



# Breathing Air and Living Underwater: Molecular Evolution of Genes Related to Antioxidant Response in Cetaceans and Pinnipeds

Giovanna Selleghin-Veiga<sup>1</sup> · Letícia Magpali<sup>1,2</sup> · Agnello Picorelli<sup>1</sup> · Felipe A. Silva<sup>1</sup> · Elisa Ramos<sup>1,3</sup> · Mariana F. Nery<sup>1</sup>

Received: 21 May 2023 / Accepted: 5 April 2024 / Published online: 12 May 2024  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

## Abstract

Cetaceans and pinnipeds are lineages of mammals that have independently returned to the aquatic environment, acquiring varying degrees of dependence on it while sharing adaptations for underwater living. Here, we focused on one critical adaptation from both groups, their ability to withstand the ischemia and reperfusion experienced during apnea diving, which can lead to the production of reactive oxygen species (ROS) and subsequent oxidative damage. Previous studies have shown that cetaceans and pinnipeds possess efficient antioxidant enzymes that protect against ROS. In this study, we investigated the molecular evolution of key antioxidant enzyme genes (CAT, GPX3, GSR, PRDX1, PRDX3, and SOD1) and the ROS-producing gene XDH, in cetaceans and pinnipeds lineages. We used the ratio of non-synonymous ( $dN$ ) to synonymous ( $dS$ ) substitutions as a measure to identify signatures of adaptive molecular evolution in these genes within and between the two lineages. Additionally, we performed protein modeling and variant impact analyzes to assess the functional consequences of observed mutations. Our findings revealed distinct selective regimes between aquatic and terrestrial mammals in five of the examined genes, including divergences within cetacean and pinniped lineages, between ancestral and recent lineages and between crown groups. We identified specific sites under positive selection unique to Cetacea and Pinnipedia, with one site showing evidence of convergent evolution in species known for their long and deep-diving capacities. Notably, many sites under adaptive selection exhibited radical changes in amino acid properties, with some being damaging mutations in human variations, but with no apparent detrimental impacts on aquatic mammals. In conclusion, our study provides insights into the adaptive changes that have occurred in the antioxidant systems of aquatic mammals throughout their evolutionary history. We observed both distinctive features within each group of Cetacea and Pinnipedia and instances of convergence. These findings highlight the dynamic nature of the antioxidant system in response to challenges of the aquatic environment and provide a foundation for further investigations into the molecular mechanisms underlying these adaptations.

**Keywords** Molecular evolution · Antioxidants · Aquatic mammals · Ischemia and reperfusion response · Reactive oxygen species

---

Handling editor: **Belinda Chang**.

✉ Giovanna Selleghin-Veiga  
giovanna.slv@gmail.com

✉ Mariana F. Nery  
mariananery@gmail.com

Letícia Magpali  
leticiamagpali@gmail.com

Agnello Picorelli  
agpicorelli@gmail.com

Felipe A. Silva  
felipebio97@gmail.com

Elisa Ramos  
elisaksramos@gmail.com

- <sup>1</sup> Laboratório de Genômica Evolutiva, Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Brazil
- <sup>2</sup> Department of Biology, Dalhousie University, Halifax, NS, Canada
- <sup>3</sup> Zoological Institute, Department of Environmental Sciences, University of Basel, Basel, Switzerland

## Introduction

Cetaceans (whales, porpoises, and dolphins) and pinnipeds (seals and sea lions) are aquatic mammals that evolved from different terrestrial ancestors about 50 and 30 million years ago (Ma), respectively (Fordyce 2018; Berta et al. 2018). However, cetaceans are exclusively aquatic while pinnipeds use both land and sea. Despite their differences in dependence on the aquatic environment, these groups share several morphological adaptations for underwater life. These include a streamlined body shape, modification of forelimbs into flippers, and a thick layer of fat for thermoregulation (Davis 2019). One striking characteristic of both cetaceans and pinnipeds is their diving behavior, which also demands many specific adaptations, especially in the physiology and metabolic homeostasis of hypoxia (Davis 2014; Blix 2018; Ponganis 2019).

Hypoxia is a condition where the concentrations of oxygen ( $O_2$ ) available for tissue usage are low, occurring in an acute form during the apnea diving periods (Ramirez et al. 2007). Among the adaptations to deal with extreme hypoxia is the vasoconstriction of peripheral tissues that redistributes blood flow to oxygen-dependent tissues, and bradycardia, which is the reduction of heart rate to maintain blood pressure (Zapol et al. 1979; McDonald and Ponganis 2013; Davis 2014). One of the main challenges of vasoconstriction is the potential for ischemia, where limited blood flow can lead to reperfusion injury, which occurs when blood flow is restored to ischemic tissue, leading to the sudden production of reactive oxygen species (ROS), such as superoxide ( $O_2^-$ ) and hydrogen peroxides ( $H_2O_2$ ) (Murphy 2009; Kalogeris et al. 2012; Hancock 2021). In normal physiological conditions, ROS plays an important role in cell signaling and metabolism. However, excessive ROS production during reperfusion, mostly from the xanthine dehydrogenase (XDH) pathway, can cause oxidative stress, leading to cellular damage (Jones 2006; Halliwell and Gutteridge 2015). This increase in ROS production is observed in aquatic mammal tissues, nonetheless, they do not present higher oxidative damage than terrestrial ones (Elsner et al. 1998; Zenteno-Savín et al. 2002; Wilhelm Filho et al. 2002).

The remarkable tolerance to oxidative stress seen in aquatic mammals is linked to their antioxidant enzyme system. This system consists of a set of proteins with catalytic sites that stabilize and remove ROS from their tissues (Halliwell and Gutteridge 2015). The main antioxidant enzymes include catalase (CAT), glutathione peroxidase (GPX), superoxide dismutase (SOD), and peroxiredoxins (PRDX). Several studies have investigated the activity of these enzymes by comparing the tissues of aquatic and terrestrial mammals, revealing the important role of these enzymes in the protection against oxidative stress (Zenteno-Savín et al.

2011; Vázquez-Medina et al. 2012; Allen and Vázquez-Medina 2019). Furthermore, the activity and expression of these enzymes vary between cetaceans and pinnipeds, suggesting that lineages may have unique responses to oxidative stress (Wilhelm Filho et al. 2002; Cantú-Medellín et al. 2011; Geßner et al. 2022). These variations may be related to specific challenges, such as different depths and duration of dives, leading to specific adaptations in their antioxidant enzyme system.

The metabolism of aquatic mammals, including the antioxidant system, has been studied for many years from a metabolic perspective to better understand how it adjusts to a drastic change in the aquatic environment. With the advance of genomic sequencing and analysis, the availability of more mammalian genomes has grown and many genes related to hypoxia adaptations have been found to be positively selected in cetacean lineages (McGowen et al. 2014; Nery et al. 2013; Tian et al. 2016; Cabrera et al. 2021). Also, convergence studies between cetaceans and pinnipeds have proven to be a valuable approach, because, despite differences in habitat and diving behavior, both groups have been subjected to similar evolutionary pressures in their management of oxygen (Zhou et al. 2015; Yuan et al. 2021). The antioxidant enzymes also have been targeted by genomic studies (Tian et al. 2018, 2019); however, many of them use a limited number of species and focus on cetaceans groups, lacking a more comprehensive and comparative approach between the Cetacea and Pinnipedia lineages, as well as a comparison between the evolution of the antioxidant system as a whole.

In this context, to better understand the adaptation of aquatic mammals to oxidative stress, our objective was to investigate the molecular evolution of genes involved in oxidative stress derived from hypoxia in breath-hold dives. We estimated rates and identified sites under positive selection specific to each group of aquatic mammals in some of the antioxidant enzymes, responsible for removing ROS, and one of the producers of ROS, the XDH enzyme. Additionally, we identify signs of positively selected convergent mutation in two deep-diving species, the Cuvier's beaked whale and the Southern elephant seal. Finally, we inferred the possible impacts of these selected mutations on overall protein structure and function.

## Material and Methods

### Genomic Data, Sequences, and Alignment

To obtain the coding sequences (CDS) of the six antioxidant genes (CAT, GPX3, GSR, SOD1, PRDX1, and PRDX3) and the xanthine dehydrogenase (XDH) gene, we retrieved the

