Contraction of the ROS Scavenging Enzyme Glutathione S-Transferase Gene Family in Cetaceans

Ran Tian,*,† Inge Seim,†,‡ Wenhua Ren,* Shixia Xu,*,1 and Guang Yang*,1

*Jiangsu Key Laboratory for Biodiversity and Biotechnology and †Integrative Biology Laboratory, College of Life Sciences, Nanjing Normal University, Nanjing, Jiangsu, 210046, China, and ‡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

ORCID IDs: 0000-0002-4564-1886 (R.T.); 0000-0001-8594-7217 (I.S.); 0000-0001-6285-6937 (G.Y.)

ABSTRACT Cetaceans are a group of marine mammals whose ancestors were adapted 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.

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 min (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 min (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). Similarly, 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 cetaceans 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 share 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 (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 species 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 cutoff to 10 and hits with the highest score, lowest *E*-value, and a length of ≥ 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 predicted GST sequences were verified by BLAST against their respective species' genome 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, and 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 facilitated 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). 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 10% burn-in.

Adaptive evolution analyses

To evaluate the positive selection of 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 criterion 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 (Woolley *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 (Jones *et al.* 2014). We also used PyMOL v2.3 (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 or 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, 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.

Data availability

The authors affirm that all data necessary for confirming the conclusions of this article are represented fully within the article and its tables and figures. Supplemental material available at FigShare: https://doi.org/10.25387/g3.7403828.

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

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.

