Supplementary Materials for

Ancient convergent losses of *Paraoxonase 1* yield potential risks for modern marine mammals

Wynn K. Meyer, Jerrica Jamison, Rebecca Richter, Stacy E. Woods, Raghavendran Partha, Amanda Kowalczyk, Charles Kronk, Maria Chikina, Robert K. Bonde, Daniel E. Crocker, Joseph Gaspard, Janet M. Lanyon, Judit Marsillach, Clement E. Furlong, and Nathan L. Clark

correspondence to: nclark@pitt.edu

This PDF file includes:

Materials and Methods
Figs. S1 to S6
Tables S2, S3, S6, S8, S10, and S11
Captions for Additional Data tables S1, S4, S5, S7, S9, and S12

Other Supplementary Materials for this manuscript includes the following:

Additional data tables S1, S4, S5, S7, S9, S10, and S13 are available as Excel files.

Materials and Methods

Scoring gene orthologs as functional or pseudogenes across eutherian mammals in the 100-way alignment

Our first goal was to assess whether the annotated sequence for each species in each gene's publicly available alignment represented a functional gene or an unprocessed pseudogene. We used the following pipeline to identify genes that displayed strong evidence of having lost function in any species, using the sequence data from the hg19 UCSC 100-way alignment (http://genome.ucsc.edu/). We scanned amino acid sequences for stop characters (Z) and nucleotide sequences for frameshifts, using the following filters to exclude putative lesions that would be unlikely to disrupt function or could be caused by issues of data quality:

1. For first and last exons, which are known to be highly variable and sometimes alternatively spliced, we excluded the whole exon from our scans if it made up no more than 10% of the entire gene sequence, or the terminal 60 bp (20 aa) otherwise.
2. We additionally excluded from our scans any exons containing more than 25% gap characters, as these might represent incorrectly identified orthologs or low quality genomic data resulting in poor quality alignments.
3. We excluded any pairs of frameshifts that were within 15 bp of each other, since these could also represent issues with alignment and/or data quality.
4. As a further check against including erroneous frameshifts caused by gaps in reference sequence data, we excluded any frameshifts greater than 8 bp in length.
5. To avoid erroneous frameshifts caused by errors in aligning exon boundaries, we excluded the first and last three bp of each exon from our frameshift scan.

For the sequence of each eutherian mammal species in the alignment for each gene, we estimated the proportion of gaps within the sequence included after the above filters were applied. When we excluded an entire internal exon, we counted all of its sequence as gap characters. Where the proportion of gaps differed between nucleotide and amino acid estimates, we chose the larger value. We then set a threshold of the proportion of gap characters above which to exclude a call at a gene for a given species in order to limit the probability of erroneously calling a pseudogene functional (i.e., the rate of false negative pseudogene calls). Specifically, we sought to minimize the following:

$$P_{false\_negative} = ((N_{pseudogenes} / \sum P_{non\text{-}missing}) \cdot \sum P_{missing}) / N_{genes},$$

where $N_{pseudogenes}$ represents the total number of pseudogenes in the included set, $\sum P_{non\text{-}missing}$ represents the sum of the proportions of non-gap characters for all genes in the included set, $\sum P_{missing}$ represents the sum of the proportions of gap characters for all genes in the included set, and N_{genes} represents the total number of genes included at this threshold across all species. In effect, this estimates the rate at which pseudogenes occur per total amount of non-missing data, and then estimates how many pseudogenes would be unobserved based on the total amount of missing data within the included gene set. We chose a threshold that we estimated would result in a false negative pseudogene once every 10 genes (i.e., at most one erroneously called functional gene in any of the 58 species every 10 genes). Any gene sequence exceeding this threshold (16%) for proportion of gap characters for a given species was excluded (not called) for that species.

Several alignments in the dataset had multiple University of California 'known gene' identification numbers (UCIDs) corresponding to a single gene symbol. When we found such

instances, we identified the UCID that had the smallest total proportion of gap characters (missing data) across all species and used only the data from the alignment for that UCID for the corresponding gene symbol; if multiple UCIDs had equivalent missing data, we excluded that gene from analysis (this situation was encountered for eight genes).

Manual validation and estimation of error rates

Our pipeline for identifying lesions and calling candidate pseudogenes may be subject to various sources of error. We performed several checks to rule out errors in orthology and to estimate error rates stemming from issues with our automated method of pseudogene calling and issues in the 100-way vertebrate alignment, and the next three sections describe these checks. We performed all of these assessments for a 'test set' comprised of the following: the top 20 genes from our analysis of marine-dependent loss, 20 other randomly selected genes, and five well-studied cases of gene non-functionalization, three of which passed filters for inclusion in our analysis.

Manual validation of orthology

To ensure that lesions identified in the test set were not the result of the mis-identification of orthologs for certain species, we first checked for errors in orthology by building an 85% consensus parsimony tree from the coding sequence alignment and determining whether any species or set of species not excluded for missing data was positioned as an outgroup to the remainder of the phylogeny; we built trees using the PHYLIP v3.696 dnapars algorithm (*34*), as implemented in the SeaView (version 4.6.1) software (*35*). In the single case in which species with predicted pseudogenes were positioned as outgroups, we verified that this was due to long branch attraction by determining that this protein was the best match in the human genome for the amino acid sequences of the species with predicted pseudogenes using blastp (*36*). We also validated that our predicted pseudogenes represented losses of function without loss of synteny (i.e., unprocessed, or unitary, pseudogenes) by comparing the genes within the region surrounding the predicted pseudogene in the genome of the species for which the gene was predicted to be a pseudogene with the genes surrounding this gene's ortholog in the human genome, using the UCSC Genome Browser (*37*).

Manual validation of automated method for calling lesions from 100-way alignment

To assess the reliability of our criteria to call lesions from the 100-way alignment, we manually validated predicted pseudogene calls and predicted functional ("intact") gene calls against the original sequence in the 100-way alignment for all genes in the test set (Table S9). We translated coding sequences to amino acids and manually checked for the presence of lesions in predicted pseudogenes and predicted functional genes that were not less than ten amino acids from the end of the gene. Using these manual checks, we identified 22 cases where a potential pseudogene may have been mis-identified as a functional gene (false negatives) and five cases where a predicted pseudogene may have been falsely called (false positives) across all non-excluded sequences for these 42 genes, leading to method-based error rates of 7.14% and 0.26% for false negatives and false positives, respectively. All false positives were in terrestrial species; one case resulted from two independent frameshifts that, when combined, brought the sequence back into frame, and another resulted from a premature stop codon encoded within the penultimate exon that fell only four amino acids from the sequence end (Table S9). It is important to note that these errors represent cases in which the filters in the automated method

may lead to inaccurate functional/pseudogene classification of sequences within the 100-way alignment, assuming that the alignment itself contains no errors. We separately estimated errors in the alignment (see next section).

Estimating error rates in pseudogene calls made from the 100-way alignment

To assess the reliability of our predicted pseudogene calls and to ensure that our strongest results were not driven by erroneous pseudogene calls within marine lineages, we further determined the rate of errors in calling predicted pseudogenes using sequences from reference genomes for all genes in the test set (Table S10).

We initially called potential pseudogenes for each species using the aligned coding sequences available in the '100-way alignment' from the UCSC Genome Browser (based on human genome version 'hg19') (*37*). For the 45 test set genes, we manually examined each lesion (stop codon or frameshift-causing deletion) in the original source genome that was used for the 100-way alignment. This validation allowed us to check for mis-called lesions resulting from the alignment process. There are other potential sources of error, including the genome assembly itself. However, newer genome assemblies were only available for nine of the 58 species, five of which were primates, making it challenging to assess errors due to genome assembly globally.

We obtained the source genome assemblies for all 57 non-human species from either UCSC or NCBI ('download_genomes.sh' in folder 'Estimating_error_rates' within our github repository). We made BLAST nucleotide databases for each species' genome (*36*). For each lesion-containing exon, we obtained the highest-scoring BLASTn hit from that species' genome ('match_lesions_with_genomes.sh') and manually inspected each apparent lesion using its coordinates and flanking sequence. Validation outcomes were tabulated for each lesion and for the resulting "potential pseudogene" call for each gene (Table S10). Those counts were used to calculate 3 statistics:

1) error rate per pseudogene call, *i.e.*, false positive pseudogenes / all pseudogene calls
2) false positive rate, *i.e.*, false positive pseudogenes / (false positive pseudogenes + true functional genes)
3) per-lesion error rate, *i.e.*, false positive lesions / all lesion calls.

We report all rates and calculations in Table S10. Generally, error rates show that these data are useful in the capacity of a screen to identify potential subjects of convergent loss of gene function. However, the error rates are suitably high (false positive rate 3.4% and per-lesion error rate 16%) that all lesion calls should be validated by independent methods, such as Sanger sequencing, as we did for *PON1*. In a single case, errors were seen to affect marine species in one of the top ten genes in our analysis; that gene, *SLC39A9*, was excluded from Table 1.

Assessing the potential for biased errors between marine and terrestrial species

To determine whether our scan for marine convergent functional loss might be biased due to a high error rate within our automated potential pseudogene calls specifically for marine species, we compared rates of pseudogene calls at presumed functional genes between marine and terrestrial species. For this comparison, we evaluated the rate at which a gene set previously identified as being highly conserved across eukaryotes contained called pseudogenes, under the assumption that these highly conserved genes have a low probability of losing function in the

mammalian species within our dataset. We obtained RefSeq protein accession numbers for the 248 core eukaryotic genes that tend to be present as single-copy genes across six diverse high-quality eukaryotic genomes (*38*), frequently used for assessing completeness of draft genomes using the CEGMA protocol (*39*). We translated these to gene symbols using the bioDBnet tool (*40*), and we further manually corrected gene symbols that differed between RefSeq and UCSC by identifying genes that uniquely overlapped across 100% of their length in the UCSC hg19 browser (http://genome.ucsc.edu/) (*37*, *41*). We then determined, for each species, the proportion of these CEGMA genes that were called as pseudogenes using our automated method, out of all CEGMA genes that were not filtered for missing data. The estimated error rates based on CEGMA pseudogenes for marine species fall within the range of those for terrestrial species, overlapping with the distribution of such error rates for terrestrial species that have similar genetic distances to reference sequence hg19 (Fig. S4).

Identifying signatures of convergent loss of gene function in marine mammals

The predicted pseudogene status results formed a gene-by-species matrix of gene presence (functional) / absence (pseudogene) / excluded (not assigned). We excluded any genes that had gene calls excluded for at least one third, or 19, of the 58 total species, reasoning that these may represent cases where data quality was poor or orthologs were incorrectly identified across multiple species. We then selected the set of genes that were designated predicted pseudogenes in at least two species and at most 29 species, since that range would be best powered to identify marine-specific loss. We ran two nested likelihood models in BayesTraits (*42*) version 3 using the remaining 9,950 gene vectors and a vector indicating which species are 'marine' and 'terrestrial'. The independent model contained two parameters – a gene loss rate (the rate at which a functional gene becomes a predicted pseudogene) and a rate for transition from terrestrial to marine status. Because our study was focused on gene loss, gene gain was not allowed; its rate was constrained to zero. Similarly, the rate for transition from a marine to terrestrial state was constrained to zero, since this transition is not observed in the placental mammalian phylogeny. This independent model contained no relationship between gene loss and marine/terrestrial state, and so it served as the null hypothesis. The dependent model, on the other hand, added another free parameter by dividing the gene loss rate into two parameters – loss rate on terrestrial and marine branches, separately. We compared these two nested models using a likelihood ratio test (LRT). Since we were interested in the evidence for higher loss on marine branches, we reversed the sign of the LRT statistic for all genes inferred to have a higher loss rate on terrestrial branches in the independent model.

The distribution of our modified LRT statistic deviates from the chi-square distribution with 1 degree of freedom, due to the effects of sample size limitations and restricted parameter ranges, as well as to the reversal of sign for genes with higher terrestrial loss rates. To estimate empirical *P*-values for each gene based on the distribution of this modified statistic under the null, we performed simulations of gene loss across the mammalian phylogeny. To recapitulate the pattern of loss for each gene, we set branch lengths to the genome-wide average amino acid distances, multiplied by the gene's inferred loss rate from the independent model of BayesTraits. We stratified the number of simulated datasets per gene based on each gene's likelihood ratio test *P*-value assuming a chi-square distribution – 10 million simulations for the genes with $P < 10^{-5}$ (genes ranked 1-3), 1 million simulations for genes with $P < 10^{-4}$ (genes ranked 4-10), 100,000 simulations for genes with $P < 10^{-3}$ (genes ranked 10-195), and 10,000 simulations for the rest. We generated simulated datasets using the 'sim.char' function in the R package 'geiger,'(*43*);

our simulation-based *P*-value, reported as "Empirical *P*-value" in Tables 1 and S1, represents the proportion of simulations with a higher modified LRT statistic than that observed for the gene of interest.

In order to estimate empirical study-wide false discovery rates (FDR), we simulated datasets matching the evolution of the 13,853 genes in our dataset with at least one pseudogene and at least one functional gene among at least 39 species with non-excluded gene status. We simulated 10,000 datasets per gene using the methods described above. We subsequently filtered the simulated datasets to include only simulated genes with a minimum of 2 pseudogenes and a maximum of 29, to create a null dataset of simulated genes subject to the same filters as the real dataset. We then used the distribution of test statistics from simulated genes to estimate the FDR in an approach similar to empirical permutation-based FDR calculations. Studies that perform permutation-based FDR calculations commonly use a modification of the Benjamini-Hochberg procedure wherein they compare observed test statistics with empirically defined null distributions obtained from repeated permutations of the data and labels, in place of the procedure's traditional comparison of observed *P*-values to a null distribution based on uniform quantiles (*44*, *45*). However, in our case permuting tip labels would frequently change the branch lengths on which functional losses could occur and modify the well-supported relationships among foreground species. In our analysis, we thus use the same modified Benjamini-Hochberg procedure to compare the observed modified LRT statistic distributions to the distribution of modified LRT statistics for simulated genes (the empirical null distribution), in place of the distribution of a permutation-based test statistic. This approach results in FDR calculations based on test statistic distributions from a null dataset more closely matching the true dataset, commonly preferred in genomic data analysis.

While our simulation approach enables the generation of an empirical null distribution based on multiple datasets preserving phylogenetic relationships among foreground species, we also compared our results to those from a single permutation wherein we selected a set of foreground lineages whose branch lengths and relationships to other foreground species were matched to those of the marine species, but which are not known for convergence in any phenotype or environment. Specifically, those species were the aardvark, alpaca, Bactrian camel, little brown bat, and David's Myotis bat (*5*). We applied our genome-wide scan for convergence to this matched foreground set using the likelihood-based methods implemented in BayesTraits, as previously described.

Considering only genes showing higher inferred marine loss rates, we see some evidence for enrichment of genes in the real dataset showing higher raw LRTs compared to the null distribution obtained from simulations or the distribution for a single matched foreground set (Fig. S5A and C). In strong contrast, genes in the real dataset with higher inferred terrestrial loss rates do not show comparable enrichment for higher LRTs relative to the simulated null distribution or matched foreground distribution (Fig. S5B and D). This suggests that there is empirical evidence for enrichment of genes showing marine-biased pseudogenization.

Functional enrichment analyses

To generate a ranked list for enrichment tests, we ranked genes in descending order by LRT statistic; we reversed the sign of the LRT statistic for genes with higher inferred loss rates on terrestrial branches than on marine branches, since in these cases large LRT would represent evidence against marine-biased loss. This ranked list was tested for functional enrichment using the Gene Ontology annotations available through the GOrilla server (*46*), and the set of the top

137 genes (representing a false discovery rate, or FDR, of 25%; see previous section) was tested for functional enrichment using the MSigDB canonical, curated, and biological process gene ontology databases (*47*) and the MGI mammalian phenotypes database (*48*), with mammalian phenotype sets built by compiling lists of gene symbols associated with each phenotype and including all genes for a given phenotype as part of the set associated with that phenotype's ancestors in the ontology (from https://bioportal.bioontology.org/ontologies/MP, last accessed June 6, 2016). To test for functional enrichment using these datasets, we performed a hypergeometric test using the set of 9,950 genes that passed inclusion filters (see above) as our background gene set. We corrected for multiple testing in these analyses using the Benjamini-Hochberg procedure (*45*).

Phylogenetic tree for analyses

For all analyses, we used the same tree topology, based on that inferred by Meredith et al. (*17*). In several cases where this tree differed from that inferred by Bininda-Emonds et al. (*49*), we chose a consensus topology based on studies that inferred the local phylogeny using focused sampling of species within the clade of interest. Specifically, we set the star-nosed mole as an outgroup to the hedgehog and shrew (*50*, *51*); the cow as an outgroup to the Tibetan antelope, sheep, and goat (*52*, *53*); and the ursids as an outgroup to mustelids and pinnipeds (*54*, *55*). For inferring date of PON1 functional loss in pinnipeds, we estimated d_N/d_S separately using mustelids and ursids as the sister clade, to demonstrate robustness of the dating to assumptions about the local topology. The full tree topology, incorporating new species added specifically for *PON1* analyses (see next section) is provided below ("Phylogenetic trees used for evolutionary inferences", tree #1). For analyses that required branch lengths (including tests for marine convergent functional loss in BayesTraits), we estimated branch lengths on this consensus species tree topology using the average branch lengths from a large set of trees as follows. We chose a set of genes in which each gene had a sequence from each of the 58 species. For each gene, we estimated branch lengths using *codeml* on the fixed tree topology with an amino acid model (*56*). We scaled the resulting trees to unit vector length, and the average of each scaled branch length across all genes present in all species became the representative branch length in the master tree.

Assessing robustness of results to variation in the phylogenetic tree

Given that the consensus tree topology used in our analyses (see previous section) may not accurately represent the evolutionary history of all genes due to incomplete lineage sorting or post-divergence gene flow, we assessed the robustness of our results to variation in the tree by performing inferences of convergence for our 20 top genes using 14 alternate trees. For these analyses, we used trees with branch lengths inferred from the concatenated sequence alignment of 10 genes, assuming the following tree topologies: two previously published mammalian supertrees inferred from multiple nuclear loci (*17*, *49*), the tree provided by UCSC and used as a guide for the 100-way alignment (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz100way/), and our original consensus tree topology (with branch lengths re-estimated using the same gene set as for the alternate topologies). We also used trees generated from single-gene alignments for a different set of 10 genes. These different trees represent a range of realistic relationships among the included species (see "Phylogenetic trees used for evolutionary inferences" below).

We obtained the two previously published trees and the UCSC tree directly from their respective sources and pruned them as necessary to create subtree topologies containing only the 58 species of interest. We randomly resolved polytomies within the Bininda-Emonds (*49*) tree using the multi2di function from the ape package in R (*57*). To provide sequence input for estimating branch lengths, we concatenated the nucleotide sequence alignments for ten randomly selected genes that were called as functional in all 58 species and met the following two restrictions: sequence alignments were required to be 500 nucleotides long and contain between 30% and 70% variable sites (BECN1, CLDN4, DNAJC5B, FBXO30, GINS4, GPR22, LPAR6, SARIA, SMPD2, and SNX16). For each of the four pre-determined topologies (two published, UCSC, and our consensus topology), we estimated branch lengths from this concatenated alignment, using *codeml* (*56*). We used a codon model with equilibrium codon frequencies calculated from the alignment, a fixed d_N/d_S ratio across branches, a neutral selection model, and estimated kappa and omega values.

To generate additional trees from single-gene alignments, we randomly selected ten different genes that were called as functional in all 58 species and whose alignments were at least 1000 nucleotides long (FICD, FUT9, GPR22, HMGCS1, LRFN5, MSL1, NUDT12, PTPRA, SGMS2, and TTC5). We estimated maximum likelihood trees from the nucleotide alignments for each of these genes using PhyML, as implemented in seaview4 (*35*, *58*). We used default PhyML settings: a GTR model, aLRT branch support, empirical nucleotide equilibrium frequencies, no invariable sites, optimized across site rate variation with four rate categories, NNI tree searching operation, and optimized tree topology with a BioNJ starting tree and five random starts.

We assessed evidence for convergent loss of function in marine species as previously described for the 20 genes with the highest LRT statistics in our original analysis, using each of the 14 trees estimated above. The resulting LRT statistics for these genes are largely consistent across all trees (Fig. S6). These results suggest that our conclusions are robust to variations in tree toplogy among gene trees, and therefore to uncertainty in our chosen consensus tree due to incomplete lineage sorting or post-divergence gene flow.

Adding *PON1* sequences for mammalian species not in the 100-way alignment

In order to provide a more complete representation of *PON1* sequences for marine and semi-aquatic species, as well as species within other clades of interest based on evolutionary rates, we first obtained the following species' publicly available *PON1* coding sequences and added them to the mammalian subset of the 100-way alignment: Brandt's bat (*Myotis brandtii*)(*59*), Canadian beaver (*Castor canadensis*)(*60*), Hawaiian monk seal (*Neomonachus schauinslandi*)(*61*), minke whale (*Balaenoptera acutorostrata*)(*62*), Natal long-fingered bat (*Miniopterus natalensis*)(*63*), polar bear (*Ursus maritimus*)(*64*), sea otter (*Enhydra lutris*)(*65*), sperm whale (*Physeter macrocephalus*) and Yangtze River dolphin (*Lipotes vexillifer*)(*66*). To address issues with the annotation of exon boundaries in sea otter, we downloaded the full gene sequence, including introns, and annotated exons manually, using the results from a discontiguous megablast of the ferret coding sequence as a guide (*67*). We obtained the predicted sequence of *PON1* for Antarctic fur seal (*Arctocephalus gazella*) by downloading genomic scaffolds (*68*) and identifying sequences orthologous to Weddell seal *PON1* coding sequence using BLAT v36x1 (*69*). We obtained publicly available RNA sequencing read data from liver for hippopotamus (*Hippopotamus amphibius*) and spotted seal (*Phoca largha*) (*70*). We derived the predicted *PON1* sequence for these species by mapping the sequencing reads to *PON1*, *PON2* and *PON3* coding sequences of the most closely related species in the 100-way alignment

(dolphin and Weddell seal for the hippopotamus and spotted seal, respectively) simultaneously using NextGenMap (*71*) and retaining the consensus sequence for reads mapped to *PON1* using SAMtools (*72*). We additionally experimentally determined the sequence of the dugong (see below). We added these species to their inferred locations in the mammalian phylogenetic tree using published phylogenetic inferences of the topology of the relevant clades (*17*, *54*, *73–76*). Table S11 lists accession numbers for all datasets used to add new species.

Estimating branch-specific d_N/d_S for *PON1* and the timing of its loss along marine lineages

We estimated branch-specific d_N/d_S, or omega (ω), across the expanded mammalian phylogeny, excluding non-eutherian mammals, using the *codeml* program in the PAML software (*77*). We used the branch model with freely varying omega (model = 1, NsSites = 0) to infer d_N/d_S across all branches separately. After estimating parameters for each branch independently, we subsequently constrained some branches to have equal rates in order to more accurately estimate rates for short branches and ran PAML with model = 2; we additionally pruned the tree for some analyses to reduce run time (see next section). To estimate the time at which *PON1*'s evolutionary rate shifted from a background functional rate (ω_f) to the rate for pseudogenic lineages (ω_p) in each marine lineage, indicative of the time of loss, we applied equation (5) from Meredith et al. (*78*). To determine the value for ω_p, we tested whether the d_N/d_S ratio was significantly different from 1, the theoretical expectation for a pseudogene, on branches fully subsequent to the inferred first appearance of genetic lesions. This value was estimated to be 0.98, which was not significantly different from 1 ($P = 0.93$); we therefore set $\omega_p = 1$. To account for differences in d_S between functional and pseudogenic lineages, we assumed that the ratio of functional to pseudogenic d_S was 0.7, based on the finding by Bustamante, Nielsen, and Hartl (*79*) that processed pseudogenes within regions of similar GC content to their parent genes accrued synonymous substitutions at a rate 70% of that of the parent genes. We estimated the background functional rate (ω_f) from the closest evolutionary lineages to the focal clade: for cetaceans, we included all bovids and the bovid ancestral branch; for sirenians, we included all other Afrotherian lineages except sirenians and the Afrotherian ancestral branch; for Phocidae, we included all Carnivora, excluding the sea otter and all pinniped lineages except for the branch ancestral to all pinnipeds (see next section).

To determine whether d_N/d_S was significantly different from 1 in some lineages and to estimate its confidence interval, we constrained d_N/d_S in the focal lineage using fix_omega = 1. We derived *P*-values for the hypothesis that $d_N/d_S = 1$ using a likelihood ratio test, comparing the likelihood with omega fixed at one to its maximum likelihood value. We derived 95% confidence intervals by running codeml for various fixed values of d_N/d_S and estimating the value at which the likelihood ratio test *P*-value would drop below 0.05.

Phylogenetic trees used for evolutionary inferences

Table S11 relates the abbreviations used in these trees to common names and data sources for all species. We used the following tree and its pruned subsets as our input to BayesTraits for the main analyses reported in the paper:

((((((((((hg19:0.005957477577,panTro4:0.006721826689):0.001382639829,gorGor3:0.007765177171):0.005572327638,ponAbe2:0.0164503644):0.002187630666,nomLeu3:0.01770384793):0.007043113559,(chlSab1:0.007693724903,((macFas5:0.001292320552,rheMac3:0.00713015786):0.002951690224,papHam1:0.005199240711):0.002049749893):0.01566263562):0.0135408115,(calJac3:0.02474184521,saiBol1:0.02096868307):0.02784675729):0.04299750653,otoG

ar3:0.108738222):0.01379370868,((((((cavPor3:0.09048639907,(chiLan1:0.05332953299,octDeg1:0.08476954109):0.01287861561):0.02118937782,hetGla2:0.08588673524):0.07432515556,speTri2:0.08896424642):0.006291577528,((((criGri1:0.04084640027,mesAur1:0.04456203524):0.02314125062,micOch1:0.06932402649):0.01947113467,(mm10:0.05273642272,rn5:0.05576007402):0.04435347588):0.08380065137,jacJac1:0.1438649666):0.04270536633):0.01663675397,(ochPri3:0.1256544445,oryCun2:0.07131655591):0.06535533418):0.009050428462,tupChi1:0.1191189141):0.003894252213):0.01425600689,(((((ailMel1:0.03854019703,((lepWed1:0.02002160645,odoRosDi:0.02064385875):0.01734764946,musFur1:0.04613997497):0.002879093616):0.009005888384,canFam3:0.05339127565):0.01185166857,felCat5:0.05020331605):0.03285617057,((((((bosTau7:0.02168740723,((capHir1:0.01157093136,oviAri3:0.01246322594):0.0049716126,panHod1:0.01522587482):0.01465511149):0.0662523666,(orcOrc1:0.006371664911,turTru2:0.01086552617):0.06014682602):0.01216198069,susScr3:0.0796745271):0.006785823323,(camFer1:0.01240650215,vicPac2:0.01096629635):0.06374554586):0.02551888691,(cerSim1:0.04977357056,equCab2:0.061454379):0.02510111297):0.00331214686,((eptFus1:0.032485466546,(myoDav1:0.02344332842,myoLuc2:0.01567729315):0.02193849809):0.09455328094,(pteAle1:0.005833353548,pteVam1:0.01611220178):0.07567400302):0.02385546003):0.002057771224):0.004845253848,(conCri1:0.1239823369,(eriEur2:0.1696142244,sorAra2:0.1934205791):0.02079474546):0.0235875333):0.01477733374):0.01915406518,((((chrAsi1:0.1017903453,echTel2:0.1749615473):0.01592632003,eleEdw1:0.1516860647):0.006610995228,oryAfe1:0.08326528894):0.008243787904,(loxAfr3:0.06812658238,triMan1:0.06198982615):0.0224994529):0.03384011363,dasNov3:0.1342602666);

We used the following trees to constrain branches for various purposes in PAML:

#1 For generating Figs. 1 and S1:

(((conCri1 #1,(eriEur2 #2,sorAra2 #3) #4) #4,((felCat5 #6,(canFam3 #7,((ailMel1 #8,ursMar1 #9) #10,((musFur1 #11,enhLut #12) #13,(((lepWed1 #14,Hawaii #14) #15,largha #16) #17,(odoRosDi #18,arcGaz #19) #20) #21) #22) #23) #24) #25,(((pteVam1 #26,pteAle1 #27) #28,(mini #29,(eptFus1 #30,(myoDav1 #31,(brandtBat #32,myoLuc2 #33) #34) #34) #36) #37) #38,((cerSim1 #39,equCab2 #40) #41,((vicPac2 #42,camFer1 #43) #44,(susScr3 #45,

((((((turTru2 #46,orcOrc1 #47) #48,lipVex #49) #50,phyCat #51) #52,Minke #53) #54,hippo #55) #56,(bosTau7 #57,(panHod1 #58,(oviAri3 #59,capHir1 #60) #60) #62) #63) #64) #65) #66) #67) #69) #69) #69,((tupChi1 #71,((ochPri3 #72,oryCun2 #73) #74,((casCan #75,(jacJac1 #76,((rn5 #77,mm10 #78) #79,(micOch1 #80,(mesAur1 #81,criGri1 #82) #83) #84) #85) #86) #87,(speTri2 #88,(hetGla2 #89,(cavPor3 #90,(chiLan1 #91,octDeg1 #92) #93) #94) #95) #96) #96) #98) #99,(otoGar3 #100,((saiBol1 #101,calJac3 #102) #103,((chlSab1 #104,(papHam1 #105,(rheMac3 #106,macFas5 #107) #107) #109) #110,(nomLeu3 #111,(ponAbe2 #112,(gorGor3 #113,(hg19 #114,panTro4 #115) #116) #116) #118) #119) #119) #121) #122) #124,(((oryAfe1 #125,((echTel2 #126,chrAsi1 #0) #117,eleEdw1 #120) #108) #108,((triMan1 #97,dugDug #123) #61,loxAfr3 #35) #68) #108,dasNov3 #70) #5);

#2 For estimating significance of cetacean ancestral branch d_N/d_S difference from 1 and confidence interval:

((((conCri1,(eriEur2,sorAra2))$1,((felCat5,(canFam3,((ailMel1,ursMar1),((musFur1,enhLut),(((lepWed1,Hawaii),largha),(odoRosDi,arcGaz))))))$2,(((pteVam1,pteAle1),(mini,(eptFus1,(myoDav1,(brandtBat,myoLuc2))))))$3,((cerSim1#4,equCab2#4)#4,((vicPac2#4,camFer1#4)#4,(susScr3#4,((((((turTru2#8,orcOrc1#8)#8,lipVex#8)#8,phyCat#8)#8,Minke#8)#12,hippo#11)#4,(bos

Tau7#4,(panHod1#4,(oviAri3#4,capHir1#4)#4)#4)#4)#4)#4)#4)#4)#2)#2)#2,((tupChi1,((ochPri3,oryCun2),((casCan,(jacJac1,((rn5,mm10),(micOch1,(mesAur1,criGri1))))),(speTri2,(hetGla2,(cavPor3,(chiLan1,octDeg1)))))))$5,(otoGar3,((saiBol1,calJac3),((chlSab1,(papHam1,(rheMac3,macFas5))),(nomLeu3,(ponAbe2,(gorGor3,(hg19,panTro4)))))))))$6,(((oryAfe1#7,((echTel2#7,chrAsi1#7)#7,eleEdw1#7)#7)#7,((triMan1,dugDug)$10,loxAfr3#7)#7)#7,dasNov3#7)#7);

#3 For estimating significance of siren ancestral branch d_N/d_S difference from 1 and confidence interval:

((((conCri1,(eriEur2,sorAra2))$1,((felCat5,(canFam3,((ailMel1,ursMar1),((musFur1,enhLut),(((lepWed1,Hawaii),largha),(odoRosDi,arcGaz))))))$2,(((pteVam1,pteAle1),(mini,(eptFus1,(myoDav1,(brandtBat,myoLuc2))))))$3,((cerSim1#4,equCab2#4)#4,((vicPac2#4,camFer1#4)#4,(susScr3#4,((((((turTru2,orcOrc1),lipVex),phyCat),Minke)$8,hippo#4)#4,(bosTau7#4,(panHod1#4,(oviAri3#4,capHir1#4)#4)#4)#4)#4)#4)#4)#4)#4)#2)#2,((tupChi1,((ochPri3,oryCun2),((casCan,(jacJac1,((rn5,mm10),(micOch1,(mesAur1,criGri1))))),(speTri2,(hetGla2,(cavPor3,(chiLan1,octDeg1)))))))$5,(otoGar3,((saiBol1,calJac3),((chlSab1,(papHam1,(rheMac3,macFas5))),(nomLeu3,(ponAbe2,(gorGor3,(hg19,panTro4)))))))))$6,(((oryAfe1#7,((echTel2#7,chrAsi1#7)#7,eleEdw1#7)#7)#7,((triMan1#9,dugDug#9)#10,loxAfr3#7)#7)#7,dasNov3#7)#7);

#4 For estimating date of loss in ancestor of Weddell seal and confidence interval:

((((conCri1,(eriEur2,sorAra2))$1,((felCat5#2,(canFam3#2,((ailMel1#2,ursMar1#2)#2,((musFur1#2,enhLut#10)#2,(((lepWed1#11,Hawaii#9)#11,largha#9)#11,(odoRosDi#9,arcGaz#9)#9)#11)#2)#2)#2)#2,(((pteVam1,pteAle1),(mini,(eptFus1,(myoDav1,(brandtBat,myoLuc2))))))$3,((cerSim1#4,equCab2#4)#4,((vicPac2#4,camFer1#4)#4,(susScr3#4,((((((turTru2,orcOrc1),lipVex),phyCat),Minke)$8,hippo#4)#4,(bosTau7#4,(panHod1#4,(oviAri3#4,capHir1#4)#4)#4)#4)#4)#4)#4)#4)#2)#2,((tupChi1,((ochPri3,oryCun2),((casCan,(jacJac1,((rn5,mm10),(micOch1,(mesAur1,criGri1))))),(speTri2,(hetGla2,(cavPor3,(chiLan1,octDeg1)))))))$5,(otoGar3,((saiBol1,calJac3),((chlSab1,(papHam1,(rheMac3,macFas5))),(nomLeu3,(ponAbe2,(gorGor3,(hg19,panTro4)))))))))$6,(((oryAfe1,((echTel2,chrAsi1),eleEdw1)),((triMan1,dugDug),loxAfr3)),dasNov3)$7);

#5 For estimating date of loss in ancestor of Hawaiian monk seal and confidence interval:

((((conCri1,(eriEur2,sorAra2))$1,((felCat5#2,(canFam3#2,((ailMel1#2,ursMar1#2)#2,((musFur1#2,enhLut#10)#2,(((lepWed1#9,Hawaii#12)#12,largha#9)#12,(odoRosDi#9,arcGaz#9)#9)#12)#2)#2)#2)#2,(((pteVam1,pteAle1),(mini,(eptFus1,(myoDav1,(brandtBat,myoLuc2))))))$3,((cerSim1#4,equCab2#4)#4,((vicPac2#4,camFer1#4)#4,(susScr3#4,((((((turTru2,orcOrc1),lipVex),phyCat),Minke)$8,hippo#4)#4,(bosTau7#4,(panHod1#4,(oviAri3#4,capHir1#4)#4)#4)#4)#4)#4)#4)#4)#2)#2,((tupChi1,((ochPri3,oryCun2),((casCan,(jacJac1,((rn5,mm10),(micOch1,(mesAur1,criGri1))))),(speTri2,(hetGla2,(cavPor3,(chiLan1,octDeg1)))))))$5,(otoGar3,((saiBol1,calJac3),((chlSab1,(papHam1,(rheMac3,macFas5))),(nomLeu3,(ponAbe2,(gorGor3,(hg19,panTro4)))))))))$6,(((oryAfe1#7,((echTel2#7,chrAsi1#7)#7,eleEdw1#7)#7)#7,((triMan1,dugDug)$11,loxAfr3#7)#7)#7,dasNov3#7)#7);