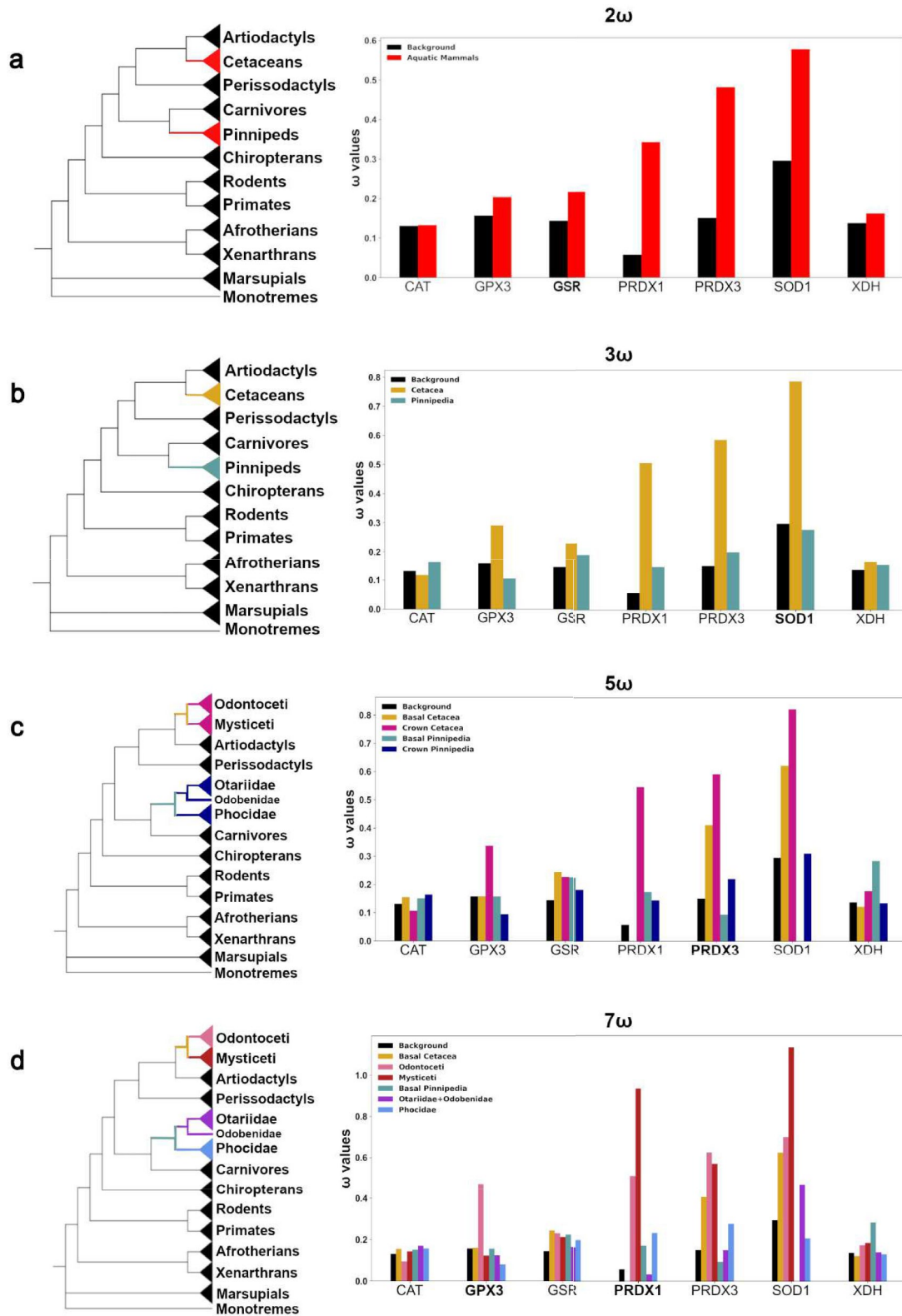
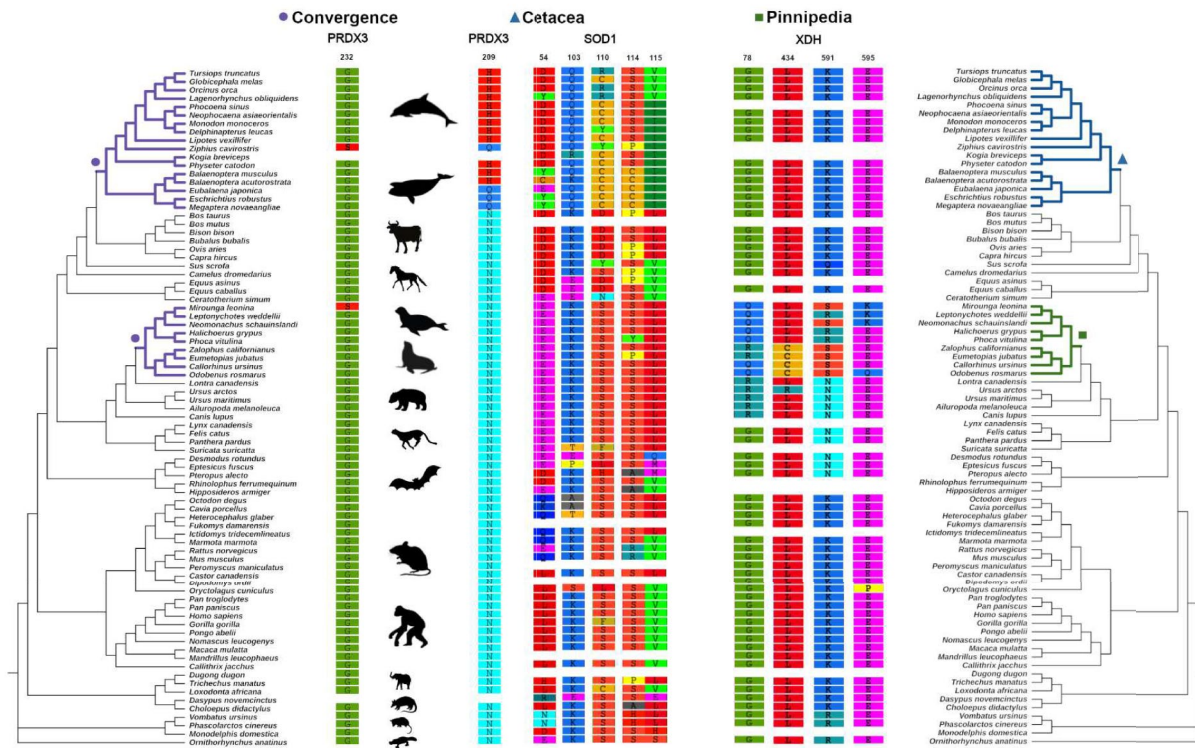


Fig. 1



**Fig. 2** Comparison of protein residues in which positive selection was identified in the BSM and complementary analysis. The residues were obtained from the protein alignment from each gene, in which it is possible to compare the amino acids from terrestrial mammals and substitutions happening in the aquatic mammals' lineages. Species

trees are provided showing all the species used in our analysis and the lineages of interest used as *foreground* in each test are highlighted using different colors and shapes. Purple circle on the left: convergence test with Cetacea+Pinnipedia marked; Blue triangle on the right: Cetacea test; Green rectangle on the right: Pinnipedia test

longest transcripts of these genes available from the NCBI database (<https://www.ncbi.nlm.nih.gov/>) for 72 species of mammals (Supplementary Table S1). In 9 target species, annotated sequences were missing or incomplete, so we used the CDS retrieved from a closely related species as a query sequence in the NCBI BLAST tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) searching directly in the genome assemblies of target species (Supplementary Table S2.1). The criteria of  $E$ -value  $< 0.05$  and percent of identity  $> 90\%$  were used to select the best results (Supplementary Table S2.2). Sometimes, because some of the genomes used as targets were fragmented, the sequence fragments returned by BLAST needed to be mapped to the query species gene, generating a concise sequence for our target species.

Nucleotide and amino acid sequences were aligned using the MUSCLE tool (Edgar 2004). Alignments were curated by removing or replacing incomplete sequences and tidying up mismatched regions. To obtain the codon alignments used

in the evolutionary analyses, we used PRANK v.170427 (Löytynoja 2014).

## Phylogenetic Inference

We reconstructed the evolutionary history of the antioxidant genes by performing a phylogenetic analysis of both nucleotide and amino acid alignments. Maximum likelihood (ML) phylogenetic reconstruction was performed using IQ-TREE v. 1.6.2 (Nguyen et al. 2015) with 1,000 bootstrap replicates to estimate branch confidence (Hoang et al. 2017), where nodes with support values  $\geq 80$  were considered robust. For Bayesian inference of nucleotide trees, we used PARTITION FINDER v. 2.1.1 (Lanfear et al. 2017) to estimate the partitioned evolutionary models using the Bayesian Information Criterion (BIC) (Supplementary Table S3). The models inferred for each gene were used in MRBAYES v. 3.2.7a (Ronquist and Huelsenbeck 2003), where the posterior probabilities of the nodes were calculated applying Markov

chain Monte Carlo (MCMC) running for 10,000,000 generations with four chains, and trees were sampled every 100 generations.

## Natural Selection Analyses

We used the *codeml* program in the PAML v. 4.4 package (Yang 2007) to estimate the evolutionary rates of genes under a maximum likelihood framework (MLE) (Goldman and Yang 1994). Codon substitution models were applied to estimate the rates of non-synonymous ( $dN$ ) and synonymous ( $dS$ ) substitutions to detect positive selection. The omega ratio— $\omega$  ( $dN/dS$ ) is the measure used to infer the strength of the selection of codon changes, in which  $\omega > 1$  indicates positive selection,  $\omega < 1$  indicates purifying selection, and  $\omega = 1$  indicates neutral evolution. We used the codon alignments retrieved by PRANK and a species tree based on a mammalian phylogeny (Upham et al. 2019) available in VertLife (<https://vertlife.org/>).

We employed three different methods in *codeml* to detect positive selection in our data: branch models (BM), branch-site models (BSM), and site models (SM) (Yang 1998; Yang and Nielsen 2002; Yang et al. 2000).

To detect selection in specific lineages, we used BM, which estimates the  $dN/dS$  for all branches in the tree as a null model or estimates separated ratios for *foreground* lineages and the *background*. We employed five branch models (Fig. 1): (a) one-ratio model ( $1\omega$ ), the null model; (b) two-ratio model ( $2\omega$ ), estimating one  $\omega$  value for all the lineages of aquatic mammals (Cetacea + Pinnipedia), as *foreground*, and one for all the terrestrial mammals, as *background* branches; (c) three-ratio model ( $3\omega$ ), distinguishing the clades within aquatic mammals in Cetacea and Pinnipedia; (d) five-ratio model ( $5\omega$ ), separating the ancestral lineage from its recent lineages inside each aquatic mammal group; and (e) seven-ratio model ( $7\omega$ ), separating the recent groups of Cetacea (Mysticeti and Odontoceti) and Pinnipedia (Phocidae and Otariidae + Odobenidae). In each model,  $\omega$  values were estimated for the branches previously marked in the phylogeny based on the species evolution.

To detect positive selection acting on a few sites along specific lineages, we used the BSM analysis. This method sorts the sites in  $\omega$  classes and only allows  $\omega_2 > 1$  in *foreground* branches (Yang and Nielsen 2002). We applied the modified version (Zhang et al. 2005) using Model A as the alternative model, where there are four classes of  $\omega$  and  $\omega_2 > 1$  is only allowed in the *foreground* branches. We performed three tests: (a) convergence, with Cetacea and Pinnipedia as *foreground* branches; (b) Cetacea, where only the group of cetaceans were marked as *foreground*; and (c) Pinnipedia, with only pinnipeds as *foreground* (Fig. 2). As a comparative test, the same analysis was performed using close relative groups: Artiodactyla, Carnivores (excluding

Pinnipedia), and both groups were marked together for convergence.