(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 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)
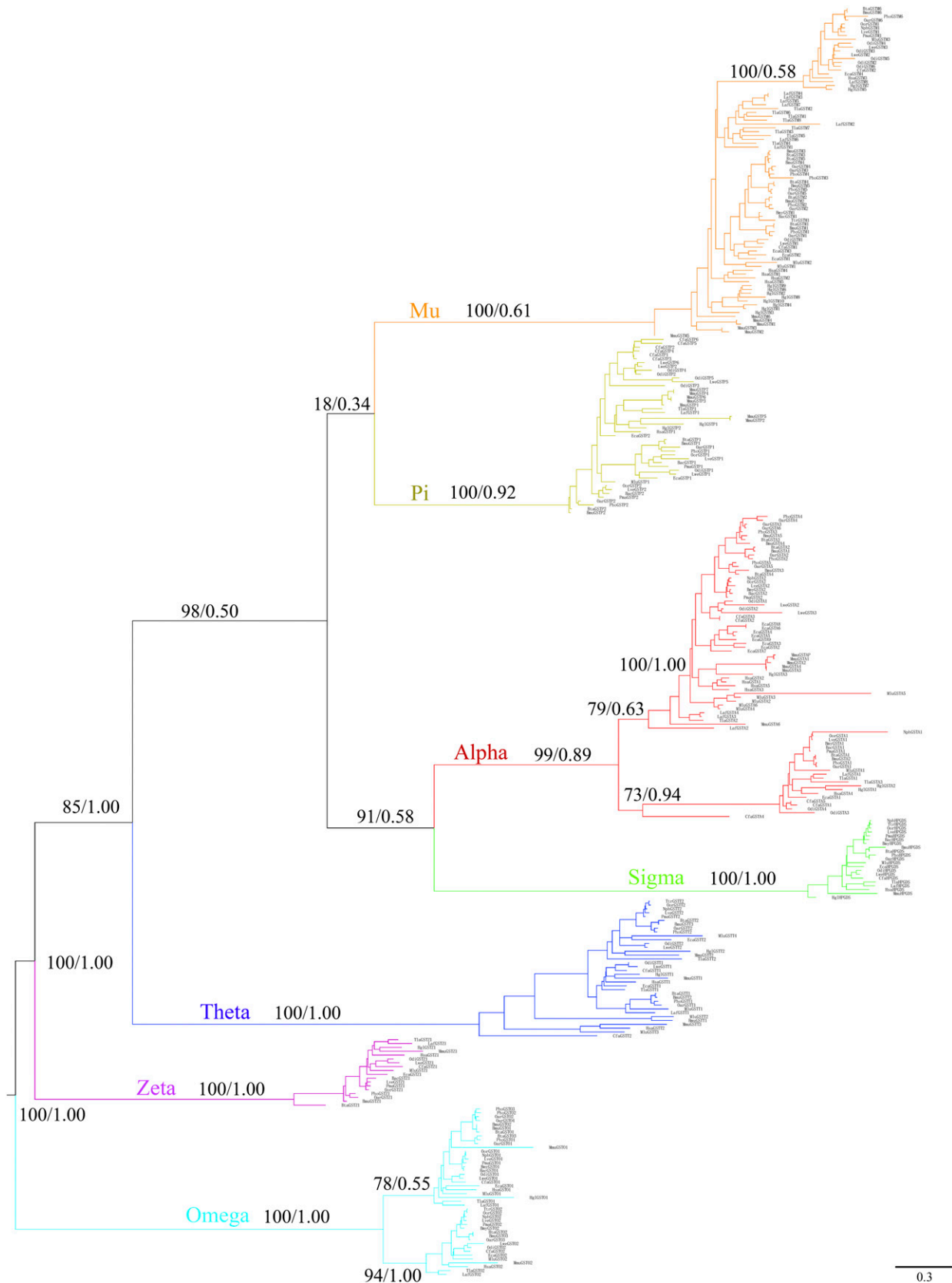
Figure 1 Phylogenetic tree of GST gene family in mammals. Maximum likelihood and Bayesian phylograms describing phylogenetic relationships among 333 intact (complete coding sequences) mammalian GSTs (21 species). Numbers on nodes correspond to maximum likelihood bootstrap support values and Bayesian posterior probabilities. The GST subclasses are indicated by different