Highly Dynamic Gene Family Evolution Suggests Changing Roles for *PON* Genes Within Metazoa

Sarah A.M. Lucas, Allie M. Graham, Jason S. Presnell, and Nathan L. Clark *

Department of Human Genetics, University of Utah

*Corresponding author. E-mail: nclark@utah.edu.

Accepted: 24 January 2023

Abstract

Change in gene family size has been shown to facilitate adaptation to different selective pressures. This includes gene duplication to increase dosage or diversification of enzymatic substrates and gene deletion due to relaxed selection. We recently found that the *PON1* gene, an enzyme with arylesterase and lactonase activity, was lost repeatedly in different aquatic mammalian lineages, suggesting that the *PON* gene family is responsive to environmental change. We further investigated if these fluctuations in gene family size were restricted to mammals and approximately when this gene family was expanded within mammals. Using 112 metazoan protein models, we explored the evolutionary history of the *PON* family to characterize the dynamic evolution of this gene family. We found that there have been multiple, independent expansion events in tardigrades, cephalochordates, and echinoderms. In addition, there have been partial gene loss events in monotremes and sea cucumbers and what appears to be complete loss in arthropods, urochordates, platyhelminths, ctenophores, and placozoans. In addition, we show the mammalian expansion to three *PON* paralogs occurred in the ancestor of all mammals after the divergence of sauropsida but before the divergence of monotremes from therians. We also provide evidence of a novel *PON* expansion within the brushtail possum. In the face of repeated expansions and deletions in the context of changing environments, we suggest a range of selective pressures, including pathogen infection and mitigation of oxidative damage, are likely influencing the diversification of this dynamic gene family across metazoa.

Key words: gene family evolution, paraoxonase (PON), phylogenetics, brushtail possum, gene loss, gene family expansion.

Significance

While the paraoxonase (PON) enzyme family has been documented to have roles in atherosclerosis, degradation of bacterial quorum sensing molecules, and potentially diving adaption, we still do not know what the enzymes' native substrates are or what selective pressures led to their maintenance and the fixation of expansions or losses in different species. By searching for orthologs in over 100 metazoan species, we discovered this family has changed in gene number more frequently than what was anticipated based on the gene number stability observed in mammals and vertebrates. This paper identifies unique family expansions for which we can next determine the selective pressures that led to their fixations and identify additional roles of this gene family.

Introduction

Gene families are groups of homologous genes found in different species. Identifying members of gene families is of interest as homologous genes frequently have similar or related functions (Pearson 2013). Genes can be related through speciation (orthologs), duplication events (paralogs), whole-genome duplication (ohnolog), horizontal gene transfer (xenolog), or hybridization (homoeolog; Koonin 2005; Glover et al. 2016; Altenhoff et al. 2019). Gene family size changes all throughout metazoa

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/gbe/article/15/2/evad011/7013737 by University of Pittsburgh user on 23 April 2024

(Fernández and Gabaldón 2020) through a variety of modes including, but not limited to, tandem duplication, retroduplication, and segmental duplication (Hahn 2009). In response to different selective pressures, a change in gene family size can become fixed within different species. In the phylogenetic setting, the frequent gain or loss of gene copies allow for examination of convergent selective pressure(s). This paper will examine the phylogenetic history of the paraoxonase (*PON*) gene family.

The *PON* gene family was named based on the discovery that one member could degrade the insecticide parathion, whose active metabolite paraoxon functions as a neurotoxic cholinesterase inhibitor (Costa et al. 1990). In 1996, it was revealed that this enzyme is part of a multigene family in humans (Primo-Parmo et al. 1996), whose genes were named *PON1*, *PON2*, and *PON3* in order of discovery. While these three genes produce protein products commonly known as serum paraoxonases, members of this protein family can also be found elsewhere. *PON1* and *PON3* are predominantly expressed extracellularly in the liver and their proteins are found on high-density lipoprotein (HDL) particles in blood serum. *PON2* is expressed intracellularly in a wide range of tissues and its protein product localizes to the endoplasmic reticulum and nuclear envelope (Ng et al. 2001; Horke et al. 2007). All three genes are in a tandem array on human chromosome 7, have approximately the same length, and contain the same number of exons. In contrast, non-mammal vertebrates like birds were revealed to only have a single *PON* gene. *PON*-like sequences have been identified in several other species including bacteria, nematodes, frogs, and mammals (Draganov and La Du 2004).

The native substrate(s) of PON proteins are still unclear. This makes it challenging to determine what selective pressure(s) resulted in the fixation and continued maintenance of these enzymes in most mammalian species (Billecke et al. 2000; Muthukrishnan et al. 2012). To address this, early physiological studies identified these proteins interact with a wide range of chemical structures. While PON1 hydrolyzed compounds such as paraoxon and lipid peroxides, the only classes of compounds which all three mammalian PONs can act upon are aromatic and aliphatic lactones (cyclic carboxylic esters; Draganov et al. 2005; Bar-Rogovsky et al. 2013), arylesters (aromatic esters; Billecke et al. 2000; Draganov et al. 2005; Khersonsky and Tawfik 2005), and homoserine lactone (HSL) which are key molecules for quorum sensing in bacteria (Draganov et al. 2005; Stoltz et al. 2007; Teiber et al. 2008). Other studies have revealed PON1 has atheroprotective—protection against plaque formation—effects (Shih et al. 1998; Tward et al. 2002), antioxidant properties through the degradation of lipoperoxides (Aviram et al. 1998), and an ability to co-regulate inflammation through an interaction with myeloperoxidase on HDL particles (Huang et al. 2013; Variji et al. 2019). Meanwhile PON2 has also been shown to have atheroprotective effects through its ability to reduce superoxide release (Ng et al. 2001; Horke et al. 2007; Altenhöfer et al. 2010; Devarajan et al. 2011) as well as anti-apoptotic properties (Krüger et al. 2015). PON3 has been associated with several diseases, but its exact functional role in those diseases have yet to be elucidated (Shih et al. 2007; Rull et al. 2012).