Finally, we performed SM to identify specific codon sites that may be under positive selection, regardless of the lineage. In this model, each site in the alignment is distributed into different  $\omega$  classes, and no *foreground* or *background* lineages are designated (Yang et al. 2000). We used the positive selection test, comparing the null model M7beta with the alternative model M8beta& $\omega$  (Supplementary Table S4).

We also performed selection analyses using the HyPhy package (Pond et al. 2005; Kosakovsky et al. 2020a). For branch-site tests, we used aBSREL and BUSTED, which estimate  $\omega$  values in each branch of the tree. While aBSREL estimates the proportion of sites under positive selection in the tested lineages (Smith et al. 2015), BUSTED is a general approach for identifying gene-wide positive selection in the tested lineage (Murrell et al. 2015). We also used RELAX to calculate the strength of selection (i.e., whether it was intensified or relaxed) in our lineages of interest (Wertheim et al. 2015).

We used contrast-FEL to test sites with different evolutionary rates between two subsets of a phylogeny evolving under different environmental conditions (Kosakovsky et al. 2020b). We also used site models like FEL (Fixed Effects Likelihood), SLAC (Single-Likelihood Ancestor Counting), and FUBAR (Fast, Unconstrained Bayesian AppRoximation), which identify PSS in a phylogeny using maximum likelihood, counting approaches, and Bayesian approaches, respectively (Kosakovsky et al. 2005; Murrell et al. 2013). These tests, together with SM in *codeml*, were used as complementary tests for sites identified in BSM models (Supplementary Table S5.1 and S5.2).

## Statistical Analysis

For the *codeml* models, to test whether  $dN$  was significantly higher than  $dS$ , we used a likelihood ratio test (LRT) comparing the likelihoods of the null ( $l_0$ ) and alternative ( $l_1$ ) models in the Eq. 2  $X(l_1 - l_0)$  (Yang and Nielsen 1998; Yang 1998). The LRT results follow a  $\chi^2$  distribution, and values with  $p < 0.05$  were considered significant, supporting the alternative hypotheses.

Although BSM and SM estimate  $\omega$  for site lineage and site only, respectively, they use the same statistical method to infer the posterior probability (pp) of each site in a  $\omega > 1$  class, the Bayes empirical Bayes (BEB). Sites with significant LRT and pp  $> 0.9$  are considered positively selected (Nielsen and Yang 1998); however, we considered only the results of BSM to infer positively selected sites (PSS) in our lineages of interest, and SM was used as a complementary test, once it does not have a lineage specificity.

## Amino Acid Properties and Protein Structure

We used the program TreeSAAP v.3.2 (Woolley et al. 2003) to evaluate the physical–chemical changes induced by non-synonymous substitutions in protein residues within a phylogenetic context. The program categorizes these amino acid mutations into eight groups based on the magnitude of their physical–chemical effects, ranging from mild to radical substitutions. We used a goodness-of-fit test, which yields a z-score, to determine whether these mutations were under positive selection. The results were treated in IMPACT-S v. 1.0.0 (Maldonado et al. 2014), generating a table with “Property by Site” where only sites with significant z-scores and substitutions in the categories 6–8 were considered as being under positive destabilizing selection.

To assess the potential impact of positively selected sites on the protein structure and function, we employed a combination of bioinformatic tools and structural modeling techniques. First, we used AlphaFold Database (Varadi et al. 2022; Jumper et al. 2021) and AlphaFold Colab (AlphaFold v2.3.1.) in the UCSF ChimeraX v. 1.5 software (Godard et al. 2018; Pettersen et al. 2021) to model the protein structures of the target species and their close relatives (*Bos taurus* for Cetacea and *Canis lupus familiaris* or *Felis catus* for Pinnipedia). We then compared the regions of the protein with high confidence (pLDDT > 90) to identify potential changes induced by positively selected mutations (Supplementary Figs S22–S28).

To further investigate the functional impact of the positively selected sites, we checked if they had known effects on human proteins. Using PolyPhen-2 (Adzhubei et al. 2010) based on the HumDiv dataset, we compared the wild-type properties of the amino acids with the mutant properties found in human variants that harbored the same mutations as the positively selected sites. PolyPhen-2 employs a Naive Bayes posterior probability score to estimate the damaging effect of amino acid replacements, classifying them into three categories of effect in the protein: benign, possibly damaging, and probably damaging. Finally, to gain additional insights into the function and evolution of the proteins, we retrieved additional information from the UniProt (<https://www.uniprot.org/>). We visualized the protein models using the UCSF ChimeraX v.1.5 software.

## Results

### Dataset and Phylogenetic Reconstruction

We retrieved the longest and most complete transcript for each gene for 81 species from 15 mammalian orders using annotated CDS sequences and BLAST tool search

(Supplementary Table S6). There was a variable number of represented species for each gene: 77 for CAT, 78 for GPX3 and PRDX1, 72 for GSR, 79 for PRDX3, 75 for SOD1, and 68 for XDH (Supplementary Table S1 and S2.2). The alignments showed conserved regions in the sequences corresponding to functional domains of the proteins, with the PRDX1 gene being the most conserved one and GSR with a high variable region at the beginning of the sequences.

The topology of the phylogenetic gene trees recovers the main mammalian groups without evident accumulation of modifications in a specific branch (Supplementary Figs S1–S21). Most mammalian families had good support values for bootstrap (> 90) and posterior probabilities (> 0.9) in the nucleotide tree, especially the Cetacea and Pinnipedia groups. Because there were mostly minor disagreements between the gene tree and species tree, we used the species tree in the PAML selection analysis for a better delimitation of the *foreground* branches and branch length. For the HyPhy analysis, we used the gene tree as recommended by the literature (Kosakovsky et al. 2005, 2020b; Murrell et al. 2013, 2015; Smith et al. 2015; Wertheim et al. 2015).

### Evolution of Antioxidant Enzymes is Accelerated in Aquatic Mammal Lineages

We performed branch analysis using different models to understand the selection patterns in the aquatic mammal branches. In all models, the *background* lineages presented similar values of  $\omega$  for most genes ( $\bar{X}_{\omega} = 0.156$ ,  $SD = 0.067$ ).

In the CAT and XDH genes, the LRT test did not find significant changes in the substitution rates of *foreground* branches (Table 1), indicating no differences between rates of evolution of *background* and *foreground* lineages. In BM, we consider the genes evolutionary pattern as the most specific significant  $\omega$  test, once the branch partitions are nested within each other in a hierarchical structure (i.e.,  $2\omega$  is the general branch partition and within it, the  $3\omega$ ,  $5\omega$  and  $7\omega$  are each more specific ones). Differences between the terrestrial and aquatic mammals— $2\omega$ —were significant in the GSR gene (Fig. 1a), suggesting that aquatic mammals (including cetaceans and pinnipeds) accumulated more mutations when compared to the terrestrial mammals included in our dataset. The  $3\omega$  fitted better for the SOD1 gene, with Cetacea presenting a higher  $\omega$  value than Pinnipedia (Fig. 1b), suggesting that aquatic mammal groups evolved at divergent rates for this gene. The  $5\omega$  better fitted the gene PRDX3, which showed a higher accumulation of non-synonymous mutations in crown lineages of aquatic mammals compared to the ancestral ones (Fig. 1c). The  $7\omega$ , which distinguishes  $\omega$  values of crown lineages of cetaceans and pinnipeds, was a better fit for the genes GPX3 and PRDX1. GPX3 acceleration is likely caused by

the higher  $\omega$  rates of the Odontoceti lineage ( $\omega=0.471$ ), while Mysticeti had lower values ( $\omega=0.122$ ). In PRDX1, all crown cetaceans accumulate mutations at higher rates than the other lineages (Odontoceti:  $\omega=0.509$ ; Mysticeti:  $\omega=0.934$ ) (Table 1 and Fig. 1d).

### Positively Selected Sites in Aquatic Mammals

For the *codeml* BSM, the genes PRDX3 (LRT = 21.47,  $p=0.0000$ ) and SOD1 (LRT = 9.19,  $p=0.0024$ ) in

Cetacea, as well as GSR (LRT = 9.22,  $p=0.0024$ ), PRDX1 (LRT = 4.78,  $p=0.0289$ ), and XDH (LRT = 13.74,  $p=0.0002$ ) in Pinnipedia as *foreground* showed a better fit under model A than the null model (Table 2). In these genes, the positively selected sites (PSS) were identified using the empirical Bayes' method (BEB) at posterior probabilities higher than 0.9, resulting in 14 PSS in PRDX3 and three PSS in SOD1 for Cetacea, as well as one PSS in PRDX1 and six PSS in XDH for Pinnipedia. No PSS with significant BEB posterior probabilities were found for GSR. For the SM,

**Table 1** Log-likelihood and omega values estimated for various lineages models in the aquatic mammals' groups retrieved by the branch model analyses using *codeml*

