2004 Incorporation of a single His residue by rational design enables thiol-ester hydrolysis by human glutathione transferase A1-1. *Proceedings of the National Academy of Sciences* 101 (36):13163-13167.

Henderson, C.J., A.G. Smith, J. Ure, K. Brown, E.J. Bacon *et al.*, 1998 Increased skin tumorigenesis in mice lacking pi class glutathione S-transferases. *Proceedings of the National Academy of Sciences* 95 (9):5275-5280.

Hermes-Lima, M., D.C. Moreira, G.A. Rivera-Ingraham, M. Giraud-Billoud, T.C. Genaro-Mattos *et al.*, 2015 Preparation for oxidative stress under hypoxia and metabolic depression: Revisiting the proposal two decades later. *Free Radical Biology & Medicine* 89:1122-1143.

Hughes, G.M., E.S. Boston, J.A. Finarelli, W.J. Murphy, D.G. Higgins *et al.*, 2018 The birth and death of olfactory receptor gene families in mammalian niche adaptation. *Molecular biology and evolution* 35 (6):1390-1406.

Kasamatsu, M., R. Kawauchi, M. Tsunokawa, K. Ueda, E. Uchida *et al.*, 2009 Comparison of serum lipid compositions, lipid peroxide, alpha-tocopherol and lipoproteins in captive marine mammals (bottlenose dolphins, spotted seals and West Indian manatees) and terrestrial mammals. *Research in Veterinary Science* 86 (2):216-222.

Khan, I., 2014 Using genomic and proteomic information to characterize the evolution of genes involved in development and adaptation in vertebrates under differential conditions of selective pressure.

Kondrashov, F.A., 2012 Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings Biological Sciences* 279 (1749):5048-5057.

Kumar, S., G. Stecher, M. Suleski, and S.B. Hedges, 2017 TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology & Evolution* 34 (7):1812.

Li, C., and R.M. Jackson, 2002 Reactive species mechanisms of cellular hypoxia-reoxygenation injury. *American Journal of Physiology-Cell Physiology* 282 (2):C227-C241.

Liu, H.-J., Z.-X. Tang, X.-M. Han, Z.-L. Yang, F.-M. Zhang *et al.*, 2015 Divergence in enzymatic activities in the soybean GST supergene family provides new insight into the evolutionary dynamics of whole-genome duplicates. *Molecular biology and evolution* 32 (11):2844-2859.

Low, W.Y., H.L. Ng, C.J. Morton, M.W. Parker, P. Batterham *et al.*, 2007 Molecular evolution of glutathione S-transferases in the genus Drosophila. *Genetics* 177 (3):1363-1375.

Lynch, M., and J.S. Conery, 2000 The evolutionary fate and consequences of duplicate genes. *Science* 290 (5494):1151-1155.

Mcilwain, C.C., and K.D. Townsend DMTew, 2006 Glutathione S-transferase polymorphisms: cancer incidence and therapy. *Oncogene* 25 (11):1639-1648.

Meyer, W.K., J. Jamison, R. Richter, S.E. Woods, R. Partha *et al.*, 2018 Ancient convergent losses of Paraoxonase 1 yield potential risks for modern marine mammals.

Monticolo, F., C. Colantuono, and M.L. Chiusano, 2017 Shaping the evolutionary tree of green plants: evidence from the GST family. *Scientific reports* 7 (1):14363.

Morel, F., C. Rauch, B. Coles, E. Le Ferrec, and A. Guillouzo, 2002 The human glutathione transferase alpha locus: genomic organization of the gene cluster and functional characterization of the genetic polymorphism in the hGSTA1 promoter. *Pharmacogenetics and Genomics* 12 (4):277-286.

Nebert, D.W., and V. Vasiliou, 2004 Analysis of the glutathione S-transferase (GST) gene family. *Human genomics* 1 (6):460.

Nei, M., and A.P. Rooney, 2005 Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 39:121-152.

Nery, M.F., J.I. Arroyo, and J.C. Opazo, 2013 Genomic organization and differential signature of positive selection in the alpha and beta globin gene clusters in two cetacean species. *Genome biology and evolution* 5 (12):2359-2367.

Nguyen, N.T.T., P. Vincens, H. Roest Crollius, and A. Louis, 2017 Genomicus 2018: karyotype evolutionary trees and on-the-fly synteny computing. *Nucleic acids research* 46 (D1):D816-D822.

Niimura, Y., A. Matsui, and K. Touhara, 2018 Acceleration of Olfactory Receptor Gene Loss in Primate Evolution: Possible Link to Anatomical Change in Sensory Systems and Dietary Transition. *Molecular biology and evolution* 35 (6):1437-1450.

Ninfali, P., and G. Aluigi, 1998 Variability of oxygen radical absorbance capacity (ORAC) in different animal species. *Free radical research* 29 (5):399-408.

Niwa, Y., . 1999 [Oxidative injury and its defense system in vivo]. *Rinsho Byori the Japanese Journal of Clinical Pathology* 47 (3):189.

Nylander, J., 2009 MrModeltest v2. Program distributed by the author. 2004. *Evolutionary Biology Centre, Uppsala University Google Scholar*.

Oakley, A.J., 2005 Glutathione transferases: new functions. *Current opinion in structural biology* 15 (6):716-723.

Panneton, W.M., 2013 The Mammalian Diving Response: An Enigmatic Reflex to Preserve Life? *Physiology* 28 (5):284.

Patskovsky, Y., L. Patskovska, S.C. Almo, and I. Listowsky, 2006 Transition state model and mechanism of nucleophilic aromatic substitution reactions catalyzed by human glutathione S-transferase M1a-1a. *Biochemistry* 45 (12):3852-3862.

Pearson, W.R., 2005 Phylogenies of glutathione transferase families. *Methods in enzymology* 401:186-204.

Philip, J., B. David, C. Hsin-Yu, F. Matthew, L. Weizhong *et al.*, 2014 InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30 (9):1236-1240.

Pond, S.L.K., and S.D. Frost, 2005a Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21 (10):2531-2533.

Pond, S.L.K., and S.D. Frost, 2005b Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Molecular biology and evolution* 22 (5):1208-1222.

Pond, S.L.K., and S.V. Muse, 2005 HyPhy: hypothesis testing using phylogenies, pp. 125-181 in *Statistical methods in molecular evolution*. Springer.

Ramirez, J.-M., L.P. Folkow, and A.S. Blix, 2007 Hypoxia tolerance in mammals and birds: from the wilderness to the clinic. *Annu. Rev. Physiol.* 69:113-143.

Raza, H., M.-A. Robin, J.-K. Fang, and G.A. Narayan, 2002 Multiple isoforms of mitochondrial glutathione S-transferases and their differential induction under oxidative stress. *Biochemical journal* 366 (Pt 1):45.

Revell, L.J., 2012 phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3 (2):217-223.

Ronquist, F., and J.P. Huelsenbeck, 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19 (12):1572-1574.

Schreer, J.F., and K.M. Kovacs, 1997 Allometry of diving capacity in air-breathing vertebrates. *Canadian Journal of Zoology* 75 (3):339-358.

Sharma, V., N. Hecker, J.G. Roscito, L. Foerster, B.E. Langer *et al.*, 2018 A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nature Communications* 9 (1):1215.

Sheehan, D., G. Meade, and V.M. Foley, 2001 Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical journal* 360 (1):1-16.

Sherratt, P.J., and J.D. Hayes, 2002 Glutathione S-transferases, pp. 319-352 in *Enzyme systems that metabolise drugs and other xenobiotics*, edited by C. Ioannides. New York, NY: John Wiley & Sons.

Simão, F.A., R.M. Waterhouse, I. Panagiotis, E.V. Kriventseva, and E.M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31 (19):3210-3212.

Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30 (9):1312-1313.

Steve, W., J. Justin, M.J. Smith, K.A. Crandall, and D.A. Mcclellan, 2003 TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19 (5):671-672.

Suchard, M.A., P. Lemey, G. Baele, D.L. Ayres, A.J. Drummond *et al.*, 2018 Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* 4 (1):vey016.

Sun, X., Z. Zhang, Y. Sun, J. Li, S. Xu *et al.*, 2017 Comparative genomics analyses of alpha-keratins reveal insights into evolutionary adaptation of marine mammals. *Frontiers in zoology* 14 (1):41.

Swanson, W.J., R. Nielsen, and Q. Yang, 2003 Pervasive adaptive evolution in mammalian fertilization proteins. *Molecular biology and evolution* 20 (1):18-20.

Thewissen, J.G., L.N. Cooper, M.T. Clementz, S. Bajpai, and B. Tiwari, 2007 Whales originated from aquatic artiodactyls in the Eocene epoch of India. *Nature* 450 (7173):1190.

Uhen, M.D., 2007 Evolution of marine mammals: back to the sea after 300 million years. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology: Advances in Integrative Anatomy and Evolutionary Biology* 290 (6):514-522.

Valavanidis, A., T. Vlahogianni, M. Dassenakis, and M. Scoullos, 2006 Molecular biomarkers of oxidative stress in aquatic organisms in relation to toxic environmental pollutants. *Ecotoxicology and environmental safety* 64 (2):178-189.

Wilhelm Filho, D., F. Sell, L. Ribeiro, M. Ghislandi, F. Carrasquedo *et al.*, 2002 Comparison between the antioxidant status of terrestrial and diving mammals. *Comparative biochemistry and physiology. Part A, Molecular & integrative physiology* 133 (3):885-892.

Wu, B., and D. Dong, 2012 Human cytosolic glutathione transferases: structure, function, and drug discovery. *Trends in pharmacological sciences* 33 (12):656-668.

Yang, Z., 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* 24 (8):1586-1591.

Yang, Z., W.S. Wong, and R. Nielsen, 2005 Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular biology and evolution* 22 (4):1107-1118.

Yim, H.-S., Y.S. Cho, X. Guang, S.G. Kang, J.-Y. Jeong *et al.*, 2014 Minke whale genome and aquatic adaptation in cetaceans. *Nature genetics* 46 (1):88.

Young, J.M., H.F. Massa, L. Hsu, and B.J. Trask, 2010 Extreme variability among mammalian V1R gene families. *Genome research* 20 (1):10-18.

Zenteno-Savín, T., J.P. Vázquez-Medina, N. Cantú-Medellín, P.J. Ponganis, and R. Elsner, 2011 *Ischemia/Reperfusion in Diving Birds and Mammals: How they Avoid Oxidative Damage*.

Zhang, J., R. Nielsen, and Z. Yang, 2005 Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular biology and evolution* 22 (12):2472-2479.

Zhou, X., D. Sun, X. Guang, S. Ma, X. Fang *et al.*, 2018 Molecular footprints of aquatic adaptation including bone mass changes in cetaceans. *Genome biology and evolution* 10 (3):967-975.

Zoran, I., C. Dana, V. Dilip, P.A. Egner, and S. Stewart, 2010 Glutathione-S-transferase A3 knockout mice are sensitive to acute cytotoxic and genotoxic effects of aflatoxin B1. *Toxicology & Applied Pharmacology* 242 (3):241-246.

LEGENDS

Figure 1.

Phylogenetic tree of GST gene family in mammals. Maximum likelihood and Bayesian phylograms describing phylogenetic relationships among 333 intact (complete coding sequences) mammalian GSTs (21 species). Numbers on nodes correspond to maximum likelihood bootstrap support values and Bayesian posterior probabilities. The GST subclasses are indicated by different colours: alpha (red), mu (orange), pi (turquoise), omega (azure), sigma (green), zeta (purple), and theta (blue).

Figure 2.

Genomic organization of GST genes in 21 mammalian species. The arrowed boxes represent genes and directions of transcription. GST genes are shown in orange, while flanking genes are indicated in green, purple, blue, and pink. Filled boxes: intact genes; empty boxes: partial gene, and empty boxes with a vertical line: pseudogenes (P). Connecting horizontal lines indicate genes on the same chromosome/genomic scaffold. Gene family sizes for ancestral states are shown along each node in the phylogenetic tree.

Figure 3.

Differences in cytosolic GST genes between mammals inhabing aquatic and terrestrial habitats. (A) All GSTs genes combined, comparing aquatic and terrestrial mammals. Green dots indicate terrestrial species; blue dots, cetacean species; grey dots, pinnipeds; purple dots, sirenians. (B) alpha-class GSTs (GSTA) genes, comparing aquatic and terrestrial mammals. Annotated as in (A). (C) Comparison of all GST genes, alpha-class GSTs (GSTA) genes, and mu-class GSTs (GSTM) genes

in ceteaceans and terrestrial mammals. Bar chart shows mean±s.e.m. Blue bars indicates cetacean species; green bars, terrestrial species. We compared each pair of distributions by phylogenetic ANOVA (phylANOVA) tests, which control for shared ancestry. The *p*-value for each test is shown above each plot.