From this examination of human PON protein substrates and disease associations, it is highly suggestive this protein family plays two important roles: degrading lactones such as ones used in bacterial quorum sensing and contributing to antioxidant activity against lipoperoxides on HDL particles. However, there is evidence that in multiple independent lineages PON1 has been turned into a pseudogene, and that the loss of function may be adaptive, or the result of a relaxation of constraint in the aquatic environment (Meyer et al. 2018). An examination of the changes in *PON* copy number across metazoa and mammals could provide more clues as to what functions this enzyme provided.

The initial evolutionary study of this family found that *PON2* was the oldest member of this family followed by *PON3* and *PON1* (Draganov and La Du 2004); however, a more recent study challenged this finding. Through the incorporation of additional *PON* sequences, additional studies determined that *PON3* diverged before *PON1* and *PON2* (Bar-Rogovsky et al. 2013). Since those papers were published, several marsupial and monotreme genomes have become available which would allow us a deeper look into the evolutionary history of this gene family in mammals to determine when it expanded in relation to the divergence of the different mammalian lineages. Additionally, both studies were limited in the number and types of genomes available to them. They were not able to investigate if these fluctuations in gene family size are restricted to mammals or how often it changed size in throughout metazoan evolution. With evidence of multiple independent expansions or deletions, we can begin to probe what functions are being selected for or against in this gene family.

Here, we explore the deep evolutionary history of the *PON* genes across 112 metazoan genomes and two choanozoan genomes. Ultimately, we determined that mammalian *PON* expansion occurred before the divergence of monotremata from the ancestor of all extant mammals (i.e., theria). In addition, we investigated the status of *PON* genes in a broad and diverse group of metazoans and found this mammalian expansion was not unique. Lastly, we highlight evidence of new specific duplications of *PON3* in the brushtail possum (*Trichosurus vulpecula*) that were followed by positively selected diversification. Overall, the contractions and expansions of the PON family suggest they are being acted upon by diverse evolutionary pressures such as combating bacterial biofilm formation or managing oxidative stress.

Downloaded from https://academic.oup.com/gbe/article/15/2/evad011/7013737 by University of Pittsburgh user on 23 April 2024

Results

PON Expansion and Contraction has Occurred Multiple Independent Times Within Metazoa

HMMER was used to identify the sequences containing an arylesterase domain. It is the only functional domain found within *PON*s and is unique to this protein family. Across 99 metazoan species and 2 choanozoa, 41 species did not contain a protein with a high confidence arylesterase domain, including species within ctenophora, placozoa, platyhelminth, urochordata, or arthropoda; however, the remaining 60 species within porifera, cnidaria, chordata, ambulacraria, spiralia, and ecdysozoa showed broad evidence of *PON* genes. All 159 resulting sequences were aligned and subjected to phylogenetic analysis. The phylogeny provides evidence of ancient duplications among closely related species (fig. 1, supplementary fig. S1, Supplementary Material online). Within Asteroidea, Echinoidea, Tardigrada, Holothuroidea, and Cephalochordata, there is evidence of ancient duplication(s), a duplication which occurred in the common ancestor of the extant species, which resulted in multiple ancient *PON* genes for each taxonomical group. In addition to ancient duplications, there are instances of more recent PON expansions. This is best demonstrated within the Cephalochordate cluster of PON sequences. Despite evidence of there being three ancient copies of *PON*s in this cluster, each of the three *Branchiostoma* species have accumulated and maintained more than three *PON* genes.

Several different modes of duplication have expanded this gene family. One of the most frequent modes of duplication observed in this tree is tandem duplication. This is easiest to view in the Mammalia, Asteroidea, Tardigrada, and Crinoidea clusters (figs. 1 and 2*A*, supplementary figs. S1, S2, and S4, Supplementary Material online). Another potential mode of duplication observed in this tree is retroduplication. One clade of *PON* genes in Cephalochordata lacks introns that were present in outgroup species, a sign of retroduplication (supplementary table S1, Supplementary Material online). Besides tandem duplication and retroduplication, there are still other modes of gene duplication. This is highlighted best by the bivalve the Great Scallop (*Pecten maximus*). While three of its nine *PON* genes are in a tandem array (i.e., <100 kb away from each other) on one of its chromosomes, the remaining six are scattered among five chromosomes and scaffolds (supplementary table S1, Supplementary Material online) giving evidence to some alternate gene duplication mechanisms such as segmental duplication, rearrangement of tandem duplicates, or horizontal gene transfer.

In addition to instances of *PON* expansion, there is also evidence of *PON* loss. Upon examining multiple species in the same taxon without a *PON* gene, this leads to the possibility of ancient losses of *PON* in the ctenophores, placozoans, urochordates, arthropods, and platyhelminths (fig. 4, supplementary fig. S3, Supplementary Material online). Outside those lineages, there are several parasitic species (i.e., *Soboliphyme baturini, Teladorsagia circumcincta, Hirudo medicinalis,* Myxozoan cnidarians), in which we were unable to identify a single *PON* gene. This was not surprising as species which evolve to become parasitic or symbiotic tend to undergo genome reduction (Wolf and Koonin 2013).

PON Expansion Occurred After Divergence of Mammals From Sauropsida

RefSeq-identified PON protein sequences from 15 tetrapod species were collected (Pruitt et al. 2005). Each of the five placental and five marsupial mammals encoded three PON proteins. Both monotreme mammals as well as the three sauropsid (reptile and birds) species each have a single copy of *PON*. While simple parsimony of gene counts would suggest that the duplications leading to three therian *PON*s occurred after divergence from monotremes, our phylogenetic analysis reveals a strong clustering of three distinct mammalian *PON* clusters (fig. 2*A*). Importantly, the monotreme *PON* genes were placed in the marsupial and placental *PON3* sequences with high bootstrap support (99.5%), instead of falling outside the *PON* gene duplications. This indicates that the mammalian *PON* gene family expanded to at least three members after the divergence of mammals from sauropsida (the ancestor to reptiles and birds) but before divergence of monotremes. This is further bolstered by the evidence of distinct clusters corresponding to the divergence of monotremes, marsupials, and placentals within each of the individual *PON* groups with 100% bootstrap support except for the marsupial *PON3* cluster which has 96% bootstrap support. The lack of additional monotreme *PON* genes strongly suggests that the ancestor of extant monotremes lost its *PON1* and *PON2* genes or the ancestor to these two genes after it diverged from the therian ancestor.