Genes	Model	$l$	LRT	$p$ values	$\omega$ values					
CAT	1 $\omega$	- 21,408.64			0.132					
	2 $\omega$	- 21,408.64	0.00	0.9465	0.131	0.133				
	3 $\omega$	- 21,407.98	1.31	0.2528	0.131	0.117	0.161			
	5 $\omega$	- 21,407.57	2.14	0.5444	0.131	0.155	0.107	0.151	0.164	
	7 $\omega$	- 21,407.13	3.01	0.6979	0.131	0.155	0.095	0.143	0.151	0.170
GPX3	1 $\omega$	- 8958.99			0.161					
	2 $\omega$	- 8957.99	2.01	0.1558	0.157	0.204				
	3 $\omega$	- 8954.22	7.53	<b>0.0061</b>	0.157	0.291	0.105			
	5 $\omega$	- 8953.04	9.89	<b>0.0195</b>	0.157	0.158	0.337	0.157	0.094	
	7 $\omega$	- 8950.10	15.77	<b>0.0075</b>	0.157	0.160	0.471	0.122	0.156	0.124
GSR	1 $\omega$	- 21,915.50			0.149					
	2 $\omega$	- 21,910.97	9.06	<b>0.0026</b>	0.144	0.217				
	3 $\omega$	- 21,910.72	0.50	0.4814	0.144	0.229	0.189			
	5 $\omega$	- 21,910.64	0.65	0.8845	0.144	0.244	0.226	0.225	0.182	
	7 $\omega$	- 21,910.55	0.84	0.9747	0.144	0.244	0.230	0.212	0.224	0.164
PRDX1	1 $\omega$	- 5545.87			0.080					
	2 $\omega$	- 5511.98	67.79	<b>0.0000</b>	0.058	0.343				
	3 $\omega$	- 5507.35	9.25	<b>0.0024</b>	0.058	0.505	0.147			
	5 $\omega$	- 5504.93	14.08	<b>0.0028</b>	0.058	0.000	0.547	0.175	0.145	
	7 $\omega$	- 5502.43	19.10	<b>0.0018</b>	0.058	0.000	0.509	0.934	0.173	0.033
PRDX3	1 $\omega$	- 10,576.33			0.172					
	2 $\omega$	- 10,549.85	52.97	<b>0.0000</b>	0.151	0.482				
	3 $\omega$	- 10,546.13	7.45	<b>0.0064</b>	0.151	0.584	0.198			
	5 $\omega$	- 10,545.74	8.22	<b>0.0417</b>	0.151	0.412	0.593	0.094	0.221	
	7 $\omega$	- 10,545.43	8.85	0.1152	0.151	0.409	0.624	0.569	0.094	0.150
SOD1	1 $\omega$	- 7357.63			0.313					
	2 $\omega$	- 7352.53	10.20	<b>0.0014</b>	0.296	0.578				
	3 $\omega$	- 7349.84	5.37	<b>0.0204</b>	0.296	0.785	0.275			
	5 $\omega$	- 7349.04	6.98	0.0725	0.296	0.623	0.823	0.000	0.311	
	7 $\omega$	- 7348.20	8.65	0.1239	0.296	0.623	0.699	1.134	0.000	0.468
XDH	1 $\omega$	- 49,196.81			0.140					
	2 $\omega$	- 49,195.23	3.15	0.0759	0.138	0.162				
	3 $\omega$	- 49,195.17	0.13	0.7188	0.138	0.165	0.155			
	5 $\omega$	- 49,192.07	6.31	0.0974	0.138	0.122	0.177	0.284	0.135	
	7 $\omega$	- 49,192.02	6.43	0.2670	0.138	0.122	0.174	0.186	0.285	0.140

The first value in the  $\omega$  value column is always referred to the *background* branches. The next values are the *foreground* branches in each model: 2 $\omega$ —Aquatic Mammals; 3 $\omega$ —Cetacea and Pinnipedia; 5 $\omega$ —stem Cetacea, Crown Cetacea, stem Pinnipedia, and Crown Pinnipedia; 7 $\omega$ —stem Cetacea, Odontoceti, Mysticeti, stem Pinnipedia, Otariidae + Odobenidae, and Phocidae. The P values in bold are the significant lineage models tested

**Table 2** Branch-site tests performed in the Paml (for each comparison of null and alternative model we have the respective LRT, P value, and sites identified) and HyPhy (aBSREL, BUSTED, and RELAX) packages using three approaches of *foreground* branches delimitation: each marine mammal clade independently (Cetacea and Pinnipedia) and the convergence test marking them both as *foreground*

Convergent	Cetacea				Pinnipedia									
	Genes	LRT	<i>p</i> values	sites (BEB > 0.900)	aBSREL	BUSTED	RELAX	LRT	<i>p</i> values	sites (BEB > 0.900)	aBSREL	BUSTED	RELAX	
CAT	0	1.0000	0.07	0.7943	461, 626	<i>p</i> > 0.05	<i>p</i> > 0.05	<i>p</i> = <b>0.0170</b> K = <b>0.10</b> relaxation	0	1.0000	–	<i>p</i> > 0.05	<i>p</i> > 0.05	<i>p</i> = 0.3500 K = 0.69
GPX3	0	1.0000	0.17	0.6767	48	<i>p</i> > 0.05	<i>p</i> > 0.05	<i>p</i> = <b>0.0140</b> K = <b>0.25</b> relaxation	0	1.0000	–	<i>p</i> > 0.05	<i>p</i> > 0.05	<i>p</i> = 0.0930 K = 37.83
GSR	0	1.0000	0	0.9972	–	<b><i>p</i> = 0.0026</b> <b>1 branch</b> <b>under</b> <b>episodic</b> <b>diversifying</b> <b>selection</b>	<i>p</i> > 0.05	<i>p</i> = 0.8600 K = 1.08	9.22	<b>0.0024</b>	–	<i>p</i> > 0.05	<i>p</i> > 0.05	<i>p</i> = 0.3950 K = 1.30
PRDX1	0.25	0.6200	8, 54, 62, 63, 89, 110, 122, 151, 184	2.01	0.1559	8, 11, 54, 62, 110, 122, 151, 153, 174, 177	<i>p</i> > 0.05	<i>p</i> = <b>0.0000</b> K = <b>0.23</b> relaxation	4.78	<b>0.0289</b>	63	<i>p</i> > 0.05	<i>p</i> > 0.05	<i>p</i> = 0.4150 K = 0.80
PRDX3	6.00	<b>0.0143</b>	121, 148, 181, 183, 208, 209, 229, 232, 237, 246, 261, 281	21.47	<b>0.0000</b>	47, 112, 121, 148, 181, 183, 208, 209, 211, 229, 237, 246, 261, 281	<i>p</i> > 0.05	<i>p</i> = <b>0.0000</b> K = <b>7.11</b> intensification	0	1.0000	–	<i>p</i> > 0.05	<i>p</i> > 0.05	<i>p</i> = 0.8230 K = 0.57
SOD1	3.30	0.0694	110	9.20	<b>0.0024</b>	54, 103, 110	<b><i>p</i> = 0.0307</b> <b>1 branch</b> <b>under</b> <b>episodic</b> <b>diversifying</b> <b>selection</b>	<i>p</i> = 0.1290 K = 1.30	0	1.0000	–	<i>p</i> > 0.05	<i>p</i> > 0.05	<i>p</i> = 0.1110 K = 0.21
XDH	1.20	0.2734	78, 431, 434, 557, 585, 595, 606, 730, 815, 862, 871, 905, 960, 962, 1265, 1445, 1457, 1485	2.09	0.1480	557, 730, 862, 871, 960, 962, 1265, 1457, 1485	<i>p</i> > 0.05	<i>p</i> = 0.7530 K = 0.95	13.74	<b>0.0002</b>	78, 431, 434, 585, 591, 595	<i>p</i> > 0.05	<i>p</i> > 0.05	<i>p</i> = 0.5710 K = 1.13

The values in bold are the *p* values for the models with significant LRT results when compared to null models



except for the PRDX1, all the other genes had a better fit for the alternative model M8 and PSS were identified, including some from BSM (Supplementary Table S4). The total number of PSS supported by at least three methods (PAML: BSM and SM and HyPhy: Contrast-FEL, FEL, SLAC, FUBAR) were GPX3 (4), GSR (1), PRDX3 (4), SOD1 (7), and XDH (10). For more details on the PSS results, see Supplementary Tables S5.1 and S5.2.

We checked whether the PSS identified in BSM analyses were exclusive to the aquatic mammal lineages. The GPX3 gene shows the same mutations in cetacean groups and many other terrestrial species. In the GSR gene, contrast-FEL, FEL, FUBAR, and SLAC support a positive selection event on site 450 (Supplementary Table S5.1 and S5.2), with an amino acid change from arginine (Arg) to serine (Ser) in three pinniped species (*Odobenus rosmarus*, *Neomonachus schauinslandi*, and *Phoca vitulina*). However, such change is also shared with *Eubalaena japonica*, *Bos taurus*, and *Bubalus bubalis*. For the PRDX3 gene, only the PSS 209 had mutations that were exclusive to cetaceans. For SOD1, five out of seven PSS had exclusive changes in cetaceans (54, 103, 110, 114, and 115). Finally, for XDH, only 4 out of 10 sites had specific changes in pinnipeds (78, 434, 591, and 595). See the amino acid substitutions for these sites in Fig. 2.

For the Hyphy branch-site methods, no evidence of positive selection was identified in the pinnipeds group. On the other hand, for cetaceans, the RELAX test identified evidence of relaxation in the genes CAT, GPX3, and PRDX1 and selection intensification in the PRDX3 gene. For the GSR gene, the aBSREL test indicated a signal of positive selection in the clade Delphinidae (*Tursiops truncatus*, *Orcinus orca*, *Lagenorhynchus obliquidens*, and *Globicephala melas*), suggesting that this group may have experienced a different selective regime during cetaceans' evolution. This test also showed a signal of positive selection in SOD1 for the *Lipotes vexillifer* lineage, but when the alignment was checked, sites with exclusive mutations for this species were not identified as PSS by any other methods.

The BSM using the close relatives groups of Artiodactyla, Carnivores (except Pinnipedia) and convergence (Artiodactyla + Carnivora) had better fitting of the alternative model A for SOD1 in Artiodactyla with 25 and 37 as PSS and XDH in Carnivora with 575, 842, 892, and 1491 as PSS (Supplementary Table S7). The PSS identified here for the genes SOD1 and XDH were not the same as the ones found for any lineage of aquatic mammals, suggesting that selection is not acting in the same sites in close groups.

## Convergent Evolution

To test for convergent selection between cetaceans and pinnipeds, we used the same *codeml* branch-site approach

but selected both aquatic mammal groups as *foreground* branches. We found evidence for positive selection on the PRDX3 gene (LRT = 6.00;  $p = 0.0143$ ), where the site 232 (PP = 0.986) changing from glycine (Gly) to serine (Ser) only in *Ziphius cavirostris* (Odontoceti, Cetacea) and *Mirounga leonina* (Phocidae, Pinnipedia) (Fig. 2).