Figure 1

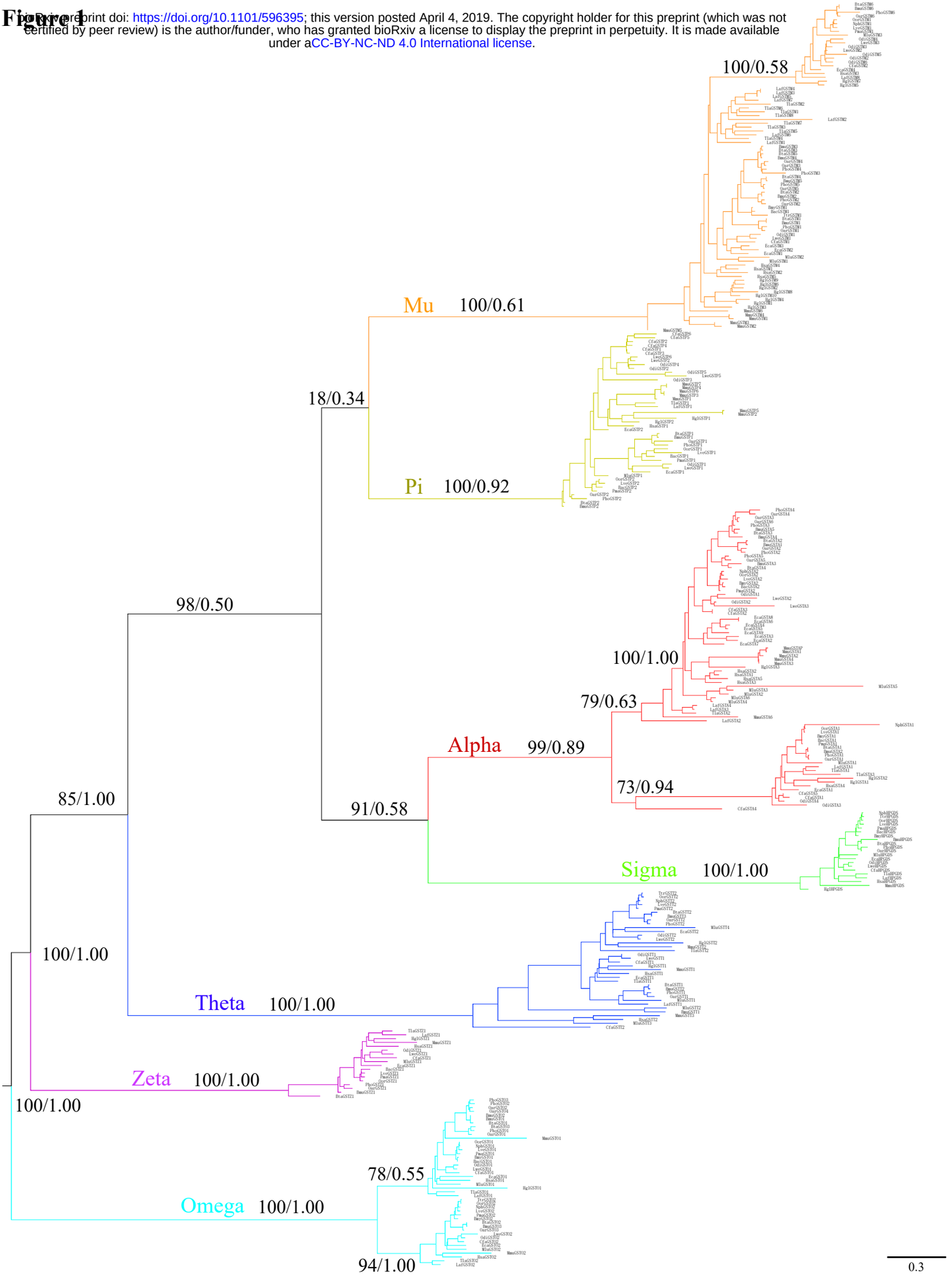


Figure 2

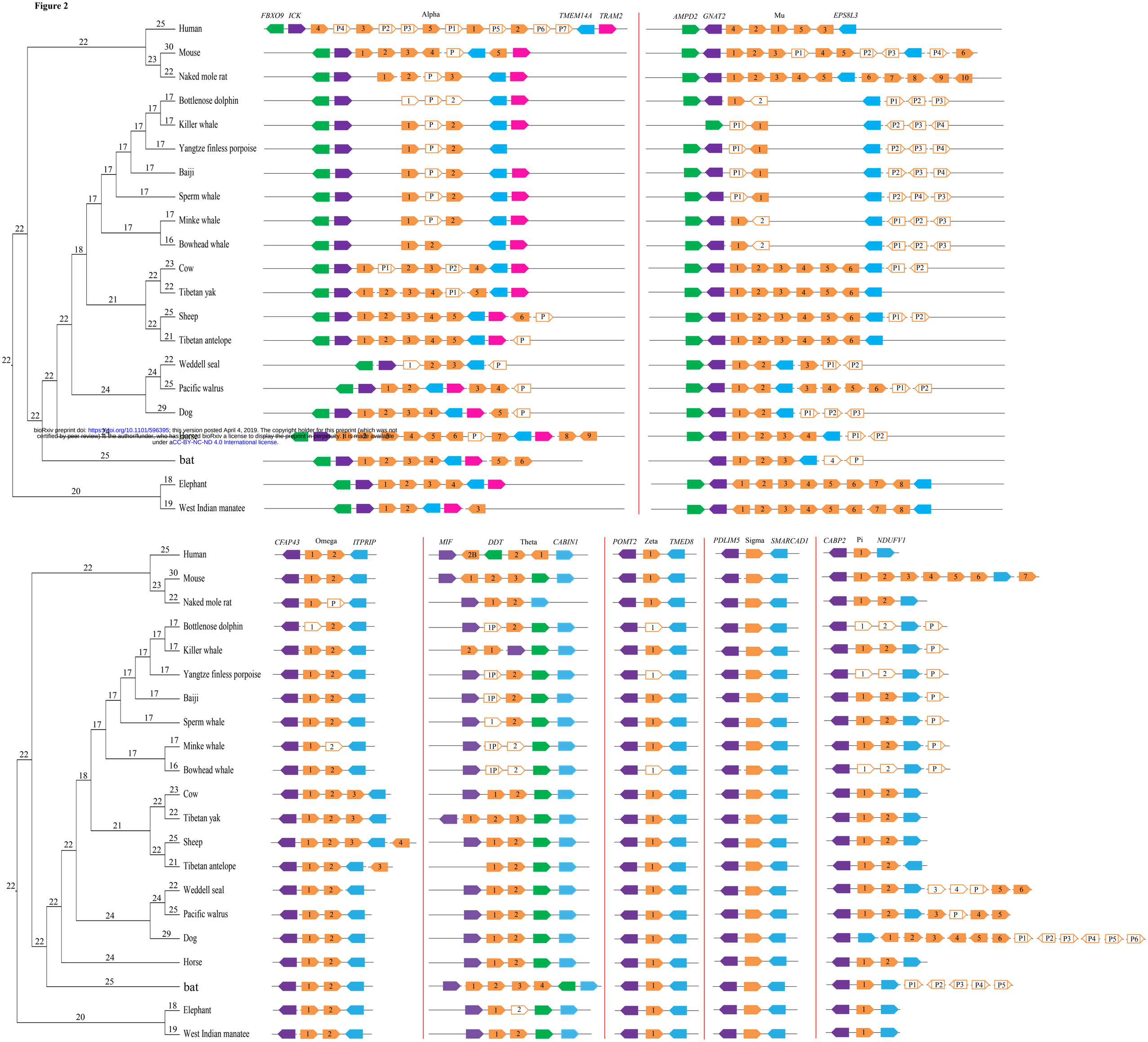
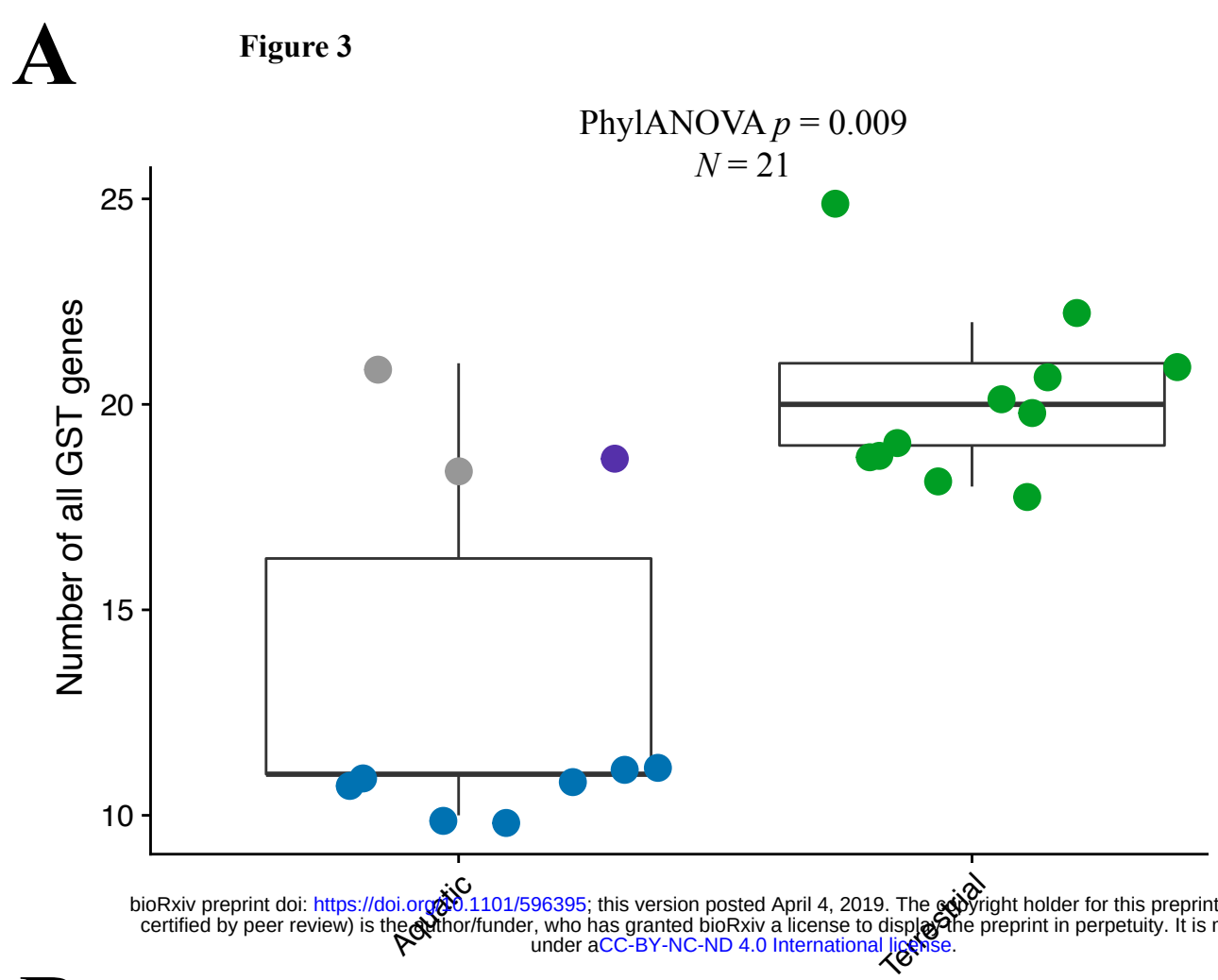
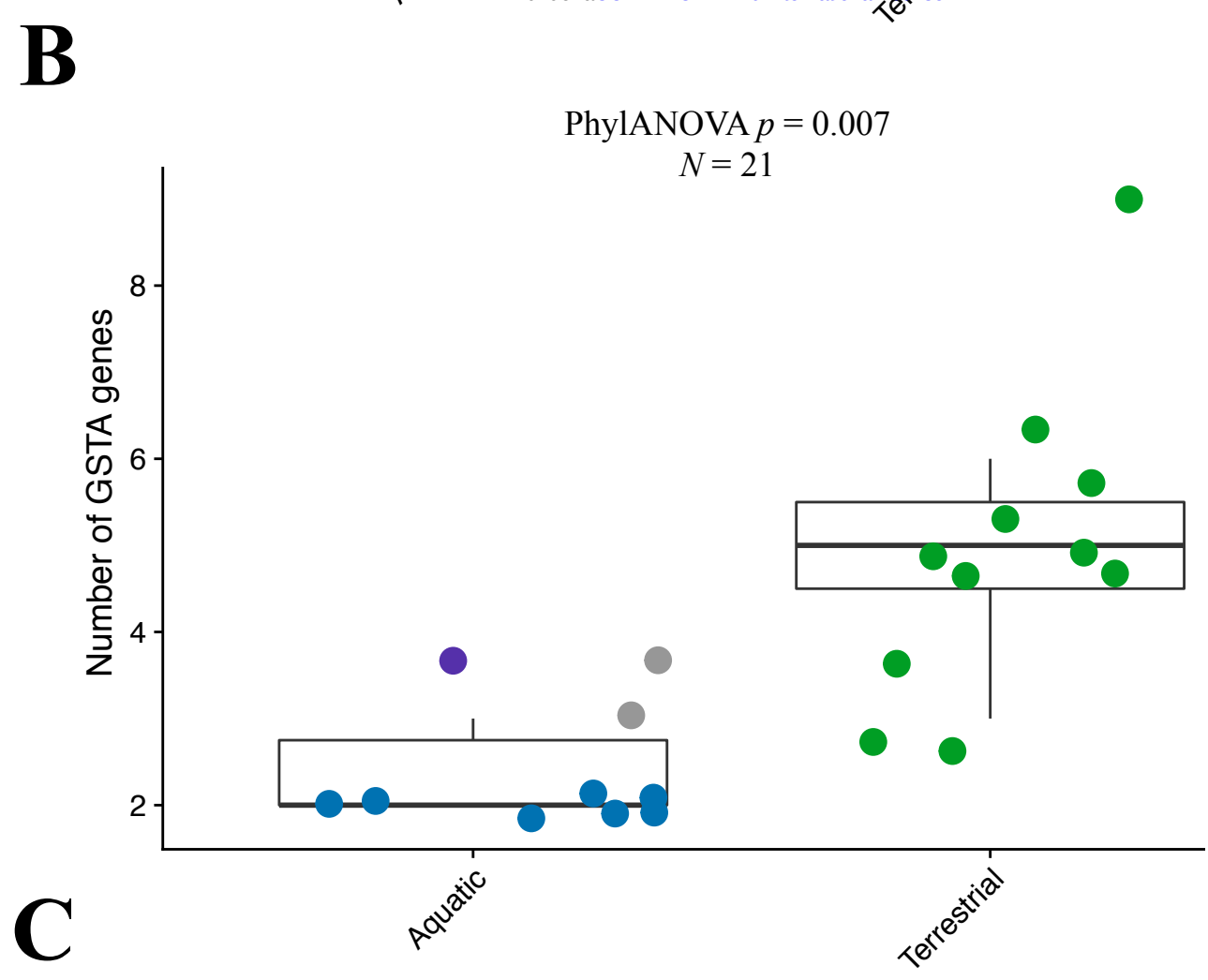