#6 For assessing robustness of Weddell seal estimates to local topology, using ursids as the outgroup to pinnipeds:

((((conCri1,(eriEur2,sorAra2))$1,((felCat5#2,(canFam3#2,((musFur1#2,enhLut#10)#2,((ailMel1#2,ursMar1#2)#2,(((lepWed1#11,Hawaii#9)#11,largha#9)#11,(odoRosDi#9,arcGaz#9)#9)#11)#2)#2)#2)#2,(((pteVam1,pteAle1),(mini,(eptFus1,(myoDav1,(brandtBat,myoLuc2))))))$3,((ce

rSim1#4,equCab2#4)#4,((vicPac2#4,camFer1#4)#4,(susScr3#4,((((((turTru2,orcOrc1),lipVex),phyCat),Minke)$8,hippo#4)#4,(bosTau7#4,(panHod1#4,(oviAri3#4,capHir1#4)#4)#4)#4)#4)#4)#4)#4)#2)#2,((tupChi1,((ochPri3,oryCun2),((casCan,(jacJac1,((rn5,mm10),(micOch1,(mesAur1,criGri1))))),(speTri2,(hetGla2,(cavPor3,(chiLan1,octDeg1)))))))$5,(otoGar3,((saiBol1,calJac3),((chlSab1,(papHam1,(rheMac3,macFas5))),(nomLeu3,(ponAbe2,(gorGor3,(hg19,panTro4)))))))))$6,(((oryAfe1,((echTel2,chrAsi1),eleEdw1)),((triMan1,dugDug),loxAfr3)),dasNov3)$7);

#7 For assessing robustness of Hawaiian monk seal estimates to local topology, using ursids as the outgroup to pinnipeds:

((((conCri1,(eriEur2,sorAra2))$1,((felCat5#2,(canFam3#2,((musFur1#2,enhLut#10)#2,((ailMel1#2,ursMar1#2)#2,(((lepWed1#9,Hawaii#12)#12,largha#9)#12,(odoRosDi#9,arcGaz#9)#9)#12)#2)#2)#2)#2)#2,(((pteVam1,pteAle1),(mini,(eptFus1,(myoDav1,(brandtBat,myoLuc2))))))$3,((cerSim1#4,equCab2#4)#4,((vicPac2#4,camFer1#4)#4,(susScr3#4,((((((turTru2,orcOrc1),lipVex),phyCat),Minke)$8,hippo#4)#4,(bosTau7#4,(panHod1#4,(oviAri3#4,capHir1#4)#4)#4)#4)#4)#4)#4)#4)#2)#2,((tupChi1,((ochPri3,oryCun2),((casCan,(jacJac1,((rn5,mm10),(micOch1,(mesAur1,criGri1))))),(speTri2,(hetGla2,(cavPor3,(chiLan1,octDeg1)))))))$5,(otoGar3,((saiBol1,calJac3),((chlSab1,(papHam1,(rheMac3,macFas5))),(nomLeu3,(ponAbe2,(gorGor3,(hg19,panTro4)))))))))$6,(((oryAfe1#7,((echTel2#7,chrAsi1#7),eleEdw1#7)#7)#7,((triMan1,dugDug)$11,loxAfr3#7)#7)#7,dasNov3#7)#7);

#8 For estimating d_N/d_S on fully pseudogenic lineages and assessing significance of its difference from 1:

(((conCri1 #1,(eriEur2 #2,sorAra2 #3) #4) #4,((felCat5 #6,(canFam3 #7,((ailMel1 #8,ursMar1 #9) #10,((musFur1 #11,enhLut #12) #13,(((lepWed1 #14,Hawaii #14) #15,largha #16) #17,(odoRosDi #18,arcGaz #19) #20) #21) #22) #23) #24) #25,(((pteVam1 #26,pteAle1 #27) #28,(mini #29,(eptFus1 #30,(myoDav1 #31,(brandtBat #32,myoLuc2 #33) #34) #34) #36) #37) #38,((cerSim1 #39,equCab2 #40) #41,((vicPac2 #42,camFer1 #43) #44,(susScr3 #45,

((((((turTru2 #127,orcOrc1 #127) #127,lipVex #127) #127,phyCat #127) #127,Minke #127) #54,hippo #55) #56,(bosTau7 #57,(panHod1 #58,(oviAri3 #59,capHir1 #60) #60) #62) #63) #64) #65) #66) #67) #69) #69) #69,((tupChi1 #71,((ochPri3 #72,oryCun2 #73) #74,((casCan #75,(jacJac1 #76,((rn5 #77,mm10 #78) #79,(micOch1 #80,(mesAur1 #81,criGri1 #82) #83) #84) #85) #86) #87,(speTri2 #88,(hetGla2 #89,(cavPor3 #90,(chiLan1 #91,octDeg1 #92) #93) #94) #95) #96) #96) #98) #99,(otoGar3 #100,((saiBol1 #101,calJac3 #102) #103,((chlSab1 #104,(papHam1 #105,(rheMac3 #106,macFas5 #107) #107) #109) #110,(nomLeu3 #111,(ponAbe2 #112,(gorGor3 #113,(hg19 #114,panTro4 #115) #116) #116) #118) #119) #119) #121) #122) #124,(((oryAfe1 #125,((echTel2 #126,chrAsi1 #0) #117,eleEdw1 #120) #108) #108,((triMan1 #127,dugDug #127) #61,loxAfr3 #35) #68) #108,dasNov3 #70) #5);

We used the following trees for assessing robustness of our results to variation in the tree used for inferences of functional loss and marine transition rates in BayesTraits:

From Bininda-Emonds et al. (*49*): ((((((((((((((((chrAsi1: 0.143350, echTel2: 0.233284): 0.025760, eleEdw1: 0.229356): 0.006898, oryAfe1: 0.111972): 0.007094, (triMan1: 0.073576, loxAfr3: 0.072415): 0.029773): 0.046373, dasNov3: 0.182288): 0.020857, ((((((((((oviAri3: 0.011186, capHir1: 0.018999): 0.009035, panHod1: 0.016435): 0.018612, bosTau7: 0.027888): 0.107579, (turTru2: 0.005355, orcOrc1: 0.002535): 0.057611): 0.015675, susScr3: 0.095112): 0.008406, (camFer1: 0.016128, vicPac2: 0.018535): 0.082116): 0.034641, (cerSim1: 0.057399,

equCab2: 0.087013): 0.036501): 0.005781, ((((musFur1: 0.066447, (odoRosDi: 0.022019, lepWed1: 0.024733): 0.026455): 0.006208, ailMel1: 0.053454): 0.016055, canFam3: 0.098719): 0.019762, felCat5: 0.088373): 0.035795): 0.001311, ((pteAle1: 0.008890, pteVam1: 0.018174): 0.116559, (eptFus1: 0.041851, (myoLuc2: 0.013929, myoDav1: 0.024245): 0.023789): 0.091639): 0.041616): 0.004263, (conCri1: 0.172800, (eriEur2: 0.259454, sorAra2: 0.232393): 0.024629): 0.025037): 0.018426): 0.015506, ((((jacJac1: 0.204239, (((criGri1: 0.048366, mesAur1: 0.065787): 0.044915, (mm10: 0.090139, rn5: 0.092226): 0.085868): 0.001762, micOch1: 0.129134): 0.132206): 0.054664, (octDeg1: 0.136000, ((hetGla2: 0.123639, chiLan1: 0.085930): 0.004342, cavPor3: 0.144027): 0.001709): 0.128634): 0.007555, speTri2: 0.131122): 0.024663, (oryCun2: 0.098849, ochPri3: 0.187411): 0.109700): 0.013394): 0.004239, tupChi1: 0.172116): 0.013486, otoGar3: 0.152967): 0.063382, (calJac3: 0.028548, saiBol1: 0.028151): 0.038638): 0.023027, (chlSab1: 0.011251, (papHam1: 0.007843, (rheMac3: 0.001701, macFas5: 0.001399): 0.002293): 0.001712): 0.028199): 0.011241, nomLeu3: 0.018425): 0.002529, ponAbe2: 0.016272): 0.008027, gorGor3: 0.006483): 0.002264, panTro4: 0.005083): 0.000018, hg19: 0.005066);

From Meredith et al. (*17*): ((((((((((dasNov3: 0.182346, ((oryAfe1: 0.112172, ((echTel2: 0.233380, chrAsi1: 0.143404): 0.025867, eleEdw1: 0.229424): 0.006654): 0.007010, (triMan1: 0.073719, loxAfr3: 0.072303): 0.029773): 0.046455): 0.020829, ((eriEur2: 0.278620, (conCri1: 0.161523, sorAra2: 0.241537): 0.017235): 0.023583, ((felCat5: 0.087603, (canFam3: 0.098721, (musFur1: 0.068036, (ailMel1: 0.055555, (lepWed1: 0.024733, odoRosDi: 0.022026): 0.028340): 0.002033): 0.017750): 0.019948): 0.036465, (((pteVam1: 0.018167, pteAle1: 0.008894): 0.116496, (eptFus1: 0.041782, (myoDav1: 0.024256, myoLuc2: 0.013918): 0.023862): 0.091789): 0.041270, ((cerSim1: 0.057702, equCab2: 0.086736): 0.036605, ((vicPac2: 0.018536, camFer1: 0.016128): 0.081868, (susScr3: 0.095240, ((turTru2: 0.005368, orcOrc1: 0.002521): 0.056741, (panHod1: 0.017222, (bosTau7: 0.045187, (oviAri3: 0.010760, capHir1: 0.019462): 0.009747): 0.000287): 0.123715): 0.016363): 0.008448): 0.034326): 0.004944): 0.002178): 0.004555): 0.018411): 0.015395, (tupChi1: 0.168696, ((ochPri3: 0.187276, oryCun2: 0.098973): 0.111064, ((jacJac1: 0.206518, ((rn5: 0.091889, mm10: 0.090465): 0.064033, (micOch1: 0.122339, (mesAur1: 0.065467, criGri1: 0.048665): 0.035515): 0.036236): 0.111902): 0.056731, (speTri2: 0.128493, (hetGla2: 0.096980, (cavPor3: 0.138094, (chiLan1: 0.070197, octDeg1: 0.121169): 0.021754): 0.030846): 0.106569): 0.006421): 0.025081): 0.008142): 0.007276): 0.014487, otoGar3: 0.152776): 0.064176, (saiBol1: 0.028181, calJac3: 0.028525): 0.038639): 0.023035, (chlSab1: 0.011252, (papHam1: 0.007844, (rheMac3: 0.001701, macFas5: 0.001399): 0.002293): 0.001712): 0.028173): 0.011251, nomLeu3: 0.018431): 0.002523, ponAbe2: 0.016274): 0.008026, gorGor3: 0.006483): 0.002264, panTro4: 0.005083): 0.005081, hg19: 0.000004);

From UCSC: ((((((((((((((((((ailMel1: 0.055565, (lepWed1: 0.024710, odoRosDi: 0.022049): 0.028265): 0.002325, musFur1: 0.067738): 0.018059, canFam3: 0.098263): 0.019934, felCat5: 0.087674): 0.035872, (cerSim1: 0.057971, equCab2: 0.086686): 0.039139): 0.001294, ((eptFus1: 0.041879, (myoDav1: 0.024254, myoLuc2: 0.013921): 0.023756): 0.091935, (pteAle1: 0.008885, pteVam1: 0.018177): 0.116286): 0.042207): 0.002249, (((((bosTau7: 0.045079, (capHir1: 0.019458, oviAri3: 0.010773): 0.009844): 0.000291, panHod1: 0.017118): 0.122727, (orcOrc1: 0.002645, turTru2: 0.005244): 0.057704): 0.018998, (camFer1: 0.016249, vicPac2: 0.018416): 0.085138): 0.001164, susScr3: 0.097354): 0.041701):

0.004128, ((conCri1: 0.161274, sorAra2: 0.241682): 0.017428, eriEur2: 0.278172): 0.023349): 0.018577, ((((chrAsi1: 0.142889, echTel2: 0.234488): 0.033522, ((eleEdw1: 0.243957, loxAfr3: 0.085550): 0.003880, triMan1: 0.088660): 0.012528): 0.005075, oryAfe1: 0.114169): 0.045829, dasNov3: 0.182524): 0.021504): 0.015602, ((((cavPor3: 0.137851, (chiLan1: 0.070284, octDeg1: 0.121058): 0.021939): 0.030825, hetGla2: 0.097341): 0.108968, (((((criGri1: 0.048691, mesAur1: 0.065429): 0.035767, micOch1: 0.122176): 0.036137, (mm10: 0.090615, rn5: 0.091697): 0.064155): 0.111607, jacJac1: 0.206193): 0.055282, speTri2: 0.129369): 0.005714): 0.024641, (ochPri3: 0.187334, oryCun2: 0.098942): 0.109851): 0.012750): 0.004414, tupChi1: 0.170891): 0.014037, otoGar3: 0.152503): 0.063931, (calJac3: 0.028554, saiBol1: 0.028149): 0.038484): 0.023179, (chlSab1: 0.011248, ((macFas5: 0.001399, rheMac3: 0.001701): 0.002292, papHam1: 0.007844): 0.001715): 0.028206): 0.011232, nomLeu3: 0.018423): 0.002531, ponAbe2: 0.016274): 0.008026, gorGor3: 0.006483): 0.002263, panTro4: 0.005083): 0.004954, hg19: 0.000131);

Our consensus topology, with branch lengths re-estimated from 10 randomly selected genes: ((((((((((dasNov3: 0.182165, ((oryAfe1: 0.112147, ((echTel2: 0.233323, chrAsi1: 0.143401): 0.025859, eleEdw1: 0.229433): 0.006652): 0.007006, (triMan1: 0.073719, loxAfr3: 0.072291): 0.029799): 0.046681): 0.020578, ((conCri1: 0.172666, (eriEur2: 0.259720, sorAra2: 0.232012): 0.024729): 0.025173, ((felCat5: 0.088291, (canFam3: 0.098671, (ailMel1: 0.053450, (musFur1: 0.066457, (lepWed1: 0.024748, odoRosDi: 0.022006): 0.026436): 0.006218): 0.016070): 0.019850): 0.035853, (((pteVam1: 0.018164, pteAle1: 0.008898): 0.116483, (eptFus1: 0.041792, (myoDav1: 0.024256, myoLuc2: 0.013918): 0.023851): 0.091834): 0.041032, ((cerSim1: 0.057793, equCab2: 0.086634): 0.036549, ((vicPac2: 0.018539, camFer1: 0.016125): 0.082116, (susScr3: 0.095269, ((turTru2: 0.005368, orcOrc1: 0.002521): 0.057664, (bosTau7: 0.027892, (panHod1: 0.016438, (oviAri3: 0.011185, capHir1: 0.018999): 0.009034): 0.018612): 0.107547): 0.015573): 0.008388): 0.034207): 0.005158): 0.002134): 0.004861): 0.018220): 0.015404, (tupChi1: 0.168739, ((ochPri3: 0.187281, oryCun2: 0.098947): 0.111092, ((jacJac1: 0.206431, ((rn5: 0.091908, mm10: 0.090430): 0.064031, (micOch1: 0.122321, (mesAur1: 0.065470, criGri1: 0.048655): 0.035511): 0.036261): 0.111967): 0.056743, (speTri2: 0.128394, (hetGla2: 0.096996, (cavPor3: 0.138075, (chiLan1: 0.070202, octDeg1: 0.121153): 0.021754): 0.030828): 0.106633): 0.006418): 0.025037): 0.008123): 0.007542): 0.014124, otoGar3: 0.152700): 0.064313, (saiBol1: 0.028183, calJac3: 0.028523): 0.038628): 0.023044, (chlSab1: 0.011251, (papHam1: 0.007844, (rheMac3: 0.001704, macFas5: 0.001400): 0.002292): 0.001712): 0.028172): 0.011251, nomLeu3: 0.018430): 0.002524, ponAbe2: 0.016274): 0.008026, gorGor3: 0.006483): 0.002263, panTro4: 0.005083): 0.000005, hg19: 0.005079);

FICD: ((((camFer1:0.0113151,vicPac2:0.00833575)1.00
:0.0680512,(susScr3:0.0849603,((bosTau7:0.0190462,(panHod1:0.0117587,(capHir1:0.0072367
2,oviAri3:0.00296763)0.95 :0.00803626)0.96 :0.0105205)1.00
:0.0611889,(orcOrc1:0.00431941,turTru2:0.00364521)1.00 :0.0474277)0.88
:0.00825514)0.89 :0.0114841)0.91 :0.0154473,(sorAra2:0.14432,eriEur2:0.174343)0.99
:0.054644)0.76 :0.0182471,(ochPri3:0.152512,oryCun2:0.0833944)1.00
:0.0620949,((speTri2:0.0931615,eleEdw1:0.166057)0.86
:0.0185115,(((oryAfe1:0.102396,(loxAfr3:0.0547178,triMan1:0.0474531)0.91
:0.0182737)0.85 :0.0116118,(echTel2:0.13055,chrAsi1:0.157965)0.00 :0.0138564)0.99

:0.0368897,(((hetGla2:0.0512119,(cavPor3:0.0504482,(chiLan1:0.0417467,octDeg1:0.0685713)0.92 :0.0137134)0.85 :0.0116289)1.00 :0.0598863,(jacJac1:0.137513,((micOch1:0.0561351,(mesAur1:0.0293882,criGri1:0.0444404)0.89 :0.0156857)0.99 :0.029273,(rn5:0.0634989,mm10:0.0663803)0.89 :0.0159785)1.00 :0.0742083)0.98 :0.0358984)0.90 :0.0115399,((tupChi1:0.0825691,((calJac3:0.0289725,saiBol1:0.0173061)1.00 :0.0232913,((chlSab1:0.00568676,(papHam1:0.00295565,(macFas5:8e-008,rheMac3:0.00097339)0.85 :0.00194624)0.88 :0.00426717)1.00 :0.0177847,(ponAbe2:0.0100221,(nomLeu3:0.0131533,(gorGor3:0.00420826,(hg19:0.00898183,panTro4:0.00298853)0.89 :0.00378587)0.97 :0.00643285)0.21 :0.00099022)0.98 :0.0122414)0.89 :0.00696588)1.00 :0.0474217)0.30 :0.00936917,(otoGar3:0.108909,(dasNov3:0.140854,((conCri1:0.11771,(cerSim1:0.0519077,equCab2:0.0449336)0.94 :0.0181295)0.00 :0.00055546,((felCat5:0.0613243,(canFam3:0.0617033,(musFur1:0.0500125,(ailMel1:0.0347766,(lepWed1:0.00967075,odoRosDi:0.022054)0.96 :0.0110034)0.83 :0.00538302)0.96 :0.0142826)0.00 :0.00430767)0.99 :0.0275869,((eptFus1:0.0214641,(myoLuc2:0.0144116,myoDav1:0.0257732)0.75 :0.0076459)1.00 :0.0625247,(pteAle1:7e-008,pteVam1:0.00286805)1.00 :0.0680928)0.95 :0.0203209)0.90 :0.0174658)0.96 :0.0258138)0.85 :0.0140837)0.81 :0.00806651)0.94 :0.0184718)0.79 :0.0183756)0.78 :0.0206549)0.93 :0.0183008);

FUT9:

(((speTri2:0.0252541,(hetGla2:0.0191224,(cavPor3:0.0414395,(chiLan1:0.0302658,octDeg1:0.0308716)0.29 :0.00322026)0.93 :0.00707553)1.00 :0.022308)0.95 :0.0119096,((tupChi1:0.0399769,((ochPri3:0.0418329,oryCun2:0.0253716)0.78 :0.0088666,(((calJac3:0.00478283,saiBol1:0.00584235)0.92 :0.00398593,(((rheMac3:0.00093944,macFas5:0.00093949)0.00 :8e-008,(papHam1:0.00282833,chlSab1:0.00093951)0.00 :8e-008)0.99 :0.00789241,(ponAbe2:0.00476144,(nomLeu3:0.00569764,(panTro4:0.00378619,(gorGor3:5e-008,hg19:0.00188882)0.73 :0.00097065)0.94 :0.00391104)0.73 :0.0009181)0.82 :0.00171276)0.87 :0.00298119)1.00 :0.0168347,otoGar3:0.0413524)0.68 :0.00194349)0.10 :0.00110366)0.77 :0.00132292,((dasNov3:0.044413,((oryAfe1:0.0175616,(chrAsi1:0.0397976,(loxAfr3:0.0105874,triMan1:0.00825529)0.88 :0.00325993)0.00 :4.545e-005)0.00 :7e-008,(eleEdw1:0.0513143,echTel2:0.0969533)0.84 :0.0104349)0.99 :0.0135642)0.95 :0.0086058,(((cerSim1:0.0188239,equCab2:0.0120704)0.95 :0.00781477,((pteAle1:0.00367297,pteVam1:0.00292656)1.00 :0.0284741,(eptFus1:0.00476038,(myoDav1:0.00758172,myoLuc2:0.007672)0.98 :0.00987068)1.00 :0.0310162)0.93 :0.00870513)0.73 :0.00101565,((susScr3:0.027518,((orcOrc1:0.00131581,turTru2:0.00242622)1.00 :0.0168157,((bosTau7:0.00104191,(panHod1:0.00477473,(capHir1:0.00335392,oviAri3:0.00133732)0.88 :0.00270829)0.92 :0.00570738)1.00 :0.0254804,(camFer1:0.00393808,vicPac2:0.00073725)1.00 :0.0178516)0.23 :0.00151601)0.28 :0.0037157)0.96 :0.00777287,((((lepWed1:0.00373807,odoRosDi:0.00967308)0.90 :0.00409797,(ailMel1:0.029386,musFur1:0.00926697)0.81 :0.00252731)0.86

:0.00266733,(felCat5:0.0249583,canFam3:0.00671789)0.73 :0.00092662)0.98
:0.0102845,(conCri1:0.0539119,sorAra2:0.0587859)0.85 :0.00698123)0.55
:0.00061409)0.76 :0.00163296)0.94 :0.00558026)0.89 :0.00555256)0.52
:0.00396628)0.43 :0.00886831,(jacJac1:0.0811645,((mm10:0.0215201,rn5:0.0264752)0.95
:0.0109563,(micOch1:0.0454886,(criGri1:0.0251423,mesAur1:0.0239087)0.89
:0.00666539)0.89 :0.00732828)1.00 :0.03702)0.62 :0.0094863,eriEur2:0.201525);

GPR22:
(jacJac1:0.0565971,((micOch1:0.032242,(criGri1:0.0248524,mesAur1:0.0369192)0.98
:0.0158816)0.72 :0.00451527,(mm10:0.0243887,rn5:0.0243697)0.99 :0.0194461)1.00
:0.039754,((speTri2:0.0316967,((ochPri3:0.0588563,oryCun2:0.0241152)1.00
:0.0213299,((tupChi1:0.0312922,(otoGar3:0.0224436,((calJac3:0.00645248,saiBol1:0.00898666
)0.91 :0.00495251,(((nomLeu3:0.00355323,ponAbe2:0.00445163)0.00 :1.1e-007,(hg19:1e-
008,(panTro4:0.00088375,gorGor3:8e-008)0.00 :1e-008)0.81 :0.00088505)0.95
:0.00423273,((macFas5:8e-008,rheMac3:0.00177079)0.82 :0.00088231,(papHam1:6e-
008,chlSab1:0.00176667)0.00 :6e-008)0.89 :0.00293661)0.69 :0.00187277)0.96
:0.00598497)0.89 :0.00353074)0.00 :1.7e-
007,((dasNov3:0.0285546,((loxAfr3:0.0210592,triMan1:0.0129913)0.64
:0.00262651,(oryAfe1:0.0345266,(chrAsi1:0.0382627,(echTel2:0.080383,eleEdw1:0.0456413)0.
75 :0.00593339)0.79 :0.00319943)0.00 :0.00050969)0.83 :0.00341471)0.99
:0.0085972,(((camFer1:0.00153295,vicPac2:0.00538669)1.00
:0.0217491,(susScr3:0.0141247,((orcOrc1:0.00179724,turTru2:0.00081614)0.99
:0.0101841,(bosTau7:0.00543224,(panHod1:0.00350998,(capHir1:0.00174224,oviAri3:0.00086
68)0.87 :0.00175798)0.92 :0.0034465)1.00 :0.0118994)0.93 :0.0064075)0.63
:0.00081167)0.96
:0.00592339,((sorAra2:0.0511974,(eriEur2:0.0391292,conCri1:0.0271244)0.07
:0.00348964)0.89 :0.00520869,(((pteAle1:0.00075575,pteVam1:0.00690898)1.00
:0.0219761,(eptFus1:0.0123169,(myoDav1:0.00536628,myoLuc2:0.00447631)0.86
:0.00314951)1.00 :0.020654)0.95
:0.00791576,((cerSim1:0.00725391,equCab2:0.0138684)0.91
:0.00343907,(felCat5:0.0335148,(canFam3:0.019121,((lepWed1:0.00792362,odoRosDi:0.00169
497)0.99 :0.00733065,(musFur1:0.0165202,ailMel1:0.0111458)0.55 :0.00038508)0.95
:0.00373955)0.00 :0.00029403)0.98 :0.00582452)0.00 :8e-008)0.74 :0.00087093)0.00
:8e-008)0.96 :0.00443706)0.87 :0.00187165)0.00 :1e-007)0.99 :0.00971806)0.00
:0.00099399,(hetGla2:0.0200643,(cavPor3:0.0157943,(chiLan1:0.0123284,octDeg1:0.0282111)
0.92 :0.00632537)0.96 :0.00941812)0.98 :0.0135757)0.98 :0.0216899);

HMGCS1: (((((ochPri3:0.0856176,oryCun2:0.0266581)0.99
:0.0212798,((jacJac1:0.0987852,((mm10:0.0372053,rn5:0.0359171)1.00
:0.0266626,(micOch1:0.0568156,(criGri1:0.0165676,mesAur1:0.0135188)0.99
:0.0151017)0.94 :0.00956395)0.99 :0.0188561)1.00
:0.0264911,(speTri2:0.0375365,(hetGla2:0.02556,(cavPor3:0.0379203,(octDeg1:0.118323,chiLa
n1:0.0204085)0.85 :0.00404669)0.92 :0.00895868)1.00 :0.0232632)0.79
:0.00204225)0.84 :0.00339879)0.94
:0.00505521,(tupChi1:0.0476103,(otoGar3:0.0499805,((calJac3:0.00574844,saiBol1:0.00456361
)1.00 :0.00930511,((chlSab1:0.00368577,(papHam1:6e-

008,(macFas5:0.00072364,rheMac3:0.00072651)0.94 :0.00219309)0.87 :0.00142437)0.98
:0.00590278,(nomLeu3:0.00803454,(hg19:0.00073058,((gorGor3:0.00217919,panTro4:6e-
008)0.93 :0.00217552,ponAbe2:0.00511685)0.00 :7e-008)0.88 :0.00155083)0.87
:0.0022914)0.98 :0.0066349)1.00 :0.0168596)0.91 :0.00619577)0.63 :0.00057843)0.96
:0.00555991,(dasNov3:0.0304368,((loxAfr3:0.0450844,triMan1:0.0165639)0.98
:0.0123257,((chrAsi1:0.0533801,echTel2:0.052989)0.62 :0.0046883,oryAfe1:0.0337237)0.58
:0.00091322)0.99 :0.0112845)0.92 :0.00522304)1.00
:0.0113655,((pteAle1:0.00388615,pteVam1:0.00398453)1.00
:0.0550374,(felCat5:0.0332636,(((ailMel1:0.0271614,musFur1:0.0337699)0.74
:0.00312041,(lepWed1:0.019165,odoRosDi:0.00831526)0.97 :0.00856045)0.40
:0.00453015,canFam3:0.0323225)0.96 :0.00830633)1.00 :0.0247181)0.54
:0.00104489,((cerSim1:0.0161512,equCab2:0.0213758)0.97
:0.00682934,(((camFer1:0.00488398,vicPac2:0.00236493)1.00
:0.0264767,(susScr3:0.0478804,((bosTau7:0.0103307,(panHod1:0.00422455,(capHir1:0.004263
36,oviAri3:0.00286926)0.86 :0.0022023)0.95 :0.00595797)1.00
:0.031872,(orcOrc1:0.00234852,turTru2:0.00123681)1.00 :0.0212617)0.90
:0.00568621)0.89 :0.00363082)0.99 :0.0111512,((conCri1:0.0443008,sorAra2:0.098017)0.91
:0.0102665,(eriEur2:0.116644,(eleEdw1:0.14555,(eptFus1:0.0163758,(myoDav1:0.01515,myoL
uc2:0.00254169)0.95 :0.0101269)1.00 :0.0592239)0.71 :0.0128595)0.17 :0.0156363)0.83
:0.00646448)0.82 :0.00224554)0.72 :0.00075171);

LRFN5: ((dasNov3:0.0318458,((loxAfr3:0.0191675,triMan1:0.0144298)0.97
:0.00627334,((eleEdw1:0.037298,oryAfe1:0.0248276)0.74
:0.00399622,(echTel2:0.0671128,chrAsi1:0.0309183)0.07 :0.00320143)0.86
:0.00235037)1.00 :0.00848744)0.84
:0.00398305,((((calJac3:0.00483367,saiBol1:0.00776163)1.00
:0.00773721,((chlSab1:0.00312383,(papHam1:0.00102598,(macFas5:1e-008,rheMac3:1e-
008)0.79 :0.00051348)0.75 :0.0004841)0.98
:0.00433112,(nomLeu3:0.00955918,(ponAbe2:0.00604416,(hg19:0.00258286,(panTro4:0.00207
407,gorGor3:0.00311947)0.72 :0.00049324)0.93 :0.00176096)0.17 :0.00042452)0.94
:0.00300433)0.82 :0.002707)1.00 :0.017587,(tupChi1:0.0379258,otoGar3:0.0985895)0.83
:0.00908717)0.83 :0.00233482,((ochPri3:0.0469995,oryCun2:0.022518)1.00
:0.0218685,(speTri2:0.0375964,((hetGla2:0.0425399,(cavPor3:0.0361735,(chiLan1:0.021009,oc
tDeg1:0.0480313)0.83 :0.00305516)0.95 :0.00689752)1.00
:0.0229595,(jacJac1:0.0788883,((mm10:0.015615,rn5:0.0237309)0.98
:0.0108368,(micOch1:0.0312104,(criGri1:0.0122702,mesAur1:0.0195854)1.00
:0.0134545)0.80 :0.00617706)1.00 :0.034339)0.97 :0.0144288)0.87 :0.00415255)0.99
:0.00781307)0.82 :0.00176424)0.99
:0.00558171,(conCri1:0.0431041,(((felCat5:0.0262974,(canFam3:0.0201341,(musFur1:0.01042
84,(ailMel1:0.0189768,(lepWed1:0.00843224,odoRosDi:0.00905721)0.93 :0.00281516)0.74
:0.00049886)0.84 :0.00181713)0.93 :0.00381748)1.00
:0.00936281,((sorAra2:0.0572675,eriEur2:0.103489)0.90
:0.00887645,((eptFus1:0.0235021,(myoDav1:0.0135051,myoLuc2:0.006397)0.98
:0.00844063)1.00 :0.0225213,(pteAle1:0.00089945,pteVam1:0.00116055)1.00
:0.0197367)0.68 :0.00337329)0.00 :6e-008)0.87
:0.00122768,((cerSim1:0.0184677,equCab2:0.0263263)0.97

:0.00619997,((camFer1:0.00404404,vicPac2:0.00416504)1.00
:0.0284819,(susScr3:0.0226796,((bosTau7:0.00718105,(panHod1:0.00201075,(capHir1:0.00206
515,oviAri3:0.00204003)0.85 :0.0010597)0.66 :0.00175357)1.00 :0.0274274,(orcOrc1:5e-
008,turTru2:0.00204089)1.00 :0.0145571)0.93 :0.00513814)0.33 :0.00102878)0.97
:0.00588712)0.06 :0.00145147)0.61 :0.00070464)0.97 :0.00510741);

MSL1:
(speTri2:0.0224281,((tupChi1:0.0121128,(((((odoRosDi:0.00109101,(canFam3:0.00536747,(mu
sFur1:1.2e-007,(ailMel1:0.00213755,(lepWed1:0.00320526,felCat5:7e-008)0.00 :7e-008)0.86
:0.00106359)0.00 :7e-008)0.31 :0.00104601)0.99 :0.00650514,((((camFer1:2.6e-
007,vicPac2:0.00106484)0.78
:0.00293806,(susScr3:0.00723797,((orcOrc1:0.00106595,turTru2:1e-008)0.92
:0.00437777,(bosTau7:0.0036306,(panHod1:1e-008,(capHir1:1e-008,oviAri3:1e-008)0.00 :1e-
008)0.83 :0.001975)0.97 :0.00813381)0.30 :0.00260788)0.71 :0.00181286)0.96
:0.00546003,(eriEur2:0.0268699,sorAra2:0.0208485)0.63 :0.00137338)0.78
:0.00156905,((cerSim1:0.00876284,(conCri1:0.0124684,equCab2:0.00682486)0.33
:0.00083457)0.75
:0.0010816,((eptFus1:0.00218162,(myoDav1:0.0110614,myoLuc2:0.00219969)0.85
:0.00207264)0.98 :0.00892742,(pteAle1:1e-008,pteVam1:1e-008)0.98 :0.00664995)0.81
:0.0020657)0.77 :0.00104746)0.00 :5e-008)0.88
:0.00215269,(dasNov3:0.022645,((eleEdw1:0.0308487,(echTel2:0.037209,(loxAfr3:0.0054376,t
riMan1:0.00661684)0.76 :0.00097929)0.74 :0.00121823)0.71
:0.00085257,(chrAsi1:0.0192974,oryAfe1:0.0132616)0.73 :0.00109194)0.95
:0.00567834)0.70 :0.00092765)0.93
:0.00323301,((calJac3:0.00437073,saiBol1:0.00102353)0.95
:0.00592161,(otoGar3:0.00936697,((rheMac3:1e-008,(papHam1:1e-008,(chlSab1:1e-
008,macFas5:1e-008)0.00 :1e-008)0.00 :1e-008)0.79
:0.00106775,(nomLeu3:0.00214049,(hg19:0.0010654,(gorGor3:0.00213634,(panTro4:1e-
008,ponAbe2:0.0010654)0.00 :7e-008)0.00 :6e-008)0.00 :6e-008)0.80 :0.00106877)0.78
:0.00158588)0.63 :0.00214705)0.78 :0.00173874)0.00 :1e-
007,(ochPri3:0.0225345,oryCun2:0.00768134)0.81 :0.00212661)0.00 :6e-008)0.78
:0.00109597,(jacJac1:0.0172782,((mesAur1:0.0082598,criGri1:0.00596939)0.83
:0.00220242,(micOch1:0.00930889,(mm10:0.0136225,rn5:0.0121802)0.99 :0.0142137)0.64
:0.00069856)1.00 :0.0202659)0.87 :0.00441366)0.00 :1.15e-
006,(hetGla2:0.0151803,(cavPor3:0.0186586,(chiLan1:0.0176824,octDeg1:0.0392829)0.82
:0.00851798)0.89 :0.00847947)1.00 :0.0184795);