To determine which of the three *PON*s diverged first, we considered three separate models in which each of the mammalian *PON*s diverged before the other two (fig. 2*B*). Using the multiple protein alignment from the first analysis, PAML determined the likelihood that the alignment would support that model. The model with the highest likelihood shows *PON1* diverging first (−8,410.56), although it was not significantly better than the model with *PON3* diverging first (−8,411.31). Thus, we cannot be certain which of those two *PON*s is ancestral to the other ($P = 0.22$); however, it can be stated with statistical significance that *PON2* did not branch off before *PON1* (−8,413.77, −8,410.56, $P = 0.0113$) nor

Downloaded from https://academic.oup.com/gbe/article/15/2/evad011/7013737 by University of Pittsburgh user on 23 April 2024

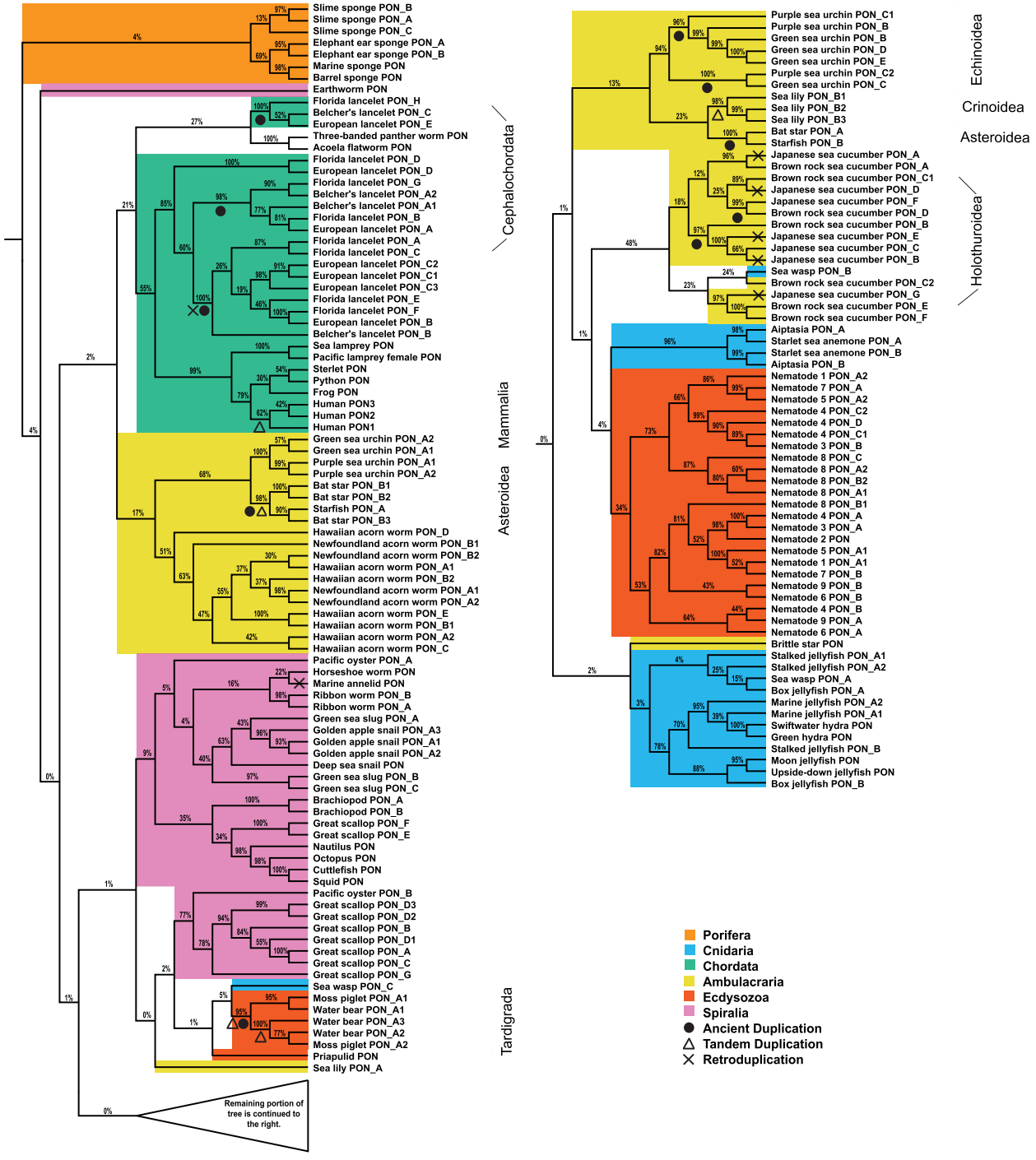


Fig. 1.—Evolution of PON in metazoans. Phylogenetic tree of PON family proteins in metazoans determined by RAxML based on multiple sequence alignment. Bootstrap support values are shown as percentages out of 1,000 bootstraps. If species have multiple *PON* genes and they are located on different chromosome/scaffold or are sufficiently far from one another, then they are indicated by different alphabet characters. If they *PON* genes are located on the same chromosome/scaffold and are within one 100 kb of another *PON* gene, this is indicated by a number. Taxa mentioned in the results section are labeled. Specific nodes highlighting instances ancient, tandem, or retroduplication are indicated by a circle, triangle, or cross, respectively, either underneath or on the branch.