Figure 3





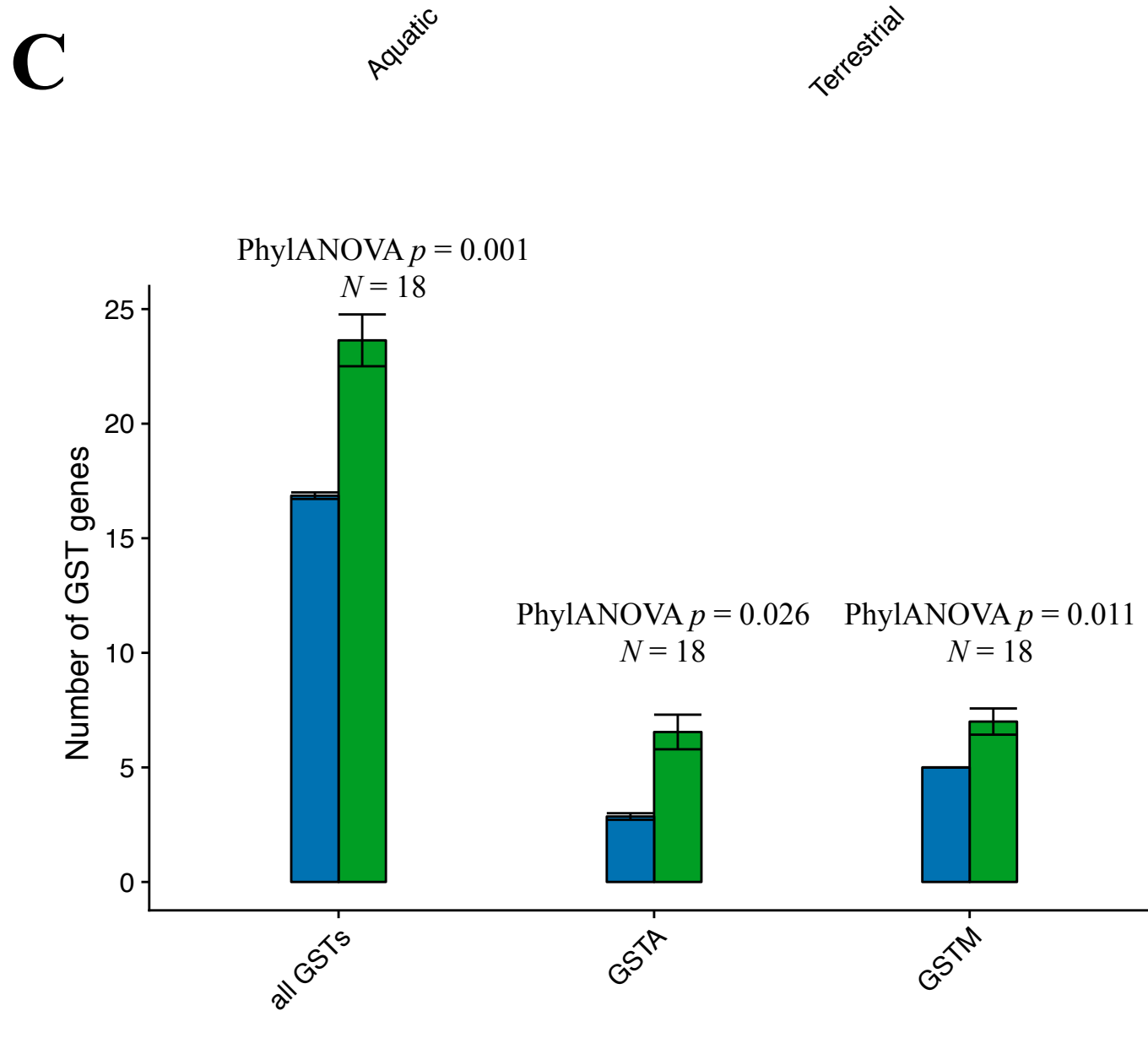


Table 1 Overview of cytosolic glutathione transferase (GST) genes in 21 mammals.

Subclass	Alpha (GSTA)	Mu (GSTM)	Theta (GSTT)	Pi (GSTP)	Zeta (GSTZ)	Omega (GSTO)	Sigma (HPGDS)	Total number of GSTs	Pseudogene proportion
Human	12(7:5:0)	5(0:5:0)	3(0:3:0)	1(0:1:0)	1(0:1:0)	2(0:2:0)	1(0:1:0)	25(7:18:0)	0.28
Mouse	6(1:5:0)	10(4:6:0)	3(0:3:0)	7(0:7:0)	1(0:1:0)	2(0:2:0)	1(0:1:0)	30(5:25:0)	0.17
Naked mole rat	4(1:3:0)	10(0:10:0)	2(0:2:0)	2(0:2:0)	1(0:1:0)	2(1:1:0)	1(0:1:0)	22(2:20:0)	0.09
Bottlenose dolphin	3(1:0:2)	5(3:1:1)	2(1:1:0)	3(1:0:2)	1(0:0:1)	2(0:1:1)	1(0:1:0)	17(6:4:7)	0.35
Killer whale	3(1:2:0)	5(4:1:0)	2(0:2:0)	3(1:2:0)	1(0:1:0)	2(0:2:0)	1(0:1:0)	17(6:11:0)	0.35
Yangtze finless porpoise	3(1:2:0)	5(4:1:0)	2(1:1:0)	3(1:0:2)	1(0:0:1)	2(0:2:0)	1(0:1:0)	17(7:7:3)	0.41
Yangtze river dolphin	3(1:2:0)	5(4:1:0)	2(1:1:0)	3(1:2:0)	1(0:1:0)	2(0:2:0)	1(0:1:0)	17(7:10:0)	0.41
Sperm whale	3(1:2:0)	5(4:1:0)	2(0:1:1)	3(1:2:0)	1(0:1:0)	2(0:2:0)	1(0:1:0)	17(6:10:1)	0.35
Minke whale	3(1:2:0)	5(3:1:1)	2(1:1:0)	3(1:2:0)	1(0:1:0)	2(0:1:1)	1(0:1:0)	17(6:9:2)	0.35
Bowhead whale	2(0:2:0)	5(3:1:1)	2(1:1:0)	3(1:0:2)	1(0:0:1)	2(0:2:0)	1(0:1:0)	16(5:7:4)	0.31
Cow	6(2:4:0)	8(2:6:0)	2(0:2:0)	2(0:2:0)	1(0:1:0)	3(0:3:0)	1(0:1:0)	23(4:19:0)	0.17
Tibetan yak	6(1:5:0)	6(0:6:0)	3(0:3:0)	2(0:2:0)	1(0:1:0)	3(0:3:0)	1(0:1:0)	22(1:21:0)	0.05
Sheep	7(1:6:0)	8(2:6:0)	2(0:2:0)	2(0:2:0)	1(0:1:0)	4(0:4:0)	1(0:1:0)	25(3:22:0)	0.12
Tibetan antelope	6(1:5:0)	6(0:6:0)	2(0:2:0)	2(0:2:0)	1(0:1:0)	3(0:3:0)	1(0:1:0)	21(1:20:0)	0.05
Weddell seal	4(1:2:1)	5(2:3:0)	2(0:2:0)	7(1:4:2)	1(0:1:0)	2(0:2:0)	1(0:1:0)	22(4:15:3)	0.18
Pacific walrus	5(1:4:0)	8(2:6:0)	2(0:2:0)	6(1:5:0)	1(0:1:0)	2(0:2:0)	1(0:1:0)	25(4:21:0)	0.16
Dog	6(1:5:0)	5(3:2:0)	2(0:2:0)	12(6:6:0)	1(0:1:0)	2(0:2:0)	1(0:1:0)	29(10:19:0)	0.34
Horse	10(1:9:0)	6(2:4:0)	2(0:2:0)	2(0:2:0)	1(0:1:0)	2(0:2:0)	1(0:1:0)	24(3:21:0)	0.13
Microbat	6(0:6:0)	5(1:3:1)	4(0:4:0)	6(5:1:0)	1(0:1:0)	2(0:2:0)	1(0:1:0)	25(6:18:1)	0.24
Florida manatee	4(0:4:0)	8(0:8:0)	2(0:2:0)	1(0:1:0)	1(0:1:0)	2(0:2:0)	1(0:1:0)	19(0:19:0)	0.00
Elephant	3(0:3:0)	8(0:8:0)	2(0:1:1)	1(0:1:0)	1(0:1:0)	2(0:2:0)	1(0:1:0)	18(0:17:1)	0.00

The number outside brackets is the number of genes in a subclass, while the numbers in the brackets, separated by a colon, indicate the number of pseudogenes, intact gene, and partial genes. The orders of cetacean are shadowed.

Table 2 Amino acid sites under positive selection detected by ML methods

Gene	Site Position	PAML	Datamonkey			TreeSAAP	
		M8[a]	SLAC[b]	FEL[c]	REL[d]	Radical Changes in AA Properties[e]	Total
GSTA1	36	√		√	√	P_α, *c*, *pHi*, α_c, *P*	5
	49	√	√	√	√	N_s, R_F, P_c, *h*, *F*, *p*, R_a, *P*	8
	96	√	√	√	√	*pK′*, *F*, R_a, H_p, H_t, *P*	4
	100	√		√		-	0
	103	√		√	√	*p*	1
	121	√	√	√	√	*pK′*, R_a, H_t	3
	208	√			√	P_α, P_c, *pK′*, *F*, *P*	5
	212	√			√	-	0
	215	√	√	√	√	B_r, R_F, *h*, *pHi*, H_{nc}, *p*, α_c, E_t	8
	222	√	√	√		*pK′*	1
GSTM1	107	√	√	√	√	N_s, R_F, *h*, *p*, *P*	5
GSTO1	23		√	√	√	B_r, B_l, R_F, P_c, *h*, *F*, *p*, E_t, *P*	9
	47		√	√	√	-	0
	69	√			√	K^0, *F*, *P*	3
	125	√		√		α_c	1
	127	√	√	√	√	*pK′*	1
	128	√	√	√	√	N_s, B_r, B_l, R_F, P_c, *h*, *F*, *p*, E_l, R_a, H_p, H_t, E_t, *P*	14
	216	√			√	N_s, B_r, R_F, P_c, *h*, *F*, *p*, K^0, H_{nc}, E_l, α_c, α_n, E_{sm}, E_t, *P*	15
	226	√			√	*pK′*	1
	227	√			√	P_α, *pHi*, *P*	3
GSTO2	14	√		√	√	P_α	1
	23		√	√	√	-	0
	43	√	√	√	√	N_s, α_n, *pHi*, B_r, R_F, *h*, H_{nc}, *p*, E_{sm}, E_t	10
	141	√			√	P_α, P_c, *P*	3
	164	√	√	√	√	N_s, P_c, *F*, R_a, *P*	5
GSTP1	12	√	√			P_c, K^0, *pHi*, α_c	4
	40		√	√	√	P_α, B_r, P_c, *h*, *F*, *p*, E_t, *P*	8
GSTP2	11	√				P_α, P_c, α_n, R_a, *P*	5
	12	√	√	√	√	K^0, *pHi*, α_c	3
	76		√	√	√	-	0
	111	√	√	√		N_s, *pHi*, E_{sm}	3
	121		√	√		*h*, *pHi*, *p*	3
GSTT2	25		√	√		α_c, P_r, E_t, *pHi*, E_{sm}	5
	80		√	√		P_α	1
	237	√		√		-	0
GSTZ1	71		√	√		B_l, P_c, *F*, α_m, *P*	5
	125		√	√	√	K^0, P_c	2

137	√		√	-	0

[a] Sites detected under selection in M8 with posterior probabilities > 80% in the BEB analyses.

[b] Codons with P values < 0.2.

[c] Codons with P values < 0.2.

[d] Codons with Bayes factors > 50.

[e] Radical changes in amino acid properties under category 6-8 were detected in TreeSAAP. Physicochemical amino acid properties available in TreeSAAP are as following: $\boldsymbol{\alpha_c}$: Power to be C-term., α-helix; $\boldsymbol{\alpha_n}$: Power to be in the N-terminal of an α-helix; $\boldsymbol{B_r}$: Buriedness; $\boldsymbol{C_a}$: Helical contact energy; $\boldsymbol{E_l}$: Long-range non-bonded energy; $\boldsymbol{E_{sm}}$: Short and medium range non-bonded energy; $\boldsymbol{E_t}$: Total non-bonding Energy; $\boldsymbol{F}$: Mean r.m.s. fluctuation displacement; $\boldsymbol{h}$:Hydropathy; $\boldsymbol{H_{nc}}$: Normal consensus hydrophobicity; $\boldsymbol{H_p}$: Surrounding hydrophobicity; $\boldsymbol{H_t}$: Thermodynamic transfer hydrophobicity; $\boldsymbol{K^o}$: Compressibility; $\boldsymbol{\mu}$: Refractive index; $\boldsymbol{M_v}$: Molecular volume; $\boldsymbol{M_w}$: Molecular weight; $\boldsymbol{N_s}$: Average number of surrounding residues; $\boldsymbol{P_\alpha}$: α- helical tendencies; $\boldsymbol{P_\beta}$: β-structure tendencies; $\boldsymbol{P_c}$: Coil tendencies; $\boldsymbol{P}$: Turn tendencies; $\boldsymbol{p}$: Polarity; $\boldsymbol{pH_i}$: Isoelectric point; $\boldsymbol{pK'}$: Equilibrium Constant of ionization for COOH; $\boldsymbol{P_r}$: Polar requirement; $\boldsymbol{R_a}$: Solvent accessible reduction ratio; $\boldsymbol{R_F}$: Chromatographic index; $\boldsymbol{V^0}$: Partial specific volume;