## Change in Amino Acid Properties and 3D Modeling

We investigated the impact of the aquatic mammal PSS mutations on protein structure. Using TreeSAAP, we identified radical changes in physicochemical properties (substitutions in the categories 6–8) for cetacean PSS found in genes GPX3 (167), SOD1 (2, 54, 103, 110, and 114) and for pinniped PSS in GSR (450) and XDH (78, 337, 591, 595, and 834) (Supplementary Table S8 and Table 3).

We then searched for known human variations matching the aquatic mammal PSS. We identified variations in the genes GPX3 (2), GSR (1), PRDX3 (1), SOD1 (2), and XDH (1) (Supplementary Table S8 and Table 3) that were not associated with diseases. For the rest of the sites, there were no known modifications or the variations were to a different amino acid in humans. However, in the SOD1 and XDH genes, certain sites had the same variations that are associated with diseases in humans. For instance, sites 54, 103, 110, 114, and 115 in SOD1 are close to sites affected by Amyotrophic lateral sclerosis 1 (ALS1) (Table 3). In the XDH gene, the substitution for a valine amino acid at site 834 was identified in some mammal species (*Ornithorhynchus anatinus*, *Oryctolagus cuniculus*, *Dipodomys ordii*, *Peromyscus maniculatus*, *Cavia porcellus*, *Octodon degus*, and *Choloepus didactylus*) and is associated in humans with Xanthinuria type II (ClinVar) or Hereditary xanthinuria type 1 (Supplementary Table S8). Most mutations were considered “Benign” by PolyPhen-2, indicating no negative effect on human proteins. However, we found sites that were “Possibly Damaging” and “Probably Damaging” in PRDX3, SOD1, and XDH (Table 3). Of note, site 232 in PRDX3 has been found in somatic mutations in cancers (CPTAC-3 project at NCI-TCGA).

In the protein structure modeling, we focused on the sites with changes exclusive to aquatic mammals in the genes PRDX3 (1), SOD1 (5), and XDH (4), in addition to the case of convergence in PRDX3. Most of the PSS did not result in differences in the protein structure compared to closely related species without the mutation (Fig. 3a). However, we observed that the PSS in SOD1 (sites 54, 103, 110, 114, and 115) are located spatially near the active and binding sites of the protein (Fig. 3c). The PSS in the XDH gene was located on the protein's surface (Fig. 3d), which could affect its solubility. We found a slight translocation of the loop where site 232 is located in the species *Ziphius cavirostris* compared to *Bos taurus* (Fig. 3a). Notably, the PSS in PRDX3 and XDH

were located in the domain regions of the protein, such as site 209 in the Thioredoxin Domain (IPR013766), site 232 in C-terminal domain (IPR019479), site 78 in 2Fe-2S ferredoxin-type iron–sulfur binding domain (IPR001041), and site 434 in FAD-binding domain, PCMH-type (IPR016166) (Fig. 3a, b, and d).

## Discussion

In this study, we examined the molecular evolutionary history of antioxidant enzymes, which play a critical role in eliminating reactive oxygen species (ROS) that can cause oxidative stress, especially during long dives. We focused on two groups of aquatic mammals, Cetacea and Pinnipedia, that share important physiological adaptations to apnea diving. We aimed to identify signs of positive selection in the genes encoding these enzymes and their potential impact on protein function. Our results revealed contrasting evolutionary rates among antioxidant genes and between the two groups of aquatic mammals, with accelerated rates of evolution compared to terrestrial mammals. We also identified positively selected sites with amino acid changes occurring exclusively in aquatic mammal lineages, including an example of convergent evolution.

The results from our branch models suggest an accumulation of non-synonymous mutations in lineages of aquatic mammals compared with their terrestrial counterparts. The  $\omega$  values for our *background* species were similar to the average estimates for terrestrial species' genomes (Yuan et al. 2021). The increase in the omega values is not uniform in the aquatic lineages, occurring at a higher rate in cetaceans, consistent with the previous study that showed higher mean  $\omega$  values in cetacean genomes. The discrepancies observed between cetaceans and pinnipeds, for example, in the case of PRDX3 and SOD1, and inside crown groups of cetaceans after the splitting of dolphins and whales, such as in GPX3 and PRDX1, may be related to their life history, once each group present specificities in physiology and metabolic responses (Janecka et al. 2012).

The XDH gene codes for the enzyme xanthine dehydrogenase, which under hypoxic conditions, is converted into xanthine oxidase (XO), producing  $H_2O_2$  in the reaction of hypoxanthine to uric acid, causing oxidative stress during submersion (Kelley et al. 2010). XO induction during voluntarily associated apneas in elephant seal pups suggests a role in the development of the antioxidant system (Vázquez-Medina et al. 2011a). Previously, unique sites were found in the XDH gene in pinnipeds, cetaceans, and other oxidant stress-tolerant species with no positive selection signals associated with them (Tian et al. 2022). In our work, we found evidence of positive selection, but the sites were not the same as in the previous work. In addition, this

same study performed an *in vitro* assay and showed that even apparently neutral mutations in cetaceans can impact enzyme activity, raising questions about how adaptive mutations impact XDH.

The SOD1 gene belongs to the superoxide dismutase family, the first line of defense against superoxides ( $O_2^-$ ) and one of the most important antioxidants. It codes the CuZn-SOD isoform found in the mitochondria (Zelko et al. 2002), which is essential in removing superoxide produced during cellular respiration and by XDH under reoxygenation (Murphy 2009; Kelley et al. 2010). This enzyme is very active in several tissues of aquatic mammals compared with terrestrial ones (Elsner et al. 1998; Wilhelm Filho et al. 2002; Vázquez-Medina et al. 2006). Our results showed an accumulation of mutations in cetaceans compared to pinnipeds and terrestrial mammals. In the cetaceans' lineage, we also identified positively selected sites with radical modifications near important parts of the protein, similar to previous studies (Tian et al. 2021a). Furthermore, the same study presented evidence of differential evolution of SOD in long/deep and short/shallow divers for cetaceans' species. Finally, although pinnipeds also present differences in diving habits among species, they might not rely on mutations in the protein sequence, but in regulatory regions that affect gene expression levels (Righetti et al. 2014; Martens et al. 2022).

The CAT gene, responsible for catalase production, did not show any signal of differential evolution between aquatic and terrestrial mammals, except for relaxation in cetacean branches. The catalase activity in aquatic mammal tissues is not significantly higher than in terrestrial mammals (Cantú-Medellín et al. 2011; Wilhelm Filho et al. 2002; Vázquez-Medina et al. 2006). Although catalase is present in many tissues, its content in the liver and kidney is known to be higher, where it will respond to the severe increase of  $H_2O_2$  (Chance et al. 1979; Michiels et al. 1994) and may be inactivated in constant concentrations of  $H_2O_2$  (Kirkman and Gaetani 2007). Accordingly, with previous works that found little divergence between the CAT sequences in different metazoan species (Hewitt and Degnan 2023), our results show that this enzyme is very conserved even in macroevolutionary changes.

The glutathione system is crucial in eliminating ROS during oxidative stress (Michiels et al. 1994). The enzyme responsible for reducing oxidized glutathione (GSSG) to GSH is the glutathione reductase (GR), which is codified by the gene GSR and known to be more active in aquatic mammals than in semi-aquatic and terrestrial mammals (Vázquez-Medina et al. 2007; Righetti et al. 2014; García-Castañeda et al. 2017). In previous studies, the GSR gene has been identified as being positively selected in dolphin lineage (Yim et al. 2014) but not in pinnipeds (Martens et al. 2022). In our findings, we showed an acceleration of aquatic mammal lineages compared to terrestrial ones, but

**Table 3** PSS physicochemical amino acid properties variations determined by TreeSAAP and impact of mutants in human proteins by PolyPhen-2 results (“PolyPhen-2 Score” and “Score Classification”)

Gene	Uniprot ID	PSS	TreeSAAP properties	Human protein position	Human → Aquatic mammal	PolyPhen-2 Score	Score Classification	Uniprot information
PRDX3	P30048 PRDX3_HUMAN	209	–	157	N → H	0.012	Benign	Variation not present in humans
					N → Q	0.028	Benign	Variation not present in humans
SOD1	P0044 SODC_HUMAN	54	Chromatographic index, Polar requirement; Polarity	180	G → S	1.000	Probably Damaging	Somatic variant
					L → Y	0	Benign	ASL close site
					L → C	0.991	Probably Damaging	
					L → E	0.005	Benign	
					L → D	0.007	Benign	
					K → Q	0.015	Benign	ASL close site
SOD1	P0044 SODC_HUMAN	103	Isoelectric_point	92	K → R	0.064	Benign	
					S → C	0.002	Benign	ASL close site
					S → R	0.568	Possibly Damaging	
					S → Y	0.014	Benign	
SOD1	P0044 SODC_HUMAN	110	Hydropathy; Chromatographic index; Helical contact area; Isoelectric point; Polarity	99	S → C	0.014	Benign	ASL close site
					S → P	0.003	Benign	
					V → I	0	Benign	ASL close site
					G → R	0.031	Benign	Variation not present in humans
XDH	P47989 XDH_HUMAN	78	Isoelectric point	35	G → Q	0.953	Possibly Damaging	Variation not present in humans
					L → C	1.000	Probably Damaging	Xanthinuria type II close site
					K → S	0.523	Possibly Damaging	Variation not present in humans
					K → R	0	Benign	Variation not present in humans
					E → Q	0.381	Benign	Variation not present in humans
					E → K	0.108	Benign	Variation not present in humans

The relative sites and amino acid substitutions used in PolyPhen-2 inference are shown in “Human protein position” and “Human → Aquatic mammal,” respectively. “Uniprot information” was retrieved from the Uniprot database (<https://www.uniprot.org/>) for the specific amino acid substitution in the respective site of the human protein sequence

the methods were not capable of identifying specific positive selected sites. GSR has likely experienced molecular adaptation in specific lineages in aquatic mammals, as is the case of the Delphinidae family.