colors: alpha (red), mu (orange), pi (turquoise), omega (azure), sigma (green), zeta (purple), and theta (blue).
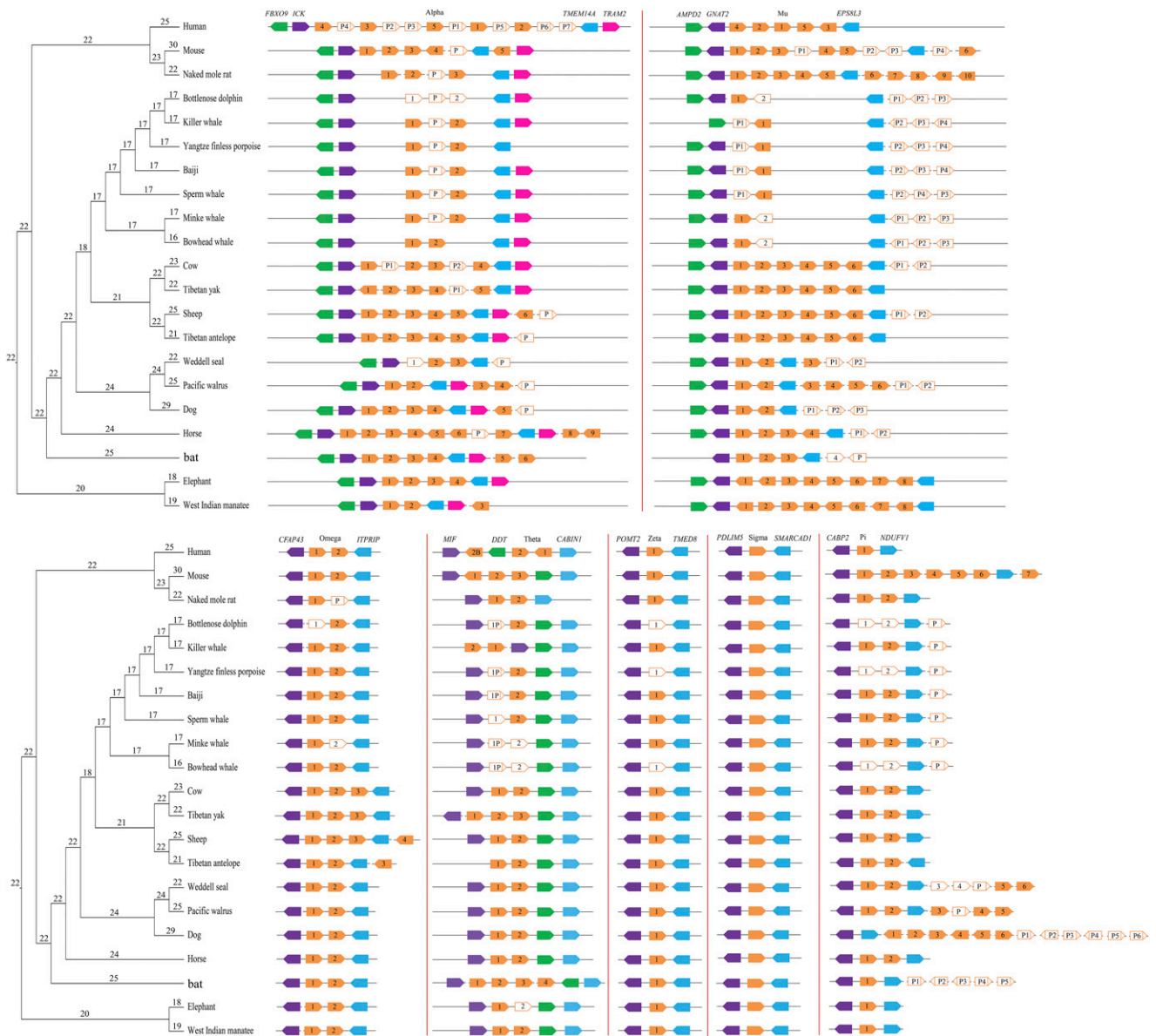
Figure 2 Genomic organization of GST genes in 21 mammalian species. The arrowed boxes represent genes and directions of transcription. GST genes are shown in orange, while flanking genes are indicated in green, purple, blue, and pink. Filled boxes: intact genes; empty boxes: partial gene, and empty boxes 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.