NUDT12:
(speTri2:0.0425823,((hetGla2:0.0357194,(octDeg1:0.0561017,(chiLan1:0.0386855,cavPor3:0.05
09025)0.55 :0.00142625)0.81 :0.00297384)0.99
:0.0145238,(jacJac1:0.0940661,((mm10:0.0473469,rn5:0.043772)1.00
:0.0467127,(micOch1:0.0530517,(criGri1:0.0274031,mesAur1:0.0325238)1.00
:0.0341697)0.98 :0.0259445)1.00 :0.0557146)0.92 :0.0151456)0.00 :1.7e-
007,(((ochPri3:0.0907288,oryCun2:0.0233144)1.00
:0.0275419,(((calJac3:0.00850664,saiBol1:0.00299127)1.00
:0.0134525,((chlSab1:0.0012899,(papHam1:0.00459331,(macFas5:0.00075802,rheMac3:1e-

008)0.78 :0.00076655)0.75 :0.00098603)0.96
:0.00560302,((panTro4:0.00151027,(hg19:0.0053391,gorGor3:0.00455136)0.00 :8e-008)0.99
:0.00536743,(nomLeu3:0.0053599,ponAbe2:0.00538995)0.75 :0.00074778)0.96
:0.00548682)0.87 :0.00615404)1.00
:0.0298959,((dasNov3:0.0431147,(chrAsi1:0.0455918,((oryAfe1:0.0376974,echTel2:0.0762268)
0.65 :0.00156073,(eleEdw1:0.0871813,(loxAfr3:0.0250531,triMan1:0.0249174)0.57
:0.00723485)0.80 :0.00298499)0.68 :0.00164909)0.99 :0.0134537)0.87
:0.0050861,((((camFer1:0.0055332,vicPac2:0.00206213)1.00
:0.0183576,(susScr3:0.0239199,((bosTau7:0.00762166,(panHod1:0.00389328,(capHir1:0.00161
154,oviAri3:0.00294361)0.81 :0.00157972)0.61 :0.0005261)1.00
:0.0226462,(orcOrc1:0.00343307,turTru2:0.00264252)1.00 :0.0170346)0.88
:0.00437814)0.72 :0.00085922)0.97
:0.00589454,(((ailMel1:0.0117042,(musFur1:0.0151096,(lepWed1:0.00899603,odoRosDi:0.007
27803)0.94 :0.00477357)0.79 :0.00128439)0.90
:0.00255516,(canFam3:0.015646,felCat5:0.0186303)0.44 :0.00078944)1.00
:0.0112148,(conCri1:0.0575939,(sorAra2:0.125381,eriEur2:0.0610922)0.82 :0.00867204)0.87
:0.00604912)0.48 :0.00105968)0.48
:0.0014471,((eptFus1:0.00842926,(myoDav1:0.0100876,myoLuc2:0.00155801)0.90
:0.00321662)1.00 :0.020784,((pteAle1:0.00182038,pteVam1:0.00344592)1.00
:0.0316241,(cerSim1:0.00859963,equCab2:0.0295157)0.98 :0.00911111)0.00
:0.00037414)0.75 :0.00097393)0.94 :0.00620674)0.99 :0.00765129)0.00 :2.8e-007)0.66
:0.00151878,(tupChi1:0.0455097,otoGar3:0.107945)0.72 :0.00499685)0.99 :0.0131177);

PTPRA:
((hetGla2:0.0268904,(cavPor3:0.0331103,(chiLan1:0.0184881,octDeg1:0.0270941)0.94
:0.0064955)0.98 :0.00875287)1.00
:0.0315438,(speTri2:0.0448732,(jacJac1:0.0552835,((micOch1:0.0276857,(criGri1:0.0167669,m
esAur1:0.0161632)0.88 :0.00408778)0.94 :0.00585956,(mm10:0.027634,rn5:0.0229902)1.00
:0.0201627)1.00 :0.0358993)0.98 :0.0148902)0.74
:0.00204476,((ochPri3:0.0499259,oryCun2:0.0352172)1.00
:0.0283439,((otoGar3:0.0488228,(((conCri1:0.0492213,(sorAra2:0.0694607,eriEur2:0.0646959)
0.84 :0.00672661)0.97
:0.0075155,(((felCat5:0.0204158,(canFam3:0.0218601,(musFur1:0.0197894,(ailMel1:0.0151697
,(lepWed1:0.00647877,odoRosDi:0.00533484)1.00 :0.00997791)0.73 :0.00072647)0.94
:0.00328284)0.97 :0.00595109)1.00
:0.0109434,((pteAle1:0.00155829,pteVam1:0.00171926)1.00
:0.0274149,(eptFus1:0.00556836,(myoDav1:0.00646891,myoLuc2:0.00429125)0.99
:0.00597579)1.00 :0.0292591)0.96 :0.0057771)0.49
:0.00108902,(((camFer1:0.00498721,vicPac2:0.00250175)1.00
:0.0301485,(susScr3:0.025868,((bosTau7:0.007496,(panHod1:0.00373799,(oviAri3:0.00278984,
capHir1:0.00093465)0.00 :4e-008)0.98 :0.00530317)1.00
:0.0221771,(orcOrc1:0.00259109,turTru2:0.00113714)1.00 :0.00937249)0.77
:0.00175597)0.57 :0.00356505)1.00
:0.0120025,(cerSim1:0.0180549,equCab2:0.0215482)1.00 :0.0110019)0.79 :0.0019811)0.83
:0.00203881)1.00
:0.00985003,(dasNov3:0.0652303,(eleEdw1:0.0623985,(oryAfe1:0.0344514,((loxAfr3:0.033804

1,triMan1:0.0150292)0.97 :0.00723591,(echTel2:0.0566443,chrAsi1:0.0492765)0.84
:0.00406868)0.63 :0.00105058)0.81 :0.00309594)1.00 :0.0145081)0.69
:0.00492782)0.93 :0.00368391)0.60
:0.00197567,(tupChi1:0.0624698,((calJac3:0.00804924,saiBol1:0.00640387)1.00
:0.0102931,((chlSab1:0.00233851,(papHam1:0.00185493,(macFas5:5e-
008,rheMac3:0.00046156)0.74 :0.00046207)0.76 :0.00044473)1.00
:0.004678,(nomLeu3:0.00950289,(ponAbe2:0.00331403,(panTro4:0.00325956,(hg19:0.0013864
9,gorGor3:0.00185452)0.00 :4e-008)0.90 :0.00134775)0.75 :0.00044318)0.89
:0.00142981)0.97 :0.0059635)1.00 :0.021963)0.67 :0.0013095)0.85 :0.00362697)0.99
:0.0112292);

SGMS2:
tupChi1:0.04833,(sorAra2:0.108439,((eriEur2:0.135287,conCri1:0.0907102)0.09
:0.0161817,(hetGla2:0.0968242,(cavPor3:0.0843947,(octDeg1:0.0681874,chiLan1:0.0499745)0.
89 :0.0194923)1.00 :0.0512568)1.00 :0.135671)0.90 :0.0274779)0.98
:0.0556303,((eptFus1:0.0265863,(myoDav1:0.023898,myoLuc2:0.0117866)0.89
:0.0156714)1.00 :0.064967,((jacJac1:0.0777415,((mm10:0.0554998,rn5:0.0350787)1.00
:0.0392444,(micOch1:0.0522111,(criGri1:0.0105598,mesAur1:0.0282327)0.97
:0.0174241)0.70 :0.00598433)1.00 :0.0698775)0.73
:0.0233108,((((bosTau7:0.00999796,(panHod1:0.00614558,(capHir1:0.00041066,oviAri3:0.007
10956)0.92 :0.00541065)0.99 :0.0134307)1.00
:0.0564218,(orcOrc1:0.00336311,turTru2:0.00324603)1.00 :0.0226077)0.79
:0.00428688,(susScr3:0.0575526,(camFer1:0.00816558,vicPac2:0.00490805)1.00
:0.0637746)0.68 :0.00492172)0.92
:0.00896535,((((speTri2:0.0526967,(ochPri3:0.0696533,oryCun2:0.0181164)1.00
:0.0474739)0.00
:0.00606093,(otoGar3:0.0871842,((ponAbe2:0.00198538,((panTro4:0.00112562,(hg19:1.1e-
007,gorGor3:0.0022532)0.00 :1.1e-007)0.92
:0.00224737,(nomLeu3:0.00681556,(chlSab1:0.00342103,(papHam1:0.00112664,(macFas5:1e-
008,rheMac3:1e-008)0.80 :0.00112912)0.92 :0.00340187)0.96 :0.00459561)0.00 :6e-
008)0.83 :0.00252821)0.98 :0.0143342,(calJac3:0.0128919,saiBol1:0.0118061)0.98
:0.0138027)0.98 :0.0223547)0.94 :0.0121142)0.82
:0.00365608,(dasNov3:0.04528,(eleEdw1:0.0765632,((chrAsi1:0.041895,(loxAfr3:0.0276129,tri
Man1:0.00748677)0.95 :0.0114076)0.00
:0.00051051,(oryAfe1:0.0389153,echTel2:0.158075)0.73 :0.00483841)0.79
:0.00344324)0.98 :0.0178768)0.73 :0.00392889)0.96
:0.0144574,((felCat5:0.0183984,(canFam3:0.0326621,(ailMel1:0.0252866,(musFur1:0.0490313,
(lepWed1:0.00698862,odoRosDi:0.0102898)0.89 :0.00632492)0.91 :0.00865389)0.83
:0.00469485)0.84 :0.00593671)1.00 :0.024701,((pteAle1:0.00443157,pteVam1:6.1e-
007)1.00 :0.0441138,(cerSim1:0.0270733,equCab2:0.0444156)0.99 :0.0254533)0.33
:0.00397169)0.84 :0.00613964)0.86 :0.0105326)0.92 :0.0240844)0.73 :0.0313321)1.00
:0.10503);

TTC5:
((((hetGla2:0.0264745,(octDeg1:0.0413466,(chiLan1:0.044299,cavPor3:0.0474303)0.84
:0.00611231)0.98 :0.0124802)1.00 :0.0206428,speTri2:0.0436478)0.82

:0.00341975,(jacJac1:0.0701968,((micOch1:0.0596696,(criGri1:0.0162028,mesAur1:0.0534992)0.97 :0.0143606)0.94 :0.0123306,(mm10:0.0328693,rn5:0.0406286)1.00 :0.0283616)1.00 :0.0537968)0.91 :0.00990701)0.99 :0.011198,(ochPri3:0.0863282,oryCun2:0.0496241)1.00 :0.0318025,(tupChi1:0.0556265,((otoGar3:0.0479033,((calJac3:0.0100017,saiBol1:0.0127578)1.00 :0.0195696,((chlSab1:0.0087537,(papHam1:0.00078188,(macFas5:1e-008,rheMac3:0.0023566)0.89 :0.0015764)0.82 :0.00242194)1.00 :0.014905,(ponAbe2:0.00647828,(nomLeu3:0.00235143,(gorGor3:0.00078348,(hg19:0.000782,panTro4:0.00157126)0.77 :0.00078521)0.98 :0.00399449)0.82 :0.00146781)0.00 :5.2e-007)0.97 :0.00790913)1.00 :0.0315007)0.76 :0.00344387,((dasNov3:0.067816,((chrAsi1:0.0546978,eleEdw1:0.0807454)0.81 :0.00481242,((echTel2:0.0964234,oryAfe1:0.0360768)0.80 :0.0056093,(loxAfr3:0.0220795,triMan1:0.0207824)0.98 :0.0100591)0.00 :7.349e-005)0.93 :0.0073919)0.92 :0.00589321,((felCat5:0.0320141,(canFam3:0.119123,((ailMel1:0.0323163,(lepWed1:0.0116698,odoRosDi:0.00446856)1.00 :0.0143813)0.05 :0.00127196,musFur1:0.0175834)0.91 :0.00560905)0.13 :0.00214188)1.00 :0.0205836,((cerSim1:0.0184871,equCab2:0.0269502)0.95 :0.00614456,(((susScr3:0.0312493,((orcOrc1:0.00248977,turTru2:0.00067239)1.00 :0.0215981,(bosTau7:0.00964364,((panHod1:0.00529606,capHir1:0.00265264)0.28 :0.00080467,oviAri3:0.00664175)0.97 :0.00718947)1.00 :0.020393)0.70 :0.00318006)0.72 :0.00112734,(camFer1:0.00483822,vicPac2:0.00232235)1.00 :0.0172551)0.99 :0.00901076,(((pteAle1:0.00068014,pteVam1:0.00281784)1.00 :0.0427225,(eptFus1:0.00796342,(myoDav1:0.00923764,myoLuc2:0.00287082)0.95 :0.00736384)1.00 :0.0267545)0.63 :0.00348544,(conCri1:0.058549,(sorAra2:0.0716877,eriEur2:0.0747044)0.65 :0.00844031)0.94 :0.0116335)0.76 :0.00246178)0.76 :0.00067702)0.73 :0.00097847)0.98 :0.00632163)0.86 :0.00254937)0.91 :0.0043503)0.70 :0.00151902);

We used the following tree for estimating evolutionary rate covariation (ERC) with PON1 in terrestrial mammals (branch lengths in this tree were not used in the analysis):

((((((((((((((ailMel1:1,(lepWed1:1,odoRosDi:1):1):1,musFur1:1):1,canFam3:1):1,felCat5:1):1,(cerSim1:1,equCab2:1):1):1,((eptFus1:1,(myoDav1:1,myoLuc2:1):1):1,(pteAle1:1,pteVam1:1):1):1):1,(((((bosTau7:1,oviAri3:2):1,panHod1:1):1,(orcOrc1:1,turTru2:1):1):1,vicPac2:2):1,susScr3:1):1):1,((conCri1:1,sorAra2:1):1,eriEur2:1):1):1,((((chrAsi1:1,echTel2:1):1,((eleEdw1:1,loxAfr3:1):1,triMan1:1):1):1,oryAfe1:1):1,dasNov3:1):2):1,((((cavPor3:1,(chiLan1:1,octDeg1:1):1):1,hetGla2:1):1,(((((criGri1:1,mesAur1:1):1,micOch1:1):1,(mm10:1,rn5:1):1):1,jacJac1:1):1,speTri2:1):1):1,(ochPri3:1,oryCun2:1):1):1):1,tupChi1:1):1,otoGar3:1):1,(calJac3:1,saiBol1:1):1):1,rheMac3:4):4,hg19:1);

Sample acquisition and permissions

We obtained blood samples for seven healthy, wild manatees from Crystal River, Florida. Blood was obtained using previously published techniques for manatee blood sample collection (*80*, *81*). We obtained blood and skin samples from three healthy, wild dugongs from Moreton Bay, Australia using published sampling techniques (*82*, *83*). We obtained blood from the fluke veins of two healthy adult bottlenose dolphins (one 10-year-old male and one 23-year-old male) in human care and from the digital vein of one healthy adult California sea lion (21-year-old

male) in human care using a butterfly needle. We obtained blood from the tarsal vein of one healthy adult walrus (23-year-old female) in human care using a 21ga 1.5" needle. We obtained blood from one healthy juvenile (<1 year old) female Northern elephant seal and one healthy adult (3-year-old) female Canadian beaver in human care. All blood collection from animals in human care took place during routine health examinations. We obtained blood during necropsy from four adult ferrets that had previously been exposed to influenza but had recovered at time of sacrifice. We obtained blood from eight healthy, wild Northern elephant seals (five males and three females) from Año Nuevo State Reserve in San Mateo County, CA, USA. Five wild-type mice (C57BL/6J strain background) were purchased from The Jackson Laboratory (Bar Harbor, MI). Five *Pon1* knockout mice (*Pon1*$^{-/-}$) were kindly provided by Drs. Lusis, Shih and Tward (UCLA, Los Angeles, CA) (*26*). Wild-type and *Pon1*$^{-/-}$ mice were anesthetized with tribromoethanol (600 mg/kg, ip; Sigma-Aldrich, St. Louis, MO) and blood extracted via cardiac puncture. Mice were housed in a centralized, AAALAC-accredited, specific pathogen free facility at the University of Washington. They were maintained at room temperature in a 12 h light-dark cycle with unlimited access to food and water.

We obtained all appropriate animal care and use permissions from the relevant research institutions and management organizations, as follows:

1. Blood samples from two bottlenose dolphins, one Canadian beaver, one California sea lion, seven manatees, one Northern elephant seal, and one walrus in human care were obtained using procedures approved by the Pittsburgh Zoo and PPG Aquarium / National Aviary IACUC (protocol # 2015-NC-001) and each sampling institution's research review committees prior to conduction. All manatee samples were collected and held by USGS under IACUC protocol #USGS-WARC-2016-03 and USFWS research permit MA791721.
2. Dugong blood and tissue samples were collected under Australian Scientific Purposes Permit no. WISP01660304, Moreton Bay Marine Park Permit no. QS2004/CVL228 and University of Queensland Animal Ethics no. ZOO/ENT/344/04/NSF/CRL, and transferred to the USGS Sirenian Project laboratory under authority of the CITES permits 08US808447/9 and 2009AU570750.
3. Ferret blood samples were obtained with approval from the University of Pittsburgh IACUC protocol #16077170.
4. Blood samples from eight wild Northern elephant seals were obtained with approval from Sonoma State University IACUC protocol #2014-48, under NMFS permit #19108.
5. All mouse experiments were approved by the Animal Care and Use Committee of the University of Washington (IACUC protocol # 2343-01), and carried out in accordance with National Research Council Guide for the Care and Use of Laboratory Animals, as adopted by the US National Institutes of Health.

Validating manatee *PON1* coding sequence and determining dugong sequence

Manatee DNA was extracted by a standard phenol-chloroform technique after extracting cells from clotted blood (*84*). Dugong DNA was extracted from two whole blood samples using the salting out procedure of Miller, Dykes, and Polesky (*85*) and from one skin sample using a standard phenol-chloroform technique. We designed primers for all exons of PON1 using Primer3Plus version 2.4.0 and the manatee genome sequence as reference. All exons were amplified using PCR, which was carried out in a 20 μL volume containing: 10 μL of 10x Taq

polymerase buffer (New England BioLabs), 0.4 μL each of 10 μM forward and reverse primers, 0.4 μL of 10 mM dNTP mix, 0.5 μL of template DNA (16 – 50 ng/μL) 16.3 μl of water, and 0.08 μL of Taq polymerase. The thermal cycler was programmed for 3 min at 95 °C for initial denaturation, then 34 cycles of 30 s at 95 °C for denaturation, 30 s at 59 °C for annealing, and 45 s at 72 °C for extension, followed by 5 min at 72 °C for the final extension, with minimal adjustments (see Table S12).

PCR products were sequenced by Sanger sequencing at the University of Pittsburgh Genomics Research Core. Sequencing Reaction Sequencing buffer and a 1:4 dilution of BigDye 3.1 were added and thermocycling performed according to ABI recommendations. Removal of unincorporated sequencing reagents was performed using CleanSeq magnetic beads according to manufacturer instructions (Agencourt). The resulting sequence files were manually inspected to confirm homozygosity at observed lesions. The raw data for each exon was aligned to the reference exon using MEGA 7.0.14, and inconsistencies and splice sites were checked manually. Individual exon sequences were then concatenated to generate a consensus coding sequence for *PON1* for dugong.

Evaluating primary function(s) of PON1 using evolutionary rate covariation (ERC)

To generate Tables S6 and S7, we assessed the extent to which genes' branch-specific rates of amino acid evolution correlated with those of PON1 by implementing an evolutionary rate covariation (ERC) analysis, as described in Clark and Aquadro (*86*). We included only non-marine species in this analysis, in order to capture patterns of co-evolution that are not primarily influenced by the loss of function of PON1. To reduce the influence of long branches on the results, we constrained our analyses to eutherian mammals and further pruned the tree to eliminate members of pairs with very short evolutionary distances (see "Constraints on PON1 phylogenetic trees for evolutionary inferences"). To reduce artifacts driven by variation in genes with high levels of missing data, we restricted our analyses to those genes with available sequences for at least 30 species. We estimated evolutionary rate covariation for all remaining 17,511 genes with *PON1*, using gene trees with branch lengths estimated as in Chikina et al. (*5*). To evaluate enrichment of functional categories within our top signals, we performed gene ontology enrichment analysis using GOrilla (*46*), focusing on the top 100 genes that showed a positive correlation in rate with the rate of *PON1*.

Assessing enzymatic activity of blood plasma against four PON1 substrates and control substrate (alkaline phosphatase)

Blood was collected in lithium heparin tubes and centrifuged at 1,500 - 10,000 x g for 10 - 15 min at 4 °C. Plasma was separated from the blood cell fraction and kept stored at -80 °C until use.

All activity assays were determined in a SPECTRAmax® PLUS Microplate Spectrophotometer (Molecular Devices, Sunnyvale, CA). The assay values were corrected for path-length using the software SoftMax Pro 5.4 (Molecular Devices). Levels of plasma arylesterase (AREase), chlorpyrifos-oxonase (CPOase), diazoxonase (DZOase) and paraoxonase (POase) activities were determined as previously described (*87*). Briefly, plasma of all the species analyzed were diluted 1/10 in dilution buffer (9 mM Tris-HCl pH 8.0, 0.9 mM $CaCl_2$) and assayed in triplicate at either 37 °C (for CPOase and POase) or at room temperature (for AREase and DZOase). Activities were expressed in U/mL (AREase, CPOase, and DZOase) or in U/L (POase), based on the molar extinction coefficients of 1.31 mM^{-1} cm^{-1} for phenol (the

hydrolysis product of phenyl acetate, AREase activity); 5.56 mM^{-1} cm^{-1} for 3,5,6-trichloropyridinol (the hydrolysis product of CPO); 3 mM^{-1} cm^{-1} for 2-isopropyl-4-methyl-6-hydroxypyrimidine (the hydrolysis product of DZO); or 18 mM^{-1} cm^{-1} for p-nitrophenol (the hydrolysis product of PO). Alkaline phosphatase was assayed in undiluted plasma of all species at 37 °C in triplicate as follows: The plasma sample, 10 µL, was added to 170 µL of 0.95 M diethanolamine pH 9.8, 0.5 mM $MgCl_2$. The assay was initiated by adding 20 µL of 112 mM p-nitrophenyl phosphate in water (*88*). The absorbance at 405 nm was followed for 4 min. Activities were expressed in U/L based on the molar extinction coefficient of 18.0 mM^{-1} cm^{-1} for p-nitrophenol (the hydrolysis product of p-nitrophenyl phosphate). The numeric results of each replicate for each substrate are provided in Table S13.

Phenyl acetate (CAS 122-79-2, 99% purity), p-nitrophenyl phosphate (CAS 333338-18-4, ≥97% purity), and other reagent chemicals were purchased from Sigma-Aldrich. Chlorpyrifos oxon (CPO; CAS 5598-15-2; 98% purity), diazoxon (DZO; CAS 962-58-3; 99% purity) and paraoxon (PO; CAS 311-45-5, 99% purity) were purchased from Chem Service Inc. (West Chester, PA).

Visualizing proximity of agricultural land to manatee habitat

We created maps using QGIS software (version 2.10.1-Pisa). We acquired federally mandated manatee protection zone information from the U.S. Fish and Wildlife Service's Environmental Conservation Online System (*89*). State mandated manatee protection zone information was obtained from the Florida Fish and Wildlife Conservation Commission (*90*). The U.S. Census Bureau TIGER products supplied datasets for the state and county boundaries and Brevard County rivers, canals, lakes, and other waterways (*91–94*).

We obtained agricultural land use data from the Florida Department of Environmental Protection, extracting level 1 land use code 2000 ("agriculture") data and further extracting agricultural land use of specific interest to a geographic exploration of the potential intersection between organophosphate pesticide application and manatee migration areas through the level 2 land use code and description, including cropland and pastureland (level 2 code 2100), nurseries and vineyards (2400), other open lands (2600), and tree crops (2200). This subset discarded agricultural land where organophosphate pesticides are less likely to be applied, including poultry, cattle, and other feeding operations (level 2 code 2300) and specialty farms (2500), which entail horse farms, wet prairies, dairies, aquaculture, tropical fish farms, sewage treatment, and other treatment ponds (*95*).

We created a map focused on Brevard County because Brevard County has been identified as a key area for manatee residence and migration, with an estimated 70% of manatees along the Atlantic coast migrating through or residing in Brevard waterways at least seasonally (*27*, *28*) (Fig. 3). To illustrate the potential for manatee interaction with pesticide water pollution from agricultural areas, we included agriculture land use of interest as described above and all waterways in Brevard County, including lakes, rivers, and canals.

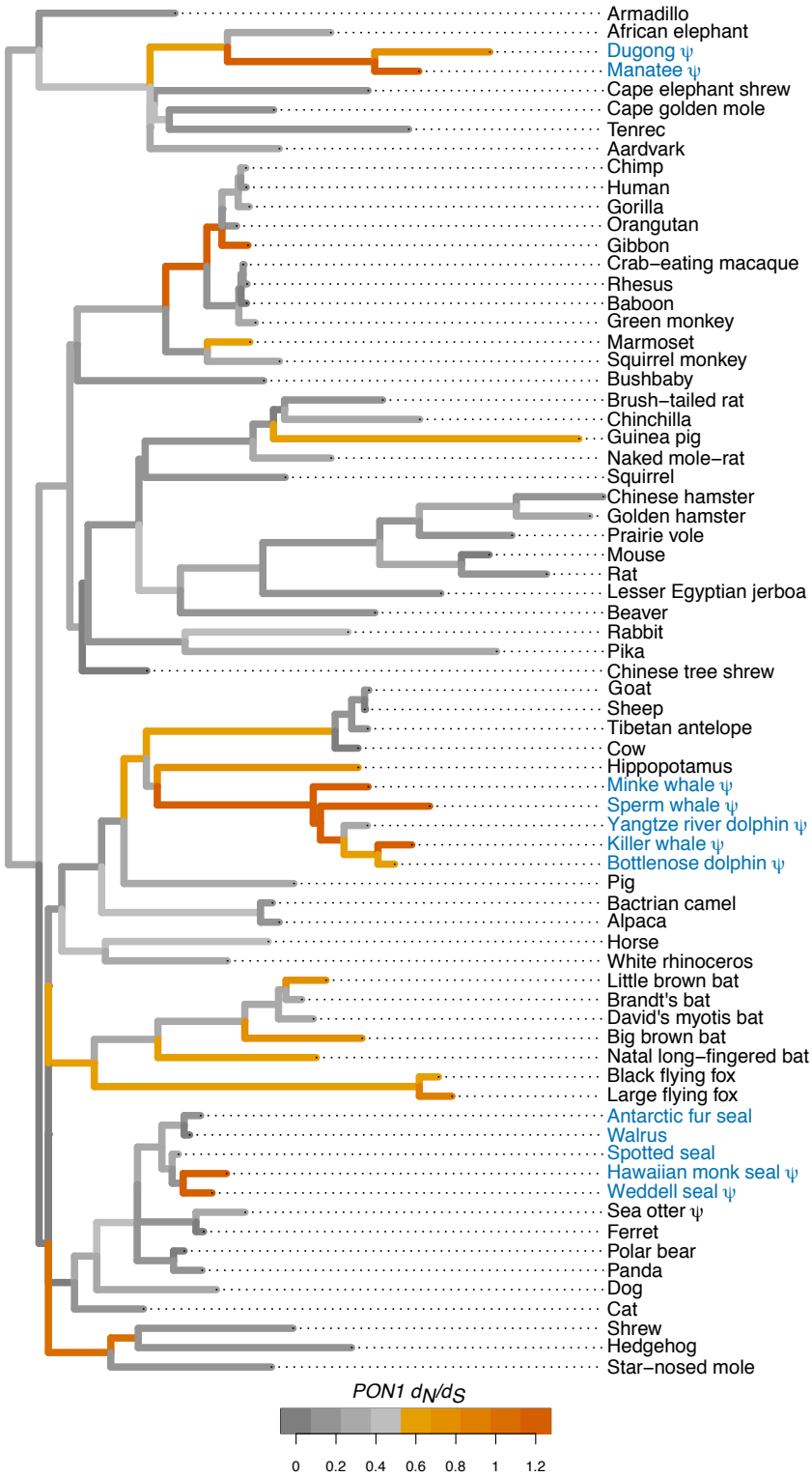

Fig. S1. Evolutionary rate of *PON1* coding sequence across the full mammalian phylogeny. Shown is the phylogeny of 71 eutherian mammals whose sequences were included for rate estimation. See Fig. 1C legend for details.

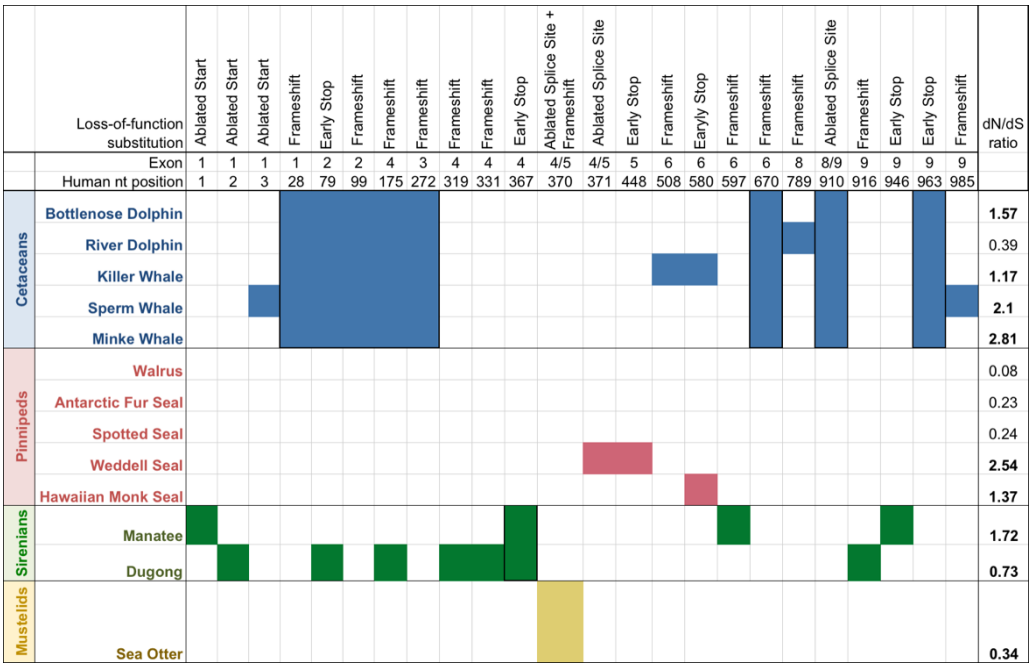

Fig. S2. Loss-of-function substitutions in PON1 across marine and semi-aquatic mammals. Colored cells indicate the observation of the relevant loss-of-function substitution (lesion) in the relevant species. Columns or sets of columns representing lesions shared across all members of a clade are outlined with wider black borders.

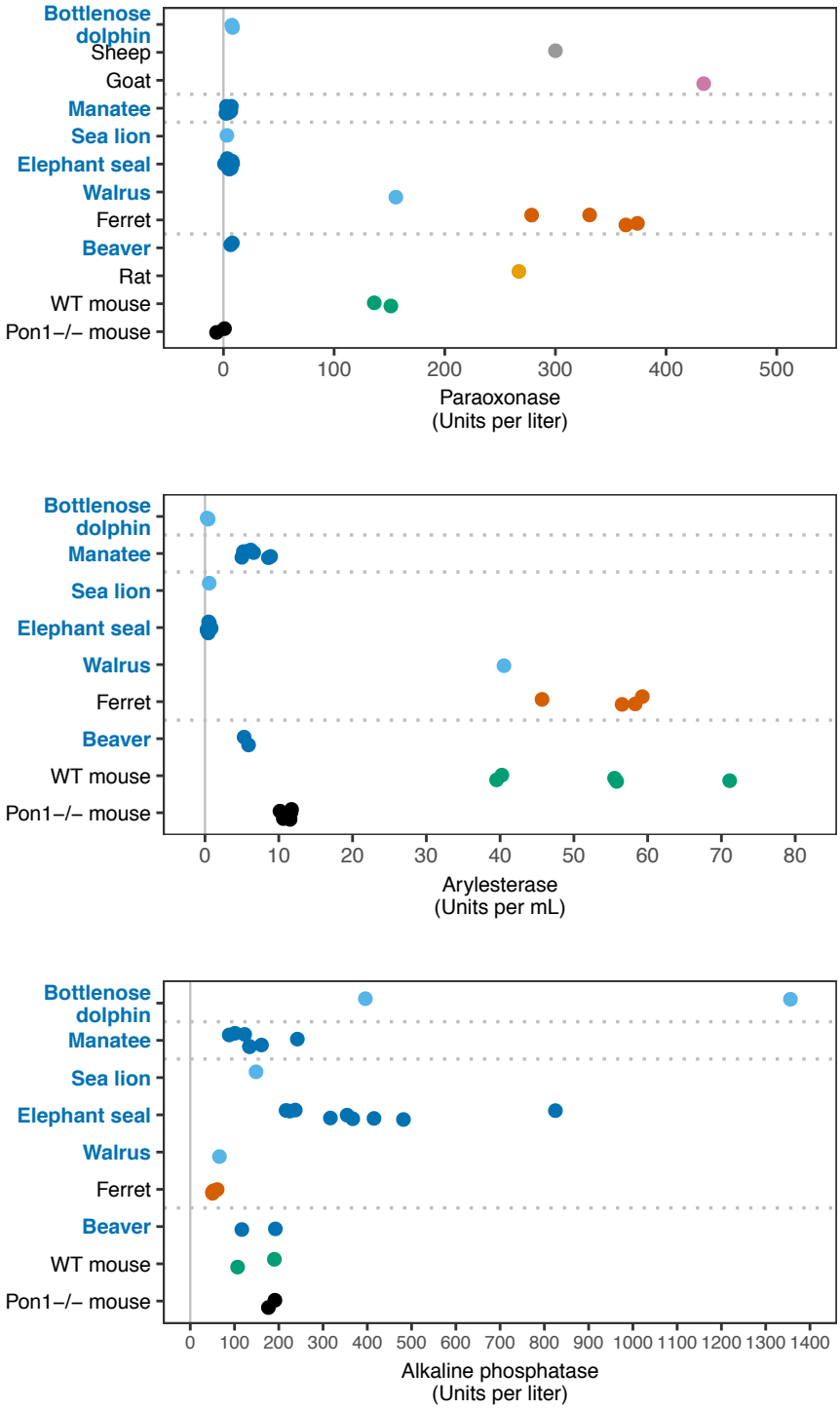

Fig. S3. Blood plasma enzymatic activity against two PON1 substrates and a non-PON1 substrate (control). Points represent rates of hydrolysis of paraoxon in μmol/min/L (top), phenyl acetate in μmol/min/mL (middle), and non-PON1 substrate alkaline phosphatase in μmol/min/L (bottom). In mouse (and potentially other species), carboxylesterase can also hydrolyze phenyl acetate (in addition to PON1). See Fig. 2 legend for additional details.