Glutathione peroxidases are also highly active in aquatic mammal tissues (Wilhelm Filho et al. 2002; Vázquez-Medina et al. 2006, 2011b; Righetti et al. 2014). This gene family includes glutathione peroxidase 3, codified by the GPX3 gene, abundant in the cytoplasm of cells. The PSS found in our work and by Tian et al. (2021b) are not exclusive to aquatic mammals. The results of RELAX, which identified a relaxation on cetacean lineages, may explain the

higher omega value in Odontoceti. Glutathione peroxidase 3 is one of 6 other forms with different cellular locations and mechanisms of reaction and although being similar to glutathione peroxidase 1 (GPX1), GPX3 is secreted in plasma, while GPX1 acts inside the cell (Brigelius-Flohé and Maiorino 2013). Previous studies found positively selected sites and an increase in gene copies for GPX1 in cetaceans (Tian et al. 2021b).

The PRDX1 gene is highly conserved among terrestrial mammals, but we found evidence of relaxation on cetacean lineages, likely responsible for the high omega value in Mysticeti ( $\omega = 0.934$ ). No positive selection sites met the criteria

of complementary tests. Together with our findings, the presence of a higher number of copies of the PRDX1 gene in cetaceans, specially Mysticeti, suggests that the presence of extra copies for this gene allows a decrease in purifying selection, enabling an accumulation in neutral mutations without impacting gene function (Yim et al. 2014; Zhou et al. 2018).

On the other hand, the PRDX3 gene shows evidence of positive selection, with an acceleration of the evolutionary rate among crown cetaceans and pinnipeds when compared to the ancestral lineages and the presence of a PSS. The results in BSM identified many PSS among *Eschrichtius robustus*, *Eubalaena japonica*, *Megaptera novaeangliae*, and *Ziphius cavirostris*, including site 209, the only PSS with exclusive changes for cetaceans. These species were also artificially grouped in the phylogenetic reconstruction for this gene, indicating that they share mutations beyond the PSS (Supplementary Figs S13–S15). Also, an intensification in selective pressure was identified in Cetacea. Interestingly, the PRDX3 gene also has a copy number expansion in cetaceans, although not as many as PRDX1 (Yim et al. 2014). Peroxiredoxins are considered a conserved protein among metazoans (Hewitt and Degnan 2023), and our data, combined with other studies, provides evidence for their adaptive evolution in cetacean oxidative stress tolerance.

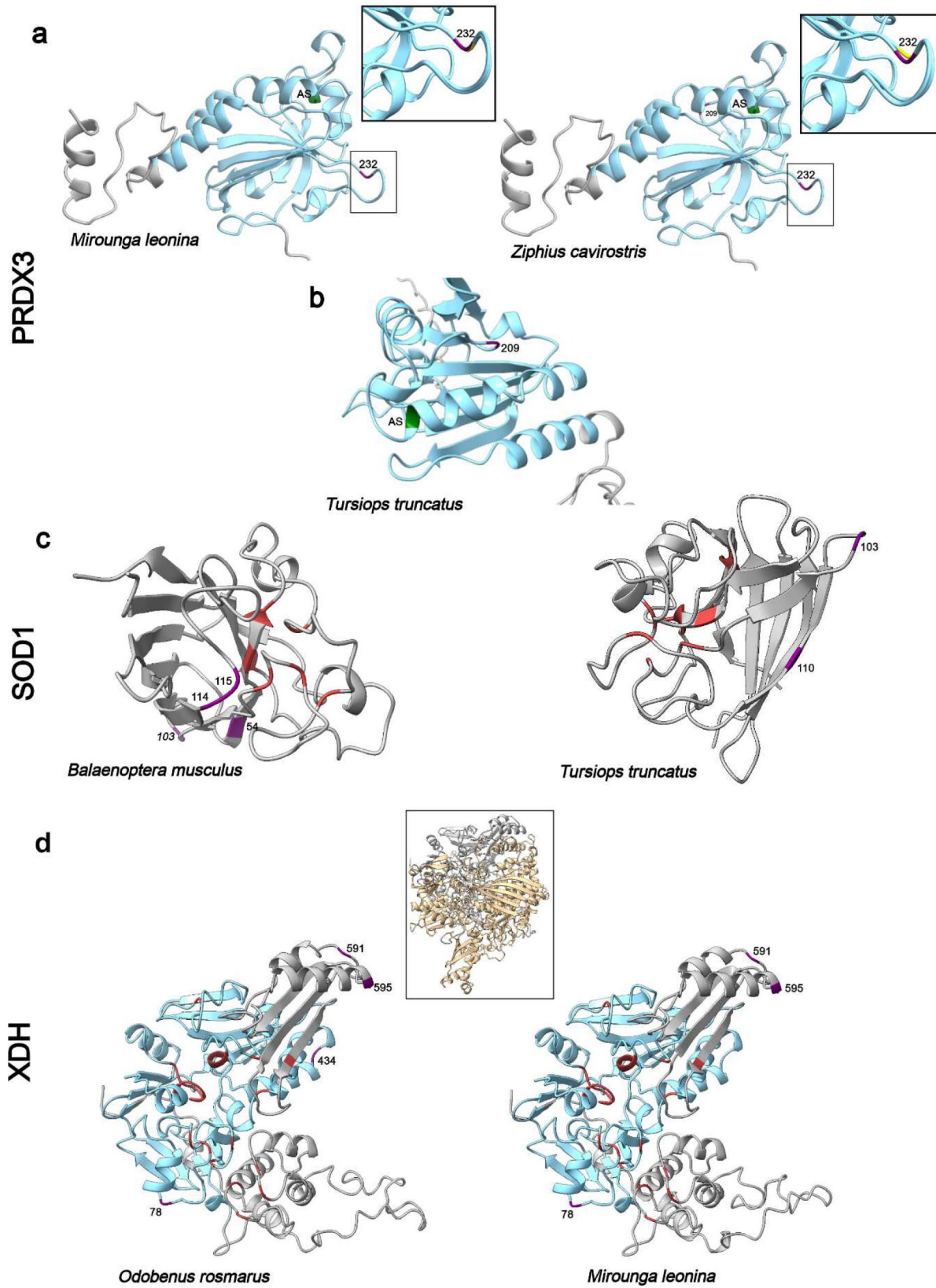
To evaluate which sites were shared exclusively between the groups of aquatic mammals and potentially adaptive, we enriched our dataset with multiple mammalian species. Although the selective pressures we identified in the aquatic mammals differed from those observed in their terrestrial counterparts, we found evidence of convergent mutations between these groups. While sites with identical amino acids are common in mammalian species and likely derived from neutral processes, adaptive modifications in species with similar phenotypic traits are rare (Foote et al. 2015; Chikina et al. 2016). Nevertheless, it is still possible that a shared mutation may be neutral in terrestrial species but confer an advantage in the aquatic environment and therefore be adaptive.

Our results suggest that some of the antioxidant genes evolved more rapidly in Cetacea than in Pinnipedia (Fig. 1), indicating that these genes did not converge in selective pressure intensity. As phenotype convergence may occur at various levels, including amino acid substitutions (Hao et al. 2019) we also examined sites at the same position in independent lineages to identify convergent parallel substitution (Zhou et al. 2015). Using this method, we detected site 232 in the PRDX3 gene as a potential target of positive selection, presenting the same amino acid substitution in both Cuvier's beaked whale (*Ziphius cavirostris*, Ziphiidae, Cetacea) and Southern elephant seal (*Mirounga leonina*, Phocidae, Pinnipedia). Besides presenting other known deep divers species in our dataset, such as sperm whale (*Physeter*

*catodon*, Physeteridae, Cetacea) and Pygmy sperm whale (*Kogia breviceps*, Kogiidae, Cetacea), the beaked whale and elephant seal have extraordinary aerobic diving limits without a significative increase in surface periods after longer dives (Hindell et al. 1992; Quick et al. 2020), spending less than 20% of their time in shallow waters (Hooker et al. 2015; Castellini and Mellish 2023).

Although the identification of site 232 was supported only by Contrast-FEL, which compares the rates of substitution between species and sites, we considered it relevant due to the evidence of intensified selection for the Cetacea clade and the high conservation of this site for other species in the alignment (Fig. 2). Furthermore, we can infer an impact on the protein structures, once it was classified as “Possibly Damaging” in human variations. The PRDX3 is part of the subgroup of peroxiredoxin containing 2-Cys as active sites, allowing them to build toroid structures, in this case, a dodecameric ring (Cao et al. 2005) This structure allows these enzymes to react with H<sub>2</sub>O<sub>2</sub> in normal concentrations and change its conformation when levels of peroxide increase, initiating a regulatory signaling pathway (Wood et al. 2003). In a bovine PRDX3, site 232 is faced in the inner part of the ring, located in the junction area between dimmers, surrounded by many hydrophobic residues that participate in the binding (Cao et al. 2005) (Supplementary Fig S29 a, b). The presence of this dodecameric structure was shown to be more active in humans compared to its dimmers, with the internal part of the ring playing a key role in controlling the enzyme activity (Cao et al. 2007; Yewdall et al. 2018). We also identified in the dodecameric bovine model (Cao et al. 2005), a hydrogen bond between Glycine (ancestral state of site 232) and Lysine (site 218 in our alignment); however, we could not compare with our models because AlphaFold did not consider the Lysine as a part of an alpha helix structure, changing its relative position and affecting the inference of hydrogen bonding by ChimeraX. For these reasons, we consider the mutation in site 232 as a strong candidate for impacting the PRDX3 quaternary structure and its consequent activity in known deep/long divers species and future studies could focus on modeling the binding region in these non-model species.