(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 (phylA-NOVA). 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 (phylA-NOVA; $F = 27.557$, $P = 0.011$).

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
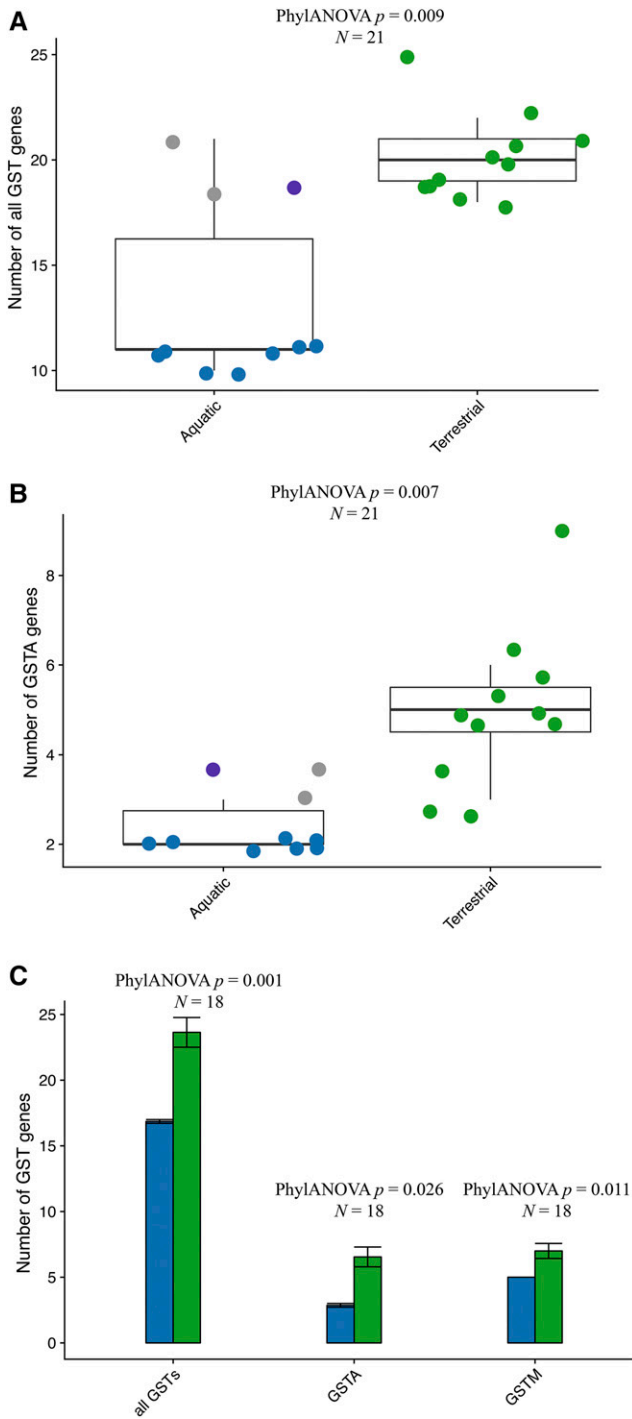
Figure 3 Differences in cytosolic GST genes between mammals inhabing aquatic and terrestrial habitats. (A) All GSTs genes combined, comparing aquatic and terrestrial mammals. Green dots indicate terrestrial species; blue dots, cetacean species; gray dots, pinnipeds; purple dots, sirenians. (B) alpha-class 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± SEM. 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.

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 (Woolley *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 may 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

Gene	Site Position	PAML M8[a]	Datamonkey SLAC[b]	FEL[c]	REL[d]	TreeSAAP 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: α_c: Power to be C-term., α-helix; α_n: Power to be in the N-terminal of an α-helix; B_r: Buriedness; C_a: Helical contact energy; E_l: Long-range non-bonded energy; E_{sm}: Short and medium range non-bonded energy; E_t: Total non-bonding Energy; F: Mean r.m.s. fluctuation displacement; h: Hydropathy; H_{nc}: Normal consensus hydrophobicity; H_p: Surrounding hydrophobicity; H_t: Thermodynamic transfer hydrophobicity; K^0: Compressibility; μ: Refractive index; M_v: Molecular volume; M_w: Molecular weight; N_s: Average number of surrounding residues; P_α: α- helical tendencies; P_β: β-structure tendencies; P_c: Coil tendencies; P: Turn tendencies; p: Polarity; pH_i: Isoelectric point; pK': Equilibrium Constant of ionization for COOH; P_r: Polar requirement; R_a: Solvent accessible reduction ratio; R_F: Chromatographic index; V^0: Partial specific volume.

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 of human *GSTM1* (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

■ **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	

(continued)

Genes	Branch-site models[a]	-lnL[b]	2ΔlnL	p-value	ω values	Positively selected site[c]
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	

[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%.

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 repertoires 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 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, while 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 are 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 toward 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 *et al.* 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 hypothesize that the cetacean homologs, in particular *GSTA1* which was intact in all seven cetaceans, are retained because of their essential role in protecting against oxidative stress in both the cytosol and mitochondria. That is, complete loss of these genes is not tolerated. This hypothesis awaits experimental proof, including an assessment of the tissue/organelle location and function of cetacean GSTs. 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 contributes 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 (Ilic *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 ceteacans adapt to an oxygen-poor aquatic environment. An ever-increasing amount of genomic research data and associated tools is likely to address this important research question.