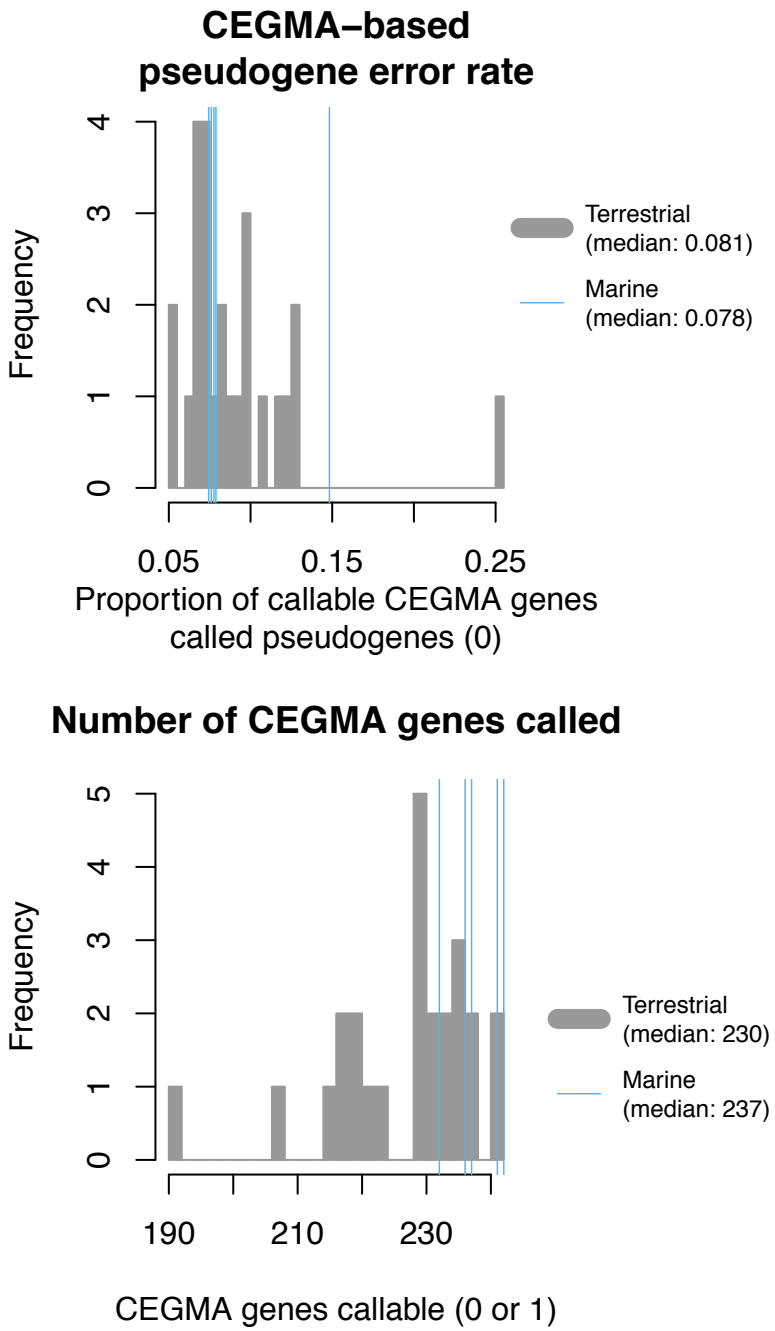

Fig. S4. Per-species pseudogene error rate as estimated from pseudogene calls among highly conserved (CEGMA) genes. Histograms of the proportion of all callable highly conserved genes in the CEGMA (*38*) dataset called as pseudogenes by our automated method for all marine species and the 25 terrestrial species whose genetic distance to reference sequence hg19 was within 0.04 of the marine range (marine range 0.220 – 0.247; terrestrial range 0.187 – 0.284) (top), and the number of CEGMA genes considered callable (not filtered for missing data) out of the 246 genes with alignments in our dataset (bottom).

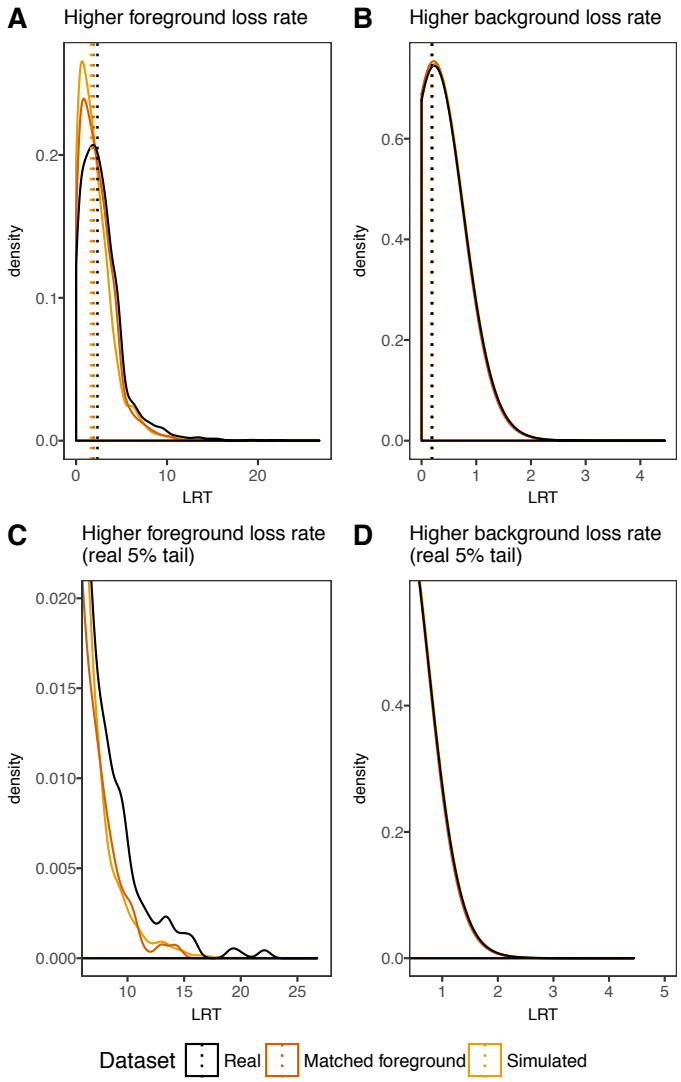

Fig. S5. Distribution of raw LRT statistics in real datasets, compared with those from simulations and a matched foreground set. Plots represent the densities of raw LRT statistics reflecting the significance of a model with different rates of gene functional loss on marine and terrestrial branches, as compared to a model where rates of loss are independent of marine/terrestrial status, for real data (in black), for a test using a set of matched foreground species (in orange), and for 10,000 simulations of each gene under the independent model (in gold). The distribution of the LRT statistic for genes with higher inferred loss rates on foreground (marine in the real dataset) branches (A and C) shows a skew towards higher values in real data as compared with the matched foreground set or simulated data; the upper 5% tail for real data corresponds to 2.64% and 2.65% of the cumulative distribution of matched foreground and simulation sets, respectively (C). This indicates a potential genome-wide signal of preferential gene loss in marine species. In contrast, the distribution of LRT statistics for genes with higher inferred loss rates on background (terrestrial in the real dataset) branches (B and D) is very similar across all three datasets; the upper 5% tail for real data corresponds to 6.19% and 6.33% of the cumulative distribution of matched foreground and simulation sets, respectively (D). Note the large difference in the ranges of the x-axes.

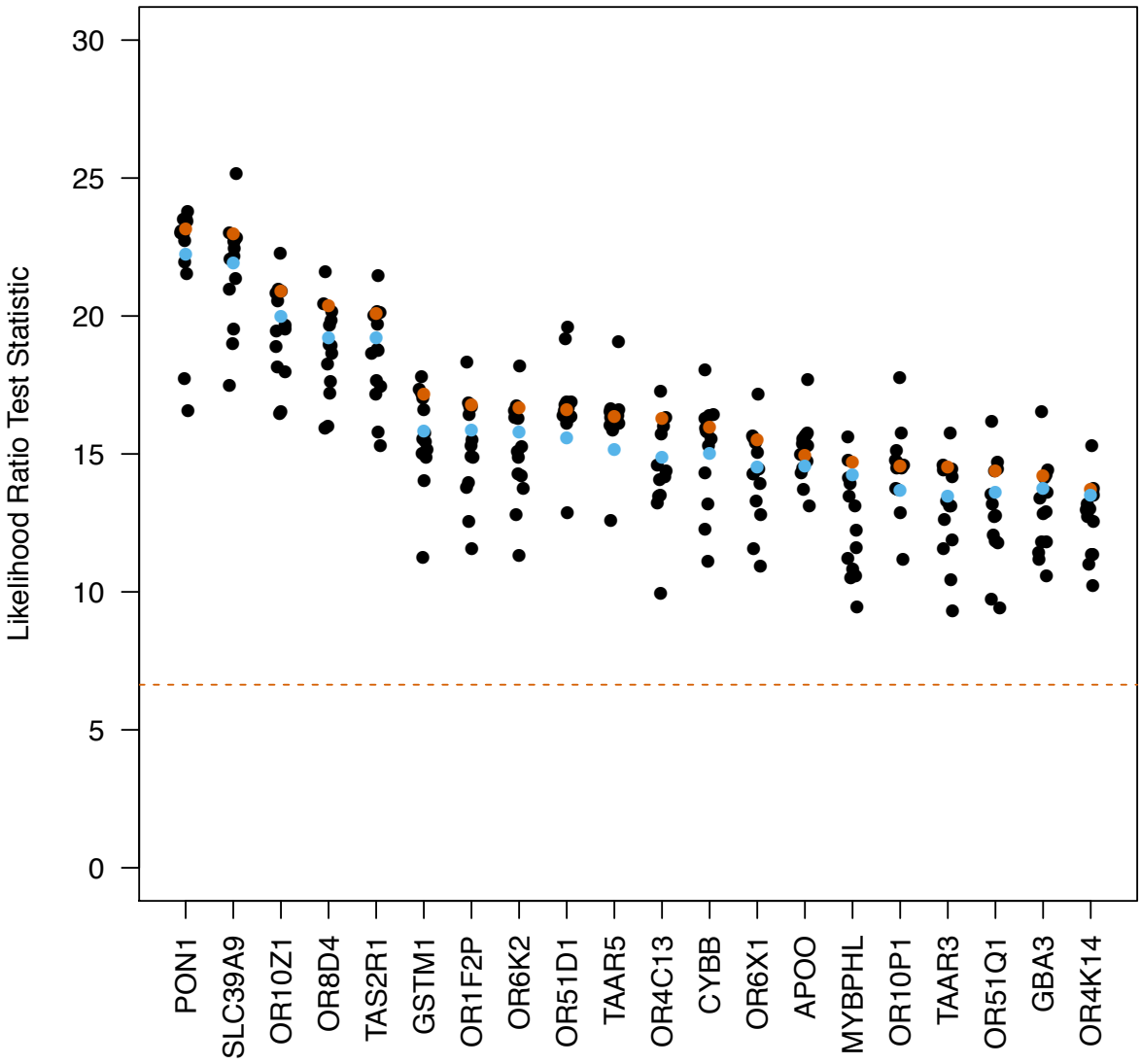

Fig. S6. Evidence for convergence of functional loss in marine species is consistent across varying phylogenetic trees for top genes. Points represent the LRT statistics estimated using the phylogenetic tree used in the main analysis (in blue) and 14 alternate phylogenetic trees, including one re-estimated from a sampled subset of genes but with the same topology as the tree used in the main analysis (in orange). The dashed orange line represents the value of the LRT statistic corresponding to a *P*-value of 0.01 for a chi-square test with one degree of freedom.

Table S2. Gene ontology enrichment for top genes lost in marine lineages. Categories displaying enrichment among genes with strong evidence for higher rate of loss in marine lineages. Gene set enrichment was assessed using the ranked list approach in the GOrilla online enrichment tool (*46*), with all genes included in the BayesTraits analysis as the background gene set. For lists of genes in each category driving the signatures of enrichment, see Table S4.

BIOLOGICAL PROCESS								
GO Term	**Description**	**P-value**	**FDR q-value**	**Enrichment**	**N**	**B**	**n**	**b**
GO:0050907	detection of chemical stimulus involved in sensory perception	3.65E-68	4.78E-64	8.59	9430	285	393	102
GO:0009593	detection of chemical stimulus	4.36E-66	2.85E-62	8.13	9430	304	393	103
GO:0050906	detection of stimulus involved in sensory perception	2.76E-63	1.20E-59	7.79	9430	314	393	102
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	1.54E-58	5.03E-55	8.4	9430	257	393	90
GO:0051606	detection of stimulus	5.11E-57	1.34E-53	6.59	9430	391	384	105
GO:0007186	G-protein coupled receptor signaling pathway	4.12E-41	8.98E-38	4.02	9430	722	393	121
GO:0007606	sensory perception of chemical stimulus	5.08E-16	9.49E-13	8.83	9430	101	275	26
GO:0007608	sensory perception of smell	2.20E-14	3.60E-11	2.72	9430	80	2380	55
GO:0050896	response to stimulus	2.74E-13	3.99E-10	1.55	9430	2803	472	218
GO:0007165	signal transduction	3.70E-12	4.84E-09	1.59	9430	2599	432	189
GO:0050912	detection of chemical stimulus involved in sensory perception of taste	7.53E-10	8.96E-07	12.04	9430	25	376	12
GO:0001580	detection of chemical stimulus involved in sensory perception of bitter taste	4.58E-09	5.00E-06	11.99	9430	23	376	11
GO:0007600	sensory perception	9.46E-08	9.53E-05	4.63	9430	302	135	20
GO:0050877	nervous system process	1.22E-06	1.14E-03	3.05	9430	486	178	28
GO:0042178	xenobiotic catabolic process	1.59E-06	1.39E-03	45.17	9430	5	167	4
GO:0070458	cellular detoxification of nitrogen compound	9.45E-06	7.73E-03	449.05	9430	2	21	2
GO:0051410	detoxification of nitrogen compound	9.45E-06	7.27E-03	449.05	9430	2	21	2
GO:0018916	nitrobenzene metabolic process	1.05E-05	7.65E-03	56.47	9430	3	167	3
GO:0021798	forebrain dorsal/ventral pattern formation	2.54E-05	1.75E-02	23.21	9430	5	325	4
GO:0003008	system process	3.59E-05	2.35E-02	2.35	9430	744	178	33
FUNCTION								
GO Term	**Description**	**P-value**	**FDR q-value**	**Enrichment**	**N**	**B**	**n**	**b**
GO:0004984	olfactory receptor activity	1.54E-58	5.64E-55	8.4	9430	257	393	90
GO:0004930	G-protein coupled receptor activity	8.01E-56	1.47E-52	5.73	9430	493	384	115
GO:0099600	transmembrane receptor activity	2.61E-41	3.18E-38	4.03	9430	737	384	121
GO:0004888	transmembrane signaling receptor activity	2.61E-41	2.39E-38	4.03	9430	737	384	121
GO:0038023	signaling receptor activity	2.36E-39	1.73E-36	3.78	9430	806	384	124
GO:0004872	receptor activity	2.48E-34	1.52E-31	3.2	9430	915	431	134
GO:0004871	signal transducer activity	3.12E-34	1.64E-31	3.08	9430	991	432	140
GO:0060089	molecular transducer activity	1.77E-33	8.10E-31	3.15	9430	931	431	134
GO:0005549	odorant binding	1.61E-14	6.57E-12	8.59	9430	61	414	23
GO:0008527	taste receptor activity	1.74E-11	6.37E-09	15.05	9430	20	376	12
GO:0033038	bitter taste receptor activity	6.39E-10	2.13E-07	15.67	9430	16	376	10
GO:0004063	aryldialkylphosphatase activity	1.06E-04	3.24E-02	9,430.00	9430	1	1	1
GO:0016765	transferase activity, transferring alkyl or aryl (other than methyl) groups	1.14E-04	3.21E-02	4.06	9430	26	1073	12
COMPONENT								
GO Term	**Description**	**P-value**	**FDR q-value**	**Enrichment**	**N**	**B**	**n**	**b**
GO:0031224	intrinsic component of membrane	2.10E-18	3.50E-15	1.83	9430	2683	362	188
GO:0016021	integral component of membrane	5.88E-18	4.91E-15	1.83	9430	2616	362	184
GO:0044425	membrane part	2.89E-13	1.61E-10	1.58	9430	3446	363	209
GO:0005886	plasma membrane	1.17E-12	4.90E-10	1.64	9430	2443	472	201
GO:0016020	membrane	4.29E-09	1.43E-06	1.32	9430	4139	473	275

Table S3. Additional gene set enrichment for top genes lost in marine lineages. This table shows categories displaying enrichment among the top 137 genes with evidence for higher rate of loss in marine lineages, corresponding to a 25% FDR. Gene set enrichment was assessed for the canonical pathways, curated pathways, and GO biological process datasets in the Molecular Signatures Database (mSigDB) (*96*) and the Mouse Genome Database (*97*) using a hypergeometric test, with all genes included in the BayesTraits analysis as the background gene set. No sets of at least three genes were found to be significant at $q < 0.05$ in the Mouse Genome Database. For lists of genes in each category driving the signatures of enrichment, see Table S5.

Canonical Pathways at mSigDB							
Category	**P-value**	**FDR q-value**	**Enrichment**	**number in background**	**fraction in background**	**number in foreground**	**fraction in foreground**
Reactome olfactory signaling pathway	1.63E-26	7.51E-25	9.3246	213	0.0475	35	0.443
KEGG olfactory transduction	1.11E-23	2.56E-22	7.7584	256	0.0571	35	0.443
Reactome signaling by GPCR	6.73E-20	1.03E-18	4.61	517	0.1153	42	0.5316
Reactome GPCR downstream signaling	9.25E-19	1.06E-17	4.7697	464	0.1035	39	0.4937
KEGG taste transduction	3.43E-03	3.16E-02	6.3052	36	0.008	4	0.0506
Curated Pathways at mSigDB							
Category	**P-value**	**FDR q-value**	**Enrichment**	**number in background**	**fraction in background**	**number in foreground**	**fraction in foreground**
Reactome olfactory signaling pathway	2.11E-29	6.95E-27	12.4883	213	0.023	35	0.2869
KEGG olfactory transduction	1.48E-26	2.44E-24	10.3906	256	0.0276	35	0.2869
Reactome signaling by GPCR	6.87E-23	7.54E-21	6.1741	517	0.0558	42	0.3443
Reactome GPCR downstream signaling	1.06E-21	8.69E-20	6.3879	464	0.05	39	0.3197
Kondo prostate cancer with H3K27Me3	2.77E-04	1.82E-02	5.4845	97	0.0105	7	0.0574
GO Biological Process at mSigDB							
Category	**P-value**	**FDR q-value**	**Enrichment**	**number in background**	**fraction in background**	**number in foreground**	**fraction in foreground**
Sensory perception of chemical stimulus	7.29E-04	3.93E-02	16.2336	19	0.0058	3	0.0938
Regulation of action potential	2.47E-03	3.93E-02	25.7031	8	0.0024	2	0.0625
Regulation of axonogenesis	2.47E-03	3.93E-02	25.7031	8	0.0024	2	0.0625
Regulation of neurogenesis	3.93E-03	3.93E-02	20.5625	10	0.003	2	0.0625
Sensory perception of taste	3.93E-03	3.93E-02	20.5625	10	0.003	2	0.0625
Neurological system process	4.11E-03	3.93E-02	3.3013	218	0.0663	7	0.2188
Response to chemical stimulus	4.17E-03	3.93E-02	3.7845	163	0.0495	6	0.1875

Table S6. Gene ontology enrichment for top genes that co-evolve with *PON1*. Gene set enrichment was performed with the 100 genes with the highest Evolutionary Rate Covariation (ERC) with *PON1* using the GOrilla online enrichment tool (*46*), with all genes included in the ERC analysis as the background set. For lists of genes in each category driving the signatures of enrichment, see Table S7.

BIOLOGICAL PROCESS								
GO Term	**Description**	**P-value**	**FDR q-value**	**Enrichment**	**N**	**B**	**n**	**b**
GO:0006629	lipid metabolic process	3.15E-10	4.68E-06	4.12	16491	1108	94	26
GO:0006631	fatty acid metabolic process	3.52E-09	2.61E-05	8.57	16491	266	94	13
GO:0044255	cellular lipid metabolic process	9.11E-08	4.51E-04	3.98	16491	881	94	20
GO:0019752	carboxylic acid metabolic process	9.65E-08	3.58E-04	4.18	16491	798	94	19
GO:0032787	monocarboxylic acid metabolic process	1.54E-07	4.58E-04	5.67	16491	433	94	14
GO:0016042	lipid catabolic process	1.78E-07	4.40E-04	7.72	16491	250	94	11
GO:0006082	organic acid metabolic process	1.80E-07	3.83E-04	3.82	16491	919	94	20
GO:0044712	single-organism catabolic process	4.23E-07	7.86E-04	4	16491	790	94	18
GO:0006635	fatty acid beta-oxidation	6.05E-07	9.97E-04	19.49	16491	54	94	6
GO:0016054	organic acid catabolic process	6.09E-07	9.03E-04	7.8	16491	225	94	10
GO:0046395	carboxylic acid catabolic process	6.09E-07	8.21E-04	7.8	16491	225	94	10
GO:0043436	oxoacid metabolic process	6.46E-07	7.99E-04	3.69	16491	903	94	19
GO:0044710	single-organism metabolic process	3.34E-06	3.81E-03	2.04	16491	3264	94	38
GO:0019395	fatty acid oxidation	3.65E-06	3.87E-03	14.42	16491	73	94	6
GO:0034440	lipid oxidation	3.96E-06	3.92E-03	14.22	16491	74	94	6
GO:0009062	fatty acid catabolic process	6.72E-06	6.23E-03	13	16491	81	94	6
GO:0044281	small molecule metabolic process	7.28E-06	6.36E-03	2.63	16491	1599	94	24
GO:0055114	oxidation-reduction process	1.87E-05	1.54E-02	3.35	16491	837	94	16
GO:0072329	monocarboxylic acid catabolic process	2.26E-05	1.77E-02	10.53	16491	100	94	6
GO:0044282	small molecule catabolic process	2.49E-05	1.85E-02	5.13	16491	342	94	10
GO:0044242	cellular lipid catabolic process	3.44E-05	2.43E-02	7.72	16491	159	94	7
FUNCTION								
GO Term	**Description**	**P-value**	**FDR q-value**	**Enrichment**	**N**	**B**	**n**	**b**
GO:0003995	acyl-CoA dehydrogenase activity	3.17E-07	1.41E-03	63.79	16491	11	94	4
GO:0016491	oxidoreductase activity	6.65E-06	1.48E-02	3.86	16491	681	94	15

Table S8. Annotated functional amino acid positions in PON1 with observed substitutions in marine or semi-aquatic lineages. Sites with substitutions in marine or semi-aquatic lineages that have been annotated from previous functional studies as important to PON1 function in one or more of the following categories: sites crucial for catalytic activity, sites within the wall of the active site, and sites that abrogate activity when experimentally substituted. Amino acid positions are based on human (hg19) sequence within alignment and may differ from those reported in the literature.

				Substitution observed in:					
Amino Acid	**Importance**	**Reference**	**Observed Substitution**	**Cetaceans**	**Sirenians**	**Pinnipeds**	**Bats**	**Semi-aquatic**	**Conservation across all other lineages**
Sites Crucial for Catalytic Activity									
E53	Binds to Catalytic Calcium	Harel et al.(*98*)	E53K	Sperm whale					100%
D169	Binds to Structural Calcium	Harel et al.(*98*)	D169G	All		Hawaiian monk seal			100%
N270	Binds to Catalytic Calcium	Harel et al. (*98*)	N270T					Sea otter	100%
Sites in the Active Site Wall (Substrate Specificity)									
I74	Active Site Wall (Phosphotriesters); Mutant Showed >20x Decrease in Phenyl Acetate and Paraoxon Catalytic Efficiency (I74A)	Harel et al.(*98*); Ben-David et al.(*99*)	I74M* I74F†			*Antarctic fur seal, Walrus	†David's Myotis bat, Black flying fox, Large flying fox		85%
H184	Active Site Wall; Mutants have Undetectable Paraoxon and Phenylacetate Activity (H184A/D/Y)	Harel et al.(*98*); Yeung et al.(*100*)	H184L					Beaver	100%
R192	Active Site Wall	Harel et al. (*98*)	R192K* R192S†		*All		*All except Big brown bat and David's myotis bat †David's myotis bat		85%
F222	Active Site Wall (Aryl Esters)	Harel et al. (*98*)	F222L					Sea otter	99%
F292	Active Site Wall (Aryl Esters); Mutant has 2% WT Phosphotriester Activity (F292A)	Harel et al.(*98*)	F292L			Antarctic fur seal, Walrus			100%
T332	Active Site Wall	Harel et al.(*98*)	T332M* T332S† T332A‡	*Yangtze River dolphin			†Big brown bat, Little brown bat	‡Sea otter	93%
Sites that Abrogate Activity When Mutated									
C42	Disulfide Velcro	Harel et al.(*98*)	C42R			Weddell seal			100%
W194	Mutants 30-50% Activity (W194A)	Josse et al. (*101*)	W194X			Hawaiian monk seal			100%
W202	Mutants 30-50% Activity (W201A)	Josse et al.(*102*)	W202C* W202L†	*All			†Big brown bat		100%
H243	Mutants <1% Activity (H243N)	Josse et al.(*102*)	H243R* H243Q†	*Minke whale	†All		*Natal long-fingered bat		100%
H246	Mutants 30-50% Activity (H245N)	Josse et al.(*102*)	H246R* H246C†		*Manatee, †Dugong		*Black flying fox, Large flying fox		100%
C284	Core Stability	Harel et al.(*98*)	C284R	All					100%
V304	Mutants No Detectable Arylesterase or Paraoxonase Activity (V304A)	Yeung et al. (*103*)	V304M					Sea otter	100%

Table S11. Abbreviations, common names, and sources for species included in *PON1* phylogenetic trees.

Abbreviation	Common_Name	Source
ailMel1	Panda	UCSC 100-way vertebrate alignment
arcGaz	Antarctic fur seal	Mapped Weddell seal exons to assembly from Dryad (doi:10.5061/dryad.599f2)
bosTau7	Cow	UCSC 100-way vertebrate alignment
brandtBat	Brandt's bat	NCBI annotated genome assembly (accession GCF_000412655.1)
calJac3	Marmoset	UCSC 100-way vertebrate alignment
camFer1	Bactrian camel	UCSC 100-way vertebrate alignment
canFam3	Dog	UCSC 100-way vertebrate alignment
capHir1	Goat	UCSC 100-way vertebrate alignment
casCan	Beaver	NCBI annotated genome assembly (accession GCA_001984765.1)
cavPor3	Guinea pig	UCSC 100-way vertebrate alignment
cerSim1	White rhinoceros	UCSC 100-way vertebrate alignment
chiLan1	Chinchilla	UCSC 100-way vertebrate alignment
chlSab1	Green monkey	UCSC 100-way vertebrate alignment
chrAsi1	Cape golden mole	UCSC 100-way vertebrate alignment
conCri1	Star-nosed mole	UCSC 100-way vertebrate alignment
criGri1	Chinese hamster	UCSC 100-way vertebrate alignment
dasNov3	Armadillo	UCSC 100-way vertebrate alignment
dugDug	Dugong	Sequencing from this study
echTel2	Tenrec	UCSC 100-way vertebrate alignment
eleEdw1	Cape elephant shrew	UCSC 100-way vertebrate alignment
enhLut	Sea otter	NCBI annotated genome assembly (accession GCA_002288905.2)
eptFus1	Big brown bat	UCSC 100-way vertebrate alignment
equCab2	Horse	UCSC 100-way vertebrate alignment
eriEur2	Hedgehog	UCSC 100-way vertebrate alignment
felCat5	Cat	UCSC 100-way vertebrate alignment
gorGor3	Gorilla	UCSC 100-way vertebrate alignment
Hawaii	Hawaiian monk seal	NCBI annotated genome assembly (accession GCA_002201575.1)
hetGla2	Naked mole-rat	UCSC 100-way vertebrate alignment
hg19	Human	UCSC 100-way vertebrate alignment
hippo	Hippopotamus	Mapped RNA-seq reads (accession SRX1164570) to dolphin
jacJac1	Lesser Egyptian jerboa	UCSC 100-way vertebrate alignment
largha	Spotted seal	Mapped RNA-seq reads (accession SRX120902) to Weddell seal
lepWed1	Weddell seal	UCSC 100-way vertebrate alignment
lipVex	Yangtze river dolphin	NCBI annotated genome assembly (accession GCA_000442215.1)
loxAfr3	African elephant	UCSC 100-way vertebrate alignment
macFas5	Crab-eating macaque	UCSC 100-way vertebrate alignment
mesAur1	Golden hamster	UCSC 100-way vertebrate alignment
micOch1	Prairie vole	UCSC 100-way vertebrate alignment
mini	Natal long-fingered bat	NCBI annotated genome assembly (accession GCF_001595765.1)
Minke	Minke whale	NCBI annotated genome assembly (accession GCA_000493695.1)
mm10	Mouse	UCSC 100-way vertebrate alignment
musFur1	Ferret	UCSC 100-way vertebrate alignment
myoDav1	David's Myotis bat	NCBI annotated genome assembly (accession GCA_000327345.1)
myoLuc2	Little brown bat	UCSC 100-way vertebrate alignment
nomLeu3	Gibbon	UCSC 100-way vertebrate alignment
ochPri3	Pika	UCSC 100-way vertebrate alignment
octDeg1	Brush-tailed rat	UCSC 100-way vertebrate alignment
odoRosDi	Walrus	UCSC 100-way vertebrate alignment
orcOrc1	Killer whale	UCSC 100-way vertebrate alignment
oryAfe1	Aardvark	UCSC 100-way vertebrate alignment
oryCun2	Rabbit	UCSC 100-way vertebrate alignment
otoGar3	Bushbaby	UCSC 100-way vertebrate alignment
oviAri3	Sheep	UCSC 100-way vertebrate alignment
panHod1	Tibetan antelope	UCSC 100-way vertebrate alignment
panTro4	Chimp	UCSC 100-way vertebrate alignment
papHam1	Baboon	UCSC 100-way vertebrate alignment
phyCat	Sperm whale	NCBI annotated genome assembly (accession GCA_000472045.1)
ponAbe2	Orangutan	UCSC 100-way vertebrate alignment
pteAle1	Black flying fox	UCSC 100-way vertebrate alignment
pteVam1	Large flying fox	UCSC 100-way vertebrate alignment
rheMac3	Rhesus	UCSC 100-way vertebrate alignment
rn5	Rat	UCSC 100-way vertebrate alignment
saiBol1	Squirrel monkey	UCSC 100-way vertebrate alignment
sorAra2	Shrew	UCSC 100-way vertebrate alignment
speTri2	Squirrel	UCSC 100-way vertebrate alignment
susScr3	Pig	UCSC 100-way vertebrate alignment
triMan1	Manatee	UCSC 100-way vertebrate alignment
tupChi1	Chinese tree shrew	UCSC 100-way vertebrate alignment
turTru2	Bottlenose dolphin	UCSC 100-way vertebrate alignment
ursMar1	Polar bear	NCBI annotated genome assembly (accession GCA_000687225.1)
vicPac2	Alpaca	UCSC 100-way vertebrate alignment

Table S12. Primers used to amplify and sequence *PON1* exons in manatee and dugong DNA samples. Except where noted, the same primers were used for both PCR and sequencing.

Primer	Sequence	Paired With	PON1 Exon	Species	PCR Annealing Temp (° C)	PCR Cycles
TML1F	ACAGCTTTCCCTTCCTTGC	TML1R	1	Both	59	34
TML1R	AGCCTGGGTCCTCCTTTCT	TML1F	1	Both	59	34
DdE2F5	CAGGTTTCTGGAACCACCTC	DdE2R5	2	Dugong	57.5	34
DdE2R5	TGAGCTACTCACTCCTCTCACAA	DdE2F5	2	Dugong	57.5	34
DdE3F2	TGAATTTCCATGAGCTTTATGTG	DdE3R2	3	Dugong	59	34
DdE3R2	CAGTTGAATGGGAAGCCACT	DdE3F2	3	Dugong	59	34
TML2F	GCCAGGAGACTTCCTGTGTG	TML2R	4	Manatee	59	34
TML2R	CCATAAAGATTAGGGCTGCAT	TML2F	4	Manatee	59	34
DdE4F	CAAGGTGAATCCGTGTGCTA	DdE4R	4	Dugong	59	34
DdE4R	GGGAAACTTGAAACCCAGAA	DdE4F	4	Dugong	59	34
DdE5F	AGACAGGGCTGACAGCTGAG	DdE5R	5	Dugong	59	34
DdE5R	TGGATTAGTCATCCTCTGGAA	DdE5F	5	Dugong	59	34
TML3_4F	GTTATGCATTTTGCTCCCAGA	TML3_4R	6	Both	59	34
TML3_4R	GGTTGATATGTTGTGGGGTTGT	TML3_4F	6	Both	59	34
DdL3_4intR	GGAATCTATTATAAAGATATCTAA	TML3_4F/TML3_4R (sequencing only)	6	Dugong	59	34
TML5_7F	TGCACTGCAAGCTCATTCTT	TML5_7R	7	Manatee	59	34
TML5_7R	CGACATCAAATGGAGGAAGG	TML5_7F	7	Manatee	59	34
DdE7F	CTCCACCGTCTCCTTTTGAA	DdE7R	7	Dugong	59	34
DdE7R	CACCCATCCCCATTAGACAA	DdE7F	7	Dugong	59	34
TML8_10F	TCCCATATCTTCCCCCTACC	TML8_10R	8	Both	59	34
TML8_10R	CCCCTAGGAACTCCTCTTGC	TML8_10F	8	Both	59	34
TML11_15F	TTGCCAGCATTTAAACACCA	TML11_15R	9	Manatee	59	34
TML11_15R	AAGGATGGGCTCACAGTTTC	TML11_15F	9	Manatee	59	34
DdE9F	GTGTGCTCACCACCTCTGTTAAA	DdE9R	9	Dugong	59	50
DdE9R	TGATCCCTCATGATGTCCAA	DdE9F	9	Dugong	59	50

Additional Data table S1 (separate file)

Presence/absence/excluded (NA) values for extant species and results of BayesTraits model comparison for gene loss (where applied) for all 9,950 genes included in the analysis and 3,904 genes excluded from analysis but whose loss rates were used for simulations.

Additional Data table S4 (separate file)

GOrilla gene ontology enrichment for top genes lost in marine lineages, with gene lists (see Table S2).

Additional Data table S5 (separate file)

Additional gene set enrichment for top genes lost in marine lineages, with gene lists (see Table S3).

Additional Data table S7 (separate file)

Gene ontology enrichment for top genes that co-evolve with *PON1*, with gene lists (see Table S6).

Additional Data table S9 (separate file)

Results from manual validation of pseudogene calls for 20 top genes, 5 genes representing known cases of pseudogenization, and 20 randomly selected genes, using manual checks of the sequences within the 100-way alignment.

Additional Data table S10 (separate file)

Results from manual validation of pseudogene calls for 20 top genes, 5 genes representing known cases of pseudogenization, and 20 randomly selected genes, using sequences from the reference genomes for all 58 species.

Additional Data table S13 (separate file)

Values from triplicate assays of hydrolysis for four PON1 substrates and alkaline phosphatase control (means plotted in Figs. 2 and S2).