**Fig. 3** Protein structure modeled using the sequences of aquatic mammals in the AlphaFold Colab (AlphaFold v2.3.1.). Light blue: Domains of the protein; Red: Binding sites; Dark green: Active sites; Purple: Positively selected sites (PSS); Yellow: PSS in the closely related species; AS—Active site identification. **a** PRDX3 gene from the two species of Cetacea and Pinnipedia with convergent PSS, 232, in evidence, comparing with the site in *Canis lupus familiaris* for the *Mirounga leonina* and *Bos taurus* for *Ziphius cavirostris*; **b** PRDX3 from *Tursiops truncatus* with PSS 209 marked; **c** SOD1 for two species of cetaceans, a Mysticeti and an Odontoceti showing the position of different PSS for each group and the PSS 103, shared between them; **d** region of *Felis catus*'s XDH protein where the pinnipeds sequences were modeled with the PSS located in Odobenidae and Phocidae



Amino acid substitutions can modify proteins' physical and chemical properties, ultimately impacting their structure and function (Betts and Russell 2003). In our study, while some sites had benign changes that did not significantly impact protein structure, others presented more radical changes that could affect protein function. For example, site 54 in SOD1 is located near the protein's binding site, and mutations in this region can influence substrate selectivity and catalytic activity (Morley and Kazlauskas 2005). The presence of radical changes in conserved regions, such as protein domains, suggests a directional Darwinian selection, where residues are pushed toward non-synonymous changes (McClellan 2013). Despite providing valuable insights into the potential effects of these mutations on protein, further experimental approaches are required to confirm our findings and investigate regulatory regions and gene expression. Notably, aquatic mammals exhibit variations in amino acid properties that do not appear to have any negative effects, whereas similar changes in humans have been linked to various diseases (Ratovitski et al. 1999; Levartovsky et al. 2000; Rebelo et al. 2021). Therefore, aquatic mammals may have evolved mechanisms to tolerate and regulate these changes, providing valuable insights for future research into human health.

## Conclusion

We found that the colonization of new aquatic environments by marine mammals and the necessity to cope with oxidative stress derived from hypoxia in dives has affected the evolution of genes related to the antioxidant system. We identified divergent acceleration in evolutionary rates between the groups of Cetacea and Pinnipedia, meaning that besides similar selective pressures, the genes evolved differently between the two lineages. In addition, genes like GPX3 and PRDX1 presented differences within the recent evolutionary history and diversification of cetaceans and pinnipeds lineages. Furthermore, we were able to infer the non-random distribution of positively selected sites, once many were close to important regions of proteins and with mutations usually associated with radical impacts in amino acid properties and negative effects in humans. Besides the majority of findings pointing to different strategies for managing ischemia and reperfusion in the aquatic mammals' lineages, nevertheless, we were able to identify a convergent evolution between two extreme divers in a site located in a binding region of PRDX3 when forming dodecameric structures. Overall, we were able to establish an extensive comparison analysis between antioxidant enzyme evolution in our two lineages of interest, pinpointing interesting sites that could be investigated with additional experimental

data, something that due to limitations of our study, we were unable to address.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00239-024-10170-3>.

**Acknowledgements** This study was funded by the São Paulo Research Foundation (FAPESP) (2020/03588-2). LM was funded by Nova Scotia Graduate Scholarship, NSERC Discovery Grant and Biology Graduate A- Fellowship. AP was funded by FAPESP training program (2021/03325-4), FAS was funded by Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) master's scholarship, and ER by FAPESP doctoral scholarship (2018/01236-1). We are grateful to the scientists that made available the gene sequences used in this study.

**Author contributions** G.S.V. and M.F.N. conceived the research hypothesis. G.S.V., L.M., A.P., and F.A.S performed the data analysis. The manuscript was written by G.S.V. and previous versions were reviewed by E.R., L.M. and M.F.N. All authors read and approved the final manuscript.

## Declarations

**Conflict of interest** The authors have no competing interests to declare that are relevant to the content of this article.

## References

- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7(4):248–249
- Allen KN, Vázquez-Medina JP (2019) Natural tolerance to ischemia and hypoxemia in diving mammals: a review. *Front Physiol* 10(September):1199
- Berta A, Churchill M, Boessenecker RW (2018) The origin and evolutionary biology of pinnipeds: seals, sea lions, and walruses. *Annu Rev Earth Planet Sci* 46(1):203–228
- Betts MJ, Russell RB (2003) Amino acid properties and consequences of substitutions. *Bioinform Genet* 317:289
- Blix AS (2018) Adaptations to deep and prolonged diving in phocid seals. *J Exp Biol*. <https://doi.org/10.1242/jeb.182972>
- Brigelius-Flohé R, Maiorino M (2013) Glutathione peroxidases. *Biochem Biophys Acta* 1830(5):3289–3303
- Cabrera AA, Bérubé M, Lopes XM, Louis M, Oosting T, Rey-Iglesia A, Rivera-León VE, Székely D, Lorenzen ED, Palsbøll PJ (2021) A genetic perspective on cetacean evolution. *Annu Rev Ecol Evol Syst* 52(1):131–151
- Cantú-Medellín N, Byrd B, Hohn A, Vázquez-Medina JP, Zenteno-Savín T (2011) Differential antioxidant protection in tissues from marine mammals with distinct diving capacities. Shallow/short vs. deep/long divers. *Comp Biochem Physiol Part A Mol Integr Physiol* 158(4):438–443
- Cao Z, Roszak AW, Gourlay LJ, Lindsay JG, Isaacs NW (2005) Bovine mitochondrial Peroxiredoxin III forms a two-ring Catenane. *Structure* 13(11):1661–1664
- Cao Z, Bhella D, Lindsay JG (2007) Reconstitution of the mitochondrial PrxIII antioxidant defence pathway: general properties and factors affecting PrxIII activity and oligomeric state. *J Mol Biol* 372(4):1022–1033

- Castellini M, Mellish JA (2023) Physiology of marine mammals: adaptations to the ocean. CRC Press, Boca Raton
- Chance B, Sies H, Boveris A (1979) Hydroperoxide metabolism in mammalian organs. *Physiol Rev* 59(3):527–605
- Chikina M, Robinson JD, Clark NL (2016) Hundreds of genes experienced convergent shifts in selective pressure in marine mammals. *Mol Biol Evol* 33(9):2182–2192
- Davis RW (2014) A review of the multi-level adaptations for maximizing aerobic dive duration in marine mammals: from biochemistry to behavior. *J Comp Physiol B Biochem Syst Environ Physiol* 184(1):23–53
- Davis RW (2019) Marine mammals: adaptations for an aquatic life. Springer, Berlin
- Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform* 5(August):113
- Elsner R, Oyasaeter S, Almaas R, Saugstad OD (1998) Diving seals, ischemia-reperfusion and oxygen radicals. *Comp Biochem Physiol Part A Mol Integr Physiol* 119(4):975–980
- Foote AD, Liu Y, Thomas GWC, Vinař T, Alföldi J, Deng J, Dugan S et al (2015) Convergent evolution of the genomes of marine mammals. *Nat Genet* 47(3):272–275
- Fordyce RE (2018) Cetacean evolution. In: Würsig B, Thewissen JGM, Kovacs KM (eds) *Encyclopedia of marine mammals*, 3rd edn. Academic Press, pp 180–185
- García-Castañeda O, Gaxiola-Robles R, Kanatous S, Zenteno-Savín T (2017) Circulating glutathione concentrations in marine, semi-aquatic, and terrestrial mammals. *Mar Mamm Sci* 33(3):738–747
- Geßner C, Krüger A, Folkow LP, Fehrle W, Mikkelsen B, Burmester T (2022) Transcriptomes suggest that pinniped and cetacean brains have a high capacity for aerobic metabolism while reducing energy-intensive processes such as synaptic transmission. *Front Mol Neurosci*. <https://doi.org/10.3389/fnmol.2022.877349>
- Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE (2018) UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci: Publ Protein Soc* 27(1):14–25
- Goldman N, Yang Z (1994) Models of DNA substitution and the discrimination of evolutionary parameters. In: *Proceedings of the XVIIth international biometrics 1*
- Halliwell JM, Gutteridge B (2015) Free radicals in biology and medicine, vol Fifth. Oxford University Press, Oxford
- Hancock JT (2021) Oxygen is instrumental for biological signaling: an overview. *Oxygen* 1(1):3–15
- Hao Y, Qu Y, Song G, Lei F (2019) Genomic insights into the adaptive convergent evolution. *Curr Genom* 20(2):81–89
- Hewitt OH, Degnan SM (2023) Antioxidant enzymes that target hydrogen peroxide are conserved across the animal kingdom, from sponges to mammals. *Sci Rep* 13(1):1–13
- Hindell MA, Slip DJ, Burton HR, Bryden MM (1992) Physiological implications of continuous, prolonged, and deep dives of the southern elephant seal (*Mirounga leonina*). *Can J Zool* 70(2):370–379
- Hoang DT, Chernomor O, Haeseler AV, Minh BQ, Vinh LS (2017) UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35(2):518–522
- Hooker SK, Fahlman A, Castellini M, Mellish J (2015) Pressure regulation. *Marine mammal physiology: requisites for ocean living*, pp 69–92
- Janecka J, Chowdhary B, Murphy W (2012) Exploring the correlations between sequence evolution rate and phenotypic divergence across the mammalian tree provides insights into adaptive evolution. *J Biosci* 37(5):897–909
- Jones DP (2006) Redefining oxidative stress. *Antioxid Redox Signal* 8(9–10):1865–1879
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K et al (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873):583–589
- Kalogeris T, Baines CP, Krenz M, Korhuis RJ (2012) Cell biology of ischemia/reperfusion injury. *Int Rev Cell Mol Biol* 298:229–317
- Kelley EE, Khoo NKH, Hundley NJ, Malik UZ, Freeman BA, Tarpey MM (2010) Hydrogen peroxide is the major oxidant product of xanthine oxidase. *Free Radical Biol Med* 48(4):493–498
- Kirkman HN, Gaetani GF (2007) Mammalian catalase: a venerable enzyme with new mysteries. *Trends Biochem Sci* 32(1):44–50
- Kosakovsky P, Sergei L, Frost SDW (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22(5):1208–1222
- Kosakovsky P, Sergei L, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD et al (2020a) HyPhy 2.5-A customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol* 37(1):295–299
- Kosakovsky P, Sergei L, Wisotsky SR, Escalante A, Magalis BR, Weaver S (2020b) Contrast-FEL—a test for differences in selective pressures at individual sites among clades and sets of branches. *Mol Biol Evol* 38(3):1184–1198
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2017) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol* 34(3):772–773
- Levartovsky D, Lagziel A, Sperling O, Liberman U, Yaron M, Hosoya T, Ichida K, Peretz H (2000) XDH gene mutation is the underlying cause of classical xanthinuria: a second report. *Kidney Int* 57(6):2215–2220
- Löytynoja A (2014) Phylogeny-aware alignment with PRANK. *Methods Mol Biol* 1079:155–170
- Maldonado E, Sunagar K, Almeida D, Vasconcelos V, Antunes A (2014) IMPACT\_S: integrated multiprogram platform to analyze and combine tests of selection. *PLoS ONE* 9(10):e96243
- Martens GA, Folkow LP, Burmester T, Geßner C (2022) Elevated antioxidant defence in the brain of deep-diving pinnipeds. *Front Physiol* 13(December):1064476
- McClellan DA (2013) Directional Darwinian Selection in Proteins. *BMC Bioinform* 14(Suppl 13):S6
- McDonald BI, Ponganis PJ (2013) Insights from venous oxygen profiles: oxygen utilization and management in diving California sea lions. *J Exp Biol* 216(Pt 17):3332–3341
- McGowen MR, Gatesy J, Wildman DE (2014) Molecular evolution tracks macroevolutionary transitions in Cetacea. *Trends Ecol Evol* 29(6):336–346
- Michiels C, Raes M, Toussaint O, Remacle J (1994) Importance of Se-glutathione peroxidase, catalase, and Cu/Zn-SOD for cell survival against oxidative stress. *Free Radical Biol Med* 17(3):235–248
- Morley KL, Kazlauskas RJ (2005) Improving enzyme properties: when are closer mutations better? *Trends Biotechnol* 23(5):231–237
- Murphy MP (2009) How mitochondria produce reactive oxygen species. *Biochem J* 417(1):1–13
- Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky SL, Scheffler K (2013) FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. *Mol Biol Evol* 30(5):1196–1205
- Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K et al (2015) Gene-wide identification of episodic selection. *Mol Biol Evol* 32(5):1365–1371
- Nery MF, Arroyo JI, Opazo JC (2013) Accelerated evolutionary rate of the myoglobin gene in long-diving whales. *J Mol Evol* 76(6):380–387