LITERATURE CITED

Apel, K., and H. Hirt, 2004 Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu. Rev. Plant Biol. 55: 373–399. https://doi.org/10.1146/annurev.arplant.55.031903.141701

Armstrong, R. N., 1997 Structure, catalytic mechanism, and evolution of the glutathione transferases. Chem. Res. Toxicol. 10: 2–18. https://doi.org/10.1021/tx960072x

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: 7698–7704. https://doi.org/10.1021/bi900895b

Bammler, T. K., C. A. Smith, and C. R. Wolf, 1994 Isolation and characterization of two mouse Pi-class glutathione S-transferase genes. Biochem. J. 298: 385–390. https://doi.org/10.1042/bj2980385

Benjamini, Y., and Y. Hochberg, 1995 Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. A Stat. Soc. 57: 289–300.

Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, J. Ostell *et al.*, 2018 GenBank. Nucleic Acids Res. 46: D41–D47. https://doi.org/10.1093/nar/gkx1094

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. J. Mol. Evol. 59: 121–132. https://doi.org/10.1007/s00239-004-2597-8

Birben, E., U. M. Sahiner, C. Sackesen, S. Erzurum, and O. Kalayci, 2012 Oxidative stress and antioxidant defense. World Allergy Organ. J. 5: 9–19. https://doi.org/10.1097/WOX.0b013e3182439613

Birney, E., M. Clamp, and R. Durbin, 2004 GeneWise and genomewise. Genome Res. 14: 988–995. https://doi.org/10.1101/gr.1865504

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. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 158: 438–443. https://doi.org/10.1016/j.cbpa.2010.11.029

Cheng, J., C. Liu, W. J. Koopman, and J. D. Mountz, 1995 Characterization of human Fas gene. Exon/intron organization and promoter region. J. Immunol. 154: 1239–1245.

Coles, B. F., and F. F. Kadlubar, 2005 Human alpha class glutathione S-transferases: genetic polymorphism, expression, and susceptibility to disease. Methods Enzymol. 401: 9–42. https://doi.org/10.1016/S0076-6879(05)01002-5

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: 717–727. https://doi.org/10.1093/carcin/bgx048

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 Evol. Biol. 10: 281. https://doi.org/10.1186/1471-2148-10-281

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: 1269–1271. https://doi.org/10.1093/bioinformatics/btl097

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 Biol. 17: 225. https://doi.org/10.1186/s12870-017-1179-z

Dröge, W., 2002 Free radicals in the physiological control of cell function. Physiol. Rev. 82: 47–95. https://doi.org/10.1152/physrev.00018.2001

Edgar, R. C., 2004 MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113. https://doi.org/10.1186/1471-2105-5-113

Elsner, R., and B. Gooden, 1983 *Diving and asphyxia: a comparative study of animals and man*, Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511735714

Frova, C., 2006 Glutathione transferases in the genomics era: new insights and perspectives. Biomol. Eng. 23: 149–169. https://doi.org/10.1016/j.bioeng.2006.05.020

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. Biochem. Pharmacol. 71: 1619–1628. https://doi.org/10.1016/j.bcp.2006.02.018

García-Castañeda, O., R. Gaxiola-Robles, S. Kanatous, and T. Zenteno-Savín, 2017 Circulating glutathione concentrations in marine, semiaquatic, and terrestrial mammals. Mar. Mamm. Sci. 33: 738–747.

Garland, T., A. W. Dickerman, C. M. Janis, and J. A. Jones, 1993 Phylogenetic Analysis of Covariance by Computer Simulation. Syst. Biol. 42: 265–292. https://doi.org/10.1093/sysbio/42.3.265

Gouy, M., S. Guindon, and O. Gascuel, 2009 SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27: 221–224. https://doi.org/10.1093/molbev/msp259

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

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. Proc. Natl. Acad. Sci. USA 101: 13163–13167. https://doi.org/10.1073/pnas.0403045101

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. Proc. Natl. Acad. Sci. USA 95: 5275–5280. https://doi.org/10.1073/pnas.95.9.5275

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 Radic. Biol. Med. 89: 1122–1143. https://doi.org/10.1016/j.freeradbiomed.2015.07.156