Supplementary Materials for

Ancient convergent losses of *Paraoxonase 1* yield potential risks for modern marine mammals

Wynn K. Meyer, Jerrica Jamison, Rebecca Richter, Stacy E. Woods, Raghavendran Partha, Amanda Kowalczyk, Charles Kronk, Maria Chikina, Robert K. Bonde, Daniel E. Crocker, Joseph Gaspard, Janet M. Lanyon, Judit Marsillach, Clement E. Furlong, and Nathan L. Clark

correspondence to: nclark@pitt.edu

This PDF file includes:

Materials and Methods
Figs. S1 to S6
Tables S2, S3, S6, S8, S10, and S11
Captions for Additional Data tables S1, S4, S5, S7, S9, and S12

Other Supplementary Materials for this manuscript includes the following:

Additional data tables S1, S4, S5, S7, S9, S10, and S13 are available as Excel files.

Materials and Methods

Scoring gene orthologs as functional or pseudogenes across eutherian mammals in the 100-way alignment

Our first goal was to assess whether the annotated sequence for each species in each gene's publicly available alignment represented a functional gene or an unprocessed pseudogene. We used the following pipeline to identify genes that displayed strong evidence of having lost function in any species, using the sequence data from the hg19 UCSC 100-way alignment (http://genome.ucsc.edu/). We scanned amino acid sequences for stop characters (Z) and nucleotide sequences for frameshifts, using the following filters to exclude putative lesions that would be unlikely to disrupt function or could be caused by issues of data quality:

1. For first and last exons, which are known to be highly variable and sometimes alternatively spliced, we excluded the whole exon from our scans if it made up no more than 10% of the entire gene sequence, or the terminal 60 bp (20 aa) otherwise.
2. We additionally excluded from our scans any exons containing more than 25% gap characters, as these might represent incorrectly identified orthologs or low quality genomic data resulting in poor quality alignments.
3. We excluded any pairs of frameshifts that were within 15 bp of each other, since these could also represent issues with alignment and/or data quality.
4. As a further check against including erroneous frameshifts caused by gaps in reference sequence data, we excluded any frameshifts greater than 8 bp in length.
5. To avoid erroneous frameshifts caused by errors in aligning exon boundaries, we excluded the first and last three bp of each exon from our frameshift scan.

For the sequence of each eutherian mammal species in the alignment for each gene, we estimated the proportion of gaps within the sequence included after the above filters were applied. When we excluded an entire internal exon, we counted all of its sequence as gap characters. Where the proportion of gaps differed between nucleotide and amino acid estimates, we chose the larger value. We then set a threshold of the proportion of gap characters above which to exclude a call at a gene for a given species in order to limit the probability of erroneously calling a pseudogene functional (i.e., the rate of false negative pseudogene calls). Specifically, we sought to minimize the following:

$$P_{false\_negative} = ((N_{pseudogenes} / \sum P_{non\text{-}missing}) \cdot \sum P_{missing}) / N_{genes},$$

where $N_{pseudogenes}$ represents the total number of pseudogenes in the included set, $\sum P_{non\text{-}missing}$ represents the sum of the proportions of non-gap characters for all genes in the included set, $\sum P_{missing}$ represents the sum of the proportions of gap characters for all genes in the included set, and N_{genes} represents the total number of genes included at this threshold across all species. In effect, this estimates the rate at which pseudogenes occur per total amount of non-missing data, and then estimates how many pseudogenes would be unobserved based on the total amount of missing data within the included gene set. We chose a threshold that we estimated would result in a false negative pseudogene once every 10 genes (i.e., at most one erroneously called functional gene in any of the 58 species every 10 genes). Any gene sequence exceeding this threshold (16%) for

proportion of gap characters for a given species was excluded (not called) for that species.

Several alignments in the dataset had multiple University of California 'known gene' identification numbers (UCIDs) corresponding to a single gene symbol. When we found such instances, we identified the UCID that had the smallest total proportion of gap characters (missing data) across all species and used only the data from the alignment for that UCID for the corresponding gene symbol; if multiple UCIDs had equivalent missing data, we excluded that gene from analysis (this situation was encountered for eight genes).

Manual validation and estimation of error rates

Our pipeline for identifying lesions and calling candidate pseudogenes may be subject to various sources of error. We performed several checks to rule out errors in orthology and to estimate error rates stemming from issues with our automated method of pseudogene calling and issues in the 100-way vertebrate alignment, and the next three sections describe these checks. We performed all of these assessments for a 'test set' comprised of the following: the top 20 genes from our analysis of marine-dependent loss, 20 other randomly selected genes, and five well-studied cases of gene non-functionalization, three of which passed filters for inclusion in our analysis.

Manual validation of orthology

To ensure that lesions identified in the test set were not the result of the mis-identification of orthologs for certain species, we first checked for errors in orthology by building an 85% consensus parsimony tree from the coding sequence alignment and determining whether any species or set of species not excluded for missing data was positioned as an outgroup to the remainder of the phylogeny; we built trees using the PHYLIP v3.696 dnapars algorithm (*34*), as implemented in the SeaView (version 4.6.1) software (*35*). In the single case in which species with predicted pseudogenes were positioned as outgroups, we verified that this was due to long branch attraction by determining that this protein was the best match in the human genome for the amino acid sequences of the species with predicted pseudogenes using blastp (*36*). We also validated that our predicted pseudogenes represented losses of function without loss of synteny (i.e., unprocessed, or unitary, pseudogenes) by comparing the genes within the region surrounding the predicted pseudogene in the genome of the species for which the gene was predicted to be a pseudogene with the genes surrounding this gene's ortholog in the human genome, using the UCSC Genome Browser (*37*).

Manual validation of automated method for calling lesions from 100-way alignment

To assess the reliability of our criteria to call lesions from the 100-way alignment, we manually validated predicted pseudogene calls and predicted functional ("intact") gene calls against the original sequence in the 100-way alignment for all genes in the test set (Table S9). We translated coding sequences to amino acids and manually checked for the presence of lesions in predicted pseudogenes and predicted functional genes that were not less than ten amino acids from the end of the gene. Using these manual checks, we identified 22 cases where a potential pseudogene may have been mis-identified as a functional gene (false negatives) and five cases where a predicted pseudogene may have

been falsely called (false positives) across all non-excluded sequences for these 42 genes, leading to method-based error rates of 7.14% and 0.26% for false negatives and false positives, respectively. All false positives were in terrestrial species; one case resulted from two independent frameshifts that, when combined, brought the sequence back into frame, and another resulted from a premature stop codon encoded within the penultimate exon that fell only four amino acids from the sequence end (Table S9). It is important to note that these errors represent cases in which the filters in the automated method may lead to inaccurate functional/pseudogene classification of sequences within the 100-way alignment, assuming that the alignment itself contains no errors. We separately estimated errors in the alignment (see next section).

Estimating error rates in pseudogene calls made from the 100-way alignment

To assess the reliability of our predicted pseudogene calls and to ensure that our strongest results were not driven by erroneous pseudogene calls within marine lineages, we further determined the rate of errors in calling predicted pseudogenes using sequences from reference genomes for all genes in the test set (Table S10).

We initially called potential pseudogenes for each species using the aligned coding sequences available in the '100-way alignment' from the UCSC Genome Browser (based on human genome version 'hg19') (*37*). For the 45 test set genes, we manually examined each lesion (stop codon or frameshift-causing deletion) in the original source genome that was used for the 100-way alignment. This validation allowed us to check for mis-called lesions resulting from the alignment process. There are other potential sources of error, including the genome assembly itself. However, newer genome assemblies were only available for nine of the 58 species, five of which were primates, making it challenging to assess errors due to genome assembly globally.

We obtained the source genome assemblies for all 57 non-human species from either UCSC or NCBI ('download_genomes.sh' in folder 'Estimating_error_rates' within our github repository). We made BLAST nucleotide databases for each species' genome (*36*). For each lesion-containing exon, we obtained the highest-scoring BLASTn hit from that species' genome ('match_lesions_with_genomes.sh') and manually inspected each apparent lesion using its coordinates and flanking sequence. Validation outcomes were tabulated for each lesion and for the resulting "potential pseudogene" call for each gene (Table S10). Those counts were used to calculate 3 statistics:

1) error rate per pseudogene call, *i.e.*, false positive pseudogenes / all pseudogene calls
2) false positive rate, *i.e.*, false positive pseudogenes / (false positive pseudogenes + true functional genes)
3) per-lesion error rate, *i.e.*, false positive lesions / all lesion calls.

We report all rates and calculations in Table S10. Generally, error rates show that these data are useful in the capacity of a screen to identify potential subjects of convergent loss of gene function. However, the error rates are suitably high (false positive rate 3.4% and per-lesion error rate 16%) that all lesion calls should be validated by independent methods, such as Sanger sequencing, as we did for *PON1*. In a single

case, errors were seen to affect marine species in one of the top ten genes in our analysis; that gene, *SLC39A9*, was excluded from Table 1.

Assessing the potential for biased errors between marine and terrestrial species

To determine whether our scan for marine convergent functional loss might be biased due to a high error rate within our automated potential pseudogene calls specifically for marine species, we compared rates of pseudogene calls at presumed functional genes between marine and terrestrial species. For this comparison, we evaluated the rate at which a gene set previously identified as being highly conserved across eukaryotes contained called pseudogenes, under the assumption that these highly conserved genes have a low probability of losing function in the mammalian species within our dataset. We obtained RefSeq protein accession numbers for the 248 core eukaryotic genes that tend to be present as single-copy genes across six diverse high-quality eukaryotic genomes (*38*), frequently used for assessing completeness of draft genomes using the CEGMA protocol (*39*). We translated these to gene symbols using the bioDBnet tool (*40*), and we further manually corrected gene symbols that differed between RefSeq and UCSC by identifying genes that uniquely overlapped across 100% of their length in the UCSC hg19 browser (http://genome.ucsc.edu/) (*37*, *41*). We then determined, for each species, the proportion of these CEGMA genes that were called as pseudogenes using our automated method, out of all CEGMA genes that were not filtered for missing data. The estimated error rates based on CEGMA pseudogenes for marine species fall within the range of those for terrestrial species, overlapping with the distribution of such error rates for terrestrial species that have similar genetic distances to reference sequence hg19 (Fig. S4).

Identifying signatures of convergent loss of gene function in marine mammals

The predicted pseudogene status results formed a gene-by-species matrix of gene presence (functional) / absence (pseudogene) / excluded (not assigned). We excluded any genes that had gene calls excluded for at least one third, or 19, of the 58 total species, reasoning that these may represent cases where data quality was poor or orthologs were incorrectly identified across multiple species. We then selected the set of genes that were designated predicted pseudogenes in at least two species and at most 29 species, since that range would be best powered to identify marine-specific loss. We ran two nested likelihood models in BayesTraits (*42*) version 3 using the remaining 9,950 gene vectors and a vector indicating which species are ‘marine’ and ‘terrestrial’. The independent model contained two parameters – a gene loss rate (the rate at which a functional gene becomes a predicted pseudogene) and a rate for transition from terrestrial to marine status. Because our study was focused on gene loss, gene gain was not allowed; its rate was constrained to zero. Similarly, the rate for transition from a marine to terrestrial state was constrained to zero, since this transition is not observed in the placental mammalian phylogeny. This independent model contained no relationship between gene loss and marine/terrestrial state, and so it served as the null hypothesis. The dependent model, on the other hand, added another free parameter by dividing the gene loss rate into two parameters – loss rate on terrestrial and marine branches, separately. We compared these two nested models using a likelihood ratio test (LRT). Since we were interested in the evidence for higher loss on marine branches, we reversed the sign of the LRT statistic for

all genes inferred to have a higher loss rate on terrestrial branches in the independent model.

The distribution of our modified LRT statistic deviates from the chi-square distribution with 1 degree of freedom, due to the effects of sample size limitations and restricted parameter ranges, as well as to the reversal of sign for genes with higher terrestrial loss rates. To estimate empirical *P*-values for each gene based on the distribution of this modified statistic under the null, we performed simulations of gene loss across the mammalian phylogeny. To recapitulate the pattern of loss for each gene, we set branch lengths to the genome-wide average amino acid distances, multiplied by the gene's inferred loss rate from the independent model of BayesTraits. We stratified the number of simulated datasets per gene based on each gene's likelihood ratio test *P*-value assuming a chi-square distribution – 10 million simulations for the genes with $P < 10^{-5}$ (genes ranked 1-3), 1 million simulations for genes with $P < 10^{-4}$ (genes ranked 4-10), 100,000 simulations for genes with $P < 10^{-3}$ (genes ranked 10-195), and 10,000 simulations for the rest. We generated simulated datasets using the 'sim.char' function in the R package 'geiger,'(*43*); our simulation-based *P*-value, reported as "Empirical *P*-value" in Tables 1 and S1, represents the proportion of simulations with a higher modified LRT statistic than that observed for the gene of interest.

In order to estimate empirical study-wide false discovery rates (FDR), we simulated datasets matching the evolution of the 13,853 genes in our dataset with at least one pseudogene and at least one functional gene among at least 39 species with non-excluded gene status. We simulated 10,000 datasets per gene using the methods described above. We subsequently filtered the simulated datasets to include only simulated genes with a minimum of 2 pseudogenes and a maximum of 29, to create a null dataset of simulated genes subject to the same filters as the real dataset. We then used the distribution of test statistics from simulated genes to estimate the FDR in an approach similar to empirical permutation-based FDR calculations. Studies that perform permutation-based FDR calculations commonly use a modification of the Benjamini-Hochberg procedure wherein they compare observed test statistics with empirically defined null distributions obtained from repeated permutations of the data and labels, in place of the procedure's traditional comparison of observed *P*-values to a null distribution based on uniform quantiles (*44*, *45*). However, in our case permuting tip labels would frequently change the branch lengths on which functional losses could occur and modify the well-supported relationships among foreground species. In our analysis, we thus use the same modified Benjamini-Hochberg procedure to compare the observed modified LRT statistic distributions to the distribution of modified LRT statistics for simulated genes (the empirical null distribution), in place of the distribution of a permutation-based test statistic. This approach results in FDR calculations based on test statistic distributions from a null dataset more closely matching the true dataset, commonly preferred in genomic data analysis.

While our simulation approach enables the generation of an empirical null distribution based on multiple datasets preserving phylogenetic relationships among foreground species, we also compared our results to those from a single permutation wherein we selected a set of foreground lineages whose branch lengths and relationships to other foreground species were matched to those of the marine species, but which are not known for convergence in any phenotype or environment. Specifically, those species

were the aardvark, alpaca, Bactrian camel, little brown bat, and David's Myotis bat (*5*). We applied our genome-wide scan for convergence to this matched foreground set using the likelihood-based methods implemented in BayesTraits, as previously described.

Considering only genes showing higher inferred marine loss rates, we see some evidence for enrichment of genes in the real dataset showing higher raw LRTs compared to the null distribution obtained from simulations or the distribution for a single matched foreground set (Fig. S5A and C). In strong contrast, genes in the real dataset with higher inferred terrestrial loss rates do not show comparable enrichment for higher LRTs relative to the simulated null distribution or matched foreground distribution (Fig. S5B and D). This suggests that there is empirical evidence for enrichment of genes showing marine-biased pseudogenization.

Functional enrichment analyses

To generate a ranked list for enrichment tests, we ranked genes in descending order by LRT statistic; we reversed the sign of the LRT statistic for genes with higher inferred loss rates on terrestrial branches than on marine branches, since in these cases large LRT would represent evidence against marine-biased loss. This ranked list was tested for functional enrichment using the Gene Ontology annotations available through the GOrilla server (*46*), and the set of the top 137 genes (representing a false discovery rate, or FDR, of 25%; see previous section) was tested for functional enrichment using the MSigDB canonical, curated, and biological process gene ontology databases (*47*) and the MGI mammalian phenotypes database (*48*), with mammalian phenotype sets built by compiling lists of gene symbols associated with each phenotype and including all genes for a given phenotype as part of the set associated with that phenotype's ancestors in the ontology (from https://bioportal.bioontology.org/ontologies/MP, last accessed June 6, 2016). To test for functional enrichment using these datasets, we performed a hypergeometric test using the set of 9,950 genes that passed inclusion filters (see above) as our background gene set. We corrected for multiple testing in these analyses using the Benjamini-Hochberg procedure (*45*).

Phylogenetic tree for analyses

For all analyses, we used the same tree topology, based on that inferred by Meredith et al. (*17*). In several cases where this tree differed from that inferred by Bininda-Emonds et al. (*49*), we chose a consensus topology based on studies that inferred the local phylogeny using focused sampling of species within the clade of interest. Specifically, we set the star-nosed mole as an outgroup to the hedgehog and shrew (*50*, *51*); the cow as an outgroup to the Tibetan antelope, sheep, and goat (*52*, *53*); and the ursids as an outgroup to mustelids and pinnipeds (*54*, *55*). For inferring date of PON1 functional loss in pinnipeds, we estimated d_N/d_S separately using mustelids and ursids as the sister clade, to demonstrate robustness of the dating to assumptions about the local topology. The full tree topology, incorporating new species added specifically for *PON1* analyses (see next section) is provided below ("Phylogenetic trees used for evolutionary inferences", tree #1). For analyses that required branch lengths (including tests for marine convergent functional loss in BayesTraits), we estimated branch lengths on this consensus species tree topology using the average branch lengths from a large set of trees as follows. We chose a set of genes in which each gene had a sequence from each of the

58 species. For each gene, we estimated branch lengths using *codeml* on the fixed tree topology with an amino acid model (*56*). We scaled the resulting trees to unit vector length, and the average of each scaled branch length across all genes present in all species became the representative branch length in the master tree.

Assessing robustness of results to variation in the phylogenetic tree

Given that the consensus tree topology used in our analyses (see previous section) may not accurately represent the evolutionary history of all genes due to incomplete lineage sorting or post-divergence gene flow, we assessed the robustness of our results to variation in the tree by performing inferences of convergence for our 20 top genes using 14 alternate trees. For these analyses, we used trees with branch lengths inferred from the concatenated sequence alignment of 10 genes, assuming the following tree topologies: two previously published mammalian supertrees inferred from multiple nuclear loci (*17*, *49*), the tree provided by UCSC and used as a guide for the 100-way alignment (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz100way/), and our original consensus tree topology (with branch lengths re-estimated using the same gene set as for the alternate topologies). We also used trees generated from single-gene alignments for a different set of 10 genes. These different trees represent a range of realistic relationships among the included species (see "Phylogenetic trees used for evolutionary inferences" below).

We obtained the two previously published trees and the UCSC tree directly from their respective sources and pruned them as necessary to create subtree topologies containing only the 58 species of interest. We randomly resolved polytomies within the Bininda-Emonds (*49*) tree using the multi2di function from the ape package in R (*57*). To provide sequence input for estimating branch lengths, we concatenated the nucleotide sequence alignments for ten randomly selected genes that were called as functional in all 58 species and met the following two restrictions: sequence alignments were required to be 500 nucleotides long and contain between 30% and 70% variable sites (BECN1, CLDN4, DNAJC5B, FBXO30, GINS4, GPR22, LPAR6, SARIA, SMPD2, and SNX16). For each of the four pre-determined topologies (two published, UCSC, and our consensus topology), we estimated branch lengths from this concatenated alignment, using *codeml* (*56*). We used a codon model with equilibrium codon frequencies calculated from the alignment, a fixed d_N/d_S ratio across branches, a neutral selection model, and estimated kappa and omega values.

To generate additional trees from single-gene alignments, we randomly selected ten different genes that were called as functional in all 58 species and whose alignments were at least 1000 nucleotides long (FICD, FUT9, GPR22, HMGCS1, LRFN5, MSL1, NUDT12, PTPRA, SGMS2, and TTC5). We estimated maximum likelihood trees from the nucleotide alignments for each of these genes using PhyML, as implemented in seaview4 (*35*, *58*). We used default PhyML settings: a GTR model, aLRT branch support, empirical nucleotide equilibrium frequencies, no invariable sites, optimized across site rate variation with four rate categories, NNI tree searching operation, and optimized tree topology with a BioNJ starting tree and five random starts.

We assessed evidence for convergent loss of function in marine species as previously described for the 20 genes with the highest LRT statistics in our original analysis, using each of the 14 trees estimated above. The resulting LRT statistics for these

genes are largely consistent across all trees (Fig. S6). These results suggest that our conclusions are robust to variations in tree toplogy among gene trees, and therefore to uncertainty in our chosen consensus tree due to incomplete lineage sorting or post-divergence gene flow.

Adding *PON1* sequences for mammalian species not in the 100-way alignment

In order to provide a more complete representation of *PON1* sequences for marine and semi-aquatic species, as well as species within other clades of interest based on evolutionary rates, we first obtained the following species' publicly available *PON1* coding sequences and added them to the mammalian subset of the 100-way alignment: Brandt's bat (*Myotis brandtii*)(*59*), Canadian beaver (*Castor canadensis*)(*60*), Hawaiian monk seal (*Neomonachus schauinslandi*)(*61*), minke whale (*Balaenoptera acutorostrata*)(*62*), Natal long-fingered bat (*Miniopterus natalensis*)(*63*), polar bear (*Ursus maritimus*)(*64*), sea otter (*Enhydra lutris*)(*65*), sperm whale (*Physeter macrocephalus*) and Yangtze River dolphin (*Lipotes vexillifer*)(*66*). To address issues with the annotation of exon boundaries in sea otter, we downloaded the full gene sequence, including introns, and annotated exons manually, using the results from a discontiguous megablast of the ferret coding sequence as a guide (*67*). We obtained the predicted sequence of *PON1* for Antarctic fur seal (*Arctocephalus gazella*) by downloading genomic scaffolds (*68*) and identifying sequences orthologous to Weddell seal *PON1* coding sequence using BLAT v36x1 (*69*). We obtained publicly available RNA sequencing read data from liver for hippopotamus (*Hippopotamus amphibius*) and spotted seal (*Phoca largha*) (*70*). We derived the predicted *PON1* sequence for these species by mapping the sequencing reads to *PON1*, *PON2* and *PON3* coding sequences of the most closely related species in the 100-way alignment (dolphin and Weddell seal for the hippopotamus and spotted seal, respectively) simultaneously using NextGenMap (*71*) and retaining the consensus sequence for reads mapped to *PON1* using SAMtools (*72*). We additionally experimentally determined the sequence of the dugong (see below). We added these species to their inferred locations in the mammalian phylogenetic tree using published phylogenetic inferences of the topology of the relevant clades (*17*, *54*, *73*–*76*). Table S11 lists accession numbers for all datasets used to add new species.

Estimating branch-specific d_N/d_S for *PON1* and the timing of its loss along marine lineages

We estimated branch-specific d_N/d_S, or omega (ω), across the expanded mammalian phylogeny, excluding non-eutherian mammals, using the *codeml* program in the PAML software (*77*). We used the branch model with freely varying omega (model = 1, NsSites = 0) to infer d_N/d_S across all branches separately. After estimating parameters for each branch independently, we subsequently constrained some branches to have equal rates in order to more accurately estimate rates for short branches and ran PAML with model = 2; we additionally pruned the tree for some analyses to reduce run time (see next section). To estimate the time at which *PON1*'s evolutionary rate shifted from a background functional rate (ω_f) to the rate for pseudogenic lineages (ω_p) in each marine lineage, indicative of the time of loss, we applied equation (5) from Meredith et al. (*78*). To determine the value for ω_p, we tested whether the d_N/d_S ratio was significantly different from 1, the theoretical expectation for a pseudogene, on branches fully subsequent to the

inferred first appearance of genetic lesions. This value was estimated to be 0.98, which was not significantly different from 1 ($P = 0.93$); we therefore set $\omega_p = 1$. To account for differences in d_S between functional and pseudogenic lineages, we assumed that the ratio of functional to pseudogenic d_S was 0.7, based on the finding by Bustamante, Nielsen, and Hartl (*79*) that processed pseudogenes within regions of similar GC content to their parent genes accrued synonymous substitutions at a rate 70% of that of the parent genes. We estimated the background functional rate (ω_f) from the closest evolutionary lineages to the focal clade: for cetaceans, we included all bovids and the bovid ancestral branch; for sirenians, we included all other Afrotherian lineages except sirenians and the Afrotherian ancestral branch; for Phocidae, we included all Carnivora, excluding the sea otter and all pinniped lineages except for the branch ancestral to all pinnipeds (see next section).

To determine whether d_N/d_S was significantly different from 1 in some lineages and to estimate its confidence interval, we constrained d_N/d_S in the focal lineage using fix_omega = 1. We derived *P*-values for the hypothesis that $d_N/d_S = 1$ using a likelihood ratio test, comparing the likelihood with omega fixed at one to its maximum likelihood value. We derived 95% confidence intervals by running codeml for various fixed values of d_N/d_S and estimating the value at which the likelihood ratio test *P*-value would drop below 0.05.

Phylogenetic trees used for evolutionary inferences

Table S11 relates the abbreviations used in these trees to common names and data sources for all species. We used the following tree and its pruned subsets as our input to BayesTraits for the main analyses reported in the paper:

((((((((((hg19:0.005957477577,panTro4:0.006721826689):0.001382639829,gorGor3:0.007765177171):0.005572327638,ponAbe2:0.0164503644):0.002187630666,nomLeu3:0.01770384793):0.007043113559,(chlSab1:0.007693724903,((macFas5:0.001292320552,rheMac3:0.00713015786):0.002951690224,papHam1:0.005199240711):0.002049749893):0.01566263562):0.0135408115,(calJac3:0.02474184521,saiBol1:0.02096868307):0.02784675729):0.04299750653,otoGar3:0.108738222):0.01379370868,((((((cavPor3:0.09048639907,(chiLan1:0.05332953299,octDeg1:0.08476954109):0.01287861561):0.02118937782,hetGla2:0.08588673524):0.07432515556,speTri2:0.08896424642):0.006291577528,((((criGri1:0.04084640027,mesAur1:0.04456203524):0.02314125062,micOch1:0.06932402649):0.01947113467,(mm10:0.05273642272,rn5:0.05576007402):0.04435347588):0.08380065137,jacJac1:0.1438649666):0.04270536633):0.01663675397,(ochPri3:0.1256544445,oryCun2:0.07131655591):0.06535533418):0.009050428462,tupChi1:0.1191189141):0.003894252213):0.01425600689,(((((ailMel1:0.03854019703,((lepWed1:0.02002160645,odoRosDi:0.02064385875):0.01734764946,musFur1:0.04613997497):0.002879093616):0.009005888384,canFam3:0.05339127565):0.01185166857,felCat5:0.05020331605):0.03285617057,((((((bosTau7:0.02168740723,((capHir1:0.01157093136,oviAri3:0.01246322594):0.0049716126,panHod1:0.01522587482):0.01465511149):0.0662523666,(orcOrc1:0.006371664911,turTru2:0.01086552617):0.06014682602):0.01216198069,susScr3:0.0796745271):0.006785823323,(camFer1:0.01240650215,vicPac2:0.01096629635):0.06374554586):0.02551888691,(cerSim1:0.04977357056,equCab2:0.061454379):0.02510111297):0.00331214686,((eptFus1:0.03248546656,(myoDav1:0.02344332842,myoLuc2:0.01567729315):0.02193849809):0.09455328094,(pteAle1:0.005833353548,pteVam1

:0.01611220178):0.07567400302):0.02385546003):0.002057771224):0.004845253848,(conCri1:0.1239823369,(eriEur2:0.1696142244,sorAra2:0.1934205791):0.02079474546):0.0235875333):0.01477733374):0.01915406518,((((chrAsi1:0.1017903453,echTel2:0.1749615473):0.01592632003,eleEdw1:0.1516860647):0.006610995228,oryAfe1:0.08326528894):0.008243787904,(loxAfr3:0.06812658238,triMan1:0.06198982615):0.02249945290):0.03384011363,dasNov3:0.1342602666);

We used the following trees to constrain branches for various purposes in PAML:

#1 For generating Figs. 1 and S1:

(((conCri1 #1,(eriEur2 #2,sorAra2 #3) #4) #4,((felCat5 #6,(canFam3 #7,((ailMel1 #8,ursMar1 #9) #10,((musFur1 #11,enhLut #12) #13,(((lepWed1 #14,Hawaii #14) #15,largha #16) #17,(odoRosDi #18,arcGaz #19) #20) #21) #22) #23) #24) #25,(((pteVam1 #26,pteAle1 #27) #28,(mini #29,(eptFus1 #30,(myoDav1 #31,(brandtBat #32,myoLuc2 #33) #34) #34) #36) #37) #38,((cerSim1 #39,equCab2 #40) #41,((vicPac2 #42,camFer1 #43) #44,(susScr3 #45,

((((((turTru2 #46,orcOrc1 #47) #48,lipVex #49) #50,phyCat #51) #52,Minke #53) #54,hippo #55) #56,(bosTau7 #57,(panHod1 #58,(oviAri3 #59,capHir1 #60) #60) #62) #63) #64) #65) #66) #67) #69) #69) #69,((tupChi1 #71,((ochPri3 #72,oryCun2 #73) #74,((casCan #75,(jacJac1 #76,((rn5 #77,mm10 #78) #79,(micOch1 #80,(mesAur1 #81,criGri1 #82) #83) #84) #85) #86) #87,(speTri2 #88,(hetGla2 #89,(cavPor3 #90,(chiLan1 #91,octDeg1 #92) #93) #94) #95) #96) #96) #98) #99,(otoGar3 #100,((saiBol1 #101,calJac3 #102) #103,((chlSab1 #104,(papHam1 #105,(rheMac3 #106,macFas5 #107) #107) #109) #110,(nomLeu3 #111,(ponAbe2 #112,(gorGor3 #113,(hg19 #114,panTro4 #115) #116) #116) #118) #119) #119) #121) #122) #124,(((oryAfe1 #125,((echTel2 #126,chrAsi1 #0) #117,eleEdw1 #120) #108) #108,((triMan1 #97,dugDug #123) #61,loxAfr3 #35) #68) #108,dasNov3 #70) #5);

#2 For estimating significance of cetacean ancestral branch d_N/d_S difference from 1 and confidence interval:

((((conCri1,(eriEur2,sorAra2))$1,((felCat5,(canFam3,((ailMel1,ursMar1),((musFur1,enhLut),(((lepWed1,Hawaii),largha),(odoRosDi,arcGaz))))))$2,(((pteVam1,pteAle1),(mini,(eptFus1,(myoDav1,(brandtBat,myoLuc2)))))$3,((cerSim1#4,equCab2#4)#4,((vicPac2#4,camFer1#4)#4,(susScr3#4,((((((turTru2#8,orcOrc1#8)#8,lipVex#8)#8,phyCat#8)#8,Minke#8)#12,hippo#11)#4,(bosTau7#4,(panHod1#4,(oviAri3#4,capHir1#4)#4)#4)#4)#4)#4)#4)#2)#2)#2,((tupChi1,((ochPri3,oryCun2),((casCan,(jacJac1,((rn5,mm10),(micOch1,(mesAur1,criGri1))))),(speTri2,(hetGla2,(cavPor3,(chiLan1,octDeg1)))))))$5,(otoGar3,((saiBol1,calJac3),((chlSab1,(papHam1,(rheMac3,macFas5))),(nomLeu3,(ponAbe2,(gorGor3,(hg19,panTro4)))))))))$6,(((oryAfe1#7,((echTel2#7,chrAsi1#7)#7,eleEdw1#7)#7)#7,((triMan1,dugDug)$10,loxAfr3#7)#7)#7,dasNov3#7)#7);

#3 For estimating significance of siren ancestral branch d_N/d_S difference from 1 and confidence interval:

((((conCri1,(eriEur2,sorAra2))$1,((felCat5,(canFam3,((ailMel1,ursMar1),((musFur1,enhLut),(((lepWed1,Hawaii),largha),(odoRosDi,arcGaz))))))$2,(((pteVam1,pteAle1),(mini,(eptFus1,(myoDav1,(brandtBat,myoLuc2)))))$3,((cerSim1#4,equCab2#4)#4,((vicPac2#4,camFer1#4)#4,(susScr3#4,((((((turTru2,orcOrc1),lipVex),phyCat),Minke)$8,hippo#4)

#4,(bosTau7#4,(panHod1#4,(oviAri3#4,capHir1#4)#4)#4)#4)#4)#4)#4)#4)#4)#2)#2,((tupChi1,((ochPri3,oryCun2),((casCan,(jacJac1,((rn5,mm10),(micOch1,(mesAur1,criGri1)))),(speTri2,(hetGla2,(cavPor3,(chiLan1,octDeg1)))))))$5,(otoGar3,((saiBol1,calJac3),((chlSab1,(papHam1,(rheMac3,macFas5))),(nomLeu3,(ponAbe2,(gorGor3,(hg19,panTro4)))))))))$6,(((oryAfe1#7,((echTel2#7,chrAsi1#7)#7,eleEdw1#7)#7)#7,((triMan1#9,dugDug#9)#10,loxAfr3#7)#7)#7,dasNov3#7)#7);

#4 For estimating date of loss in ancestor of Weddell seal and confidence interval:

((((conCri1,(eriEur2,sorAra2))$1,((felCat5#2,(canFam3#2,((ailMel1#2,ursMar1#2)#2,((musFur1#2,enhLut#10)#2,(((lepWed1#11,Hawaii#9)#11,largha#9)#11,(odoRosDi#9,arcGaz#9)#9)#11)#2)#2)#2)#2,(((pteVam1,pteAle1),(mini,(eptFus1,(myoDav1,(brandtBat,myoLuc2)))))$3,((cerSim1#4,equCab2#4)#4,((vicPac2#4,camFer1#4)#4,(susScr3#4,((((turTru2,orcOrc1),lipVex),phyCat),Minke)$8,hippo#4)#4,(bosTau7#4,(panHod1#4,(oviAri3#4,capHir1#4)#4)#4)#4)#4)#4)#4)#4)#4)#2)#2,((tupChi1,((ochPri3,oryCun2),((casCan,(jacJac1,((rn5,mm10),(micOch1,(mesAur1,criGri1))))),(speTri2,(hetGla2,(cavPor3,(chiLan1,octDeg1)))))))$5,(otoGar3,((saiBol1,calJac3),((chlSab1,(papHam1,(rheMac3,macFas5))),(nomLeu3,(ponAbe2,(gorGor3,(hg19,panTro4)))))))))$6,(((oryAfe1,((echTel2,chrAsi1),eleEdw1)),((triMan1,dugDug),loxAfr3)),dasNov3)$7);

#5 For estimating date of loss in ancestor of Hawaiian monk seal and confidence interval:

((((conCri1,(eriEur2,sorAra2))$1,((felCat5#2,(canFam3#2,((ailMel1#2,ursMar1#2)#2,((musFur1#2,enhLut#10)#2,(((lepWed1#9,Hawaii#12)#12,largha#9)#12,(odoRosDi#9,arcGaz#9)#9)#12)#2)#2)#2)#2,(((pteVam1,pteAle1),(mini,(eptFus1,(myoDav1,(brandtBat,myoLuc2)))))$3,((cerSim1#4,equCab2#4)#4,((vicPac2#4,camFer1#4)#4,(susScr3#4,((((turTru2,orcOrc1),lipVex),phyCat),Minke)$8,hippo#4)#4,(bosTau7#4,(panHod1#4,(oviAri3#4,capHir1#4)#4)#4)#4)#4)#4)#4)#4)#4)#2)#2,((tupChi1,((ochPri3,oryCun2),((casCan,(jacJac1,((rn5,mm10),(micOch1,(mesAur1,criGri1))))),(speTri2,(hetGla2,(cavPor3,(chiLan1,octDeg1)))))))$5,(otoGar3,((saiBol1,calJac3),((chlSab1,(papHam1,(rheMac3,macFas5))),(nomLeu3,(ponAbe2,(gorGor3,(hg19,panTro4)))))))))$6,(((oryAfe1#7,((echTel2#7,chrAsi1#7)#7,eleEdw1#7)#7)#7,((triMan1,dugDug)$11,loxAfr3#7)#7)#7,dasNov3#7)#7);