Downloaded from https://academic.oup.com/gbe/article/15/2/evad011/7013737 by University of Pittsburgh user on 23 April 2024

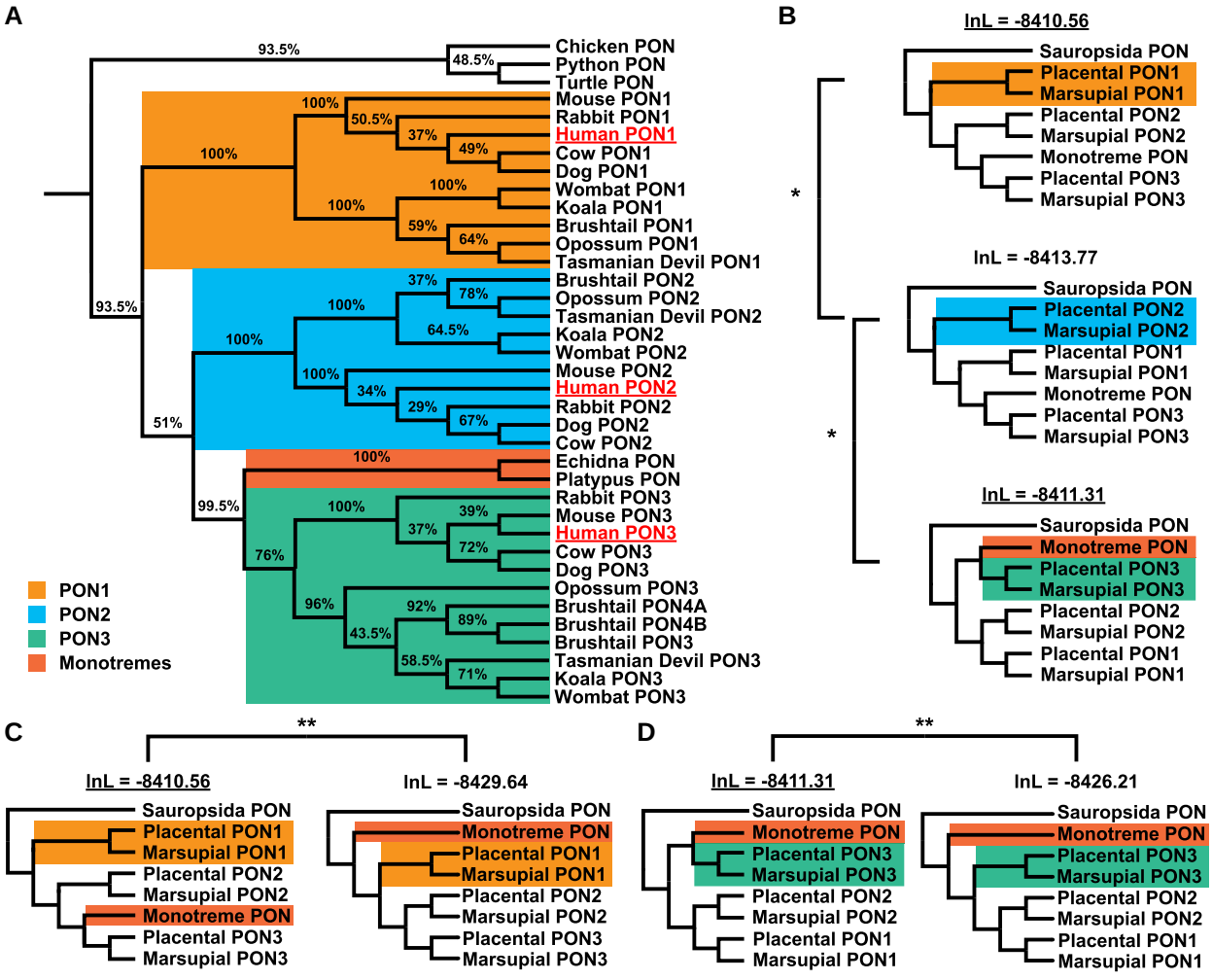


Fig. 2.—Evolution of PON in tetrapods. (*A*) Phylogenetic tree of PON family proteins in tetrapods determined by PhyML based on multiple sequence alignment. Bootstrap support values are shown as percentages out of 200. Brushtail PON4 was split into two separate genes based on RNA-seq evidence. Human sequences were underlined to help orient the reader. (*B*) Phylogenetic tree models represent the scenario in which each of the PONs is ancestral compared with the other two. *$P<0.05$. The underlined log likelihood indicates the significantly better model. (*C*, *D*) Two sets of models compare the placement of the monotremes either basally along with birds and lizards (i.e., sauropsida) or within PON3 with either PON1 being ancestral (*C*) or PON3 being ancestral (*D*). ** $P<$ 1e-8. The underlined log likelihood indicates the significantly better model. Rooted and unrooted trees produced the same log likelihood score.

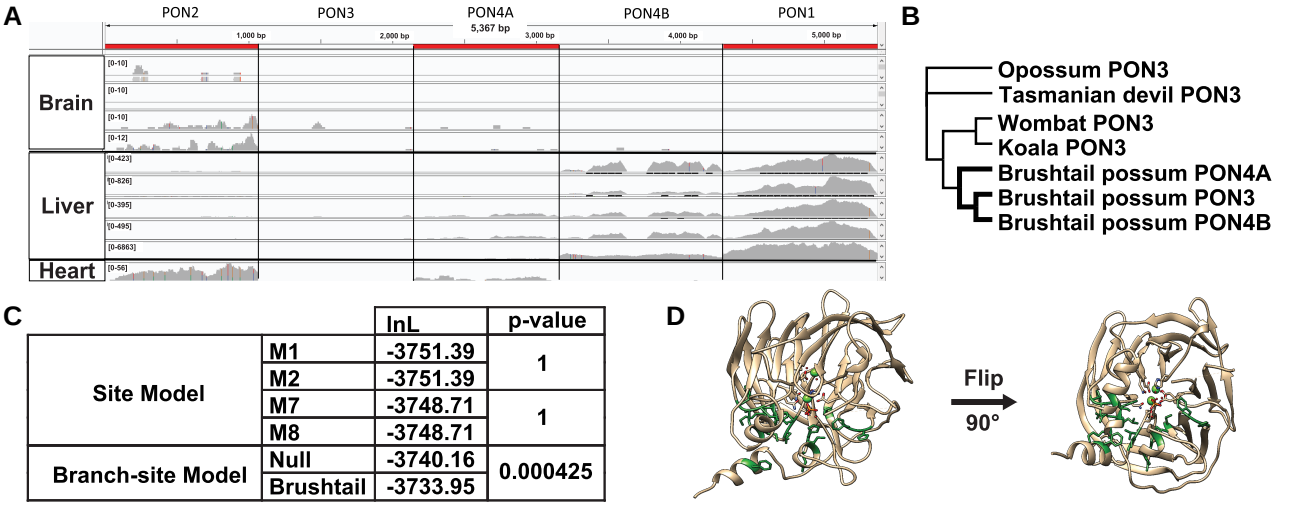
did *PON2* diverge before *PON3* (−8,413.77, −8,411.31, $P=0.0267$; fig. 2*B*).