- Nguyen LT, Schmidt HA, Haeseler AV, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32(1):268–274
- Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148(3):929–936
- Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE (2021) UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci: Publ Protein Soc* 30(1):70–82
- Pond PLK, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21(5):676–679
- Pongonis PJ (2019) State of the art review: from the seaside to the bedside: insights from comparative diving physiology into respiratory, sleep and critical care. *Thorax* 74(5):512–518
- Quick NJ, Cioffi WR, Shearer JM, Fahlman A, Read AJ (2020) Extreme diving in mammals: first estimates of behavioural aerobic dive limits in Cuvier's beaked whales. *J Exp Biol*. <https://doi.org/10.1242/jeb.222109>
- Ramirez JM, Folkow LP, Blix AS (2007) Hypoxia tolerance in mammals and birds: from the wilderness to the clinic. *Annu Rev Physiol* 69:113–143
- Ratovitski T, Corson LB, Strain J, Wong P, Cleveland DW, Culotta VC, Borchelt DR (1999) Variation in the biochemical/biophysical properties of mutant superoxide dismutase 1 enzymes and the rate of disease progression in familial amyotrophic lateral sclerosis kindreds. *Hum Mol Genet* 8(8):1451–1460
- Rebello AP, Eidhof I, Cintra VP, Guillot-Noel L, Pereira CV, Timmann D, Traschütz A et al (2021) Biallelic loss-of-function variations in PRDX3 cause cerebellar Ataxia. *Brain: J Neurol* 144(5):1467–1481
- Righetti BPH, Simões-Lopes PC, Uhart MM, Wilhelm Filho D (2014) Relating diving behavior and antioxidant status: insights from oxidative stress biomarkers in the blood of two distinct divers, *Mirounga Leonina* and *Arctocephalus Australis*. *Comp Biochem Physiol Part A Mol Integr Physiol* 173C(July):1–6
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572–1574
- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL (2015) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol* 32(5):1342–1353
- Tian R, Wang Z, Niu X, Zhou K, Xu S, Yang G (2016) Evolutionary genetics of hypoxia tolerance in cetaceans during diving. *Genome Biol Evol* 8(3):827–839
- Tian R, Seim I, Ren W, Xu S, Yang G (2018) Comparative genomics reveals contraction in cytosolic glutathione transferase genes in cetaceans: implications for oxidative stress adaptation. [bioRxiv. https://doi.org/10.1101/485615](https://doi.org/10.1101/485615)
- Tian R, Seim I, Ren W, Xu S, Yang G (2019) Contraction of the ROS scavenging enzyme glutathione S-transferase gene family in cetaceans. *G3: Genes Genom Genet* 9(7):2303–2315
- Tian R, Geng Y, Guo H, Yang C, Seim I, Yang G (2021a) Comparative analysis of the superoxide dismutase gene family in cetartiodactyla. *J Evol Biol* 34(7):1046–1060
- Tian R, Geng Y, Yang Y, Seim I, Yang G (2021b) Oxidative stress drives divergent evolution of the glutathione peroxidase (GPX) gene family in mammals. *Integr Zool* 16(5):696–711
- Tian R, Yang C, Chai SM, Guo H, Seim I, Yang G (2022) Evolutionary impacts of purine metabolism genes on mammalian oxidative stress adaptation. *Zool Res* 43(2):241–254
- Upham NS, Esselstyn JA, Jetz W (2019) Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol* 17(12):e3000494
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D et al (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 50(D1):D439–D444
- Vázquez-Medina JP, Zenteno-Savín T, Elsner R (2006) Antioxidant enzymes in ringed seal tissues: potential protection against dive-associated ischemia/reperfusion. *Comp Biochem Physiol Toxicol Pharmacol: CBP* 142(3–4):198–204
- Vázquez-Medina JP, Zenteno-Savín T, Elsner R (2007) Glutathione protection against dive-associated ischemia/reperfusion in ringed seal tissues. *J Exp Mar Biol Ecol* 345(2):110–118
- Vázquez-Medina JP, Zenteno-Savín T, Tift MS, Forman HJ, Crocker DE, Ortiz RM (2011a) Apnea stimulates the adaptive response to oxidative stress in elephant seal pups. *J Exp Biol* 214(Pt 24):4193–4200
- Vázquez-Medina JP, Soñanez-Organis JG, Burns JM, Zenteno-Savín T, Ortiz JM (2011b) Antioxidant capacity develops with maturation in the deep-diving hooded seal. *J Exp Biol* 214(Pt 17):2903–2910
- Vázquez-Medina JP, Zenteno-Savín T, Elsner R, Ortiz RM (2012) Coping with physiological oxidative stress: a review of antioxidant strategies in seals. *J Comp Physiol B Biochem Syst Environ Physiol* 182(6):741–750
- Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K (2015) RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol* 32(3):820–832
- Wilhelm Filho D, Sell F, Ribeiro L, Ghislandi M, Carrasquedo F, Fraga CG, Wallauer JP, Simões-Lopes PC, Uhart MM (2002) Comparison between the antioxidant status of terrestrial and diving mammals. *Comp Biochem Physiol Part A Mol Integr Physiol* 133(3):885–892
- Wood ZA, Poole LB, Karplus PA (2003) Peroxiredoxin evolution and the regulation of hydrogen peroxide signaling. *Science* 300(5619):650–653
- Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003) TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19(5):671–672
- Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15(5):568–573
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591
- Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol* 46(4):409–418
- Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19(6):908–917
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155(1):431–449
- Yewdall NA, Peskin AV, Hampton MB, Goldstone DC, Pearce FG, Gerrard JA (2018) Quaternary structure influences the peroxidase activity of peroxiredoxin 3. *Biochem Biophys Res Commun* 497(2):558–563
- Yim HS, Cho YS, Guang X, Kang SG, Jeong JY, Cha SS, Oh HM et al (2014) Minke whale genome and aquatic adaptation in cetaceans. *Nat Genet* 46(1):88–92
- Yuan Y, Zhang Y, Zhang P, Liu C, Wang J, Gao H, Hoelzel AR et al (2021) Comparative genomics provides insights into the aquatic adaptations of mammals. *Proc Natl Acad Sci USA*. <https://doi.org/10.1073/pnas.2106080118>
- Zapol WM, Liggins GC, Schneider RC, Qvist J, Snider MT, Creasy RK, Hochachka PW (1979) Regional blood flow during simulated



- diving in the conscious weddell seal. *J Appl Physiol: Respir Environ Exerc Physiol* 47(5):968–973
- Zelko IN, Mariani TJ, Folz RJ (2002) Superoxide dismutase multi-gene family: a comparison of the CuZn-SOD (SOD1), Mn-SOD (SOD2), and EC-SOD (SOD3) gene structures, evolution, and expression. *Free Radical Biol Med* 33(3):337–349
- Zenteno-Savín T, Hernández C, Elsner R (2002) Diving seals: are they a model for coping with oxidative stress? *Comp Biochem Physiol Toxicol Pharmacol: CBP* 133(4):527–536
- Zenteno-Savín T, Vázquez-Medina JP, Cantú-Medellín N, Ponganis PJ, Elsner R (2011) Ischemia/reperfusion in diving birds and mammals: how they avoid oxidative damage. In: *Oxidative stress in aquatic ecosystems*. John Wiley & Sons, Ltd, Chichester, pp 178–89
- Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22(12):2472–2479
- Zhou X, Seim I, Gladyshev VN (2015) Convergent evolution of marine mammals is associated with distinct substitutions in common genes. *Sci Rep* 5(November):16550
- Zhou X, Sun D, Guang X, Ma S, Fang X, Mariotti M, Nielsen R, Gladyshev VN, Yang G (2018) Molecular footprints of aquatic adaptation including bone mass changes in cetaceans. *Genome Biol Evol* 10(3):967–975

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.