#6 For assessing robustness of Weddell seal estimates to local topology, using ursids as the outgroup to pinnipeds:

((((conCri1,(eriEur2,sorAra2))$1,((felCat5#2,(canFam3#2,((musFur1#2,enhLut#10)#2,((ailMel1#2,ursMar1#2)#2,(((lepWed1#11,Hawaii#9)#11,largha#9)#11,(odoRosDi#9,arcGaz#9)#9)#11)#2)#2)#2)#2,(((pteVam1,pteAle1),(mini,(eptFus1,(myoDav1,(brandtBat,myoLuc2)))))$3,((cerSim1#4,equCab2#4)#4,((vicPac2#4,camFer1#4)#4,(susScr3#4,((((turTru2,orcOrc1),lipVex),phyCat),Minke)$8,hippo#4)#4,(bosTau7#4,(panHod1#4,(oviAri3#4,capHir1#4)#4)#4)#4)#4)#4)#4)#4)#4)#2)#2,((tupChi1,((ochPri3,oryCun2),((casCan,(jacJac1,((rn5,mm10),(micOch1,(mesAur1,criGri1))))),(speTri2,(hetGla2,(cavPor3,(chiLan1,octDeg1)))))))$5,(otoGar3,((saiBol1,calJac3),((chlSab1,(papHam1,(rheMac3,macFas5))),(nomLeu3,(ponAbe2,(gorGor3,(hg19,panTro4)))))))))$6,(((oryAfe1,((echTel2,chrAsi1),eleEdw1)),((triMan1,dugDug),loxAfr3)),dasNov3)$7);

#7 For assessing robustness of Hawaiian monk seal estimates to local topology, using ursids as the outgroup to pinnipeds:

((((conCri1,(eriEur2,sorAra2))$1,((felCat5#2,(canFam3#2,((musFur1#2,enhLut#10)#2,((ailMel1#2,ursMar1#2)#2,(((lepWed1#9,Hawaii#12)#12,largha#9)#12,(odoRosDi#9,arcGaz#9)#9)#12)#2)#2)#2)#2,(((pteVam1,pteAle1),(mini,(eptFus1,(myoDav1,(brandtBat,myoLuc2)))))$3,((cerSim1#4,equCab2#4)#4,((vicPac2#4,camFer1#4)#4,(susScr3#4,((((turTru2,orcOrc1),lipVex),phyCat),Minke)$8,hippo#4)#4,(bosTau7#4,(panHod1#4,(oviAri3#4,capHir1#4)#4)#4)#4)#4)#4)#4)#4)#4)#2)#2,((tupChi1,((ochPri3,oryCun2),((casCan,(jacJac1,((rn5,mm10),(micOch1,(mesAur1,criGri1))))),(speTri2,(hetGla2,(cavPor3,(chiLan1,octDeg1)))))))$5,(otoGar3,((saiBol1,calJac3),((chlSab1,(papHam1,(rheMac3,macFas5))),(nomLeu3,(ponAbe2,(gorGor3,(hg19,panTro4)))))))))$6,(((oryAfe1#7,((echTel2#7,chrAsi1#7),eleEdw1#7)#7)#7,((triMan1,dugDug)$11,loxAfr3#7)#7)#7,dasNov3#7)#7);

#8 For estimating d_N/d_S on fully pseudogenic lineages and assessing significance of its difference from 1:

(((conCri1 #1,(eriEur2 #2,sorAra2 #3) #4) #4,((felCat5 #6,(canFam3 #7,((ailMel1 #8,ursMar1 #9) #10,((musFur1 #11,enhLut #12) #13,(((lepWed1 #14,Hawaii #14) #15,largha #16) #17,(odoRosDi #18,arcGaz #19) #20) #21) #22) #23) #24) #25,(((pteVam1 #26,pteAle1 #27) #28,(mini #29,(eptFus1 #30,(myoDav1 #31,(brandtBat #32,myoLuc2 #33) #34) #34) #36) #37) #38,((cerSim1 #39,equCab2 #40) #41,((vicPac2 #42,camFer1 #43) #44,(susScr3 #45,

((((((turTru2 #127,orcOrc1 #127) #127,lipVex #127) #127,phyCat #127) #127,Minke #127) #54,hippo #55) #56,(bosTau7 #57,(panHod1 #58,(oviAri3 #59,capHir1 #60) #60) #62) #63) #64) #65) #66) #67) #69) #69) #69,((tupChi1 #71,((ochPri3 #72,oryCun2 #73) #74,((casCan #75,(jacJac1 #76,((rn5 #77,mm10 #78) #79,(micOch1 #80,(mesAur1 #81,criGri1 #82) #83) #84) #85) #86) #87,(speTri2 #88,(hetGla2 #89,(cavPor3 #90,(chiLan1 #91,octDeg1 #92) #93) #94) #95) #96) #96) #98) #99,(otoGar3 #100,((saiBol1 #101,calJac3 #102) #103,((chlSab1 #104,(papHam1 #105,(rheMac3 #106,macFas5 #107) #107) #109) #110,(nomLeu3 #111,(ponAbe2 #112,(gorGor3 #113,(hg19 #114,panTro4 #115) #116) #116) #118) #119) #119) #121) #122) #124,(((oryAfe1 #125,((echTel2 #126,chrAsi1 #0) #117,eleEdw1 #120) #108) #108,((triMan1 #127,dugDug #127) #61,loxAfr3 #35) #68) #108,dasNov3 #70) #5);

We used the following trees for assessing robustness of our results to variation in the tree used for inferences of functional loss and marine transition rates in BayesTraits:

From Bininda-Emonds et al. (*49*): ((((((((((((((chrAsi1: 0.143350, echTel2: 0.233284): 0.025760, eleEdw1: 0.229356): 0.006898, oryAfe1: 0.111972): 0.007094, (triMan1: 0.073576, loxAfr3: 0.072415): 0.029773): 0.046373, dasNov3: 0.182288): 0.020857, ((((((((((oviAri3: 0.011186, capHir1: 0.018999): 0.009035, panHod1: 0.016435): 0.018612, bosTau7: 0.027888): 0.107579, (turTru2: 0.005355, orcOrc1: 0.002535): 0.057611): 0.015675, susScr3: 0.095112): 0.008406, (camFer1: 0.016128, vicPac2: 0.018535): 0.082116): 0.034641, (cerSim1: 0.057399, equCab2: 0.087013): 0.036501): 0.005781, ((((musFur1: 0.066447, (odoRosDi: 0.022019, lepWed1: 0.024733): 0.026455): 0.006208, ailMel1: 0.053454): 0.016055, canFam3: 0.098719): 0.019762, felCat5: 0.088373): 0.035795): 0.001311, ((pteAle1: 0.008890, pteVam1: 0.018174): 0.116559, (eptFus1: 0.041851, (myoLuc2: 0.013929, myoDav1: 0.024245):

0.023789): 0.091639): 0.041616): 0.004263, (conCri1: 0.172800, (eriEur2: 0.259454, sorAra2: 0.232393): 0.024629): 0.025037): 0.018426): 0.015506, ((((jacJac1: 0.204239, (((criGri1: 0.048366, mesAur1: 0.065787): 0.044915, (mm10: 0.090139, rn5: 0.092226): 0.085868): 0.001762, micOch1: 0.129134): 0.132206): 0.054664, (octDeg1: 0.136000, ((hetGla2: 0.123639, chiLan1: 0.085930): 0.004342, cavPor3: 0.144027): 0.001709): 0.128634): 0.007555, speTri2: 0.131122): 0.024663, (oryCun2: 0.098849, ochPri3: 0.187411): 0.109700): 0.013394): 0.004239, tupChi1: 0.172116): 0.013486, otoGar3: 0.152967): 0.063382, (calJac3: 0.028548, saiBol1: 0.028151): 0.038638): 0.023027, (chlSab1: 0.011251, (papHam1: 0.007843, (rheMac3: 0.001701, macFas5: 0.001399): 0.002293): 0.001712): 0.028199): 0.011241, nomLeu3: 0.018425): 0.002529, ponAbe2: 0.016272): 0.008027, gorGor3: 0.006483): 0.002264, panTro4: 0.005083): 0.000018, hg19: 0.005066);

From Meredith et al. (*17*): ((((((((((dasNov3: 0.182346, ((oryAfe1: 0.112172, ((echTel2: 0.233380, chrAsi1: 0.143404): 0.025867, eleEdw1: 0.229424): 0.006654): 0.007010, (triMan1: 0.073719, loxAfr3: 0.072303): 0.029773): 0.046455): 0.020829, ((eriEur2: 0.278620, (conCri1: 0.161523, sorAra2: 0.241537): 0.017235): 0.023583, ((felCat5: 0.087603, (canFam3: 0.098721, (musFur1: 0.068036, (ailMel1: 0.055555, (lepWed1: 0.024733, odoRosDi: 0.022026): 0.028340): 0.002033): 0.017750): 0.019948): 0.036465, (((pteVam1: 0.018167, pteAle1: 0.008894): 0.116496, (eptFus1: 0.041782, (myoDav1: 0.024256, myoLuc2: 0.013918): 0.023862): 0.091789): 0.041270, ((cerSim1: 0.057702, equCab2: 0.086736): 0.036605, ((vicPac2: 0.018536, camFer1: 0.016128): 0.081868, (susScr3: 0.095240, ((turTru2: 0.005368, orcOrc1: 0.002521): 0.056741, (panHod1: 0.017222, (bosTau7: 0.045187, (oviAri3: 0.010760, capHir1: 0.019462): 0.009747): 0.000287): 0.123715): 0.016363): 0.008448): 0.034326): 0.004944): 0.002178): 0.004555): 0.018411): 0.015395, (tupChi1: 0.168696, ((ochPri3: 0.187276, oryCun2: 0.098973): 0.111064, ((jacJac1: 0.206518, ((rn5: 0.091889, mm10: 0.090465): 0.064033, (micOch1: 0.122339, (mesAur1: 0.065467, criGri1: 0.048665): 0.035515): 0.036236): 0.111902): 0.056731, (speTri2: 0.128493, (hetGla2: 0.096980, (cavPor3: 0.138094, (chiLan1: 0.070197, octDeg1: 0.121169): 0.021754): 0.030846): 0.106569): 0.006421): 0.025081): 0.008142): 0.007276): 0.014487, otoGar3: 0.152776): 0.064176, (saiBol1: 0.028181, calJac3: 0.028525): 0.038639): 0.023035, (chlSab1: 0.011252, (papHam1: 0.007844, (rheMac3: 0.001701, macFas5: 0.001399): 0.002293): 0.001712): 0.028173): 0.011251, nomLeu3: 0.018431): 0.002523, ponAbe2: 0.016274): 0.008026, gorGor3: 0.006483): 0.002264, panTro4: 0.005083): 0.005081, hg19: 0.000004);

From UCSC: ((((((((((((((((((ailMel1: 0.055565, (lepWed1: 0.024710, odoRosDi: 0.022049): 0.028265): 0.002325, musFur1: 0.067738): 0.018059, canFam3: 0.098263): 0.019934, felCat5: 0.087674): 0.035872, (cerSim1: 0.057971, equCab2: 0.086686): 0.039139): 0.001294, ((eptFus1: 0.041879, (myoDav1: 0.024254, myoLuc2: 0.013921): 0.023756): 0.091935, (pteAle1: 0.008885, pteVam1: 0.018177): 0.116286): 0.042207): 0.002249, (((((bosTau7: 0.045079, (capHir1: 0.019458, oviAri3: 0.010773): 0.009844): 0.000291, panHod1: 0.017118): 0.122727, (orcOrc1: 0.002645, turTru2: 0.005244): 0.057704): 0.018998, (camFer1: 0.016249, vicPac2: 0.018416): 0.085138): 0.001164, susScr3: 0.097354): 0.041701): 0.004128, ((conCri1: 0.161274, sorAra2: 0.241682):

0.017428, eriEur2: 0.278172): 0.023349): 0.018577, ((((chrAsi1: 0.142889, echTel2: 0.234488): 0.033522, ((eleEdw1: 0.243957, loxAfr3: 0.085550): 0.003880, triMan1: 0.088660): 0.012528): 0.005075, oryAfe1: 0.114169): 0.045829, dasNov3: 0.182524): 0.021504): 0.015602, ((((cavPor3: 0.137851, (chiLan1: 0.070284, octDeg1: 0.121058): 0.021939): 0.030825, hetGla2: 0.097341): 0.108968, (((((criGri1: 0.048691, mesAur1: 0.065429): 0.035767, micOch1: 0.122176): 0.036137, (mm10: 0.090615, rn5: 0.091697): 0.064155): 0.111607, jacJac1: 0.206193): 0.055282, speTri2: 0.129369): 0.005714): 0.024641, (ochPri3: 0.187334, oryCun2: 0.098942): 0.109851): 0.012750): 0.004414, tupChi1: 0.170891): 0.014037, otoGar3: 0.152503): 0.063931, (calJac3: 0.028554, saiBol1: 0.028149): 0.038484): 0.023179, (chlSab1: 0.011248, ((macFas5: 0.001399, rheMac3: 0.001701): 0.002292, papHam1: 0.007844): 0.001715): 0.028206): 0.011232, nomLeu3: 0.018423): 0.002531, ponAbe2: 0.016274): 0.008026, gorGor3: 0.006483): 0.002263, panTro4: 0.005083): 0.004954, hg19: 0.000131);

Our consensus topology, with branch lengths re-estimated from 10 randomly selected genes: ((((((((((dasNov3: 0.182165, ((oryAfe1: 0.112147, ((echTel2: 0.233323, chrAsi1: 0.143401): 0.025859, eleEdw1: 0.229433): 0.006652): 0.007006, (triMan1: 0.073719, loxAfr3: 0.072291): 0.029799): 0.046681): 0.020578, ((conCri1: 0.172666, (eriEur2: 0.259720, sorAra2: 0.232012): 0.024729): 0.025173, ((felCat5: 0.088291, (canFam3: 0.098671, (ailMel1: 0.053450, (musFur1: 0.066457, (lepWed1: 0.024748, odoRosDi: 0.022006): 0.026436): 0.006218): 0.016070): 0.019850): 0.035853, (((pteVam1: 0.018164, pteAle1: 0.008898): 0.116483, (eptFus1: 0.041792, (myoDav1: 0.024256, myoLuc2: 0.013918): 0.023851): 0.091834): 0.041032, ((cerSim1: 0.057793, equCab2: 0.086634): 0.036549, ((vicPac2: 0.018539, camFer1: 0.016125): 0.082116, (susScr3: 0.095269, ((turTru2: 0.005368, orcOrc1: 0.002521): 0.057664, (bosTau7: 0.027892, (panHod1: 0.016438, (oviAri3: 0.011185, capHir1: 0.018999): 0.009034): 0.018612): 0.107547): 0.015573): 0.008388): 0.034207): 0.005158): 0.002134): 0.004861): 0.018220): 0.015404, (tupChi1: 0.168739, ((ochPri3: 0.187281, oryCun2: 0.098947): 0.111092, ((jacJac1: 0.206431, ((rn5: 0.091908, mm10: 0.090430): 0.064031, (micOch1: 0.122321, (mesAur1: 0.065470, criGri1: 0.048655): 0.035511): 0.036261): 0.111967): 0.056743, (speTri2: 0.128394, (hetGla2: 0.096996, (cavPor3: 0.138075, (chiLan1: 0.070202, octDeg1: 0.121153): 0.021754): 0.030828): 0.106633): 0.006418): 0.025037): 0.008123): 0.007542): 0.014124, otoGar3: 0.152700): 0.064313, (saiBol1: 0.028183, calJac3: 0.028523): 0.038628): 0.023044, (chlSab1: 0.011251, (papHam1: 0.007844, (rheMac3: 0.001704, macFas5: 0.001400): 0.002292): 0.001712): 0.028172): 0.011251, nomLeu3: 0.018430): 0.002524, ponAbe2: 0.016274): 0.008026, gorGor3: 0.006483): 0.002263, panTro4: 0.005083): 0.000005, hg19: 0.005079);

FICD: ((((camFer1:0.0113151,vicPac2:0.00833575)1.00
:0.0680512,(susScr3:0.0849603,((bosTau7:0.0190462,(panHod1:0.0117587,(capHir1:0.0
0723672,oviAri3:0.00296763)0.95 :0.00803626)0.96 :0.0105205)1.00
:0.0611889,(orcOrc1:0.00431941,turTru2:0.00364521)1.00 :0.0474277)0.88
:0.00825514)0.89 :0.0114841)0.91
:0.0154473,(sorAra2:0.14432,eriEur2:0.174343)0.99 :0.054644)0.76
:0.0182471,(ochPri3:0.152512,oryCun2:0.0833944)1.00
:0.0620949,((speTri2:0.0931615,eleEdw1:0.166057)0.86

:0.0185115,(((oryAfe1:0.102396,(loxAfr3:0.0547178,triMan1:0.0474531)0.91
:0.0182737)0.85 :0.0116118,(echTel2:0.13055,chrAsi1:0.157965)0.00
:0.0138564)0.99
:0.0368897,(((hetGla2:0.0512119,(cavPor3:0.0504482,(chiLan1:0.0417467,octDeg1:0.06
85713)0.92 :0.0137134)0.85 :0.0116289)1.00
:0.0598863,(jacJac1:0.137513,((micOch1:0.0561351,(mesAur1:0.0293882,criGri1:0.044
4404)0.89 :0.0156857)0.99 :0.029273,(rn5:0.0634989,mm10:0.0663803)0.89
:0.0159785)1.00 :0.0742083)0.98 :0.0358984)0.90
:0.0115399,((tupChi1:0.0825691,((calJac3:0.0289725,saiBol1:0.0173061)1.00
:0.0232913,((chlSab1:0.00568676,(papHam1:0.00295565,(macFas5:8e-
008,rheMac3:0.00097339)0.85 :0.00194624)0.88 :0.00426717)1.00
:0.0177847,(ponAbe2:0.0100221,(nomLeu3:0.0131533,(gorGor3:0.00420826,(hg19:0.00
898183,panTro4:0.00298853)0.89 :0.00378587)0.97 :0.00643285)0.21
:0.00099022)0.98 :0.0122414)0.89 :0.00696588)1.00 :0.0474217)0.30
:0.00936917,(otoGar3:0.108909,(dasNov3:0.140854,((conCri1:0.11771,(cerSim1:0.0519
077,equCab2:0.0449336)0.94 :0.0181295)0.00
:0.00055546,((felCat5:0.0613243,(canFam3:0.0617033,(musFur1:0.0500125,(ailMel1:0.
0347766,(lepWed1:0.00967075,odoRosDi:0.022054)0.96 :0.0110034)0.83
:0.00538302)0.96 :0.0142826)0.00 :0.00430767)0.99
:0.0275869,((eptFus1:0.0214641,(myoLuc2:0.0144116,myoDav1:0.0257732)0.75
:0.0076459)1.00 :0.0625247,(pteAle1:7e-008,pteVam1:0.00286805)1.00
:0.0680928)0.95 :0.0203209)0.90 :0.0174658)0.96 :0.0258138)0.85
:0.0140837)0.81 :0.00806651)0.94 :0.0184718)0.79 :0.0183756)0.78
:0.0206549)0.93 :0.0183008);

FUT9:
(((speTri2:0.0252541,(hetGla2:0.0191224,(cavPor3:0.0414395,(chiLan1:0.0302658,octD
eg1:0.0308716)0.29 :0.00322026)0.93 :0.00707553)1.00 :0.022308)0.95
:0.0119096,((tupChi1:0.0399769,((ochPri3:0.0418329,oryCun2:0.0253716)0.78
:0.0088666,(((calJac3:0.00478283,saiBol1:0.00584235)0.92
:0.00398593,(((rheMac3:0.00093944,macFas5:0.00093949)0.00 :8e-
008,(papHam1:0.00282833,chlSab1:0.00093951)0.00 :8e-008)0.99
:0.00789241,(ponAbe2:0.00476144,(nomLeu3:0.00569764,(panTro4:0.00378619,(gorGo
r3:5e-008,hg19:0.00188882)0.73 :0.00097065)0.94 :0.00391104)0.73
:0.0009181)0.82 :0.00171276)0.87 :0.00298119)1.00
:0.0168347,otoGar3:0.0413524)0.68 :0.00194349)0.10 :0.00110366)0.77
:0.00132292,((dasNov3:0.044413,((oryAfe1:0.0175616,(chrAsi1:0.0397976,(loxAfr3:0.0
105874,triMan1:0.00825529)0.88 :0.00325993)0.00 :4.545e-005)0.00 :7e-
008,(eleEdw1:0.0513143,echTel2:0.0969533)0.84 :0.0104349)0.99 :0.0135642)0.95
:0.0086058,(((cerSim1:0.0188239,equCab2:0.0120704)0.95
:0.00781477,((pteAle1:0.00367297,pteVam1:0.00292656)1.00
:0.0284741,(eptFus1:0.00476038,(myoDav1:0.00758172,myoLuc2:0.007672)0.98
:0.00987068)1.00 :0.0310162)0.93 :0.00870513)0.73
:0.00101565,((susScr3:0.027518,((orcOrc1:0.00131581,turTru2:0.00242622)1.00
:0.0168157,((bosTau7:0.00104191,(panHod1:0.00477473,(capHir1:0.00335392,oviAri3:
0.00133732)0.88 :0.00270829)0.92 :0.00570738)1.00

:0.0254804,(camFer1:0.00393808,vicPac2:0.00073725)1.00 :0.0178516)0.23
:0.00151601)0.28 :0.0037157)0.96
:0.00777287,((((lepWed1:0.00373807,odoRosDi:0.00967308)0.90
:0.00409797,(ailMel1:0.029386,musFur1:0.00926697)0.81 :0.00252731)0.86
:0.00266733,(felCat5:0.0249583,canFam3:0.00671789)0.73 :0.00092662)0.98
:0.0102845,(conCri1:0.0539119,sorAra2:0.0587859)0.85 :0.00698123)0.55
:0.00061409)0.76 :0.00163296)0.94 :0.00558026)0.89 :0.00555256)0.52
:0.00396628)0.43
:0.00886831,(jacJac1:0.0811645,((mm10:0.0215201,rn5:0.0264752)0.95
:0.0109563,(micOch1:0.0454886,(criGri1:0.0251423,mesAur1:0.0239087)0.89
:0.00666539)0.89 :0.00732828)1.00 :0.03702)0.62 :0.0094863,eriEur2:0.201525);

GPR22:
(jacJac1:0.0565971,((micOch1:0.032242,(criGri1:0.0248524,mesAur1:0.0369192)0.98
:0.0158816)0.72 :0.00451527,(mm10:0.0243887,rn5:0.0243697)0.99
:0.0194461)1.00
:0.039754,((speTri2:0.0316967,((ochPri3:0.0588563,oryCun2:0.0241152)1.00
:0.0213299,((tupChi1:0.0312922,(otoGar3:0.0224436,((calJac3:0.00645248,saiBol1:0.00
898666)0.91 :0.00495251,(((nomLeu3:0.00355323,ponAbe2:0.00445163)0.00 :1.1e-
007,(hg19:1e-008,(panTro4:0.00088375,gorGor3:8e-008)0.00 :1e-008)0.81
:0.00088505)0.95 :0.00423273,((macFas5:8e-008,rheMac3:0.00177079)0.82
:0.00088231,(papHam1:6e-008,chlSab1:0.00176667)0.00 :6e-008)0.89
:0.00293661)0.69 :0.00187277)0.96 :0.00598497)0.89 :0.00353074)0.00 :1.7e-
007,((dasNov3:0.0285546,((loxAfr3:0.0210592,triMan1:0.0129913)0.64
:0.00262651,(oryAfe1:0.0345266,(chrAsi1:0.0382627,(echTel2:0.080383,eleEdw1:0.045
6413)0.75 :0.00593339)0.79 :0.00319943)0.00 :0.00050969)0.83
:0.00341471)0.99 :0.0085972,(((camFer1:0.00153295,vicPac2:0.00538669)1.00
:0.0217491,(susScr3:0.0141247,((orcOrc1:0.00179724,turTru2:0.00081614)0.99
:0.0101841,(bosTau7:0.00543224,(panHod1:0.00350998,(capHir1:0.00174224,oviAri3:0
.0008668)0.87 :0.00175798)0.92 :0.0034465)1.00 :0.0118994)0.93
:0.0064075)0.63 :0.00081167)0.96
:0.00592339,((sorAra2:0.0511974,(eriEur2:0.0391292,conCri1:0.0271244)0.07
:0.00348964)0.89 :0.00520869,(((pteAle1:0.00075575,pteVam1:0.00690898)1.00
:0.0219761,(eptFus1:0.0123169,(myoDav1:0.00536628,myoLuc2:0.00447631)0.86
:0.00314951)1.00 :0.020654)0.95
:0.00791576,((cerSim1:0.00725391,equCab2:0.0138684)0.91
:0.00343907,(felCat5:0.0335148,(canFam3:0.019121,((lepWed1:0.00792362,odoRosDi:0
.00169497)0.99 :0.00733065,(musFur1:0.0165202,ailMel1:0.0111458)0.55
:0.00038508)0.95 :0.00373955)0.00 :0.00029403)0.98 :0.00582452)0.00 :8e-
008)0.74 :0.00087093)0.00 :8e-008)0.96 :0.00443706)0.87 :0.00187165)0.00
:1e-007)0.99 :0.00971806)0.00
:0.00099399,(hetGla2:0.0200643,(cavPor3:0.0157943,(chiLan1:0.0123284,octDeg1:0.02
82111)0.92 :0.00632537)0.96 :0.00941812)0.98 :0.0135757)0.98 :0.0216899);

HMGCS1: (((((ochPri3:0.0856176,oryCun2:0.0266581)0.99
:0.0212798,((jacJac1:0.0987852,((mm10:0.0372053,rn5:0.0359171)1.00

:0.0266626,(micOch1:0.0568156,(criGri1:0.0165676,mesAur1:0.0135188)0.99
:0.0151017)0.94 :0.00956395)0.99 :0.0188561)1.00
:0.0264911,(speTri2:0.0375365,(hetGla2:0.02556,(cavPor3:0.0379203,(octDeg1:0.11832
3,chiLan1:0.0204085)0.85 :0.00404669)0.92 :0.00895868)1.00 :0.0232632)0.79
:0.00204225)0.84 :0.00339879)0.94
:0.00505521,(tupChi1:0.0476103,(otoGar3:0.0499805,((calJac3:0.00574844,saiBol1:0.00
456361)1.00 :0.00930511,((chlSab1:0.00368577,(papHam1:6e-
008,(macFas5:0.00072364,rheMac3:0.00072651)0.94 :0.00219309)0.87
:0.00142437)0.98
:0.00590278,(nomLeu3:0.00803454,(hg19:0.00073058,((gorGor3:0.00217919,panTro4:6
e-008)0.93 :0.00217552,ponAbe2:0.00511685)0.00 :7e-008)0.88 :0.00155083)0.87
:0.0022914)0.98 :0.0066349)1.00 :0.0168596)0.91 :0.00619577)0.63
:0.00057843)0.96
:0.00555991,(dasNov3:0.0304368,((loxAfr3:0.0450844,triMan1:0.0165639)0.98
:0.0123257,((chrAsi1:0.0533801,echTel2:0.052989)0.62
:0.0046883,oryAfe1:0.0337237)0.58 :0.00091322)0.99 :0.0112845)0.92
:0.00522304)1.00 :0.0113655,((pteAle1:0.00388615,pteVam1:0.00398453)1.00
:0.0550374,(felCat5:0.0332636,(((ailMel1:0.0271614,musFur1:0.0337699)0.74
:0.00312041,(lepWed1:0.019165,odoRosDi:0.00831526)0.97 :0.00856045)0.40
:0.00453015,canFam3:0.0323225)0.96 :0.00830633)1.00 :0.0247181)0.54
:0.00104489,((cerSim1:0.0161512,equCab2:0.0213758)0.97
:0.00682934,(((camFer1:0.00488398,vicPac2:0.00236493)1.00
:0.0264767,(susScr3:0.0478804,((bosTau7:0.0103307,(panHod1:0.00422455,(capHir1:0.
00426336,oviAri3:0.00286926)0.86 :0.0022023)0.95 :0.00595797)1.00
:0.031872,(orcOrc1:0.00234852,turTru2:0.00123681)1.00 :0.0212617)0.90
:0.00568621)0.89 :0.00363082)0.99
:0.0111512,((conCri1:0.0443008,sorAra2:0.098017)0.91
:0.0102665,(eriEur2:0.116644,(eleEdw1:0.14555,(eptFus1:0.0163758,(myoDav1:0.0151
5,myoLuc2:0.00254169)0.95 :0.0101269)1.00 :0.0592239)0.71 :0.0128595)0.17
:0.0156363)0.83 :0.00646448)0.82 :0.00224554)0.72 :0.00075171);

LRFN5: ((dasNov3:0.0318458,((loxAfr3:0.0191675,triMan1:0.0144298)0.97
:0.00627334,((eleEdw1:0.037298,oryAfe1:0.0248276)0.74
:0.00399622,(echTel2:0.0671128,chrAsi1:0.0309183)0.07 :0.00320143)0.86
:0.00235037)1.00 :0.00848744)0.84
:0.00398305,((((calJac3:0.00483367,saiBol1:0.00776163)1.00
:0.00773721,((chlSab1:0.00312383,(papHam1:0.00102598,(macFas5:1e-
008,rheMac3:1e-008)0.79 :0.00051348)0.75 :0.0004841)0.98
:0.00433112,(nomLeu3:0.00955918,(ponAbe2:0.00604416,(hg19:0.00258286,(panTro4:
0.00207407,gorGor3:0.00311947)0.72 :0.00049324)0.93 :0.00176096)0.17
:0.00042452)0.94 :0.00300433)0.82 :0.002707)1.00
:0.017587,(tupChi1:0.0379258,otoGar3:0.0985895)0.83 :0.00908717)0.83
:0.00233482,((ochPri3:0.0469995,oryCun2:0.022518)1.00
:0.0218685,(speTri2:0.0375964,((hetGla2:0.0425399,(cavPor3:0.0361735,(chiLan1:0.02
1009,octDeg1:0.0480313)0.83 :0.00305516)0.95 :0.00689752)1.00
:0.0229595,(jacJac1:0.0788883,((mm10:0.015615,rn5:0.0237309)0.98

:0.0108368,(micOch1:0.0312104,(criGri1:0.0122702,mesAur1:0.0195854)1.00
:0.0134545)0.80 :0.00617706)1.00 :0.034339)0.97 :0.0144288)0.87
:0.00415255)0.99 :0.00781307)0.82 :0.00176424)0.99
:0.00558171,(conCri1:0.0431041,(((felCat5:0.0262974,(canFam3:0.0201341,(musFur1:0
.0104284,(ailMel1:0.0189768,(lepWed1:0.00843224,odoRosDi:0.00905721)0.93
:0.00281516)0.74 :0.00049886)0.84 :0.00181713)0.93 :0.00381748)1.00
:0.00936281,((sorAra2:0.0572675,eriEur2:0.103489)0.90
:0.00887645,((eptFus1:0.0235021,(myoDav1:0.0135051,myoLuc2:0.006397)0.98
:0.00844063)1.00 :0.0225213,(pteAle1:0.00089945,pteVam1:0.00116055)1.00
:0.0197367)0.68 :0.00337329)0.00 :6e-008)0.87
:0.00122768,((cerSim1:0.0184677,equCab2:0.0263263)0.97
:0.00619997,((camFer1:0.00404404,vicPac2:0.00416504)1.00
:0.0284819,(susScr3:0.0226796,((bosTau7:0.00718105,(panHod1:0.00201075,(capHir1:0
.00206515,oviAri3:0.00204003)0.85 :0.0010597)0.66 :0.00175357)1.00
:0.0274274,(orcOrc1:5e-008,turTru2:0.00204089)1.00 :0.0145571)0.93
:0.00513814)0.33 :0.00102878)0.97 :0.00588712)0.06 :0.00145147)0.61
:0.00070464)0.97 :0.00510741);

MSL1:
(speTri2:0.0224281,((tupChi1:0.0121128,(((((odoRosDi:0.00109101,(canFam3:0.005367
47,(musFur1:1.2e-007,(ailMel1:0.00213755,(lepWed1:0.00320526,felCat5:7e-008)0.00
:7e-008)0.86 :0.00106359)0.00 :7e-008)0.31 :0.00104601)0.99
:0.00650514,((((camFer1:2.6e-007,vicPac2:0.00106484)0.78
:0.00293806,(susScr3:0.00723797,((orcOrc1:0.00106595,turTru2:1e-008)0.92
:0.00437777,(bosTau7:0.0036306,(panHod1:1e-008,(capHir1:1e-008,oviAri3:1e-
008)0.00 :1e-008)0.83 :0.001975)0.97 :0.00813381)0.30 :0.00260788)0.71
:0.00181286)0.96 :0.00546003,(eriEur2:0.0268699,sorAra2:0.0208485)0.63
:0.00137338)0.78
:0.00156905,((cerSim1:0.00876284,(conCri1:0.0124684,equCab2:0.00682486)0.33
:0.00083457)0.75
:0.0010816,((eptFus1:0.00218162,(myoDav1:0.0110614,myoLuc2:0.00219969)0.85
:0.00207264)0.98 :0.00892742,(pteAle1:1e-008,pteVam1:1e-008)0.98
:0.00664995)0.81 :0.0020657)0.77 :0.00104746)0.00 :5e-008)0.88
:0.00215269,(dasNov3:0.022645,((eleEdw1:0.0308487,(echTel2:0.037209,(loxAfr3:0.00
54376,triMan1:0.00661684)0.76 :0.00097929)0.74 :0.00121823)0.71
:0.00085257,(chrAsi1:0.0192974,oryAfe1:0.0132616)0.73 :0.00109194)0.95
:0.00567834)0.70 :0.00092765)0.93
:0.00323301,((calJac3:0.00437073,saiBol1:0.00102353)0.95
:0.00592161,(otoGar3:0.00936697,((rheMac3:1e-008,(papHam1:1e-008,(chlSab1:1e-
008,macFas5:1e-008)0.00 :1e-008)0.00 :1e-008)0.79
:0.00106775,(nomLeu3:0.00214049,(hg19:0.0010654,(gorGor3:0.00213634,(panTro4:1e
-008,ponAbe2:0.0010654)0.00 :7e-008)0.00 :6e-008)0.00 :6e-008)0.80
:0.00106877)0.78 :0.00158588)0.63 :0.00214705)0.78 :0.00173874)0.00 :1e-
007,(ochPri3:0.0225345,oryCun2:0.00768134)0.81 :0.00212661)0.00 :6e-008)0.78
:0.00109597,(jacJac1:0.0172782,((mesAur1:0.0082598,criGri1:0.00596939)0.83
:0.00220242,(micOch1:0.00930889,(mm10:0.0136225,rn5:0.0121802)0.99