Table 3 Selective pattern analyzed by Branch-site model

Genes	Branch-site models[a]	-lnL[b]	2ΔlnL	*p*-value	ω values	Positively selected site[c]
GSTA1	Teminal branch of Lwe					
	ma	4319.495			ω0 = 0.092 ω1 = 1.0 ω2 = 338.762	45-0.962*, 48-0.875, 52-0.875, 59-0.800, 64-0.853, 70-0.848, 184-0.864, 185-0.998**, 186-0.895, 201-0.827
	ma0	4343.658	48.325	<0.001	ω0 = 0.079 ω1 = 1.0 ω2 = 1.0	
GSTA4	Teminal branch of Hsa					
	ma	2757.115			ω0 = 0.102 ω1 = 1.0 ω2 = 999	42-0.947, 185-0.961*
	ma0	2762.892	11.555	<0.001	ω0 = 0.096 ω1 = 1.0 ω2 = 1.0	
GSTM1	Teminal branch of Odi					
	ma	3851.226			ω0 = 0.104 ω1 = 1.0 ω2 = 24.372	35 D 0.996**, 44 W 0.870, 82 H 0.869, 94 R 0.860, 166 R 0.873, 215 N 0.989*,
	ma0	3855.022	7.592	0.006	ω0 = 0.104 ω1 = 1.0 ω2 = 1.0	
	LCA of Lwe and Odi					
	ma	3856.774			ω0 = 0.113 ω1 = 1.0 ω2 = 88.802	93 I 0.943
	ma0	3860.499	7.450	0.006	ω0 = 0.111 ω1 = 1.0 ω2 = 1.0	
GSTT1	Teminal branch of Laf					
	ma	3407.906			ω0 = 0.12 ω1 = 1.0 ω2 = 999.0	172-0.943, 219-0.939,
	ma0	3411.690	7.565	0.006	ω0 = 0.119 ω1 = 1.0 ω2 = 1.0	
	Teminal branch of Mlu					
	ma	3408.749			ω0 = 0.126 ω1 = 1.0 ω2 = 999	218-0.925,
	ma0	3412.490	7.483	0.006	ω0 = 0.12 ω1 = 1.0 ω2 =1.0	
GSTT2	Teminal branch of Bta					
	Ma	3764.990			ω0 = 0.104 ω1 = 1.0, ω2 = 999	237-0.876, 239-0.987*
	Ma0	3780.489	30.999	<0.001	ω0 = 0.106, ω1 = 1.0, ω2 = 1	
	Teminal branch of Eca					
	Ma	3765.226			ω0 = 0.102 ω1 = 1.0, ω2 = 999	18- 0.977*, 19- 0.971*, 21-0.964*, 24-0.895, 26-0.926, 30-0.936, 143-0.941
	Ma0	3775.418	20.384	<0.001	ω0 = 0.098 ω1 = 1.0, ω2 = 1	
	Teminal branch of Tla					
	ma	3770.135			ω0 = 0.096 ω1 = 1.0 ω2 = 999	13-0.936, 28-0.953*, 89-0.922, 93-0.986*, 190-0.866

	ma0	3776.142	12.014	<0.001	ω0 = 0.091 ω1 = 1.0 ω2 = 1.0	
GSTP2	Teminal branch of Oar					
	ma	2704.119			ω0 = 0.069 ω1 = 1.0 ω2 = 999	113-0.989*, 114-0.961*, 205-0.804, 210-0.951*
	ma0	2717.934	27.631	<0.001	ω0 = 0.063 ω1 = 1.0 ω2 = 1.0	
	LCA of Lwe and Odi					
	ma	2711.299			ω0 = 0.059 ω1 = 1.0 ω2 = 10.047	106-0.890, 135-0.953*, 144-0.869, 174-0.987*, 208-0.878,
	ma0	2714.800	7.002	0.008	ω0 = 0.059 ω1 = 1.0 ω2 = 1.0	
	Teminal branch of Pho					
	ma	2707.268			ω0 = 0.056 ω1 = 1.0 ω2 = 15.788	3-0.814, 4-0.997**, 6-0.995**, 7-0.952*, 8-0.808
	ma0	2711.251	7.965	0.005	ω0 = 0.055 ω1 = 1.0 ω2 = 1.0	
	Teminal branch of Pma					
	ma	2710.120			ω0 = 0.062 ω1 = 1.0 ω2 = 108.24	41-0.933, 110-0.908, 117-0.941, 210-0.911
	ma0	2717.793	15.347	<0.001	ω0 = 0.062 ω1 = 1.0 ω2 = 1.0	
GSTO1	Teminal branch of Hgl					
	ma	4450.655			ω0 = 0.065 ω1 = 1.0 ω2 = 18.7	11-0.988*, 21-0.873, 24-0.853, 195-0.974*, 213-0.936
	ma0	4454.959	8.608	0.003	ω0 = 0.061 ω1 = 1.0 ω2 = 1.0	
	Teminal branch of Ttr					
	ma	4453.983			ω0 = 0.082 ω1 = 1.0 ω2 = 999	47-0.991**,
	ma0	4458.928	9.890	0.002	ω0 = 0.082 ω1 = 1.0 ω2 = 1.0	
GSTZ1	Teminal branch of Laf					
	ma	3360.900			ω0 = 0.077 ω1 = 1.0 ω2 = 34.477	14-0.815, 181-0.819, 183-0.826
	ma0	3365.554	9.306	0.002	ω0 = 0.074 ω1 = 1.0 ω2 = 1.0	
HPGDS	LCA of Cetartiodactyla					
	ma	2891.856			ω0 = 0.105 ω1 = 1.0 ω2 = 163.254	183-0.980*,
	ma0	2896.024	8.335	0.004	ω0 = 0.106 ω1 = 1.0 ω2 = 1.0	

Note:

[a] Bta: Bos taurus; Eca: Equus caballus; Hgl: Heterocephalus glaber; Has: Homo sapiens; Lwe: Leptonychotes weddellii; Laf: Loxodonta africana; Mmu: Mus musculus; Mlu: Myotis lucifugus; Nph: Neophocaena phocaenoides; Odi: Odobenus rosmarus divergens; Oor: Orcinus orca; Oar: Ovis aries; Pho: Pantholops hodgsonii; Pma: Physeter macrocephalus; Tla: Trichechus manatus latirostris; Ttr: Tursiops truncatus

[b] lnL is the log-likelihood score.

[c] Codons with posterior probabilities (pp) >80% in the BEB analyses. * pp >95%, ** pp >99%.

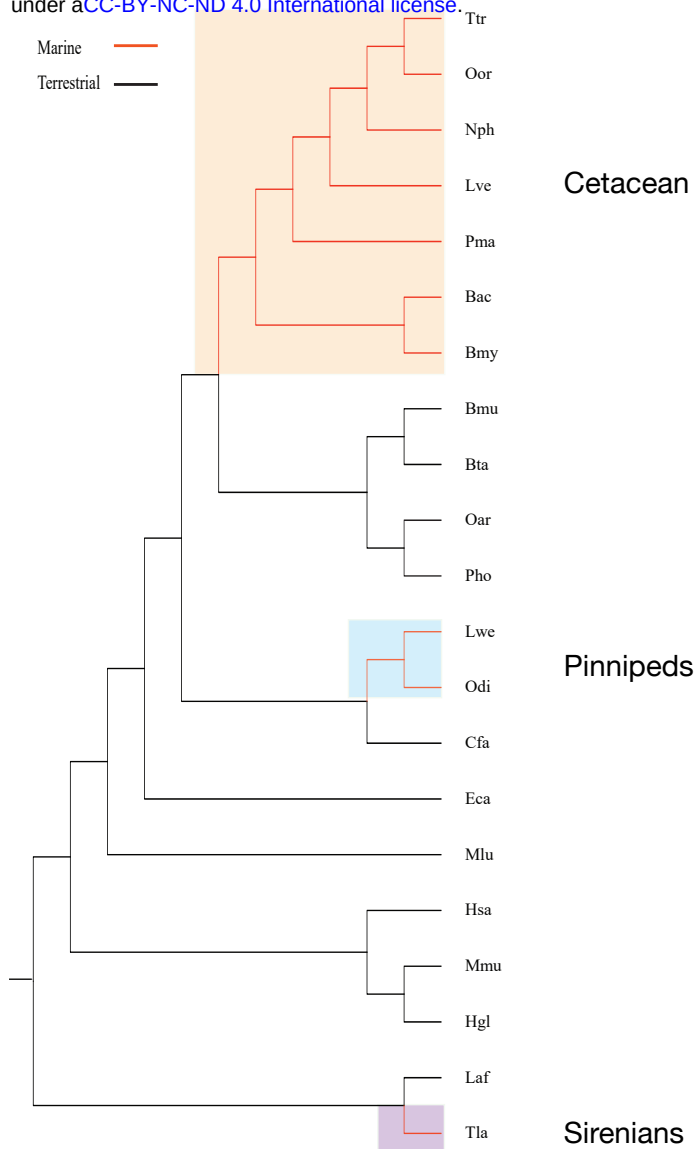


Figure S1 Mammalian species phylogeny (Douzery et al. 2014) indicating each of the clade partitions tested by the branch-site and clade models.

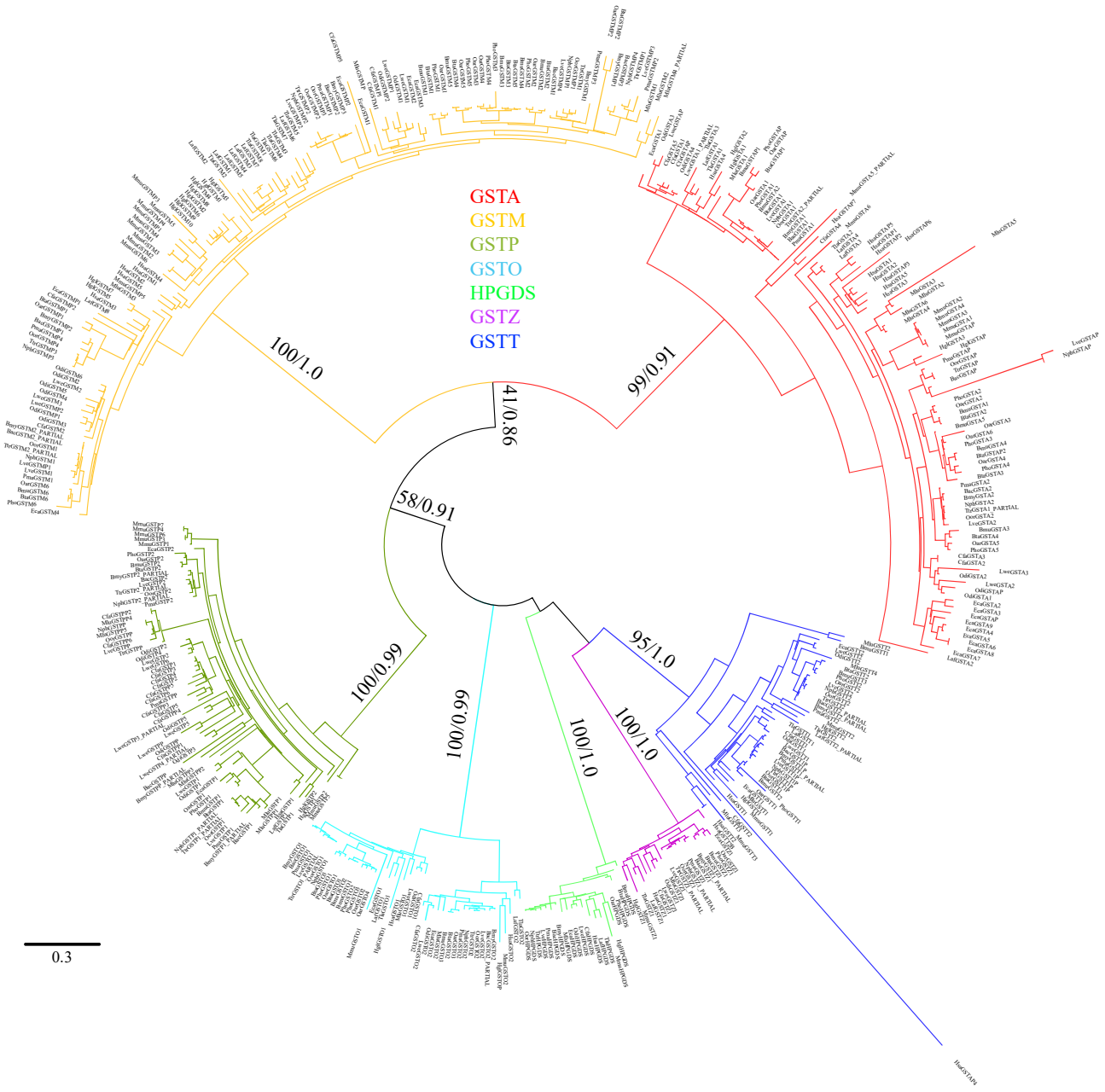


Figure S2 Phylogenetic tree of the GST gene family in mammals. Shown is the maximum likelihood and Bayesian tree built using RaxML and Mrbayes methods with the multiple alignments of 448 GST nucleotide sequences from 21 mammalian species. The bootstrap values and posterior probability are shown at nodes. The clades of alpha, mu, pi, omega, sigma, zeta, theta are shown in red, orange, turquoise, azure, green, purple, and blue, respectively.