To confirm that the monotreme *PONs* belong with the ancestral *PON3* group, two additional models were created. Both models clustered the monotremes *PONs* with the sauropsida (birds and reptiles) *PONs* instead of mammalian *PON3*, but one of them had *PON1* diverged first (fig. 2*C*) while the other had *PON3* diverge first (fig. 2*D*). Regardless of whether *PON1* or *PON3* diverged before the other, the models in which the monotremes' *PONs* belong to the mammalian *PON3* group were strongly preferred ($P=6.5\times10^{-10}$ and 4.81×10^{-8}). This indicates they lost either *PON1* and *PON2* separately or the ancestor to both *PON1* and *PON2*.

Case Study: Recent *PON* Expansion Under Positive Selection in Brushtail Possum

In addition to the three *PONs* that were expected to be found in the brushtail possum, a marsupial mammal, BLAST identified another locus (XP_036615431.1) in between the brushtail *PON3* and *PON1*. Upon closer examination this single locus in brushtail contained what could be two new *PON* genes. We therefore divided this locus into two halves, each forming a complete arylesterase domain with typical *PON* gene structure, to form *PON4A* and *PON4B*, since RNA-seq reads show they are transcribed independently (fig. 3*A*). PON4A is comprised of annotated residues 1–337 of XP_036615431.1, and PON4B consists of residues 338–717, although this study suggests the

Downloaded from https://academic.oup.com/gbe/article/15/2/evad011/7013737 by University of Pittsburgh user on 23 April 2024



		lnL	p-value
Site Model	M1	-3751.39	1
	M2	-3751.39	
	M7	-3748.71	1
	M8	-3748.71	
Branch-site Model	Null	-3740.16	0.000425
	Brushtail	-3733.95	

Fig. 3.—Brushtail PON3 and recent PON4*A*/B under positive selection. (*A*) Screenshot from IGV showing a mapped reads to a concatenated transcriptome for all brushtail possum *PON* genes. Tissue samples include four brain, five liver, and one heart sample. The order of the concatenated *PON* genes is listed above the screenshot. Order was chosen to mimic the chromosomal order found in the brushtail genome. (*B*) Phylogenetic tree of marsupial *PON3* genes used for PAML and BUSTED analysis shown. The bolded branches indicate the foreground sequences tested for positive selection. (*C*) Log likelihood values were determined by PAML. The marsupial tree in figure 3*B* was used in the marsupial analysis (*D*) 3*D* protein image of rabbit PON1 (PDB:1V04). Corresponding sites under positive selection in brushtail are highlighted and show their atomic structure. Residues which are part of the active site also show their atomic structure.

gene model should be updated as two separate genes. Based on sequence similarity, both *PON4A* and *PON4B* seemed to be expansions from the brushtail PON3 gene (fig. 2*A*).

To verify that *PON4A* and *PON4B* were real and not the result of an assembly error, brushtail possum RNA-seq reads were mapped to the five *PON* mRNA sequences (fig. 3*A*). Ten RNA-seq samples were available with four from the brushtail brain, five from the liver, and one from the heart. In the brain, there was a low level of *PON2* expression detected. As anticipated, there was robust expression of *PON1* in the liver as observed in other therian mammals. To our surprise, there was no *PON3* expression in the liver in contrast to PON3's liver expression in other therian mammals. Instead, there seemed to be expression of an isoform of *PON4A* in the liver. In the heart, there was noted expression of *PON2* as was expected, and interestingly, there was also low expression of *PON4B*. This supports a hypothesis that PONs play a role in heart maintenance (Li et al. 2018).

Given the unexpected lack of expression of *PON3* and tissue-specific expression of PON4A and PON4B, we next looked to see if there was evidence of positive selection associated with this recent expansion and diversification of brushtail *PON3* into *PON4A* and *PON4B*. We first used CODEML from the PAML package to compare sites models, M1 versus M2 and M7 versus M8, to test if there was positive selection within the entire marsupial PON3 clade (fig. 3*B* and *C*), and we found no evidence for it. We then tested if there was evidence of episodic positive selection associated with the brushtail lineage and its gene duplications compared with the rest of the marsupial *PON3* genes using branch-site models. Indeed, we observed that the branch-site model allowing positive selection in the brushtail *PON*s fit the data significantly better than the null model ($P = 0.000425$), and it was estimated that 11.5% of the positions in brushtail *PON3* sequences evolved under positive selection. The subsequent Bayes Empirical Bayes (BEB) analysis revealed 19 of the 352 positions under positive selection with a posterior probability exceeding 0.5. Sixteen of the sites were mapped to the rabbit PON1 structure (PDB 1V04) so we could visualize where within the *PON* protein the positive selection was occurring. We observed the residues to be clustered around the catalytic active site (fig. 3*D*) and we determined that these sites under positive selection are clustered together more closely than would be expected by chance (permutation $P < 1e\text{-}6$). These changes near the active site could have increased the specificity of these enzymes for a specific yet unidentified substrate(s) which enhanced the fitness of this species.