:0.0142137)0.64 :0.00069856)1.00 :0.0202659)0.87 :0.00441366)0.00 :1.15e-006,(hetGla2:0.0151803,(cavPor3:0.0186586,(chiLan1:0.0176824,octDeg1:0.0392829)0.82 :0.00851798)0.89 :0.00847947)1.00 :0.0184795);

NUDT12:
(speTri2:0.0425823,((hetGla2:0.0357194,(octDeg1:0.0561017,(chiLan1:0.0386855,cavPor3:0.0509025)0.55 :0.00142625)0.81 :0.00297384)0.99
:0.0145238,(jacJac1:0.0940661,((mm10:0.0473469,rn5:0.043772)1.00
:0.0467127,(micOch1:0.0530517,(criGri1:0.0274031,mesAur1:0.0325238)1.00
:0.0341697)0.98 :0.0259445)1.00 :0.0557146)0.92 :0.0151456)0.00 :1.7e-007,(((ochPri3:0.0907288,oryCun2:0.0233144)1.00
:0.0275419,(((calJac3:0.00850664,saiBol1:0.00299127)1.00
:0.0134525,((chlSab1:0.0012899,(papHam1:0.00459331,(macFas5:0.00075802,rheMac3:1e-008)0.78 :0.00076655)0.75 :0.00098603)0.96
:0.00560302,((panTro4:0.00151027,(hg19:0.0053391,gorGor3:0.00455136)0.00 :8e-008)0.99 :0.00536743,(nomLeu3:0.0053599,ponAbe2:0.00538995)0.75
:0.00074778)0.96 :0.00548682)0.87 :0.00615404)1.00
:0.0298959,((dasNov3:0.0431147,(chrAsi1:0.0455918,((oryAfe1:0.0376974,echTel2:0.0762268)0.65
:0.00156073,(eleEdw1:0.0871813,(loxAfr3:0.0250531,triMan1:0.0249174)0.57
:0.00723485)0.80 :0.00298499)0.68 :0.00164909)0.99 :0.0134537)0.87
:0.0050861,((((camFer1:0.0055332,vicPac2:0.00206213)1.00
:0.0183576,(susScr3:0.0239199,((bosTau7:0.00762166,(panHod1:0.00389328,(capHir1:0.00161154,oviAri3:0.00294361)0.81 :0.00157972)0.61 :0.0005261)1.00
:0.0226462,(orcOrc1:0.00343307,turTru2:0.00264252)1.00 :0.0170346)0.88
:0.00437814)0.72 :0.00085922)0.97
:0.00589454,(((ailMel1:0.0117042,(musFur1:0.0151096,(lepWed1:0.00899603,odoRosDi:0.00727803)0.94 :0.00477357)0.79 :0.00128439)0.90
:0.00255516,(canFam3:0.015646,felCat5:0.0186303)0.44 :0.00078944)1.00
:0.0112148,(conCri1:0.0575939,(sorAra2:0.125381,eriEur2:0.0610922)0.82
:0.00867204)0.87 :0.00604912)0.48 :0.00105968)0.48
:0.0014471,((eptFus1:0.00842926,(myoDav1:0.0100876,myoLuc2:0.00155801)0.90
:0.00321662)1.00 :0.020784,((pteAle1:0.00182038,pteVam1:0.00344592)1.00
:0.0316241,(cerSim1:0.00859963,equCab2:0.0295157)0.98 :0.00911111)0.00
:0.00037414)0.75 :0.00097393)0.94 :0.00620674)0.99 :0.00765129)0.00 :2.8e-007)0.66 :0.00151878,(tupChi1:0.0455097,otoGar3:0.107945)0.72 :0.00499685)0.99
:0.0131177);

PTPRA:
((hetGla2:0.0268904,(cavPor3:0.0331103,(chiLan1:0.0184881,octDeg1:0.0270941)0.94
:0.0064955)0.98 :0.00875287)1.00
:0.0315438,(speTri2:0.0448732,(jacJac1:0.0552835,((micOch1:0.0276857,(criGri1:0.0167669,mesAur1:0.0161632)0.88 :0.00408778)0.94
:0.00585956,(mm10:0.027634,rn5:0.0229902)1.00 :0.0201627)1.00 :0.0358993)0.98
:0.0148902)0.74 :0.00204476,((ochPri3:0.0499259,oryCun2:0.0352172)1.00
:0.0283439,((otoGar3:0.0488228,(((conCri1:0.0492213,(sorAra2:0.0694607,eriEur2:0.06

46959)0.84 :0.00672661)0.97
:0.0075155,(((felCat5:0.0204158,(canFam3:0.0218601,(musFur1:0.0197894,(ailMel1:0.0
151697,(lepWed1:0.00647877,odoRosDi:0.00533484)1.00 :0.00997791)0.73
:0.00072647)0.94 :0.00328284)0.97 :0.00595109)1.00
:0.0109434,((pteAle1:0.00155829,pteVam1:0.00171926)1.00
:0.0274149,(eptFus1:0.00556836,(myoDav1:0.00646891,myoLuc2:0.00429125)0.99
:0.00597579)1.00 :0.0292591)0.96 :0.0057771)0.49
:0.00108902,(((camFer1:0.00498721,vicPac2:0.00250175)1.00
:0.0301485,(susScr3:0.025868,((bosTau7:0.007496,(panHod1:0.00373799,(oviAri3:0.00
278984,capHir1:0.00093465)0.00 :4e-008)0.98 :0.00530317)1.00
:0.0221771,(orcOrc1:0.00259109,turTru2:0.00113714)1.00 :0.00937249)0.77
:0.00175597)0.57 :0.00356505)1.00
:0.0120025,(cerSim1:0.0180549,equCab2:0.0215482)1.00 :0.0110019)0.79
:0.0019811)0.83 :0.00203881)1.00
:0.00985003,(dasNov3:0.0652303,(eleEdw1:0.0623985,(oryAfe1:0.0344514,((loxAfr3:0.
0338041,triMan1:0.0150292)0.97
:0.00723591,(echTel2:0.0566443,chrAsi1:0.0492765)0.84 :0.00406868)0.63
:0.00105058)0.81 :0.00309594)1.00 :0.0145081)0.69 :0.00492782)0.93
:0.00368391)0.60
:0.00197567,(tupChi1:0.0624698,((calJac3:0.00804924,saiBol1:0.00640387)1.00
:0.0102931,((chlSab1:0.00233851,(papHam1:0.00185493,(macFas5:5e-
008,rheMac3:0.00046156)0.74 :0.00046207)0.76 :0.00044473)1.00
:0.004678,(nomLeu3:0.00950289,(ponAbe2:0.00331403,(panTro4:0.00325956,(hg19:0.0
0138649,gorGor3:0.00185452)0.00 :4e-008)0.90 :0.00134775)0.75
:0.00044318)0.89 :0.00142981)0.97 :0.0059635)1.00 :0.021963)0.67
:0.0013095)0.85 :0.00362697)0.99 :0.0112292);

SGMS2:
tupChi1:0.04833,(sorAra2:0.108439,((eriEur2:0.135287,conCri1:0.0907102)0.09
:0.0161817,(hetGla2:0.0968242,(cavPor3:0.0843947,(octDeg1:0.0681874,chiLan1:0.049
9745)0.89 :0.0194923)1.00 :0.0512568)1.00 :0.135671)0.90 :0.0274779)0.98
:0.0556303,((eptFus1:0.0265863,(myoDav1:0.023898,myoLuc2:0.0117866)0.89
:0.0156714)1.00
:0.064967,((jacJac1:0.0777415,((mm10:0.0554998,rn5:0.0350787)1.00
:0.0392444,(micOch1:0.0522111,(criGri1:0.0105598,mesAur1:0.0282327)0.97
:0.0174241)0.70 :0.00598433)1.00 :0.0698775)0.73
:0.0233108,((((bosTau7:0.00999796,(panHod1:0.00614558,(capHir1:0.00041066,oviAri
3:0.00710956)0.92 :0.00541065)0.99 :0.0134307)1.00
:0.0564218,(orcOrc1:0.00336311,turTru2:0.00324603)1.00 :0.0226077)0.79
:0.00428688,(susScr3:0.0575526,(camFer1:0.00816558,vicPac2:0.00490805)1.00
:0.0637746)0.68 :0.00492172)0.92
:0.00896535,((((speTri2:0.0526967,(ochPri3:0.0696533,oryCun2:0.0181164)1.00
:0.0474739)0.00
:0.00606093,(otoGar3:0.0871842,((ponAbe2:0.00198538,((panTro4:0.00112562,(hg19:1.
1e-007,gorGor3:0.0022532)0.00 :1.1e-007)0.92
:0.00224737,(nomLeu3:0.00681556,(chlSab1:0.00342103,(papHam1:0.00112664,(macF

as5:1e-008,rheMac3:1e-008)0.80 :0.00112912)0.92 :0.00340187)0.96
:0.00459561)0.00 :6e-008)0.83 :0.00252821)0.98
:0.0143342,(calJac3:0.0128919,saiBol1:0.0118061)0.98 :0.0138027)0.98
:0.0223547)0.94 :0.0121142)0.82
:0.00365608,(dasNov3:0.04528,(eleEdw1:0.0765632,((chrAsi1:0.041895,(loxAfr3:0.027
6129,triMan1:0.00748677)0.95 :0.0114076)0.00
:0.00051051,(oryAfe1:0.0389153,echTel2:0.158075)0.73 :0.00483841)0.79
:0.00344324)0.98 :0.0178768)0.73 :0.00392889)0.96
:0.0144574,((felCat5:0.0183984,(canFam3:0.0326621,(ailMel1:0.0252866,(musFur1:0.0
490313,(lepWed1:0.00698862,odoRosDi:0.0102898)0.89 :0.00632492)0.91
:0.00865389)0.83 :0.00469485)0.84 :0.00593671)1.00
:0.024701,((pteAle1:0.00443157,pteVam1:6.1e-007)1.00
:0.0441138,(cerSim1:0.0270733,equCab2:0.0444156)0.99 :0.0254533)0.33
:0.00397169)0.84 :0.00613964)0.86 :0.0105326)0.92 :0.0240844)0.73
:0.0313321)1.00 :0.10503);

TTC5:
((((hetGla2:0.0264745,(octDeg1:0.0413466,(chiLan1:0.044299,cavPor3:0.0474303)0.84
:0.00611231)0.98 :0.0124802)1.00 :0.0206428,speTri2:0.0436478)0.82
:0.00341975,(jacJac1:0.0701968,((micOch1:0.0596696,(criGri1:0.0162028,mesAur1:0.0
534992)0.97 :0.0143606)0.94 :0.0123306,(mm10:0.0328693,rn5:0.0406286)1.00
:0.0283616)1.00 :0.0537968)0.91 :0.00990701)0.99
:0.011198,(ochPri3:0.0863282,oryCun2:0.0496241)1.00
:0.0318025,(tupChi1:0.0556265,((otoGar3:0.0479033,((calJac3:0.0100017,saiBol1:0.012
7578)1.00 :0.0195696,((chlSab1:0.0087537,(papHam1:0.00078188,(macFas5:1e-
008,rheMac3:0.0023566)0.89 :0.0015764)0.82 :0.00242194)1.00
:0.014905,(ponAbe2:0.00647828,(nomLeu3:0.00235143,(gorGor3:0.00078348,(hg19:0.0
00782,panTro4:0.00157126)0.77 :0.00078521)0.98 :0.00399449)0.82
:0.00146781)0.00 :5.2e-007)0.97 :0.00790913)1.00 :0.0315007)0.76
:0.00344387,((dasNov3:0.067816,((chrAsi1:0.0546978,eleEdw1:0.0807454)0.81
:0.00481242,((echTel2:0.0964234,oryAfe1:0.0360768)0.80
:0.0056093,(loxAfr3:0.0220795,triMan1:0.0207824)0.98 :0.0100591)0.00 :7.349e-
005)0.93 :0.0073919)0.92
:0.00589321,((felCat5:0.0320141,(canFam3:0.119123,((ailMel1:0.0323163,(lepWed1:0.0
116698,odoRosDi:0.00446856)1.00 :0.0143813)0.05
:0.00127196,musFur1:0.0175834)0.91 :0.00560905)0.13 :0.00214188)1.00
:0.0205836,((cerSim1:0.0184871,equCab2:0.0269502)0.95
:0.00614456,(((susScr3:0.0312493,((orcOrc1:0.00248977,turTru2:0.00067239)1.00
:0.0215981,(bosTau7:0.00964364,((panHod1:0.00529606,capHir1:0.00265264)0.28
:0.00080467,oviAri3:0.00664175)0.97 :0.00718947)1.00 :0.020393)0.70
:0.00318006)0.72 :0.00112734,(camFer1:0.00483822,vicPac2:0.00232235)1.00
:0.0172551)0.99 :0.00901076,(((pteAle1:0.00068014,pteVam1:0.00281784)1.00
:0.0427225,(eptFus1:0.00796342,(myoDav1:0.00923764,myoLuc2:0.00287082)0.95
:0.00736384)1.00 :0.0267545)0.63
:0.00348544,(conCri1:0.058549,(sorAra2:0.0716877,eriEur2:0.0747044)0.65
:0.00844031)0.94 :0.0116335)0.76 :0.00246178)0.76 :0.00067702)0.73

:0.00097847)0.98 :0.00632163)0.86 :0.00254937)0.91 :0.0043503)0.70 :0.00151902);

We used the following tree for estimating evolutionary rate covariation (ERC) with PON1 in terrestrial mammals (branch lengths in this tree were not used in the analysis):

((((((((((((ailMel1:1,(lepWed1:1,odoRosDi:1):1):1,musFur1:1):1,canFam3:1):1,felCat5:1):1,(cerSim1:1,equCab2:1):1):1,((eptFus1:1,(myoDav1:1,myoLuc2:1):1):1,(pteAle1:1,pteVam1:1):1):1):1,(((((bosTau7:1,oviAri3:2):1,panHod1:1):1,(orcOrc1:1,turTru2:1):1):1,vicPac2:2):1,susScr3:1):1):1,((conCri1:1,sorAra2:1):1,eriEur2:1):1):1,((((chrAsi1:1,echTel2:1):1,((eleEdw1:1,loxAfr3:1):1,triMan1:1):1):1,oryAfe1:1):1,dasNov3:1):2):1,((((cavPor3:1,(chiLan1:1,octDeg1:1):1):1,hetGla2:1):1,(((((criGri1:1,mesAur1:1):1,micOch1:1):1,(mm10:1,rn5:1):1):1,jacJac1:1):1,speTri2:1):1):1,(ochPri3:1,oryCun2:1):1):1):1,tupChi1:1):1,otoGar3:1):1,(calJac3:1,saiBol1:1):1):1,rheMac3:4):4,hg19:1);

Sample acquisition and permissions

We obtained blood samples for seven healthy, wild manatees from Crystal River, Florida. Blood was obtained using previously published techniques for manatee blood sample collection (*80*, *81*). We obtained blood and skin samples from three healthy, wild dugongs from Moreton Bay, Australia using published sampling techniques (*82*, *83*). We obtained blood from the fluke veins of two healthy adult bottlenose dolphins (one 10-year-old male and one 23-year-old male) in human care and from the digital vein of one healthy adult California sea lion (21-year-old male) in human care using a butterfly needle. We obtained blood from the tarsal vein of one healthy adult walrus (23-year-old female) in human care using a 21ga 1.5" needle. We obtained blood from one healthy juvenile (<1 year old) female Northern elephant seal and one healthy adult (3-year-old) female Canadian beaver in human care. All blood collection from animals in human care took place during routine health examinations. We obtained blood during necropsy from four adult ferrets that had previously been exposed to influenza but had recovered at time of sacrifice. We obtained blood from eight healthy, wild Northern elephant seals (five males and three females) from Año Nuevo State Reserve in San Mateo County, CA, USA. Five wild-type mice (C57BL/6J strain background) were purchased from The Jackson Laboratory (Bar Harbor, MI). Five *Pon1* knockout mice (*Pon1*$^{-/-}$) were kindly provided by Drs. Lusis, Shih and Tward (UCLA, Los Angeles, CA) (*26*). Wild-type and *Pon1*$^{-/-}$ mice were anesthetized with tribromoethanol (600 mg/kg, ip; Sigma-Aldrich, St. Louis, MO) and blood extracted via cardiac puncture. Mice were housed in a centralized, AAALAC-accredited, specific pathogen free facility at the University of Washington. They were maintained at room temperature in a 12 h light-dark cycle with unlimited access to food and water.

We obtained all appropriate animal care and use permissions from the relevant research institutions and management organizations, as follows:

1. Blood samples from two bottlenose dolphins, one Canadian beaver, one California sea lion, seven manatees, one Northern elephant seal, and one walrus in human care were obtained using procedures approved by the Pittsburgh Zoo and PPG Aquarium / National Aviary IACUC (protocol # 2015-NC-001) and each sampling institution's research review committees prior to conduction. All manatee samples were collected and held by USGS

under IACUC protocol #USGS-WARC-2016-03 and USFWS research permit MA791721.

2. Dugong blood and tissue samples were collected under Australian Scientific Purposes Permit no. WISP01660304, Moreton Bay Marine Park Permit no. QS2004/CVL228 and University of Queensland Animal Ethics no. ZOO/ENT/344/04/NSF/CRL, and transferred to the USGS Sirenian Project laboratory under authority of the CITES permits 08US808447/9 and 2009AU570750.
3. Ferret blood samples were obtained with approval from the University of Pittsburgh IACUC protocol #16077170.
4. Blood samples from eight wild Northern elephant seals were obtained with approval from Sonoma State University IACUC protocol #2014-48, under NMFS permit #19108.
5. All mouse experiments were approved by the Animal Care and Use Committee of the University of Washington (IACUC protocol # 2343-01), and carried out in accordance with National Research Council Guide for the Care and Use of Laboratory Animals, as adopted by the US National Institutes of Health.

Validating manatee *PON1* coding sequence and determining dugong sequence

Manatee DNA was extracted by a standard phenol-chloroform technique after extracting cells from clotted blood (*84*). Dugong DNA was extracted from two whole blood samples using the salting out procedure of Miller, Dykes, and Polesky (*85*) and from one skin sample using a standard phenol-chloroform technique. We designed primers for all exons of PON1 using Primer3Plus version 2.4.0 and the manatee genome sequence as reference. All exons were amplified using PCR, which was carried out in a 20 μL volume containing: 10 μL of 10x Taq polymerase buffer (New England BioLabs), 0.4 μL each of 10 μM forward and reverse primers, 0.4 μL of 10 mM dNTP mix, 0.5 μL of template DNA (16 – 50 ng/μL) 16.3 μl of water, and 0.08 μL of Taq polymerase. The thermal cycler was programmed for 3 min at 95 °C for initial denaturation, then 34 cycles of 30 s at 95 °C for denaturation, 30 s at 59 °C for annealing, and 45 s at 72 °C for extension, followed by 5 min at 72 °C for the final extension, with minimal adjustments (see Table S12).

PCR products were sequenced by Sanger sequencing at the University of Pittsburgh Genomics Research Core. Sequencing Reaction Sequencing buffer and a 1:4 dilution of BigDye 3.1 were added and thermocycling performed according to ABI recommendations. Removal of unincorporated sequencing reagents was performed using CleanSeq magnetic beads according to manufacturer instructions (Agencourt). The resulting sequence files were manually inspected to confirm homozygosity at observed lesions. The raw data for each exon was aligned to the reference exon using MEGA 7.0.14, and inconsistencies and splice sites were checked manually. Individual exon sequences were then concatenated to generate a consensus coding sequence for *PON1* for dugong.

Evaluating primary function(s) of PON1 using evolutionary rate covariation (ERC)

To generate Tables S6 and S7, we assessed the extent to which genes' branch-specific rates of amino acid evolution correlated with those of PON1 by implementing an evolutionary rate covariation (ERC) analysis, as described in Clark and Aquadro (*86*). We included only non-marine species in this analysis, in order to capture patterns of co-evolution that are not primarily influenced by the loss of function of PON1. To reduce the influence of long branches on the results, we constrained our analyses to eutherian mammals and further pruned the tree to eliminate members of pairs with very short evolutionary distances (see "Constraints on PON1 phylogenetic trees for evolutionary inferences"). To reduce artifacts driven by variation in genes with high levels of missing data, we restricted our analyses to those genes with available sequences for at least 30 species. We estimated evolutionary rate covariation for all remaining 17,511 genes with *PON1*, using gene trees with branch lengths estimated as in Chikina et al. (*5*). To evaluate enrichment of functional categories within our top signals, we performed gene ontology enrichment analysis using GOrilla (*46*), focusing on the top 100 genes that showed a positive correlation in rate with the rate of *PON1*.

Assessing enzymatic activity of blood plasma against four PON1 substrates and control substrate (alkaline phosphatase)

Blood was collected in lithium heparin tubes and centrifuged at 1,500 - 10,000 x g for 10 - 15 min at 4 °C. Plasma was separated from the blood cell fraction and kept stored at -80 °C until use.

All activity assays were determined in a SPECTRAmax® PLUS Microplate Spectrophotometer (Molecular Devices, Sunnyvale, CA). The assay values were corrected for path-length using the software SoftMax Pro 5.4 (Molecular Devices). Levels of plasma arylesterase (AREase), chlorpyrifos-oxonase (CPOase), diazoxonase (DZOase) and paraoxonase (POase) activities were determined as previously described (*87*). Briefly, plasma of all the species analyzed were diluted 1/10 in dilution buffer (9 mM Tris-HCl pH 8.0, 0.9 mM $CaCl_2$) and assayed in triplicate at either 37 °C (for CPOase and POase) or at room temperature (for AREase and DZOase). Activities were expressed in U/mL (AREase, CPOase, and DZOase) or in U/L (POase), based on the molar extinction coefficients of 1.31 mM^{-1} cm^{-1} for phenol (the hydrolysis product of phenyl acetate, AREase activity); 5.56 mM^{-1} cm^{-1} for 3,5,6-trichloropyridinol (the hydrolysis product of CPO); 3 mM^{-1} cm^{-1} for 2-isopropyl-4-methyl-6-hydroxypyrimidine (the hydrolysis product of DZO); or 18 mM^{-1} cm^{-1} for p-nitrophenol (the hydrolysis product of PO). Alkaline phosphatase was assayed in undiluted plasma of all species at 37 °C in triplicate as follows: The plasma sample, 10 µL, was added to 170 µL of 0.95 M diethanolamine pH 9.8, 0.5 mM $MgCl_2$. The assay was initiated by adding 20 µL of 112 mM p-nitrophenyl phosphate in water (*88*). The absorbance at 405 nm was followed for 4 min. Activities were expressed in U/L based on the molar extinction coefficient of 18.0 mM^{-1} cm^{-1} for p-nitrophenol (the hydrolysis product of p-nitrophenyl phosphate). The numeric results of each replicate for each substrate are provided in Table S13.

Phenyl acetate (CAS 122-79-2, 99% purity), p-nitrophenyl phosphate (CAS 333338-18-4, ≥97% purity), and other reagent chemicals were purchased from Sigma-Aldrich. Chlorpyrifos oxon (CPO; CAS 5598-15-2; 98% purity), diazoxon (DZO; CAS 962-58-3; 99% purity) and paraoxon (PO; CAS 311-45-5, 99% purity) were purchased from Chem Service Inc. (West Chester, PA).

Visualizing proximity of agricultural land to manatee habitat

We created maps using QGIS software (version 2.10.1-Pisa). We acquired federally mandated manatee protection zone information from the U.S. Fish and Wildlife Service's Environmental Conservation Online System (*89*). State mandated manatee protection zone information was obtained from the Florida Fish and Wildlife Conservation Commission (*90*). The U.S. Census Bureau TIGER products supplied datasets for the state and county boundaries and Brevard County rivers, canals, lakes, and other waterways (*91–94*).

We obtained agricultural land use data from the Florida Department of Environmental Protection, extracting level 1 land use code 2000 ("agriculture") data and further extracting agricultural land use of specific interest to a geographic exploration of the potential intersection between organophosphate pesticide application and manatee migration areas through the level 2 land use code and description, including cropland and pastureland (level 2 code 2100), nurseries and vineyards (2400), other open lands (2600), and tree crops (2200). This subset discarded agricultural land where organophosphate pesticides are less likely to be applied, including poultry, cattle, and other feeding operations (level 2 code 2300) and specialty farms (2500), which entail horse farms, wet prairies, dairies, aquaculture, tropical fish farms, sewage treatment, and other treatment ponds (*95*).

We created a map focused on Brevard County because Brevard County has been identified as a key area for manatee residence and migration, with an estimated 70% of manatees along the Atlantic coast migrating through or residing in Brevard waterways at least seasonally (*27*, *28*) (Fig. 3). To illustrate the potential for manatee interaction with pesticide water pollution from agricultural areas, we included agriculture land use of interest as described above and all waterways in Brevard County, including lakes, rivers, and canals.

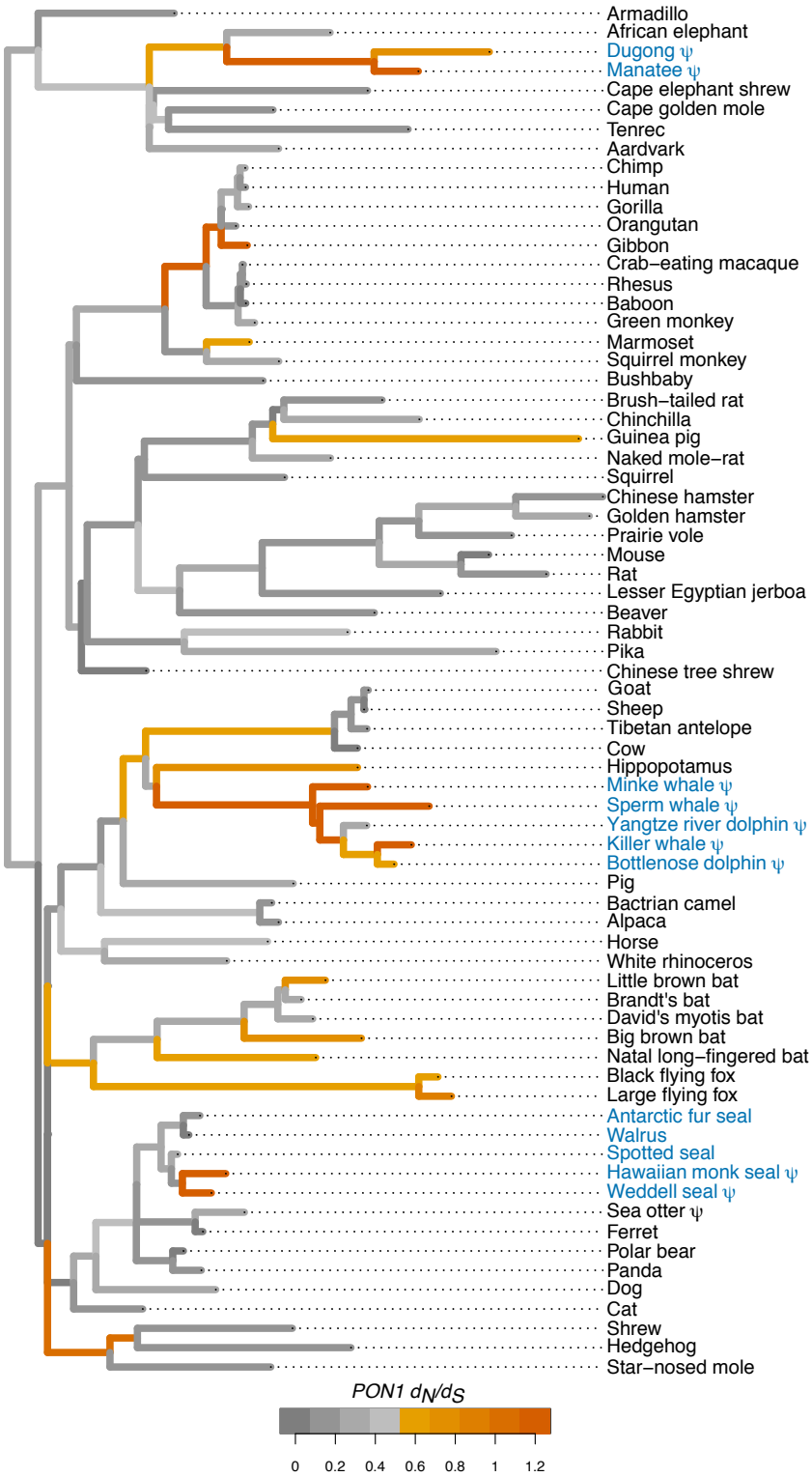

Fig. S1. Evolutionary rate of *PON1* coding sequence across the full mammalian phylogeny. Shown is the phylogeny of 71 eutherian mammals whose sequences were included for rate estimation. See Fig. 1C legend for details.

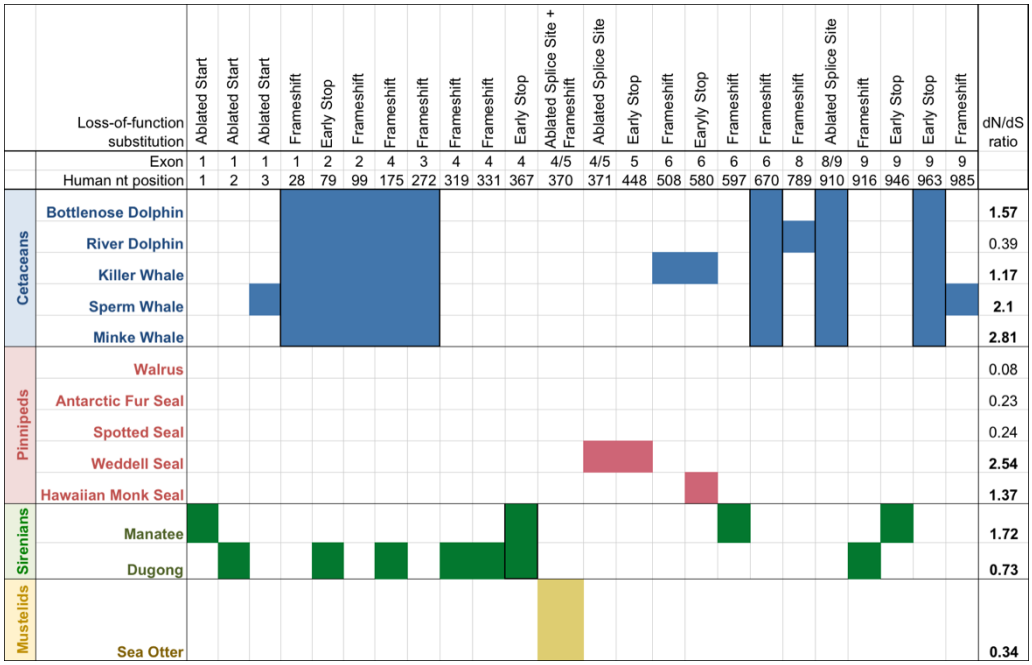

Fig. S2. Loss-of-function substitutions in PON1 across marine and semi-aquatic mammals. Colored cells indicate the observation of the relevant loss-of-function substitution (lesion) in the relevant species. Columns or sets of columns representing lesions shared across all members of a clade are outlined with wider black borders.

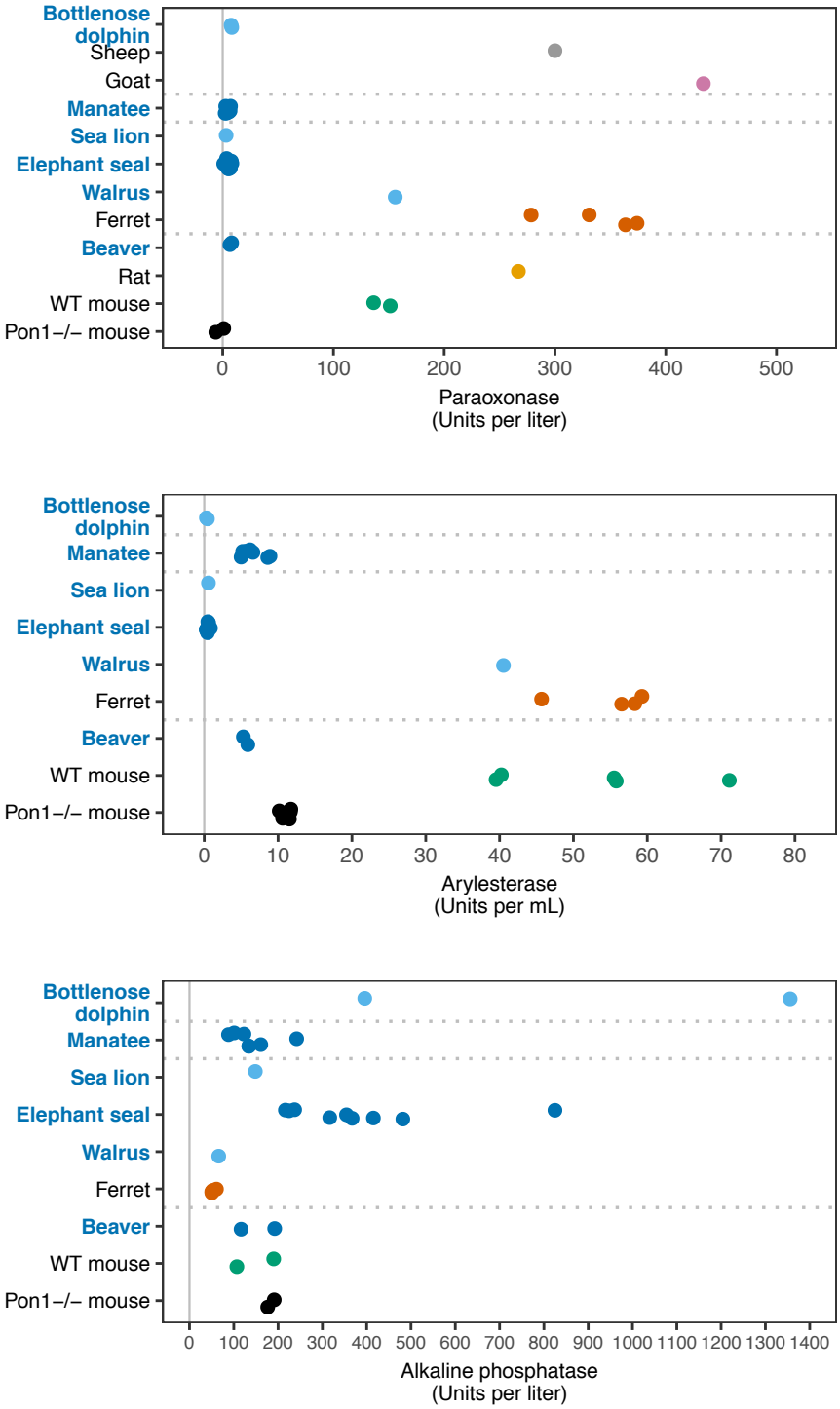

Fig. S3. Blood plasma enzymatic activity against two PON1 substrates and a non-PON1 substrate (control). Points represent rates of hydrolysis of paraoxon in μmol/min/L (top), phenyl acetate in μmol/min/mL (middle), and non-PON1 substrate alkaline phosphatase in μmol/min/L (bottom). In mouse (and potentially other species), carboxylesterase can also hydrolyze phenyl acetate (in addition to PON1). See Fig. 2 legend for additional details.