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. Mol. Biol. Evol. 35: 1390–1406. https://doi.org/10.1093/molbev/msy028

Ilic, Z., D. Crawford, D. Vakharia, P. A. Egner, and S. Sell, 2010 Glutathione-S-transferase A3 knockout mice are sensitive to acute cytotoxic and genotoxic effects of aflatoxin B1. Toxicol. Appl. Pharmacol. 242: 241–246 (erratum: Toxicol. Appl. Pharmacol. 245: 280). https://doi.org/10.1016/j.taap.2009.10.008

Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Li et al., 2014 InterProScan 5: genome-scale protein function classification. Bioinformatics 30: 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

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. Res. Vet. Sci. 86: 216–222. https://doi.org/10.1016/j.rvsc.2008.07.006

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. PhD Thesis, available at https://repositorio-aberto.up.pt

Kondrashov, F. A., 2012 Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc. Biol. Sci. 279: 5048–5057. https://doi.org/10.1098/rspb.2012.1108

Kumar, S., G. Stecher, M. Suleski, and S. B. Hedges, 2017 TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol. Biol. Evol. 34: 1812–1819. https://doi.org/10.1093/molbev/msx116

Li, C., and R. M. Jackson, 2002 Reactive species mechanisms of cellular hypoxia-reoxygenation injury. Am. J. Physiol. Cell Physiol. 282: C227–C241. https://doi.org/10.1152/ajpcell.00112.2001

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. Mol. Biol. Evol. 32: 2844–2859. https://doi.org/10.1093/molbev/msv156

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: 1363–1375. https://doi.org/10.1534/genetics.107.075838

Lynch, M., and J. S. Conery, 2000 The evolutionary fate and consequences of duplicate genes. Science 290: 1151–1155. https://doi.org/10.1126/science.290.5494.1151

McIlwain, C. C., D. M. Townsend, and K. D. Tew, 2006 Glutathione S-transferase polymorphisms: cancer incidence and therapy. Oncogene 25: 1639–1648. https://doi.org/10.1038/sj.onc.1209373

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. Science 361: 591–594. https://doi.org/10.1126/science.aap7714

Monticolo, F., C. Colantuono, and M. L. Chiusano, 2017 Shaping the evolutionary tree of green plants: evidence from the GST family. Sci. Rep. 7: 14363. https://doi.org/10.1038/s41598-017-14316-w

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. Pharmacogenet. Genomics 12: 277–286. https://doi.org/10.1097/00008571-200206000-00003

Nebert, D. W., and V. Vasiliou, 2004 Analysis of the glutathione S-transferase (GST) gene family. Hum. Genomics 1: 460. https://doi.org/10.1186/1479-7364-1-6-460

Nei, M., and A. P. Rooney, 2005 Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39: 121–152. https://doi.org/10.1146/annurev.genet.39.073003.112240

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 Biol. Evol. 5: 2359–2367. https://doi.org/10.1093/gbe/evt176

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 Res. 46: D816–D822. https://doi.org/10.1093/nar/gkx1003

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. Mol. Biol. Evol. 35: 1437–1450. https://doi.org/10.1093/molbev/msy042

Ninfali, P., and G. Aluigi, 1998   Variability of oxygen radical absorbance capacity (ORAC) in different animal species. Free Radic. Res. 29: 399–408. https://doi.org/10.1080/10715769800300441

Niwa, Y., 1999   [Oxidative injury and its defense system in vivo] Rinsho Byori the Japanese Journal of Clinical Pathology 47: 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. Curr. Opin. Struct. Biol. 15: 716–723. https://doi.org/10.1016/j.sbi.2005.10.005

Panneton, W. M., 2013   The Mammalian Diving Response: An Enigmatic Reflex to Preserve Life? Physiology (Bethesda) 28: 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: 3852–3862. https://doi.org/10.1021/bi051823+

Pearson, W. R., 2005   Phylogenies of glutathione transferase families. Methods Enzymol. 401: 186–204. https://doi.org/10.1016/S0076-6879(05)01012-8

Pond, S. L. K., and S. D. Frost, 2005a   Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21: 2531–2533. https://doi.org/10.1093/bioinformatics/bti320

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. Mol. Biol. Evol. 22: 1208–1222. https://doi.org/10.1093/molbev/msi105

Pond, S. L. K., and S. V. Muse, 2005   HyPhy: hypothesis testing using phylogenies, pp. 125–181 in *Statistical Methods in Molecular Evolution*. Springer New York, NY.