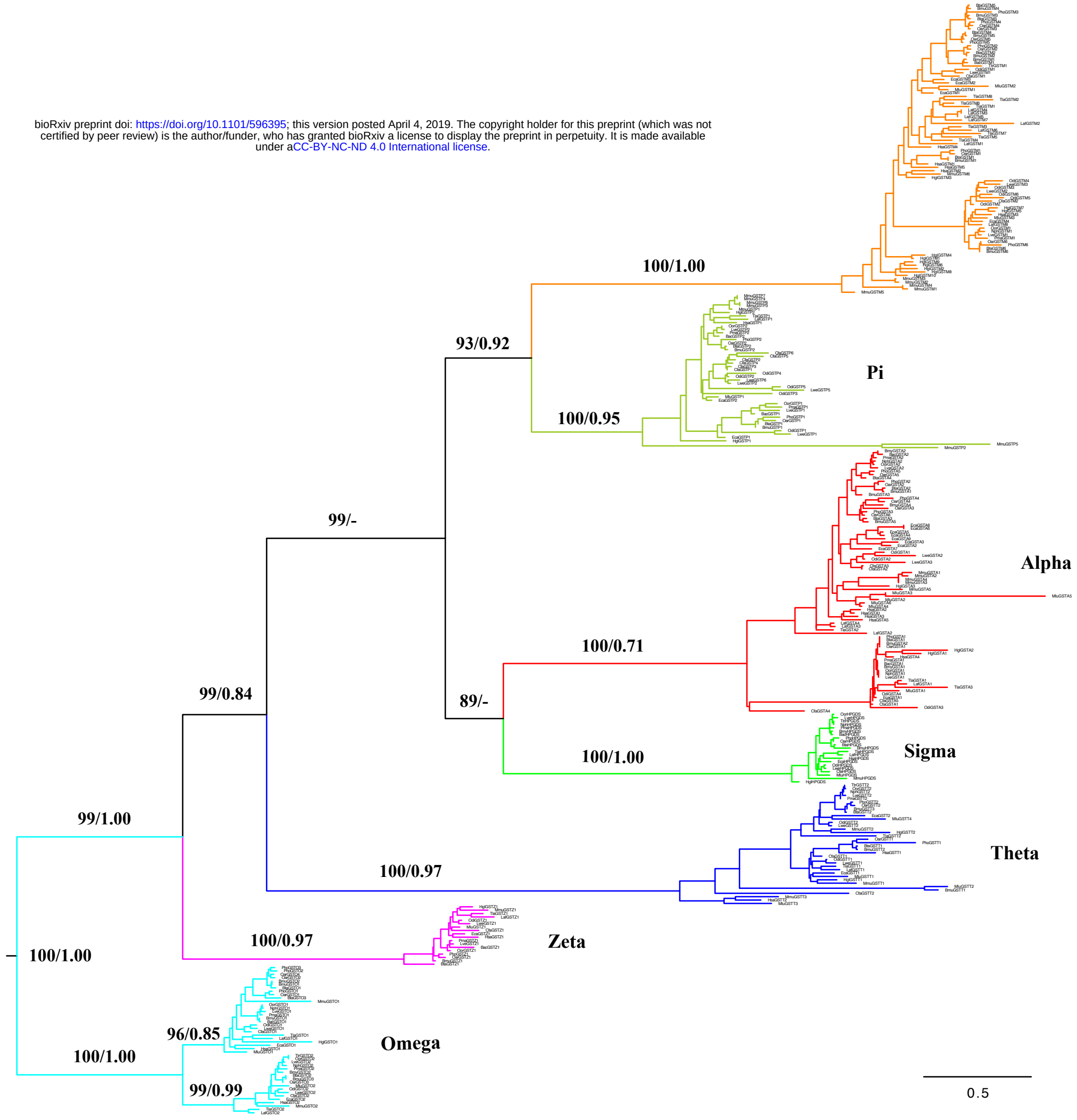


Figure S3 Phylogenetic tree of GST gene family based on 333 mammalian protein sequences. Numbers of the nodes correspond to maximum likelihood bootstrap support values and Bayesian posterior probabilities. The clades of alpha, mu, pi, omega, sigma, zeta, theta are shown in red, orange, turquoise, azure, green, purple, and blue, respectively.

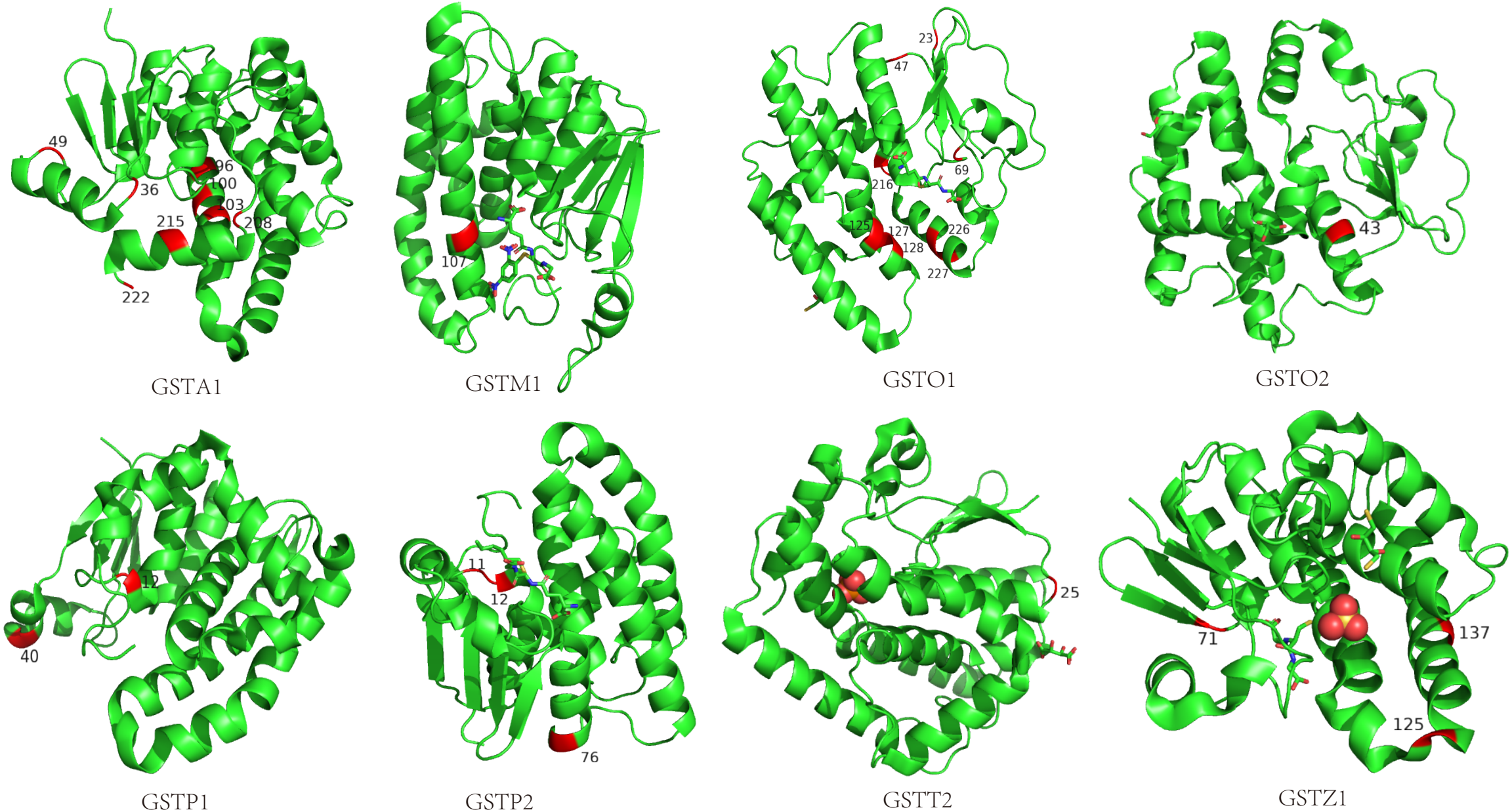
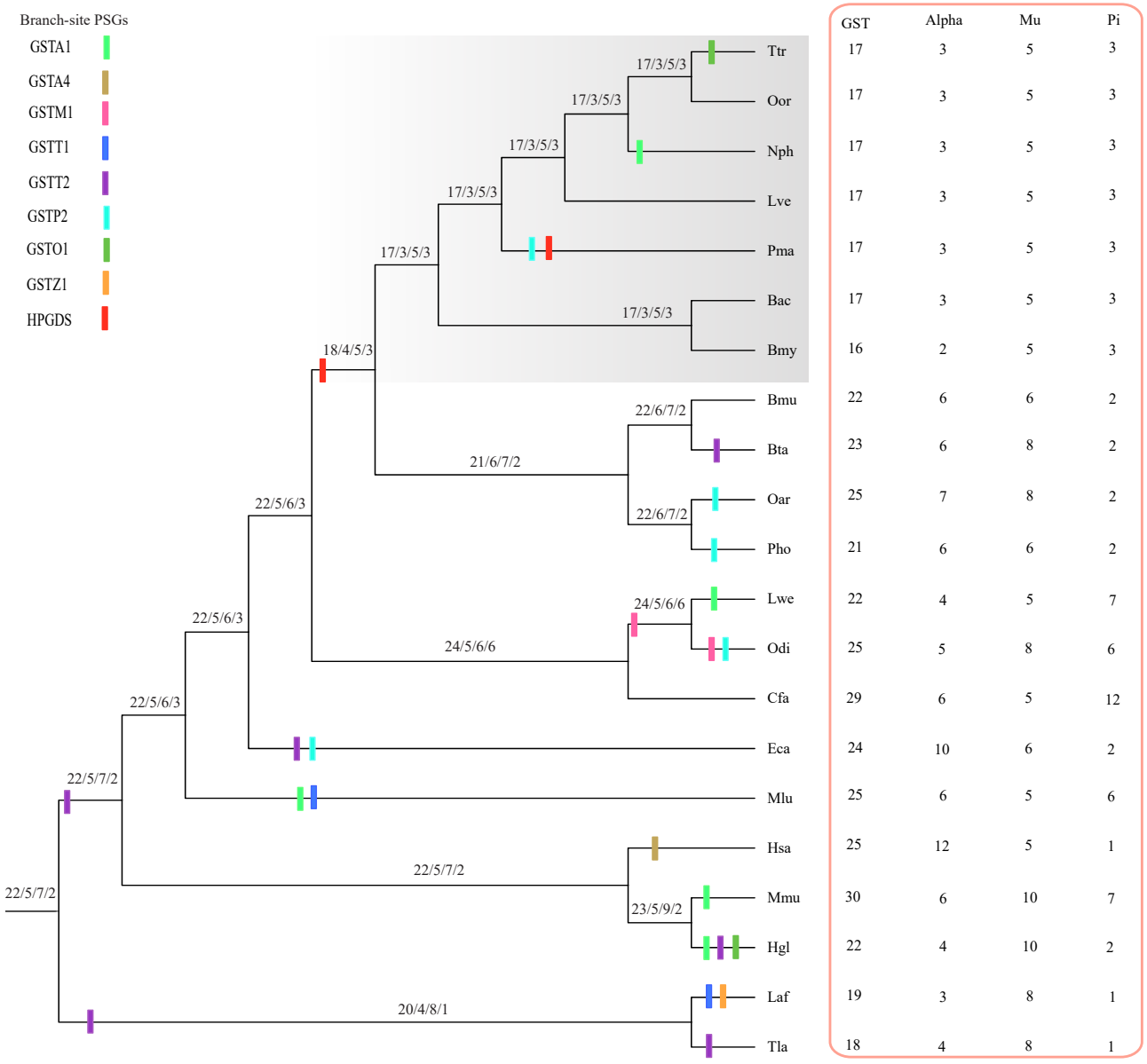


Figure S4 Positively selected sites are shown in crystal structure with red. The crystal structures of GSTA1 (1gsd), GSTM1 (1XW6), GSTO1 (5V3Q), GSTO2 (3Q18), GSTP1 (5X79), GSTP2 (P46425), GSTT2 (4MPF), GSTZ1 (2CZ2) were taken from the Protein Data Bank (http:// www.rcsb.org/pdb).



GST	Alpha	Mu	Pi
17	3	5	3
17	3	5	3
17	3	5	3
17	3	5	3
17	3	5	3
17	3	5	3
16	2	5	3
22	6	6	2
23	6	8	2
25	7	8	2
21	6	6	2
22	4	5	7
25	5	8	6
29	6	5	12
24	10	6	2
25	6	5	6
25	12	5	1
30	6	10	7
22	4	10	2
19	3	8	1
18	4	8	1

Figure S5 Evidence for lineage-specific positive selection in mammalian GSTs. Inset shows the gene number of GSTs, alpha, mu and pi class inferred to have occurred gain and loss along mammalian lineages.

Common name	Scientific name	Chromosome/Contig/Scaffold
Bottlenose dolphin	*Tursiops truncatus*	scaffold_2665
		scaffold_2665
		scaffold_2665
		scaffold_108534
		scaffold_108534
		scaffold_346
		scaffold_92999
		scaffold_114085
		scaffold_1888
		scaffold_1888
		scaffold_2747
		scaffold_376
		scaffold_112770
		scaffold_3218
		scaffold_3218
		scaffold_2443
		scaffold_2281
Killer whale	*Orcinus orca*	NW_004438672.1
		NW_004438672.1
		NW_004438672.1
		NW_004438417.1
		NW_004438417.1
		NW_004438540.1
		NW_004438596.1
		NW_004438773.1
		NW_004438429.1
		NW_004438429.1
		NW_004438509.1
		NW_004438509.1
		NW_004438491.1
		NW_004438489.1
		NW_004438489.1
		NW_004438415.1
		NW_004438503.1
Yangtze finless porpoise	*Neophocaena phocaenoides*	scaffold962
		scaffold962
		scaffold962
		scaffold35
		scaffold35
		scaffold243
		scaffold414
		scaffold424
		scaffold240
		scaffold240
		scaffold43
		scaffold43
		scaffold371

		scaffold110
		scaffold110
		scaffold49
		scaffold7
Baiji	*Lipotes vexillifer*	NW_006778796.1
		NW_006778796.1
		NW_006778796.1
		NW_006774564.1
		NW_006774564.1
		NW_006784589.1
		NW_006776528.1
		NW_006787679.1
		NW_006798383.1
		NW_006798383.1
		NW_006776536.1
		NW_006776536.1
		NW_006794929.1
		NW_006777435.1
		NW_006777435.1
		NW_006787038.1
		NW_006775603.1
Sperm whale	*Physeter macrocephalus*	NW_006724412.1
		NW_006724412.1
		NW_006724412.1
		NW_006713392.1
		NW_006713392.1
		NW_006713123.1
		NW_006714483.1
		NW_006717948.1
		NW_006714478.1
		NW_006714478.1
		NW_006724202.1
		NW_006724202.1
		NW_006716229.1
		NW_006724217.1
		NW_006724217.1
		NW_006713461.1
		NW_006712764.1
Minke whale	*Balaenoptera acutorostrata*	NW_006732678.1
		NW_006732678.1
		NW_006732678.1
		NW_006725576.1
		NW_006725576.1
		NW_006725654.1
		NW_006727021.1
		NW_006727464.1
		NW_006725776.1
		NW_006725776.1