As an additional method to test for positive selection occurring within the brushtail species, we ran Branch-site Unrestricted Statistical Test for Episodic Diversification (BUSTED, http://www.datamonkey.org/busted). While it found that the unconstrained model (logL = −3,726.8) with the brushtail *PON3*, *PON4A*, and *PON4B* as the foreground sequences fit the data better than the null model (logL = −3,728.1), it did not reject the null at an alpha of

Downloaded from https://academic.oup.com/gbe/article/15/2/evad011/7013737 by University of Pittsburgh user on 23 April 2024

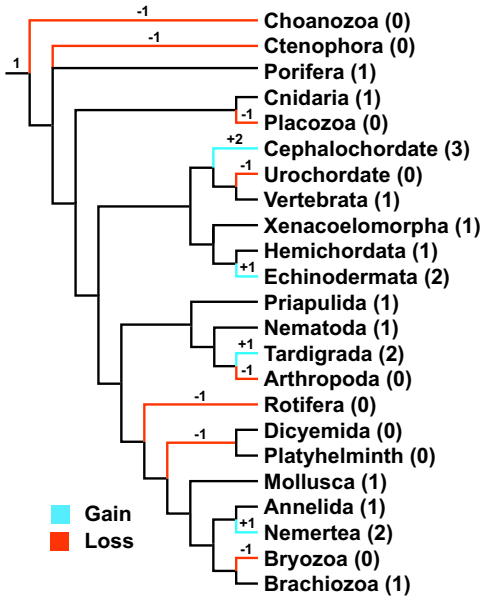


Fig. 4.—Overview of observed changes in the number of *PON* genes throughout metazoa. Phylogenetic tree represent the approximate time when changes in the number of *PON* genes occurred in metazoan broadly (Laumer et al. 2019; Philippe et al. 2019; Kapli and Telford 2020). More detailed phylogenetic tree is available as supplementary fig. S3, Supplementary Material online. The relative timing of the changes is indicated by the number and sign above the branch. Number in between parenthesis after the phylum name indicates the number of ancestral genes which could be detected for the terminal branch. The lack of changes noted in the deeper portions of this tree are not meant to indicate that there was no change in gene copy number in those branches. Rather, it is merely a reflection of the limitation of this study to probe those branches. Ctenophora and Porifera are shown as a polytomy as their exact relationship has not yet been elucidated (Laumer et al. 2019; Li et al. 2021; Redmond and McLysaght 2021). Placement of dicyemida is not well resolved at the time of this publication (Zverkov et al. 2019).

0.05 ($P = 0.142$), so inferences of positive selection should be treated with some caution. A major difference between these models and those of CODEML is that BUSTED accommodates variation in rates of synonymous site divergence. Additionally, multinucleotide substitutions can lead to false-positive results in PAML branch-site tests (Venkat et al. 2018).

Discussion

Through this phylogenetic analysis of the *PON* gene family, we see the changes in *PON* copy number are not restricted to just mammals and cannot be explained as the result of the whole-genome duplication in vertebrates and teleosts, allowing for future investigation into common selective pressures which favor the expansion or reduction in the number of PON members. While most mammals have three *PON* genes, the process of *PON1* becoming a pseudogene within diving mammals has raised the question of when this gene family expanded during evolutionary time. With the genomes of two monotremes, we concluded that the mammalian *PON* expansion occurred before the divergence of monotremata from theria, in the ancestral lineage leading to all mammals. This prompted a more extensive investigation of *PON* genes throughout all metazoa which revealed that there have been multiple independent expansions and contractions of *PON* throughout metazoa. Finally, a closer investigation of the brushtail possum genome revealed that there has been a local expansion of *PON3* within that species associated with positively selected amino acid changes and rapid divergence in expression patterns across tissues.

In contrast to previous findings, the results presented in this paper suggest that *PON1* or *PON3* diverged before *PON2* (fig. 2). The previous study was perhaps limited by the use of four placental mammals (Draganov and La Du 2004). In a more recent study, which used six placental mammals, *PON3* was identified as likely being ancestral to *PON1* and *PON2* (Bar-Rogovsky et al. 2013). In this study comprised of five placental mammals, five marsupials, and two monotremes, we are unable to conclude if *PON1* or *PON3* diverged first; however, we can conclude that *PON2* did not diverge first. Because the monotreme sequences cluster better with *PON3* instead of diverging from the ancestral branch leading to all mammals, this informs us that the *PON* family expanded before monotremes diverged. Given the lack of *PON1* and *PON2* in the otherwise contiguous monotreme assemblies this strongly suggests that ancestral monotremes had *PON1* and *PON2* but then lost them.

Throughout the metazoan tree, there are multiple examples of *PON* expansion and several independent suggestions of *PON* loss (fig. 4, supplementary fig. S3, Supplementary Material online). While this paper attempted to highlight what could be gleaned from the gene tree, the authors wish to caution against the overinterpretation of our results. The deeper nodes of this tree are poorly supported. Some of the low bootstrap support is likely the result of the inclusion of non-bilaterian metazoans whose phylogeny has been and continues to be notoriously difficult to resolve (Rokas et al. 2003; Nosenko et al. 2013; King and Rokas 2017; Pandey and Braun 2020). Additionally, some caution should be used in assemblies with low scaffold and/or contig N50s. These fragmented assemblies could give rise to uncollapsed gene models resulting in the appearance of a false duplication. Species whose scaffold N50 or contig N50 were <100 kb were specifically not mentioned within the Results section (supplementary table S1, Supplementary Material online). The species most likely to suffer from a false duplication due to a fragmented assembly are nematodes 5–7, stalked jellyfish, elephant ear sponge, sea wasp, slime sponge, and

Downloaded from https://academic.oup.com/gbe/article/15/2/evad011/7013737 by University of Pittsburgh user on 23 April 2024

the starlet sea anemone. Given the promiscuous nature of these enzymes, it can be difficult to determine what evolutionary pressures favored expansion and reduction of the *PON* gene family. There could be different functions of these enzymes which are favored in each independent instance. One possible selective pressure is a change in response to oxidative stress since it has been proposed that *PON1* at least has a role in mitigating oxidative damage to lipids (Aviram et al. 2000). Oxidative stress management is different in aquatic environments compared with living on land because diving mammals need to tolerate repeated diving-induced ischemia and reperfusion (Allen and Vázquez-Medina 2019). Given the convergent loss of a functional *PON1* gene in aquatic mammals, this suggests that its loss is beneficial for increasing marine mammals' tolerance of repeated ischemia and reperfusion (Meyer et al. 2018). Although it is not clear why losing an enzyme that is thought to mitigate the effects of oxidative stress would be lost in species that encounter increased oxidative stress. Resistance to bacteria is another possible selection pressure behind the dynamic evolution of the *PON* family. One way bacteria progress as an infection is through the construction of a biofilm which is mediated through quorum sensing (Smith and Iglewski 2003). A common signaling molecule bacteria use for quorum sensing is HSL which PON2 can degrade (Smith and Iglewski 2003; Bar-Rogovsky et al. 2013). This could potentially explain the large *PON* expansions observed within cephalochordates, ambulacraria, and bivalves. Members within these taxonomical groups feed primarily by filtering nutrients from water. Inhibiting biofilm formation would be important for these species so that it does not inhibit extraction of nutrients and sustenance from the water. Another possible explanation for the large *PON* expansion observed in these filter-feeding species is they are the frequent recipients of horizontal transfer from bacteria given their close contact with bacteria (Bernard 1989; Sagane et al. 2010; Grevskott et al. 2017; Olanrewaju et al. 2019).