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. https://doi.org/10.1146/annurev.physiol.69.031905.163111

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. Biochem. J. 366: 45–55. https://doi.org/10.1042/bj20020533

Revell, L. J., 2012   phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3: 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Ronquist, F., and J. P. Huelsenbeck, 2003   MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574. https://doi.org/10.1093/bioinformatics/btg180

Schreer, J. F., and K. M. Kovacs, 1997   Allometry of diving capacity in air-breathing vertebrates. Can. J. Zool. 75: 339–358. https://doi.org/10.1139/z97-044

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. Nat. Commun. 9: 1215. https://doi.org/10.1038/s41467-018-03667-1

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. Biochem. J. 360: 1–16. https://doi.org/10.1042/bj3600001

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 Ioannides, C., John Wiley & Sons, New York, NY. https://doi.org/10.1002/0470846305.ch9

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: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Stamatakis, A., 2014   RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

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 Evol. 4: vey016. https://doi.org/10.1093/ve/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. Front. Zool. 14: 41. https://doi.org/10.1186/s12983-017-0225-x

Swanson, W. J., R. Nielsen, and Q. Yang, 2003   Pervasive adaptive evolution in mammalian fertilization proteins. Mol. Biol. Evol. 20: 18–20. https://doi.org/10.1093/oxfordjournals.molbev.a004233

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: 1190–1194. https://doi.org/10.1038/nature06343

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: 514–522. https://doi.org/10.1002/ar.20545

Valavanidis, A., T. Vlahogianni, M. Dassenakis, and M. Scoullos, 2006   Molecular biomarkers of oxidative stress in aquatic organisms in relation to toxic environmental pollutants. Ecotoxicol. Environ. Saf. 64: 178–189. https://doi.org/10.1016/j.ecoenv.2005.03.013

Wilhelm Filho, D., F. Sell, L. Ribeiro, M. Ghislandi, F. Carrasquedo *et al.*, 2002   Comparison between the antioxidant status of terrestrial and diving mammals. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 133: 885–892. https://doi.org/10.1016/S1095-6433(02)00253-2

Woolley, S., J. Johnson, M. J. Smith, K. A. Crandall, and D. A. Mcclellan, 2003   TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics 19: 671–672. https://doi.org/10.1093/bioinformatics/btg043

Wu, B., and D. Dong, 2012   Human cytosolic glutathione transferases: structure, function, and drug discovery. Trends Pharmacol. Sci. 33: 656–668. https://doi.org/10.1016/j.tips.2012.09.007

Yang, Z., 2007   PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24: 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang, Z., W. S. Wong, and R. Nielsen, 2005   Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22: 1107–1118. https://doi.org/10.1093/molbev/msi097

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. Nat. Genet. 46: 88–92. https://doi.org/10.1038/ng.2835

Young, J. M., H. F. Massa, L. Hsu, and B. J. Trask, 2010   Extreme variability among mammalian V1R gene families. Genome Res. 20: 10–18. https://doi.org/10.1101/gr.098913.109

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*. Wiley-Blackwell, Hoboken, NJ, 2012: 178–190.

Zhang, J., R. Nielsen, and Z. Yang, 2005   Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22: 2472–2479. https://doi.org/10.1093/molbev/msi237

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 Biol. Evol. 10: 967–975. https://doi.org/10.1093/gbe/evy062

Communicating editor: A. Rokas