		NW_006725687.1
		NW_006725687.1
		NW_006725687.1
		NW_006733789.1
		NW_006733789.1
		NW_006727797.1
		NW_006727686.1
Bowhead whale	*Balaena mysticetus*	scaffold_90
		scaffold_90
		scaffold_663
		scaffold_663
		scaffold_750
		scaffold_1063
		scaffold_3148
		scaffold_2004
		scaffold_2004
		scaffold_331
		scaffold_331
		scaffold_479
		scaffold_2197
		scaffold_2197
		scaffold_101
		scaffold_1495
Cow	*Bos taurus*	chr23
		chr23
		chr23
		chr23
		chr23
		chr23
		chr3
		chr3
		chr3
		chr3
		chr3
		chr3
		chr12
		chr29
		chr26
		chr26
		chr26
		chr29
		chr29
		chr17
		chr17
		chr10
		chr6
Tibetan yak	*Bos mutus*	NW_005397583.1
		NW_005394842.1

		NW_005392861.1
		NW_005392861.1
		NW_005395320.1
		NW_005395251.1
		NW_005393277.1
		NW_005393277.1
		NW_005393277.1
		NW_005393277.1
		NW_005393277.1
		NW_005393277.1
		NW_005394722.1
		NW_005394722.1
		NW_005394722.1
		NW_005394695.1
		NW_005394695.1
		NW_005394378.1
		NW_005394378.1
		NW_005394378.1
		NW_005398357.1
		NW_005395035.1
Sheep	*Ovis aries*	chr20
		chr20
		chr20
		chr20
		chr20
		chr7
		chr17
		chr1
		chr1
		chr1
		chr1
		chr1
		chr1
		chr10
		chr21
		chr22
		chr22
		chr22
		chr7
		chr21
		chr21
		chr17
		chr17
		chr7
		chr6
Tibetan antelope	*Pantholops hodgsonii*	NW_005806736.1
		NW_005806736.1
		NW_005806736.1

		NW_005806736.1
		NW_005806736.1
		NW_005804725.1
		NW_005813453.1
		NW_005813453.1
		NW_005813453.1
		NW_005818795.1
		NW_005818795.1
		NW_005818795.1
		NW_005813409.1
		NW_005813409.1
		NW_005810114.1
		NW_005810996.1
		NW_005810996.1
		NW_005808332.1
		NW_005808332.1
		NW_005811569.1
		NW_005816419.1
Weddell seal	*Leptonychotes weddellii*	NW_006385838.1
		NW_006385838.1
		NW_006385838.1
		NW_006383120.1
		NW_006384601.1
		NW_006384601.1
		NW_006383076.1
		NW_006383055.1
		NW_006387651.1
		NW_006384846.1
		NW_006384846.1
		NW_006384721.1
		NW_006384721.1
		NW_006383048.1
		NW_006383096.1
		NW_006383397.1
		NW_006383098.1
		NW_006383313.1
		NW_006385210.1
		NW_006385279.1
		NW_006386919.1
		NW_006383499.1
Pacific walrus	*Odobenus rosmarus divergen*	NW_004450762.1
		NW_004450762.1
		NW_004450402.1
		NW_004450402.1
		NW_004450567.1
		NW_004450591.1
		NW_004450591.1
		NW_004450581.1

		NW_004450838.1
		NW_004451511.1
		NW_004452022.1
		NW_004450279.1
		NW_004450392.1
		NW_004452741.1
		NW_004452741.1
		NW_004450289.1
		NW_004450289.1
		NW_004450288.1
		NW_004450854.1
		NW_004451294.1
		NW_004451716.1
		NW_004450381.1
		NW_004450381.1
		NW_004450358.1
		NW_004450264.1
Dog	*Canis lupus familiaris*	NC_006594.3
		NC_006594.3
		NC_006594.3
		NC_006594.3
		NC_006613.3
		NC_006615.3
		NC_006588.3
		NC_006588.3
		NC_006587.3
		NC_006604.3
		NC_006621.3
		NC_006610.3
		NC_006610.3
		NC_006621.3
		NC_006591.3
		NC_006596.3
		NC_006606.3
		NC_006614.3
		NC_006611.3
		NC_006583.3
		NC_006585.3
		NC_006586.3
		NC_006598.3
		NC_006607.3
		NC_006614.3
		NC_006608.3
		NC_006608.3
		NC_006590.3
		NC_006614.3
Horse	*Equus caballus*	NC_009163.2
		NC_009163.2

		NC_009163.2
		NC_009163.2
		NC_009163.2
		NC_009163.2
		NC_009163.2
		NC_009163.2
		NW_001868821.1
		NW_001868821.1
		NC_009148.2
		NC_009148.2
		NC_009148.2
		NC_009148.2
		NC_009145.2
		NC_009175.2
		NC_009144.2
		NC_009144.2
		NC_009155.2
		NC_009155.2
		NC_009151.2
		NC_009151.2
		NC_009167.2
		NC_009146.2
Microbat	*Myotis lucifugus*	NW_005871159.1
		NW_005871159.1
		NW_005871159.1
		NW_005871159.1
		NW_005874859.1
		NW_005877976.1
		NW_005871149.1
		NW_005871149.1
		NW_005871149.1
		NW_005871063.1
		NW_005871253.1
		NW_005871210.1
		NW_005871210.1
		NW_005871463.1
		NW_005871050.1
		NW_005871163.1
		NW_005871172.1
		NW_005871244.1
		NW_005871369.1
		NW_005871838.1
		NW_005871838.1
		NW_005871838.1
		NW_005872469.1
		NW_005871082.1
		NW_005871060.1
Mouse	*Mus musculus*	NC_000075.6

		NC_000075.6
		NC_000075.6
		NC_000075.6
		NC_000067.6
		NC_000067.6
		NC_000069.6
		NC_000069.6
		NC_000069.6
		NC_000069.6
		NC_000069.6
		NC_000069.6
		NC_000069.6
		NC_000069.6
		NC_000071.6
		NC_000084.6
		NC_000085.6
		NC_000085.6
		NW_004058054.1
		NW_004058054.1
		NW_004058054.1
		NC_000085.6
		NC_000085.6
		NC_000085.6
		NC_000067.6
		NC_000076.6
		NC_000076.6
		NC_000076.6
		NC_000078.6
		NC_000072.6
Naked mole rat	*Heterocephalus glaber*	NW_004624850.1
		NW_004624812.1
		NW_004624855.1
		NW_004624855.1
		NW_004624772.1
		NW_004624772.1
		NW_004624772.1
		NW_004624772.1
		NW_004624772.1
		NW_004624750.1
		NW_004624767.1
		NW_004624774.1
		NW_004624842.1
		NW_004624834.1
		NW_004624831.1
		NW_004624831.1
		NW_004624767.1
		NW_004624767.1
		NW_004624747.1

		NW_004624747.1
		NW_004624734.1
		NW_004624757.1
Elephant	*Loxodonta africana*	NW_003573420.1
		NW_003573420.1
		NW_003573420.1
		NW_003573420.1
		NW_003573431.1
		NW_003573431.1
		NW_003573431.1
		NW_003573431.1
		NW_003573431.1
		NW_003573431.1
		NW_003573431.1
		NW_003573431.1
		NW_003573430.1
		NW_003573430.1
		NW_003573491.1
		NW_003573488.1
		NW_003573488.1
		NW_003573429.1
		NW_003573450.1
West Indian manatee	*richechus manatus latirostri*	NW_004444151.1
		NW_004444151.1
		NW_004443985.1
		NW_004444140.1
		NW_004444140.1
		NW_004444140.1
		NW_004444140.1
		NW_004444140.1
		NW_004444140.1
		NW_004444233.1
		NW_004444233.1
		NW_004443943.1
		NW_004443943.1
		NW_004444049.1
		NW_004444093.1
		NW_004444093.1
		NW_004443946.1
		NW_004444114.1

Start	End	Strand
311957	308621	-
323589	320240	-
365839	350600	-
142914	127426	-
108041	111068	+
105104	104464	-
45936	46298	+
72325	71941	-
6045	15048	+
21876	38103	+
129465	131857	+
2083	4910	+
740825	741266	+
135610	139308	+
146147	151431	+
102162	108973	+
137274	107764	-
1221243	1206859	-
1180635	1174939	-
1168761	1160286	-
24678998	24694191	+
24713421	24710381	-
7480061	7479492	-
958767	958339	-
130169	129767	-
24706574	24697563	-
24690780	24674458	-
4448204	4450601	+
4483629	4486050	+
8429355	8428914	-
10946841	10955289	+
10957036	10962087	+
15570460	15576931	+
5284422	5255322	-
522101	536474	+
563215	568025	+
580387	588805	+
1224711	1211402	-
1191666	1194690	+
1443818	1443269	-
14473	14871	+
3599881	3599308	-
6785674	6797085	+
6803844	6820196	+
5282935	5280557	-
5253133	5250711	-
5763048	5762700	-

7828812	7837219	+
7838961	7844460	+
2744885	2738459	-
14924249	14952794	+
4860723	4846298	-
4820755	4817404	-
4802593	4794171	-
2070163	2054712	-
2035418	2038526	+
592194	591645	-
703471	704366	+
821529	821179	-
4402723	4413866	+
4421640	4437971	+
646732	644202	-
611193	608777	-
851929	851488	-
840363	848536	+
850277	855345	+
4023665	4031942	+
4883267	4853702	-
354722	368852	+
390996	397095	+
407367	415823	+
616825	619259	+
636916	633811	-
415302	415879	+
116362	116913	+
29297	29702	+
170393	159571	-
153291	136621	-
818297	815914	-
784002	781564	-
84194	84622	+
795454	791797	-
786260	781181	-
431432	437872	+
245841	272640	+
7666132	7651799	-
7613441	7607478	-
7598491	7590280	-
10463283	10446972	-
10427044	10430079	+
1932901	1933235	+
698721	699305	+
223664	224298	+
1179887	1168416	-
1161625	1144988	-

19989283	19992315	+
20026301	20028694	+
20026301	20028694	+
7597041	7604904	+
7606646	7611709	+
13609812	13616393	+
4852036	4879651	+
846508	832505	-
775362	767188	-
271424	286063	+
304206	302002	-
610199	609559	-
10084	9692	-
32108	31524	-
170118	158941	-
152206	136077	-
1305203	1307243	+
1341287	1343724	+
103440	103852	+
210071	217989	+
219734	224809	+
540634	539031	-
489513	464472	-
24976300	24963702	-
24948084	24935596	-
24909300	24895016	-
24883511	24875681	-
24862758	24851483	-
24831484	24823888	-
33880574	33874472	-
33846022	33837428	-
33834835	33824883	-
33816372	33806213	-
33800618	33783445	-
33768057	33770309	+
675500	675785	+
33609765	33610289	+
25088529	25097647	+
25103880	25119551	+
25060499	25073427	+
46054893	46057276	+
46087414	46089928	+
73300188	73307534	+
73309974	73315068	+
89596890	89604430	+
31710456	31727197	+
56986	71271	+
15320	2628	-

8027	406	-
38475	27113	-
87342	86694	-
1532	16762	+
1391542	1385424	-
1356954	1348386	-
1345755	1335759	-
1327289	1316963	-
1311397	1294237	-
1278860	1281112	+
1004151	991347	-
976231	968448	-
962214	946296	-
1086544	1088911	+
1116966	1119482	+
127733	122000	-
116726	109453	-
106794	101635	-
19343	11806	-
1308044	1291382	-
25013859	25001642	-
24946521	24932542	-
24922967	24917485	-
24906447	24895654	-
24879381	24871712	-
97264259	97264927	+
15414895	15414358	-
86103902	86112509	+
86140140	86148981	+
86170117	86171539	+
86179804	86193370	+
86199012	86221997	+
86237856	86234710	-
997316	997545	+
32268417	32268913	+
23878095	23890371	+
23908434	23916771	+
23924028	23940757	+
76460329	76459604	-
44692848	44694961	+
44720826	44722766	+
70578936	70591655	+
70594129	70598917	+
85489219	85496611	+
30292032	30307578	+
629020	641168	+
697691	713107	+
723371	731434	+

741291	751741	+
775760	783351	+
793985	793434	-
66278	57644	-
29814	18626	-
18322	5367	-
1041069	1027793	-
1022556	986950	-
984316	987440	+
451428	462614	+
480039	488584	+
1245158	1245883	-
48264	46323	-
21376	19212	-
316943	321939	+
324386	329212	+
2719971	2712704	-
216701	201134	-
305583	296718	-
236389	230552	-
219883	213365	-
967442	968058	+
214171	209427	-
193067	195133	+
2224242	2223587	-
4821099	4820464	-
43548	44174	+
436159	446635	+
453828	457464	+
312068	309571	-
265722	263400	-
3123681	3124093	+
2013316	2013868	+
1558457	1559002	+
1376235	1376717	+
1332058	1332685	+
55326	62180	+
408631	419968	+
124401	129489	+
513792	503015	-
1451958	1458679	+
1470524	1477605	+
413394	414012	+
1389559	1406160	+
1248547	1248037	-
1282626	1280044	-
1261217	1263286	+
98233	98891	+

1407507	1406847	-
17869	17209	-
206327	205672	-
1621489	1620829	-
2930837	2930193	-
28119	37830	+
46253	63251	+
2812747	2810251	-
2781737	2779406	-
755632	756169	+
673946	674463	+
513589	514208	+
327592	327110	-
1371481	1381075	+
1383246	1390117	+
1578537	1572123	-
6612860	6630880	+
20442169	20427198	-
20372165	20367121	-
20357539	20349412	-
20333798	20320167	-
24100800	24101455	+
5363648	5364301	+
42207758	42203033	-
42188462	42191175	+
47893887	47893236	-
16989142	16988940	-
39227313	39227796	+
16569762	16579141	+
16587096	16604844	+
10616521	10617153	+
47025688	47025056	-
6515364	6514732	-
36339801	36340427	+
22975143	22975727	+
29944972	29944340	-
35710423	35709816	-
21291142	21291645	+
58672643	58672024	-
46015846	46016470	+
45285751	45285131	-
18743649	18743097	-
28545597	28535133	-
28532552	28526421	-
50136796	50143047	+
17189253	17172560	-
50613793	50600283	-
50518164	50508694	-