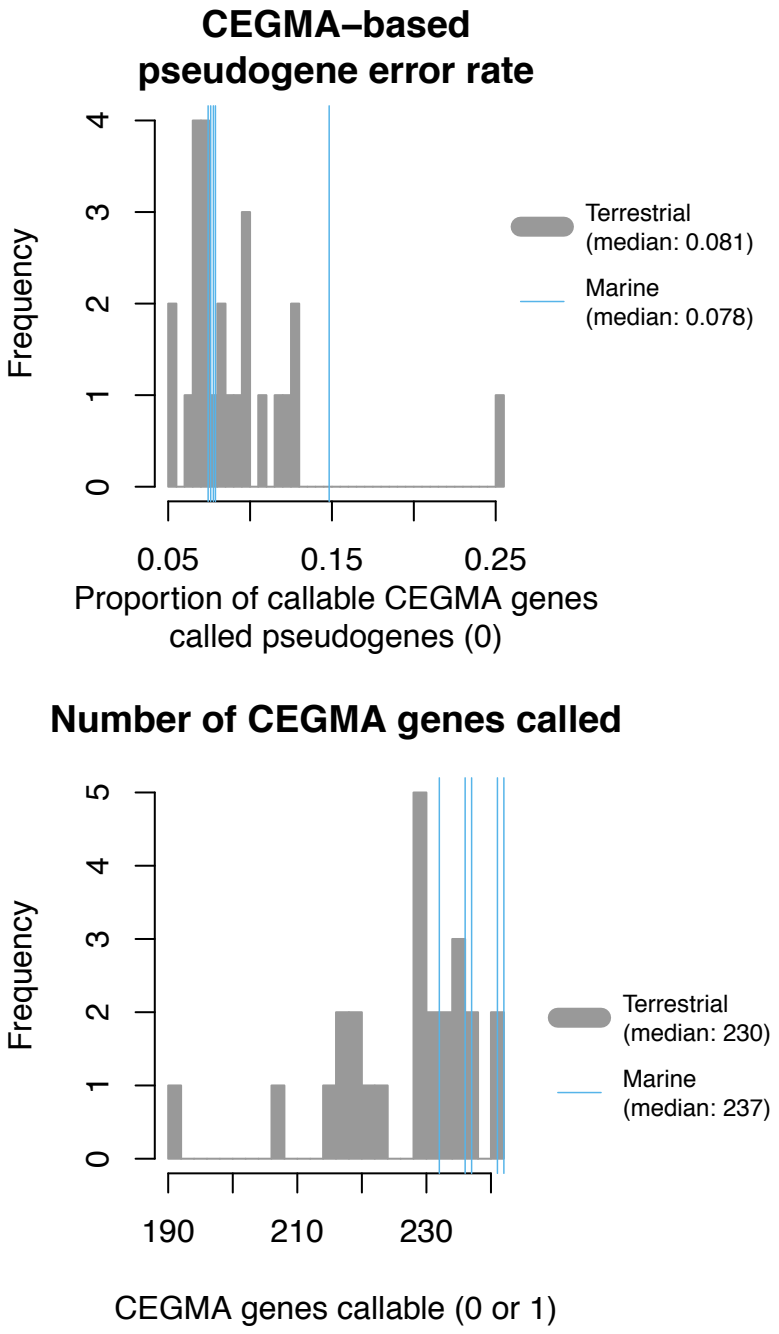

Fig. S4. Per-species pseudogene error rate as estimated from pseudogene calls among highly conserved (CEGMA) genes. Histograms of the proportion of all callable highly conserved genes in the CEGMA (*38*) dataset called as pseudogenes by our automated method for all marine species and the 25 terrestrial species whose genetic distance to reference sequence hg19 was within 0.04 of the marine range (marine range 0.220 – 0.247; terrestrial range 0.187 – 0.284) (top), and the number of CEGMA genes considered callable (not filtered for missing data) out of the 246 genes with alignments in our dataset (bottom).

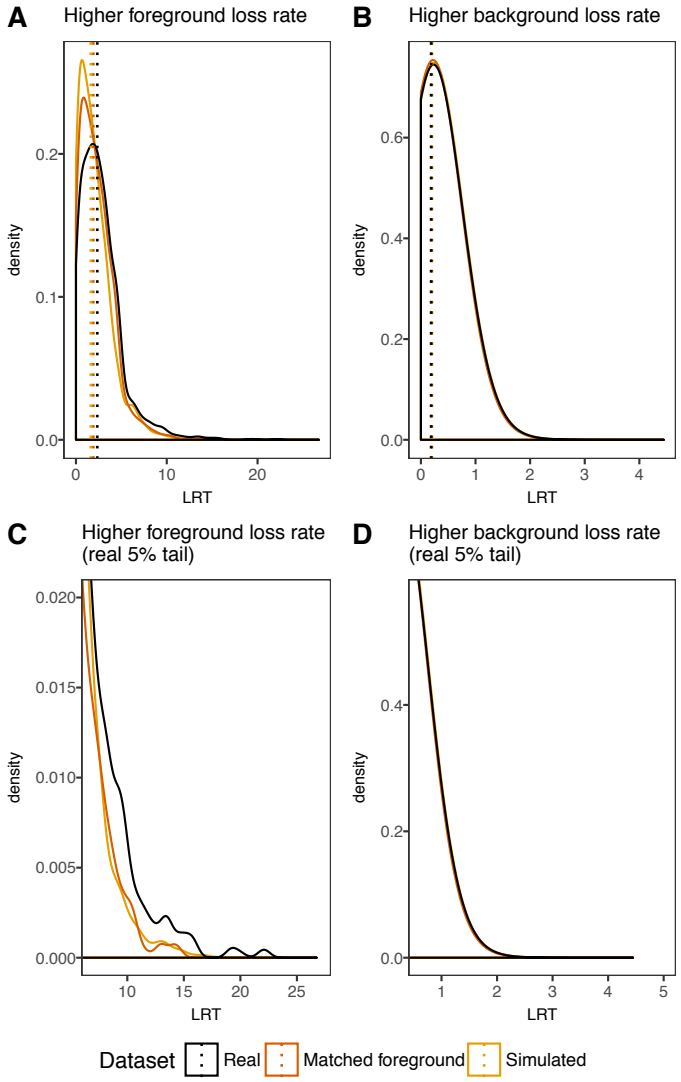

Fig. S5. Distribution of raw LRT statistics in real datasets, compared with those from simulations and a matched foreground set. Plots represent the densities of raw LRT statistics reflecting the significance of a model with different rates of gene functional loss on marine and terrestrial branches, as compared to a model where rates of loss are independent of marine/terrestrial status, for real data (in black), for a test using a set of matched foreground species (in orange), and for 10,000 simulations of each gene under the independent model (in gold). The distribution of the LRT statistic for genes with higher inferred loss rates on foreground (marine in the real dataset) branches (A and C) shows a skew towards higher values in real data as compared with the matched foreground set or simulated data; the upper 5% tail for real data corresponds to 2.64% and 2.65% of the cumulative distribution of matched foreground and simulation sets, respectively (C). This indicates a potential genome-wide signal of preferential gene loss in marine species. In contrast, the distribution of LRT statistics for genes with higher inferred loss rates on background (terrestrial in the real dataset) branches (B and D) is very similar across all three datasets; the upper 5% tail for real data corresponds to 6.19%

and 6.33% of the cumulative distribution of matched foreground and simulation sets, respectively (D). Note the large difference in the ranges of the x-axes.

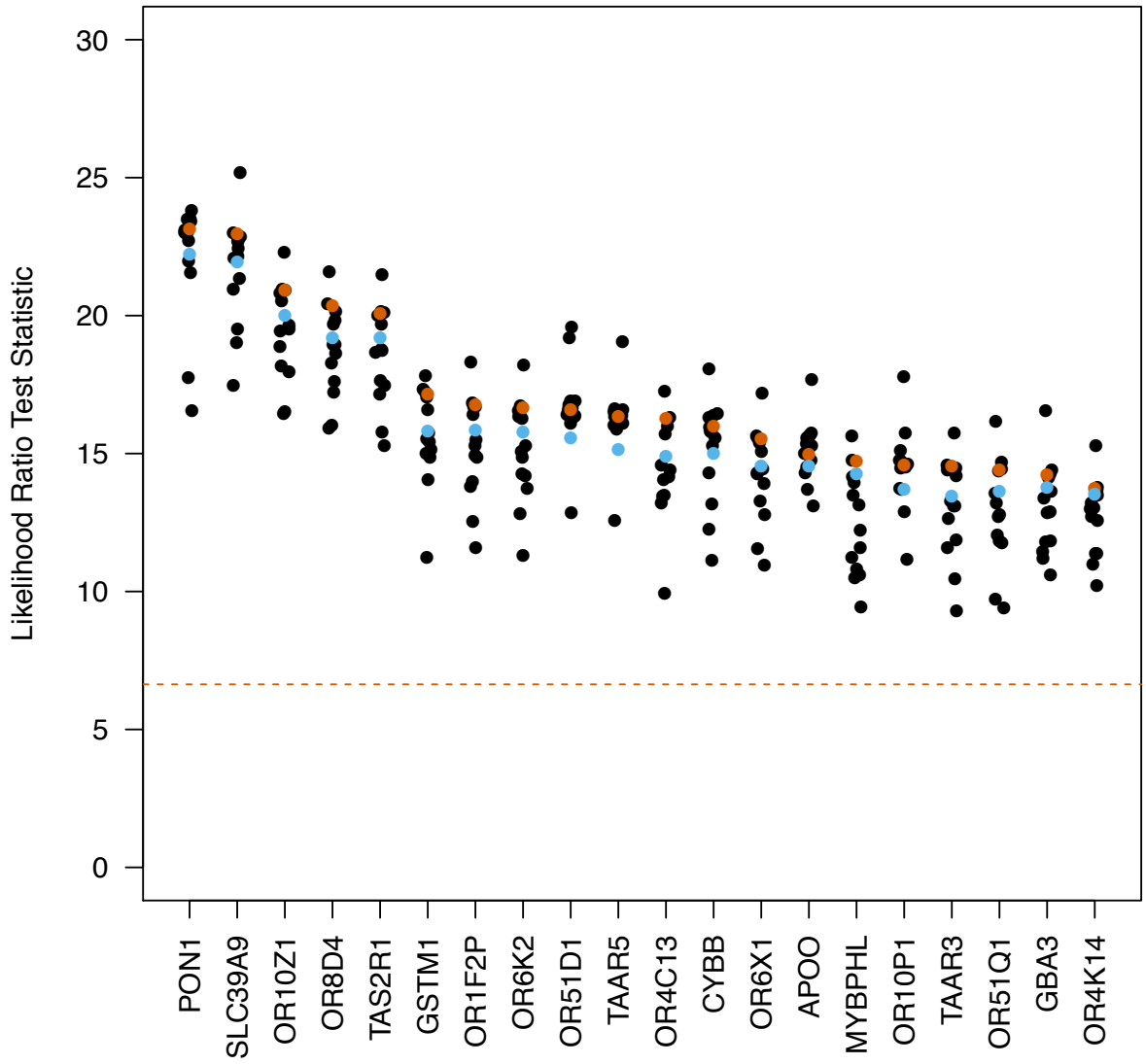

Fig. S6. Evidence for convergence of functional loss in marine species is consistent across varying phylogenetic trees for top genes. Points represent the LRT statistics estimated using the phylogenetic tree used in the main analysis (in blue) and 14 alternate phylogenetic trees, including one re-estimated from a sampled subset of genes but with the same topology as the tree used in the main analysis (in orange). The dashed orange line represents the value of the LRT statistic corresponding to a *P*-value of 0.01 for a chi-square test with one degree of freedom.

Table S2. Gene ontology enrichment for top genes lost in marine lineages. Categories displaying enrichment among genes with strong evidence for higher rate of loss in marine lineages. Gene set enrichment was assessed using the ranked list approach in the GOrilla online enrichment tool (*46*), with all genes included in the BayesTraits analysis as the background gene set. For lists of genes in each category driving the signatures of enrichment, see Table S4.

BIOLOGICAL PROCESS								
GO Term	**Description**	**P-value**	**FDR q-value**	**Enrichment**	**N**	**B**	**n**	**b**
GO:0050907	detection of chemical stimulus involved in sensory perception	3.65E-68	4.78E-64	8.59	9430	285	393	102
GO:0009593	detection of chemical stimulus	4.36E-66	2.85E-62	8.13	9430	304	393	103
GO:0050906	detection of stimulus involved in sensory perception	2.76E-63	1.20E-59	7.79	9430	314	393	102
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	1.54E-58	5.03E-55	8.4	9430	257	393	90
GO:0051606	detection of stimulus	5.11E-57	1.34E-53	6.59	9430	391	384	105
GO:0007186	G-protein coupled receptor signaling pathway	4.12E-41	8.98E-38	4.02	9430	722	393	121
GO:0007606	sensory perception of chemical stimulus	5.08E-16	9.49E-13	8.83	9430	101	275	26
GO:0007608	sensory perception of smell	2.20E-14	3.60E-11	2.72	9430	80	2380	55
GO:0050896	response to stimulus	2.74E-13	3.99E-10	1.55	9430	2803	472	218
GO:0007165	signal transduction	3.70E-12	4.84E-09	1.59	9430	2599	432	189
GO:0050912	detection of chemical stimulus involved in sensory perception of taste	7.53E-10	8.96E-07	12.04	9430	25	376	12
GO:0001580	detection of chemical stimulus involved in sensory perception of bitter taste	4.58E-09	5.00E-06	11.99	9430	23	376	11
GO:0007600	sensory perception	9.46E-08	9.53E-05	4.63	9430	302	135	20
GO:0050877	nervous system process	1.22E-06	1.14E-03	3.05	9430	486	178	28
GO:0042178	xenobiotic catabolic process	1.59E-06	1.39E-03	45.17	9430	5	167	4
GO:0070458	cellular detoxification of nitrogen compound	9.45E-06	7.73E-03	449.05	9430	2	21	2
GO:0051410	detoxification of nitrogen compound	9.45E-06	7.27E-03	449.05	9430	2	21	2
GO:0018916	nitrobenzene metabolic process	1.05E-05	7.65E-03	56.47	9430	3	167	3
GO:0021798	forebrain dorsal/ventral pattern formation	2.54E-05	1.75E-02	23.21	9430	5	325	4
GO:0003008	system process	3.59E-05	2.35E-02	2.35	9430	744	178	33
FUNCTION								
GO Term	**Description**	**P-value**	**FDR q-value**	**Enrichment**	**N**	**B**	**n**	**b**
GO:0004984	olfactory receptor activity	1.54E-58	5.64E-55	8.4	9430	257	393	90
GO:0004930	G-protein coupled receptor activity	8.01E-56	1.47E-52	5.73	9430	493	384	115
GO:0099600	transmembrane receptor activity	2.61E-41	3.18E-38	4.03	9430	737	384	121
GO:0004888	transmembrane signaling receptor activity	2.61E-41	2.39E-38	4.03	9430	737	384	121
GO:0038023	signaling receptor activity	2.36E-39	1.73E-36	3.78	9430	806	384	124
GO:0004872	receptor activity	2.48E-34	1.52E-31	3.2	9430	915	431	134
GO:0004871	signal transducer activity	3.12E-34	1.64E-31	3.08	9430	991	432	140
GO:0060089	molecular transducer activity	1.77E-33	8.10E-31	3.15	9430	931	431	134
GO:0005549	odorant binding	1.61E-14	6.57E-12	8.59	9430	61	414	23
GO:0008527	taste receptor activity	1.74E-11	6.37E-09	15.05	9430	20	376	12
GO:0033038	bitter taste receptor activity	6.39E-10	2.13E-07	15.67	9430	16	376	10
GO:0004063	aryldialkylphosphatase activity	1.06E-04	3.24E-02	9,430.00	9430	1	1	1
GO:0016765	transferase activity, transferring alkyl or aryl (other than methyl) groups	1.14E-04	3.21E-02	4.06	9430	26	1073	12
COMPONENT								
GO Term	**Description**	**P-value**	**FDR q-value**	**Enrichment**	**N**	**B**	**n**	**b**
GO:0031224	intrinsic component of membrane	2.10E-18	3.50E-15	1.83	9430	2683	362	188
GO:0016021	integral component of membrane	5.88E-18	4.91E-15	1.83	9430	2616	362	184
GO:0044425	membrane part	2.89E-13	1.61E-10	1.58	9430	3446	363	209
GO:0005886	plasma membrane	1.17E-12	4.90E-10	1.64	9430	2443	472	201
GO:0016020	membrane	4.29E-09	1.43E-06	1.32	9430	4139	473	275

Table S3. Additional gene set enrichment for top genes lost in marine lineages. This table shows categories displaying enrichment among the top 137 genes with evidence for higher rate of loss in marine lineages, corresponding to a 25% FDR. Gene set enrichment was assessed for the canonical pathways, curated pathways, and GO biological process datasets in the Molecular Signatures Database (mSigDB) (*96*) and the Mouse Genome Database (*97*) using a hypergeometric test, with all genes included in the BayesTraits analysis as the background gene set. No sets of at least three genes were found to be significant at $q < 0.05$ in the Mouse Genome Database. For lists of genes in each category driving the signatures of enrichment, see Table S5.

Canonical Pathways at mSigDB							
Category	**P-value**	**FDR q-value**	**Enrichment**	**number in background**	**fraction in background**	**number in foreground**	**fraction in foreground**
Reactome olfactory signaling pathway	1.63E-26	7.51E-25	9.3246	213	0.0475	35	0.443
KEGG olfactory transduction	1.11E-23	2.56E-22	7.7584	256	0.0571	35	0.443
Reactome signaling by GPCR	6.73E-20	1.03E-18	4.61	517	0.1153	42	0.5316
Reactome GPCR downstream signaling	9.25E-19	1.06E-17	4.7697	464	0.1035	39	0.4937
KEGG taste transduction	3.43E-03	3.16E-02	6.3052	36	0.008	4	0.0506
Curated Pathways at mSigDB							
Category	**P-value**	**FDR q-value**	**Enrichment**	**number in background**	**fraction in background**	**number in foreground**	**fraction in foreground**
Reactome olfactory signaling pathway	2.11E-29	6.95E-27	12.4883	213	0.023	35	0.2869
KEGG olfactory transduction	1.48E-26	2.44E-24	10.3906	256	0.0276	35	0.2869
Reactome signaling by GPCR	6.87E-23	7.54E-21	6.1741	517	0.0558	42	0.3443
Reactome GPCR downstream signaling	1.06E-21	8.69E-20	6.3879	464	0.05	39	0.3197
Kondo prostate cancer with H3K27Me3	2.77E-04	1.82E-02	5.4845	97	0.0105	7	0.0574
GO Biological Process at mSigDB							
Category	**P-value**	**FDR q-value**	**Enrichment**	**number in background**	**fraction in background**	**number in foreground**	**fraction in foreground**
Sensory perception of chemical stimulus	7.29E-04	3.93E-02	16.2336	19	0.0058	3	0.0938
Regulation of action potential	2.47E-03	3.93E-02	25.7031	8	0.0024	2	0.0625
Regulation of axonogenesis	2.47E-03	3.93E-02	25.7031	8	0.0024	2	0.0625
Regulation of neurogenesis	3.93E-03	3.93E-02	20.5625	10	0.003	2	0.0625
Sensory perception of taste	3.93E-03	3.93E-02	20.5625	10	0.003	2	0.0625
Neurological system process	4.11E-03	3.93E-02	3.3013	218	0.0663	7	0.2188
Response to chemical stimulus	4.17E-03	3.93E-02	3.7845	163	0.0495	6	0.1875

Table S6. Gene ontology enrichment for top genes that co-evolve with *PON1*. Gene set enrichment was performed with the 100 genes with the highest Evolutionary Rate Covariation (ERC) with *PON1* using the GOrilla online enrichment tool (*46*), with all genes included in the ERC analysis as the background set. For lists of genes in each category driving the signatures of enrichment, see Table S7.

BIOLOGICAL PROCESS								
GO Term	**Description**	**P-value**	**FDR q-value**	**Enrichment**	**N**	**B**	**n**	**b**
GO:0006629	lipid metabolic process	3.15E-10	4.68E-06	4.12	16491	1108	94	26
GO:0006631	fatty acid metabolic process	3.52E-09	2.61E-05	8.57	16491	266	94	13
GO:0044255	cellular lipid metabolic process	9.11E-08	4.51E-04	3.98	16491	881	94	20
GO:0019752	carboxylic acid metabolic process	9.65E-08	3.58E-04	4.18	16491	798	94	19
GO:0032787	monocarboxylic acid metabolic process	1.54E-07	4.58E-04	5.67	16491	433	94	14
GO:0016042	lipid catabolic process	1.78E-07	4.40E-04	7.72	16491	250	94	11
GO:0006082	organic acid metabolic process	1.80E-07	3.83E-04	3.82	16491	919	94	20
GO:0044712	single-organism catabolic process	4.23E-07	7.86E-04	4	16491	790	94	18
GO:0006635	fatty acid beta-oxidation	6.05E-07	9.97E-04	19.49	16491	54	94	6
GO:0016054	organic acid catabolic process	6.09E-07	9.03E-04	7.8	16491	225	94	10
GO:0046395	carboxylic acid catabolic process	6.09E-07	8.21E-04	7.8	16491	225	94	10
GO:0043436	oxoacid metabolic process	6.46E-07	7.99E-04	3.69	16491	903	94	19
GO:0044710	single-organism metabolic process	3.34E-06	3.81E-03	2.04	16491	3264	94	38
GO:0019395	fatty acid oxidation	3.65E-06	3.87E-03	14.42	16491	73	94	6
GO:0034440	lipid oxidation	3.96E-06	3.92E-03	14.22	16491	74	94	6
GO:0009062	fatty acid catabolic process	6.72E-06	6.23E-03	13	16491	81	94	6
GO:0044281	small molecule metabolic process	7.28E-06	6.36E-03	2.63	16491	1599	94	24
GO:0055114	oxidation-reduction process	1.87E-05	1.54E-02	3.35	16491	837	94	16
GO:0072329	monocarboxylic acid catabolic process	2.26E-05	1.77E-02	10.53	16491	100	94	6
GO:0044282	small molecule catabolic process	2.49E-05	1.85E-02	5.13	16491	342	94	10
GO:0044242	cellular lipid catabolic process	3.44E-05	2.43E-02	7.72	16491	159	94	7
FUNCTION								
GO Term	**Description**	**P-value**	**FDR q-value**	**Enrichment**	**N**	**B**	**n**	**b**
GO:0003995	acyl-CoA dehydrogenase activity	3.17E-07	1.41E-03	63.79	16491	11	94	4
GO:0016491	oxidoreductase activity	6.65E-06	1.48E-02	3.86	16491	681	94	15

Table S8. Annotated functional amino acid positions in PON1 with observed substitutions in marine or semi-aquatic lineages. Sites with substitutions in marine or semi-aquatic lineages that have been annotated from previous functional studies as important to PON1 function in one or more of the following categories: sites crucial for catalytic activity, sites within the wall of the active site, and sites that abrogate activity when experimentally substituted. Amino acid positions are based on human (hg19) sequence within alignment and may differ from those reported in the literature.

				Substitution observed in:					
Amino Acid	**Importance**	**Reference**	**Observed Substitution**	**Cetaceans**	**Sirenians**	**Pinnipeds**	**Bats**	**Semi-aquatic**	**Conservation across all other lineages**
Sites Crucial for Catalytic Activity									
E53	Binds to Catalytic Calcium	Harel et al.(*98*)	E53K	Sperm whale					100%
D169	Binds to Structural Calcium	Harel et al.(*98*)	D169G	All		Hawaiian monk seal			100%
N270	Binds to Catalytic Calcium	Harel et al. (*98*)	N270T					Sea otter	100%
Sites in the Active Site Wall (Substrate Specificity)									
I74	Active Site Wall (Phosphotriesters); Mutant Showed >20x Decrease in Phenyl Acetate and Paraoxon Catalytic Efficiency (I74A)	Harel et al.(*98*); Ben-David et al.(*99*)	I74M* I74F†			*Antarctic fur seal, Walrus	†David's Myotis bat, Black flying fox, Large flying fox		85%
H184	Active Site Wall; Mutants have Undetectable Paraoxon and Phenylacetate Activity (H184A/D/Y)	Harel et al.(*98*); Yeung et al.(*100*)	H184L					Beaver	100%
R192	Active Site Wall	Harel et al. (*98*)	R192K* R192S†		*All		*All except Big brown bat and David's myotis bat †David's myotis bat		85%
F222	Active Site Wall (Aryl Esters)	Harel et al. (*98*)	F222L					Sea otter	99%
F292	Active Site Wall (Aryl Esters); Mutant has 2% WT Phosphotriester Activity (F292A)	Harel et al.(*98*)	F292L			Antarctic fur seal, Walrus			100%
T332	Active Site Wall	Harel et al.(*98*)	T332M* T332S† T332A‡	*Yangtze River dolphin			†Big brown bat, Little brown bat	‡Sea otter	93%
Sites that Abrogate Activity When Mutated									
C42	Disulfide Velcro	Harel et al.(*98*)	C42R			Weddell seal			100%
W194	Mutants 30-50% Activity (W194A)	Josse et al. (*101*)	W194X			Hawaiian monk seal			100%
W202	Mutants 30-50% Activity (W201A)	Josse et al.(*102*)	W202C* W202L†	*All			†Big brown bat		100%
H243	Mutants <1% Activity (H243N)	Josse et al.(*102*)	H243R* H243Q†	*Minke whale	†All		*Natal long-fingered bat		100%
H246	Mutants 30-50% Activity (H245N)	Josse et al.(*102*)	H246R* H246C†		*Manatee, †Dugong		*Black flying fox, Large flying fox		100%
C284	Core Stability	Harel et al.(*98*)	C284R	All					100%
V304	Mutants No Detectable Arylesterase or Paraoxonase Activity (V304A)	Yeung et al. (*103*)	V304M					Sea otter	100%

Table S11. Abbreviations, common names, and sources for species included in *PON1* phylogenetic trees.

Abbreviation	Common_Name	Source
ailMel1	Panda	UCSC 100-way vertebrate alignment
arcGaz	Antarctic fur seal	Mapped Weddell seal exons to assembly from Dryad (doi:10.5061/dryad.599f2)
bosTau7	Cow	UCSC 100-way vertebrate alignment
brandtBat	Brandt's bat	NCBI annotated genome assembly (accession GCF_000412655.1)
calJac3	Marmoset	UCSC 100-way vertebrate alignment
camFer1	Bactrian camel	UCSC 100-way vertebrate alignment
canFam3	Dog	UCSC 100-way vertebrate alignment
capHir1	Goat	UCSC 100-way vertebrate alignment
casCan	Beaver	NCBI annotated genome assembly (accession GCA_001984765.1)
cavPor3	Guinea pig	UCSC 100-way vertebrate alignment
cerSim1	White rhinoceros	UCSC 100-way vertebrate alignment
chiLan1	Chinchilla	UCSC 100-way vertebrate alignment
chlSab1	Green monkey	UCSC 100-way vertebrate alignment
chrAsi1	Cape golden mole	UCSC 100-way vertebrate alignment
conCri1	Star-nosed mole	UCSC 100-way vertebrate alignment
criGri1	Chinese hamster	UCSC 100-way vertebrate alignment
dasNov3	Armadillo	UCSC 100-way vertebrate alignment
dugDug	Dugong	Sequencing from this study
echTel2	Tenrec	UCSC 100-way vertebrate alignment
eleEdw1	Cape elephant shrew	UCSC 100-way vertebrate alignment
enhLut	Sea otter	NCBI annotated genome assembly (accession GCA_002288905.2)
eptFus1	Big brown bat	UCSC 100-way vertebrate alignment
equCab2	Horse	UCSC 100-way vertebrate alignment
eriEur2	Hedgehog	UCSC 100-way vertebrate alignment
felCat5	Cat	UCSC 100-way vertebrate alignment
gorGor3	Gorilla	UCSC 100-way vertebrate alignment
Hawaii	Hawaiian monk seal	NCBI annotated genome assembly (accession GCA_002201575.1)
hetGla2	Naked mole-rat	UCSC 100-way vertebrate alignment
hg19	Human	UCSC 100-way vertebrate alignment
hippo	Hippopotamus	Mapped RNA-seq reads (accession SRX1164570) to dolphin
jacJac1	Lesser Egyptian jerboa	UCSC 100-way vertebrate alignment
largha	Spotted seal	Mapped RNA-seq reads (accession SRX120902) to Weddell seal
lepWed1	Weddell seal	UCSC 100-way vertebrate alignment
lipVex	Yangtze river dolphin	NCBI annotated genome assembly (accession GCA_000442215.1)
loxAfr3	African elephant	UCSC 100-way vertebrate alignment
macFas5	Crab-eating macaque	UCSC 100-way vertebrate alignment
mesAur1	Golden hamster	UCSC 100-way vertebrate alignment
micOch1	Prairie vole	UCSC 100-way vertebrate alignment
mini	Natal long-fingered bat	NCBI annotated genome assembly (accession GCF_001595765.1)
Minke	Minke whale	NCBI annotated genome assembly (accession GCA_000493695.1)
mm10	Mouse	UCSC 100-way vertebrate alignment
musFur1	Ferret	UCSC 100-way vertebrate alignment
myoDav1	David's Myotis bat	NCBI annotated genome assembly (accession GCA_000327345.1)
myoLuc2	Little brown bat	UCSC 100-way vertebrate alignment
nomLeu3	Gibbon	UCSC 100-way vertebrate alignment
ochPri3	Pika	UCSC 100-way vertebrate alignment
octDeg1	Brush-tailed rat	UCSC 100-way vertebrate alignment
odoRosDi	Walrus	UCSC 100-way vertebrate alignment
orcOrc1	Killer whale	UCSC 100-way vertebrate alignment
oryAfe1	Aardvark	UCSC 100-way vertebrate alignment
oryCun2	Rabbit	UCSC 100-way vertebrate alignment
otoGar3	Bushbaby	UCSC 100-way vertebrate alignment
oviAri3	Sheep	UCSC 100-way vertebrate alignment
panHod1	Tibetan antelope	UCSC 100-way vertebrate alignment
panTro4	Chimp	UCSC 100-way vertebrate alignment
papHam1	Baboon	UCSC 100-way vertebrate alignment
phyCat	Sperm whale	NCBI annotated genome assembly (accession GCA_000472045.1)
ponAbe2	Orangutan	UCSC 100-way vertebrate alignment
pteAle1	Black flying fox	UCSC 100-way vertebrate alignment
pteVam1	Large flying fox	UCSC 100-way vertebrate alignment
rheMac3	Rhesus	UCSC 100-way vertebrate alignment
rn5	Rat	UCSC 100-way vertebrate alignment
saiBol1	Squirrel monkey	UCSC 100-way vertebrate alignment
sorAra2	Shrew	UCSC 100-way vertebrate alignment
speTri2	Squirrel	UCSC 100-way vertebrate alignment
susScr3	Pig	UCSC 100-way vertebrate alignment
triMan1	Manatee	UCSC 100-way vertebrate alignment
tupChi1	Chinese tree shrew	UCSC 100-way vertebrate alignment
turTru2	Bottlenose dolphin	UCSC 100-way vertebrate alignment
ursMar1	Polar bear	NCBI annotated genome assembly (accession GCA_000687225.1)
vicPac2	Alpaca	UCSC 100-way vertebrate alignment

Table S12. Primers used to amplify and sequence *PON1* exons in manatee and dugong DNA samples. Except where noted, the same primers were used for both PCR and sequencing.

Primer	Sequence	Paired With	PON1 Exon	Species	PCR Annealing Temp (° C)	PCR Cycles
TML1F	ACAGCTTTCCCTTCCTTGC	TML1R	1	Both	59	34
TML1R	AGCCTGGGTCCTCCTTTCT	TML1F	1	Both	59	34
DdE2F5	CAGGTTTCTGGAACCACCTC	DdE2R5	2	Dugong	57.5	34
DdE2R5	TGAGCTACTCACTCCTCTCACAA	DdE2F5	2	Dugong	57.5	34
DdE3F2	TGAATTTCCATGAGCTTTATGTG	DdE3R2	3	Dugong	59	34
DdE3R2	CAGTTGAATGGGAAGCCACT	DdE3F2	3	Dugong	59	34
TML2F	GCCAGGAGACTTCCTGTGTG	TML2R	4	Manatee	59	34
TML2R	CCATAAAGATTAGGGCTGCAT	TML2F	4	Manatee	59	34
DdE4F	CAAGGTGAATCCGTGTGCTA	DdE4R	4	Dugong	59	34
DdE4R	GGGAAACTTGAAACCCAGAA	DdE4F	4	Dugong	59	34
DdE5F	AGACAGGGCTGACAGCTGAG	DdE5R	5	Dugong	59	34
DdE5R	TGGATTAGTCATCCTCTGGAA	DdE5F	5	Dugong	59	34
TML3_4F	GTTATGCATTTTGCTCCCAGA	TML3_4R	6	Both	59	34
TML3_4R	GGTTGATATGTTGTGGGGTTGT	TML3_4F	6	Both	59	34
DdL3_4intR	GGAATCTATTATAAAGATATCTAA	TML3_4F/TML3_4R (sequencing only)	6	Dugong	59	34
TML5_7F	TGCACTGCAAGCTCATTCTT	TML5_7R	7	Manatee	59	34
TML5_7R	CGACATCAAATGGAGGAAGG	TML5_7F	7	Manatee	59	34
DdE7F	CTCCACCGTCTCCTTTTGAA	DdE7R	7	Dugong	59	34
DdE7R	CACCCATCCCCATTAGACAA	DdE7F	7	Dugong	59	34
TML8_10F	TCCCATATCTTCCCCCTACC	TML8_10R	8	Both	59	34
TML8_10R	CCCCTAGGAACTCCTCTTGC	TML8_10F	8	Both	59	34
TML11_15F	TTGCCAGCATTTAAACACCA	TML11_15R	9	Manatee	59	34
TML11_15R	AAGGATGGGCTCACAGTTTC	TML11_15F	9	Manatee	59	34
DdE9F	GTGTGCTCACCACCTCTGTTAAA	DdE9R	9	Dugong	59	50
DdE9R	TGATCCCTCATGATGTCCAA	DdE9F	9	Dugong	59	50

Additional Data table S1 (separate file)

Presence/absence/excluded (NA) values for extant species and results of BayesTraits model comparison for gene loss (where applied) for all 9,950 genes included in the analysis and 3,904 genes excluded from analysis but whose loss rates were used for simulations.

Additional Data table S4 (separate file)

GOrilla gene ontology enrichment for top genes lost in marine lineages, with gene lists (see Table S2).

Additional Data table S5 (separate file)

Additional gene set enrichment for top genes lost in marine lineages, with gene lists (see Table S3).

Additional Data table S7 (separate file)

Gene ontology enrichment for top genes that co-evolve with *PON1*, with gene lists (see Table S6).

Additional Data table S9 (separate file)

Results from manual validation of pseudogene calls for 20 top genes, 5 genes representing known cases of pseudogenization, and 20 randomly selected genes, using manual checks of the sequences within the 100-way alignment.

Additional Data table S10 (separate file)

Results from manual validation of pseudogene calls for 20 top genes, 5 genes representing known cases of pseudogenization, and 20 randomly selected genes, using sequences from the reference genomes for all 58 species.

Additional Data table S13 (separate file)

Values from triplicate assays of hydrolysis for four PON1 substrates and alkaline phosphatase control (means plotted in Figs. 2 and S2).