While two selective pressures have been offered as explanation for *PON* expansions, neither quite explains the specific duplication of *PON3* in the brushtail possum. While an increased bacterial burden in the brushtail possum could be an explanation, an expansion of *PON2* would be more likely as PON2 is more efficient at hydrolyzing HSLs compared with PON3, at least in other theria. This suggests there are more selective pressures related to *PON3* which promoted the fixation and divergence of these duplications in the brushtail possum genome. Further sequencing of other brushtail possum subspecies and related species could determine when this expansion took place and provide clues as to what selective pressures favored it. Additionally, RNA-seq experiments in additional tissues are needed to determine where each *PON* gene is being expressed. Brushtail *PON4B* and *PON1* were expressed in the liver while *PON2* and *PON4A* are observed in the heart. It is not clear in what tissue, if any, brushtail *PON3* is expressed.

From these results, it is fair to propose that this promiscuous family of enzymes plays an ever-changing role depending on lineage. While we can theorize why this gene family expanded in some lineages and contracted in others, additional experiments are needed to test these hypotheses. Certainly, the expansion of *PON3* within the brushtail possum hints that there are still other explanations waiting to be discovered.

Materials and Methods

Identification of Metazoan *PON* Family Members

The 101 species used in this analysis (see supplementary table S1, Supplementary Material online) were sampled by multiple research groups under variable conditions, and hence likely vary in the completeness of their gene content. Our approach to minimize false-negative findings (i.e., false losses) was to examine multiple species within each group when possible, and to conservatively claim loss of a gene only when it was not detected in all species examined within the respective group.

PON genes were identified using a combination of HMMR searches and phylogenetic verification. The *PON* family of genes are characterized by the presence of an arylesterase domain (Primo-Parmo et al. 1996; Rodrigo et al. 1997); thus we used *hmmsearch* in HMMR 3.0 (Eddy 1998) to search proteins for motifs that matched PFAM profile for arylesterase (PF01731.21)—this model was created using 1,047 known *PON* sequences from 511 species across Eukaryotes. *PON* sequences were identified based on maximum full sequence *e*-value and maximum the best domain *e*-value of 1e-6 regardless of the number of domains present. If multiple *PON*s were identified within the same species, *PON*s identified on different chromosomes or on the same chromosome/scaffold, but >100 kb away, are designated by a different alphabet character. *PON*s on the same chromosome/scaffold and within 100 kb are given the same alphabet character and a different number.

We then used PASTA (v.1.8.6) (Mirarab et al. 2015) to generate a multiple sequence alignment using default parameters except –mask-gappy-sites = 6. The alignment was trimmed using Clipkit's smart-gap mode and default parameters (Steenwyk et al. 2020). Smart model selection (SMS) determined that Le–Gascuel (LG) with a gamma distribution was the best model using Akaike information criterion (AIC; Lefort et al. 2017). RAxML generated the optimal phylogenetic tree from twenty random starting trees and using the protgammaauto option. Then using the LG (Le and Gascuel 2008)+G amino acid substitution model, 1,000 bootstraps were performed. Trees were visualized using FigTree (version 1.4.4, http://tree.bio.ed.ac.uk/

Downloaded from https://academic.oup.com/gbe/article/15/2/evad011/7013737 by University of Pittsburgh user on 23 April 2024

software/figtree/) and modified in Adobe Illustrator (2020). The ambulacraria tree was produced using the same sequences identified and methods used for the metazoan analysis with the exception that only 200 bootstraps were performed.

Identification of Mammalian *PON* Family Members

To identify mammalian *PON* family members, sequences (see supplementary table S2, Supplementary Material online) were queried against human PON1 (NP_000437.3), PON2 (NP_000296.2), and PON3 (NP_000931.1). Additionally, some caution should be (Altschul et al. 1990) with a minimum query coverage of 90% and minimum percent identity of 50%. BLASTP was used for the mammalian PONs as it was more straightforward and was sufficient for the level of divergence within mammals. Additional criteria for identification were that the placental mammals had to have at least one of their PON proteins curated in RefSeq, and each protein must come from a unique chromosomal locus. In the case where multiple isoforms were available, the longest sequence was used.

Sequences identified using the criteria above were then aligned using webPRANK (https://www.ebi.ac.uk/goldman-srv/webprank/; Löytynoja and Goldman 2010), and regions of high divergence or single species–specific indels were trimmed manually using AliView (version 1.27; Larsson 2014). Phylogenetic analysis of this alignment was done using PhyML (version 3.3.20190909, http://www.atgc-montpellier.fr/phyml/; Guindon et al. 2010) using SPR tree improvement and 3 random starting trees with 200 bootstraps. SMS (Lefort et al. 2017) determined the Jones–Taylor–Thornton (JTT) substitution model (Jones et al. 1992) with a gamma distribution (G; parameter = 1.29) and 0.068 proportion of sites being invariable (I) was the best model using AIC (Akaike 1974). Trees were visualized using FigTree and modified in Adobe Illustrator (2020).

Model Comparison: *PON* Family Duplication

To determine which of the three mammalian *PON*s was the first to diverge, three different tree models were generated based upon the tetrapod species tree. We then tested which of the three models best fit the tetrapod multiple sequence alignment using the CODEML program in PAML (version 4.9) (Z. Yang 2007). The JTT substitution model was used. The log-likelihood scores produced by CODEML were compared with determine statistical significance using a likelihood-ratio test and a χ^2 distribution with one degree of freedom (Huelsenbeck and Crandall 1997).