50487044	50477334	-
50466964	50457443	-
50425513	50435062	+
50412744	50415980	+
50404376	50397989	-
50357534	50347226	-
45491	35918	-
3515	13344	+
58459595	58455084	-
58445589	58441277	-
58434064	58429383	-
58419880	58422750	+
30601798	30601567	-
36840484	36840937	+
26116965	26105075	-
26096903	26079859	-
27260358	27262826	+
27292142	27294513	+
1495423	1486925	-
1484501	1481257	-
22681192	22687981	+
44363667	44382892	+
3515522	3499246	-
3450904	3439723	-
3424106	3417889	-
3409828	3395147	-
9173	7116	-
7704	1450	-
3351868	3346651	-
3343931	3339573	-
3322537	3324482	+
14597569	14597346	-
1678680	1678914	+
2221889	2230767	+
2236096	2245923	+
493751	491544	-
12198862	12198238	-
3161039	3161638	+
1233995	1234515	+
301132	300608	-
760261	759724	-
6114	1824	-
12582	10517	-
17424	15467	-
62194	59921	-
557470	563918	+
13746368	13727246	-
78232240	78242534	+

78258712	78269017	+
78294806	78305417	+
78341916	78331127	-
21298085	21285660	-
21180470	21187381	+
108044660	108041280	-
107986103	107982094	-
107968828	107964249	-
107953113	107950823	-
107943311	107939273	-
107931654	107926933	-
107911243	107906818	-
107895975	107906809	+
116633456	116632968	-
31820406	31819907	-
47855250	47855295	+
47872024	47886158	+
44858	43569	-
27707	26368	-
23413	21490	-
4058843	4057554	-
4041692	4040353	-
4037398	4035475	-
192074452	192073822	-
75798490	75784039	-
75780958	75774874	-
75831550	75834665	+
87157891	87163808	+
65120088	65138169	-
209282	201602	-
3646021	3645549	-
3302222	3310199	+
3351836	3375651	+
5392582	5396955	+
5400979	5405232	+
5429695	5434959	+
5461203	5463095	+
5498546	5495564	-
26837350	26837945	+
13304464	13304825	+
4583747	4583256	-
2956455	2955963	-
5701224	5700725	-
3342650	3351581	+
3357150	3373889	+
19010532	19008726	-
18960795	18958523	-
10282951	10270398	-

10268675	10264744	-
24583078	24575540	-
22139937	22115107	-
110637850	110656980	+
110667282	110719976	+
110731912	110822141	+
110834372	110845583	+
54482289	54477898	-
54510433	54506642	-
54571514	54575857	+
54598612	54602981	+
54636462	54641237	+
54666120	54670979	+
54745605	54748006	+
54773319	54769878	-
14217645	14210212	-
14197023	14171479	-
2012884	2010206	-
785550	791586	+
795589	800223	+
53390822	53398050	+
445038	475256	+
3638375	3612779	-
3496309	3484635	-
7165220	7165849	+
208445	212830	+
116426	112032	-
91890	87211	-
67311	61995	-
47396	49625	+
7848	12162	+
1392365	1336976	-
1336581	1332005	-
30764473	30774766	+
30800234	30819706	+
3828061	3825476	-
2105202	2113238	+
2115024	2121536	+
30222713	30230037	+
3164438	3192486	+

Gene name
GSTA1 PARTIAL
GSTAP
GSTA2 PARTIAL
GSTM1
GSTM2 PARTIAL
GSTMP1
GSTMP2
GSTMP3
GSTO1 PARTIAL
GSTO2
GSTP1 PARTIAL
GSTP2 PARTIAL
GSTPP
GSTT1P
GSTT2
GSTZ1 PARTIAL
HPGDS
GSTA1
GSTAP
GSTA2
GSTMP1
GSTM1
GSTMP2
GSTMP3
GSTMP4
GSTO1
GSTO2
GSTP1
GSTP2
GSTPP
GSTT1P
GSTT2
GSTZ1
HPGDS
GSTA1
GSTAP
GSTA2
GSTMP1
GSTM1
GSTMP2
GSTMP3
GSTMP4
GSTO1
GSTO2
GSTP1 PARTIAL
GSTP2 PARTIAL
GSTPP

GSTT1P
GSTT2
GSTZ1 PARTIAL
HPGDS
GSTA1
GSTAP
GSTA2
GSTMP1
GSTM1
GSTMP2
GSTMP3
GSTMP4
GSTO1
GSTO2
GSTP1
GSTP2
GSTPP
GSTT1P
GSTT2
GSTZ1
HPGDS
GSTA1
GSTAP
GSTA2
GSTMP1
GSTM1
GSTMP2
GSTMP3
GSTMP4
GSTO1
GSTO2
GSTP1
GSTP2
GSTPP
GSTT1 PARTIAL
GSTT2
GSTZ1
HPGDS
GSTA1
GSTAP
GSTA2
GSTM1
GSTM2 PARTIAL
GSTMP1
GSTMP2
GSTMP3
GSTO1
GSTO2 PARTIAL

GSTP1
GSTP2
GSTPP
GSTT1P
GSTT2 PARTIAL
GSTZ1
HPGDS
GSTA1
GSTA2
GSTM1
GSTM2 PARTIAL
GSTMP1
GSTMP2
GSTMP3
GSTO1
GSTO2
GSTP1 PARTIAL
GSTP2 PARTIAL
GSTPP
GSTT1P
GSTT2 PARTIAL
GSTZ1 PARTIAL
HPGDS
GSTA1
GSTAP1
GSTA2
GSTA3
GSTAP2
GSTA4
GSTM1
GSTM2
GSTM3
GSTM4
GSTM5
GSTM6
GSTMP1
GSTMP2
GSTO1
GSTO2
GSTO3
GSTP1
GSTP2
GSTT1
GSTT2
GSTZ1
HPGDS
GSTA1
GSTA2

GSTA3
GSTA4
GSTAP1
GSTA5
GSTM1
GSTM2
GSTM3
GSTM4
GSTM5
GSTM6
GSTO1
GSTO2
GSTO3
GSTP1
GSTP2
GSTT1
GSTT2
GSTT3
GSTZ1
HPGDS
GSTA1
GSTA2
GSTA3
GSTA4
GSTA5
GSTA6
GSTAP
GSTM1
GSTM2
GSTM3
GSTM4
GSTM5
GSTM6
GSTMP1
GSTMP2
GSTO1
GSTO2
GSTO3
GSTO4
GSTP1
GSTP2
GSTT1
GSTT2
GSTZ1
HPGDS
GSTA1
GSTA2
GSTA3

GSTA4
GSTA5
GSTAP
GSTM1
GSTM2
GSTM3
GSTM4
GSTM5
GSTM6
GSTO1
GSTO2
GSTO3
GSTP1
GSTP2
GSTT1
GSTT2
GSTZ1
HPGDS
GSTA1 PARTIAL
GSTA2
GSTA3
GSTAP
GSTM1
GSTM2
GSTM3
GSTMP1
GSTMP2
GSTO1
GSTO2
GSTP1
GSTP2
GSTP3 PARTIAL
GSTP4 PARTIAL
GSTPP
GSTP5
GSTP6
GSTT1
GSTT2
GSTZ1
HPGDS
GSTA1
GSTA2
GSTA3
GSTA4
GSTAP
GSTM1
GSTM2
GSTM3

GSTM4
GSTM5
GSTM6
GSTMP1
GSTMP2
GSTO1
GSTO2
gstp1
gstp2
gstp3
gstpp
gstp4
gstp5
GSTT1
GSTT2
GSTZ1
HPGDS
GSTA1
GSTA2
GSTA3
GSTA4
GSTA5
GSTAP
GSTM1
GSTM2
GSTMP1
GSTMP2
GSTMP3
GSTO1
GSTO2
gstp1
gstp2
gstp3
gstp4
gstp5
GSTP6
GSTPP1
GSTPP2
GSTPP3
GSTPP4
GSTPP5
GSTPP6
GSTT1
GSTT2
GSTZ1
HPGDS
GSTA1
GSTA2

GSTA3
GSTA4
GSTA5
GSTA6
GSTAP
GSTA7
GSTA8
GSTA9
GSTM1
GSTM2
GSTM3
GSTM4
GSTMP1
GSTMP2
GSTO1
GSTO2
GSTP1
GSTP2
GSTT1
GSTT2
GSTZ1
HPGDS
GSTA1
GSTA2
GSTA3
GSTA4
GSTA5
GSTA6
GSTM1
GSTM2
GSTM3
GSTM4 PARTIAL
GSTMP
GSTO1
GSTO2
GSTP1
GSTPP1
GSTPP2
GSTPP3
GSTPP4
GSTPP5
GSTT1
GSTT2
GSTT3
GSTT4
GSTZ1
HPGDS
GSTA1

GSTA2
GSTA3
GSTA4
GSTAP
GSTA5
GSTM1
GSTM2
GSTM3
GSTMP1
GSTM4
GSTM5
GSTMP2
GSTMP3
GSTMP4
GSTM6
GSTO1
GSTO2
gstp1
gstp2
gstp3
gstp4
gstp5
gstp6
gstp7
GSTT1
GSTT2
GSTT3
gstz1
HPGDS
GSTA1
GSTA2
GSTAP
GSTA3
GSTM1
GSTM2
GSTM3
GSTM4
GSTM5
GSTM6
GSTM7
GSTM8
GSTM9
GSTM10
GSTO1
GSTOP
GSTP1
GSTP2
GSTT1

GSTT2
GSTZ1
HPGDS
GSTA1
GSTA2
GSTA3
GSTA4
GSTM1
GSTM2
GSTM3
GSTM4
GSTM5
GSTM6
GSTM7
GSTM8
GSTO1
GSTO2
GSTP1
GSTT1
GSTT2_PARTIAL
GSTZ1
HPGDS
GSTA1
GSTA2
GSTA3
GSTM1
GSTM2
GSTM3
GSTM4
GSTM5
GSTM6
GSTM7
GSTM8
GSTO1
GSTO2
GSTP1
GSTT1
GSTT2
GSTZ1
HPGDS

Table S2 Genomic information of species used in this study

Common name	Latin name	Sequencing technology	Sequence coverage (×)	Scaffold N50	NCBI Genomes version	RefSeq Assembly Accession	Genome resource	Assembly Completeness in BUSCO (%)
Bottlenose dolphin	*Tursiops truncatus*	Sanger, 454 FLX, Illumina HiSeq	30	116,287	Ttru_1.4	GCF_001922835.1	NCBI	92.9
Killer whale	*Orcinus orca*	Illumina HiSeq	200	12,735,091	Oorc_1.1	GCF_000331955.2	NCBI	99.3
Yangtze river dolphin	*Lipotes vexillifer*	Illumina HiSeq	115	2,419,148	Lipotes_vexillifer_v1	GCF_000442215.1	NCBI	98.8
Yangtze finless porpoise	*Neophocaena asiaeorientalis*	Illumina HiSeq	106	6,341,296	Neophocaena_asiaeorientalis_v1	GCF_003031525.1	NCBI	99.1
Sperm whale	*Physeter macrocephalus*	Illumina HiSeq	75	427,290	Physeter_macrocephalus_2.0.2	GCF_000472045.1	NCBI	99.5
Minke whale	*Balaena acutorostrata*	Illumina HiSeq	92	12,843,668	BalAcu1.0	GCF_000493695.1	NCBI	99.1
Bowhead whale	*Balaena mysticetus*	Illumina HiSeq	154	877,000	-	-	http://www.bowhead-whale.org	74.6
Cow	*Bos taurus*	Sanger, PacBio RS II	19	6,806,220	Btau_5.0.1	GCA_000003205.6	NCBI	99.5
Tibetan yak	*Bos mutus*	Illumina HiSeq,	130	1,407,960	BosGru_v2.0	GCF_000298355.1	NCBI	98.1

Sheep	*Ovis aries*	Illumina GA Illumina GAII, 454, PacBio RSII	166	100,009,711	Oar_v4.0	GCF_000298735.2	NCBI	99.7
Tibetan antelope	*Pantholops hodgsonii*	Illumina GA	67	2,772,860	PHO1.0	GCF_000400835.1	NCBI	98.6
Dog	*Canis lupus familiaris*	Sanger	7	45,876,610	CanFam3.1	GCF_000002285.3	NCBI	99.2
Weddell seal	*Leptonychotes weddellii*	Illumina HiSeq	82	904,031	LepWed1.0	GCF_000349705.1	NCBI	87.3
Pacific walrus	*Odobenus rosmarus divergens*	Illumina HiSeq	200	2,616,778	Oros_1.0	GCF_000321225.1	NCBI	99.6
Horse	*Equus caballus*	Sanger	6.8	46,749,900	EquCab2.0	GCF_000002305.2	NCBI	99.7
Microbat	*Myotis lucifugus*	Sanger	7	4,293,315	Myoluc2.0	GCF_000147115.1	NCBI	97.4
Mouse	*Mus musculus*	Sanger	-	52,589,046	GRCm38.p4	GCF_000001635.24	NCBI	99.9
Naked mole rat	*Heterocephalus glaber*	Illumina HiSeq	90	20,532,749	HetGla_female_1.0	GCF_000247695.1	NCBI	99.3
African bush elephant	*Loxodonta africana*	Sanger, ABI	7	46,401,353	Loxafr3.0	GCF_000001905.1	NCBI	98.7
Florida manatee	*Trichechus manatus latirostris*	Illumina HiSeq	150	14,442,683	TriManLat1.0	GCF_000243295.1	NCBI	98.1

Table S3 Amino acid sequence similarity (identity) between and within cytosolic GST subclass in seven cetaceans

HPGDS										
GSTA1	25.0%									
GSTA2	21.4%	56.0%								
GSTM1	25.5%	25.2%	20.0%							
GSTO1	17.2%	19.1%	17.2%	15.7%						
GSTO2	17.6%	16.6%	15.6%	15.3%	62.3%					
GSTP1	24.5%	30.7%	27.2%	31.1%	15.7%	20.8%				
GSTP2	23.8%	30.9%	27.9%	30.8%	18.9%	21.2%	82.0%			
GSTT1	15.4%	21.1%	17.8%	18.6%	16.6%	18.9%	21.5%	22.3%		
GSTZ1	19.9%	21.3%	21.3%	15.6%	21.2%	21.0%	21.1%	21.0%	25.5%	
	HPGDS	GSTA1	GSTA2	GSTM1	GSTO1	GSTO2	GSTP1	GSTP2	GSTT1	GSTZ1

Table S4 Positive selection detected in mammals by site models in PAML

Gene	Model	lnL[a]	2ΔlnL	*p*	Parameter	Positive selected sites[b]
GSTA1	M8	4353.569			ω =1.819	36-0.942, 37-0.870, 38-0.854, 45-0.871, 49-0.861, 96-0.952*, 100-0.845, 103-0.988*, 107-0.889, 112-0.972*, 121-0.853, 166-0.844, 208-0.831, 212-0.975*, 215-0.806, 222-0.805
	M8a	4360.512	6.943	0.008	ω =1	
GSTM1	M8	3852.094			ω =2.141	107-0.987*, 164-0.939, 204-0.917, 208-0.991**, 210-0.929
	M8a	3857.390	10.592	0.001	ω =1	
GSTO1	M8	4455.405			ω =1.942	69-0.949, 125-0.907, 127-0.963*, 128-0.995**, 133-0.958*, 147-0.958*, 211-0.930, 216-0.927, 217-0.831,219-0.866, 222-0.853, 223-0.841, 226-0.835, 227-0.884
	M8a	4447.314	16.182	<0.001	ω =2.809	
GSTO2	M8	3316.001			ω =5.410	14-0.989*, 43-0.921, 141-0.945
	M8a	3319.950	7.898	0.005	ω =1	
GSTP1	M8	3617.554			ω =4.046	12-0.837, 43-0.995**
	M8a	3621.743	8.378	0.004	ω =1	
GSTP2	M8	2700.044			ω =3.412	11-0.826, 12 -0.998**, 111-0.969*, 118-0.948, 205-0.993**, 210-0.999**
	M8a	2712.339	24.590	<0.001	ω =1	
GSTT2	M8	3755.867			ω =12.498	237-0.991**
	M8a	3761.854	11.974	<0.001	ω =1	
GSTZ1	M8	3360.687			ω =3.081	99-0.893, 137-0.947, 193-0.889
	M8a	3349.357	11.330	<0.001	ω =1	

Note: [a] lnL is the log-likelihood score.

[b] Codons with posterior probabilities (pp) >80% in the BEB analyses. * pp >95%, ** pp >99%.

Table S5 Identification of the domain location of each positively selected sites.

Gene	Positively selected sites[a]	Residue annotation from InterProScan	Functional Information
GSTA1	36	36	C-terminal domain interface
	49	49	GSH binding site
	96	96	C-terminal dimer interface (polypeptide binding site)
	100	100	C-terminal dimer interface (polypeptide binding site)
	103	103	C-terminal dimer interface (polypeptide binding site)
	208	208	substrate binding pocket
	215	215	substrate binding pocket
	222	222	substrate binding pocket
GSTM1	107	107	substrate binding pocket
GSTO1	23	29	C-terminal domain interface
	47	44	C-terminal domain interface
	69	71	GSH binding site
	125	125	substrate binding pocket
	127	128	substrate binding pocket
	128	128	substrate binding pocket
	216	215	N-terminal domain interface
	226	225	N-terminal domain interface
	227	229	N-terminal domain interface
GSTO2	43	43	C-terminal domain interface
GSTP1	12	12	C-terminal domain interface
	40	39	GSH binding site
GSTP2	11	11	C-terminal domain interface
	12	12	C-terminal domain interface
	76	75	dimer interface
GSTT2	25	23	C-terminal domain interface
GSTZ1	71	71	putative dimer interface
	125	123	dimer interface
	137	135	dimer interface

Note: [a] Positive selected sites are located or close to the residue annotation from InterProScan.

Table S6 Results from Clade model C (CmC) test for divergent partitioned by habitats.

Gene	Model and Partition[a]	np[b]	-ln*L*[c]	Parameters		Null	LRT	df	*p*
				ω0/ω1	ω2/ωd				
GSTA1	M2a_rel	45	4358.113	0.117/1	2.390				
	CmC: Ceta	46	4360.012	0.040/1	0.276	M2a_rel	3.798		0.050
					Ceta: 0.371				
	CmC: Pinn	46	4341.526	0.088/1	0.091	M2a_rel	33.174		<0.001
					Pinn: 5.682				
	CmC: Sire	46	4360.126	0.046/1	0.302	M2a_rel	4.026		0.044
					Sire: 0.322				
	CmC: Mari	46	4349.410	0.022/1	0.151	M2a_rel	17.406	1	<0.001
					Mari: 0.667				
	CmC: Ceta/Pinn+Sire	47	4347.092	0.075/1	0.070	M2a_rel	22.042	2	<0.001
					Ceta/Pinn: 1.416	Mari	4.636	1	0.031
					Sire: 0.0001				
	CmC: Ceta+Pinn+Sire	48	4339.168	0.093/1	0.039	M2a_rel	37.89	3	<0.001
					Ceta: 0.472	Mari	20.484	2	<0.001
					Pinn: 7.115	Ceta/Pinn+Sire	15.848	1	<0.001
					Sire: 0.0001				

GSTA4	M2a_rel	43	2756.675	0.021/1	0.359				
	CmC: Ceta	44	2752.474	0.024/1	0.402	M2a_rel	8.402	1	0.004
					Ceta: 0.039				
	CmC: Pinn	44	2753.047	0.001/1	0.032	M2a_rel	7.256	1	0.007
					Pinn: 0.000				
	CmC: Mari	44	2748.316	0.020/1	0. 434	M2a_rel	16.718	1	<0.001
					Mari: 0.035				
	CmC: Ceta/Pinn+Sire	45	2748.076	0.019/1	0.425	M2a_rel	17.198	2	<0.001
					Ceta/Pinn: 0.026	Mari	0.24	1	0.624
					Sire: 0.103				
	CmC: Ceta+Pinn+Sire	46	2747.692	0.014/1	0.411	M2a_rel	17.966	3	<0.001
					Ceta: 0.044	Mari	1.248	2	0.535
					Pinn: 0.0001	Ceta/Pinn+Sire	0.768	1	0.381
					Sire: 0.106				
GSTM1	M2a_rel	37	3856.569	0.069/1	0.407				
	CmC: Ceta	38	3853.307	0.063/1	0.340	M2a_rel	6.524	1	0.011
					Ceta: 1.071				
	CmC: Pinn	38	3849.571	0.062/1	0.256	M2a_rel	13.996	1	<0.001
					Pinn: 1.028				

	CmC: Mari	38	3847.817	0.058/1	0.181	M2a_rel	15.704	1	<0.001
					Mari: 0.824				
	CmC: Ceta/Pinn+Sire	39	3847.717	0.058/1	0.188	M2a_rel	17.704	2	<0.001
					Ceta/Pinn: 0.895	Mari	0.2	1	0.655
					Sire: 0.691				
	CmC: Ceta+Pinn+Sire	40	3847.459	0.060/1	0.181	M2a_rel	18.22	3	<0.001
					Ceta: 0.755	Mari	0.716	2	0.699
					Pinn: 1.143	Ceta/Pinn+Sire	0.516	1	0.773
					Sire: 0.718				
GSTM3	M2a_rel	41	2774.645	0.025/1	0.407				
	CmC: Ceta	42	2772.569	0.043/1	0.235	M2a_rel	4.152	1	0.042
					Ceta: 0.661				
	CmC: Pinn	42	2761.181	0.000/1	0.132	M2a_rel	26.928	1	<0.001
					Pinn: 0.824				
	CmC: Mari	42	2758.621	0.000/1	0.104	M2a_rel	32.048	1	<0.001
					Mari: 0.566				
	CmC: Ceta+Pinn+Sire	43	2757.128	0.000/1	0.106	M2a_rel	35.034	2	<0.001
					Ceta: 0.368	Mari	2.986	1	0.083
					Pinn: 0.804				

					Sire: -				
GSTT1	M2a_rel	33	3410.240	0.026/1	0.268				
	CmC: Ceta	34	3406.695	0.022/1	0.248	M2a_rel	7.090	1	0.008
					Ceta: 4.118				
	CmC: Pinn	34	3408.060	0.014/1	0.232	M2a_rel	4.360	1	0.037
					Pinn: 0.597				
	CmC: Mari	34	3404.419	0.015/1	0.221	M2a_rel	11.642	1	<0.001
					Mari: 0.727				
	CmC: Ceta/Pinn+Sire	35	3404.414	0.015/1	0.221	M2a_rel	11.652	2	0.002
					Ceta/Pinn: 0.740	Mari	0.010	1	0.920
					Sire: 0.687				
	CmC: Ceta+Pinn+Sire	36	3402.884	0.014/1	0.214	M2a_rel	14.712	3	<0.001
					Ceta: 4.122	Mari	3.07	2	0.080
					Pinn: 0.533	Ceta/Pinn+Sire	3.06	1	0.080
					Sire: 0.724				

Note:

[a] Partitions for diet, habitat and living habitats are explained in figure S1

[b] np: number of parameters

[c] ln L: In likehood

[d] Mari: Marine mammals; Ceta: Cetaceans; Pinn: Pinnipeds; Sire: Sirenians; Ceta/Pinni: Cetaceans/Pinnipeds

Table S7 Likelihood ratio tests of various models on the selective pressures on pseudogenes in the cytosolic GST subclass in cetacean lineages

Genes	Models	ω	-lnL	np	Models compared	2Δ(lnL)	p-value
GSTAP	All branches have one ω (A)	ω = 0.262	1614.650	27			
	All branches have one ω = 1 (B)	ω = 1	1640.199	26	B vs. A	51.098	<0.001
	Cetacean branches with pseudogenes have ω2, Cetacean branches with intact genes have ω1 (C)	ω1 = 0.022 ω2 = 0.331	1607.909	28	A vs. C	13.482	0.001
	Cetacean branches with pseudogenes have ω2 = 1, Cetacean branches with intact genes have ω1 (D)	ω1 = 0.023 ω2 = 1	1622.859	27	D vs. C	29.900	<0.001
	Each branch has its own ω (E)	Variable ω	1598.614	51	C vs. E	18.590	0.725
GSTMP1	All branches have one ω (A)	ω = 0.645	2731.702	27			
	All branches have one ω = 1 (B)	ω = 1	2737.880	26	B vs. A	12.356	<0.001
	Cetacean branches with pseudogenes have ω2, Cetacean branches with intact genes have ω1 (C)	ω1 = 0.511 ω2 = 0.667	2731.410	28	A vs. C	0.584	0.445
	Cetacean branches with pseudogenes have ω2 = 1, Cetacean branches with intact genes have ω1 (D)	ω1 = 0.524 ω2 = 1	2736.020	27	D vs. C	9.220	0.002
	Each branch has its own ω (E)	Variable ω	2711.784	51	C vs. E	39.252	0.019
GSTMP2	All branches have one ω (A)	ω = 1.116	1929.245	27			
	All branches have one ω = 1 (B)	ω = 1	1929.438	26	B vs. A	0.386	0.534
	Cetacean branches with pseudogenes have ω2, Cetacean branches with intact genes have ω1 (C)	ω1 = 0.635 ω2 = 1.355	1927.385	28	A vs. C	3.720	0.054

	Cetacean branches with pseudogenes have ω2 = 1, Cetacean branches with intact genes have ω1 (D)	ω1 = 0.627 ω2 = 1	1928.479	27	D vs. C	2.188	0.139
	Each branch has its own ω (E)	Variable ω	1908.534	51	C vs. E	37.702	0.027
GSTMP3	All branches have one ω (A)	ω = 0.724	1615.475	25			
	All branches have one ω = 1 (B)	ω = 1	1616.706	24	B vs. A	2.462	0.117
	Cetacean branches with pseudogenes have ω2, Cetacean branches with intact genes have ω1 (C)	ω1 = 0.502 ω2 = 0.868	1614.572	26	A vs. C	1.806	0.179
	Cetacean branches with pseudogenes have ω2 = 1, Cetacean branches with intact genes have ω1 (D)	ω1 = 0.507 ω2 = 1	1614.734	25	D vs. C	0.324	0.569
	Each branch has its own ω (E)	Variable ω	1605.964	47	C vs. E	17.216	0.751
GSTMP4	All branches have one ω (A)	ω = 0.348	2113.165	21			
	All branches have one ω = 1 (B)	ω = 1	2137.373	20	B vs. A	48.416	<0.001
	Cetacean branches with pseudogenes have ω2, Cetacean branches with intact genes have ω1 (C)	ω1 = 0.411 ω2 = 0.335	2113.009	22	A vs. C	0.312	0.576
	Cetacean branches with pseudogenes have ω2 = 1, Cetacean branches with intact genes have ω1 (D)	ω1 = 0.446 ω2 = 1	2134.604	21	D vs. C	43.190	<0.001
	Each branch has its own ω (E)	Variable ω	2097.915	39	C vs. E	30.188	0.036
GSTPP	All branches have one ω (A)	ω = 0.331	2676.809	28			
	All branches have one ω = 1 (B)	ω = 1	2718.359	27	B vs. A	83.100	<0.001
	Cetacean branches with pseudogenes have ω2, Cetacean branches with intact genes have ω1 (C)	ω1 = 0.222 ω2 = 0.431	2673.058	29	A vs. C	7.502	0.006

	Cetacean branches with pseudogenes have ω2 = 1, Cetacean branches with intact genes have ω1 (D)	ω1 = 0.216 ω2 = 1	2686.023	28	D vs. C	25.930	<0.001
	Each branch has its own ω (E)	Variable ω	2651.512	53	C vs. E	43.092	0.014
GSTTP	All branches have one ω (A)	ω = 0.252	1995.490	26			
	All branches have one ω = 1 (B)	ω = 1	2034.266	25	B vs. A	77.552	<0.001
	Cetacean branches with pseudogenes have ω2, Cetacean branches with intact genes have ω1 (C)	ω1 = 0.094 ω2 = 0.852	1981.855	27	A vs. C	27.270	<0.001
	Cetacean branches with pseudogenes have ω2 = 1, Cetacean branches with intact genes have ω1 (D)	ω1 = 0.095 ω2 = 1	1981.961	26	D vs. C	0.212	0.645
	Each branch has its own ω (E)	Variable ω	1974.393	49	C vs. E	14.924	0.897