To confirm that the monotreme *PON* was not the result of merging of two *PON* genes, multiple sequence alignments of the individual exons and grouped exons as determined by NCBI gene were generated using PRANK. To determine where the monotreme sequence clustered, PhyML was used to generate a maximum likelihood tree (tree improvement: SPR, number of random starting tree: 3, perform bootstrap: 200). Different nucleotide substitution models were used depending on what SMS determined.

RNA-seq Mapping

To verify that the *PON* expansion in the brushtail possum *(T. vulpecula)* is real we investigate whether RNA-seq samples could map to that region. Samples (see supplementary table S3, Supplementary Material online) were downloaded from NCBI-SRA using sra-toolkit (v2.10.0, https://github.com/ncbi/sra-tools) (Anon). Reads were trimmed using trim-galore (0.4.4, cutadapt v1.14 https://github.com/FelixKrueger/TrimGalore; Martin 2011; Krueger 2012), and had quality assessment done by fastqc (0.11.4, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/; Andrews 2010). Using BWA-MEM (v. 2020_03_19; Van der Auwera et al. 2013), the reads were mapped to manually concatenated *PON* mRNA (XM_036759253.1, XM_036759536.1, XM_036759540.1, and XM_036761169.1) and visualized by Integrated Genome Viewer (IGV; Thorvaldsdóttir et al. 2013; v2.9.2).

Testing for Positive Selection

To determine if the brushtail PON3, PON4A, and PON4B proteins were experiencing positive selection, RefSeq mRNA, excluding the stop codons and untranslated regions, of the marsupial PON3 proteins were acquired from NCBI nucleotide. The nucleotide sequences were aligned using PRANK and spurious sequences, UTRs, and stop codons were manually trimmed. We used a likelihood-ratio test to compare the M1 and M2 models and M7 and M8 models within the CODEML package in PAML to determine if positive selection was detected across all branches (options: Model = 0, NSsites = 1 2 7 8; Yang and Swanson 2002; Yang 2007). M1 is a neutral model which allows for two classes of sites ($0 <= \omega <= 1$ and $\omega = 1$) while M2 adds a third class to the M1 model which allows for the detection of positive selection ($\omega > 1$). M7 is also a neutral model; however, rather than having two discrete classes, the null distribution is represented as a beta distribution ($0 < \omega < 1$). M8 builds on M7 by adding a class to detect position selection ($\omega > 1$). The beta distribution allows the null model to be more flexible and better represent the data for codons under negative selection or neutral evolution. We also tested if positive selection was occurring on specific branches involving the duplication within the brushtail PON proteins using PAML's branch-site model test 2 using the same multiple sequence alignment from the previous positive selection tests (options: Model = 2, NSsites = 2, fix_kappa = 0, kappa = 2, omega = 1, fix_alpha = 1, alpha = 0; Zhang et al. 2005). All sites which

Downloaded from https://academic.oup.com/gbe/article/15/2/evad011/7013737 by University of Pittsburgh user on 23 April 2024

were identified by the BEB analysis were taken to be under positive selection (Yang et al. 2005). Additionally, another branch-site analysis was also done using BUSTED (https://www.datamonkey.org/analyses; Murrell et al. 2015) using the same alignment as the PAML analysis. The three branches leading toward the brushtail *PON* sequences as well as the internal node connecting *PON3* and *PON4B* were selected as being in the foreground.

Protein Modeling

To model where the predicted positively selected sites were located, an amino acid multiple sequence alignment between the marsupial and rabbit PON3 proteins was generated using webPRANK (Löytynoja and Goldman 2010). Positively selected sites were then mapped onto the rabbit serum paraoxonase (protein data bank: 1V04; Harel et al. 2004). The protein was visualized using Chimera 1.13.1 (Pettersen et al. 2004). Visually, it appeared that the sites experiencing positive selection appeared to be clustered near the active site of PON1. To statistically confirm if this was indeed the case, we used GETAREA to identify solvent exposed surface residues in the crystal structure (Fraczkiewicz and Braun 1998) and used that information to run random permutation simulations to statistically determine if these residues are clustered (Clark et al. 2007).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged. The computational resources used were partially funded by the NIH Shared Instrumentation Grant 1S10OD021644-01A1. This work was supported by the National Institutes of Health (grant numbers 3T32DK007115 to A.M.G., K99GM144774 to A.M.G., R01 HG009299 to A.M.G. and N.L.C., and R01 EY030546 to S.A.M.L., J.S.P., and N.L.C.). The authors thank the rest of the Clark lab, Paige Haffener, and Chris Stringham for thoughtful suggestions and useful discussions.

Data Availability

The protein models underlying this article are individually available at a variety of repositories Zenodo (Copley et al. 2018; Kvist et al. 2019; schultz and Francis 2020), National Center for Biotechnology Information (NCBI), Ensembl, SIMRBASE (Data), OIST Marine Genome Projects (Data), Github (Ryan), Harvard Dataverse (Qingxiang), Plos One (Delroisse et al. 2016), Google Drive (Data), Neurobase (Data), Bitbucket, Dryad (Data), Ephybase (Data), Reefgenomics (Liew et al. 2016), GigaDB (Data), National Genomics Data Center (NGDC) (Chen et al. 2021), Figshare (Data), PeerJ (Jin et al. 2020), Planmine (Rozanski et al. 2019), NHGRI (Data), and the Ryan Lab website (Ryan). All protein models have their assembly information listed in supplementary table S1, Supplementary Material online. In addition, they are designated as having come from NCBI, Ensembl or with URL listed. All newick trees and multiple sequence alignments underlying this article are available at the Clark website https://clark.genetics.utah.edu/software-data-and-collaborators/. The protein crystal structure is available at https://www.rcsb.org/structure/1V04.

Literature Cited

Akaike H. 1974. A new look at the statistical model identification. IEEE Trans Automat Contr. 19:716–723.

Allen KN, Vázquez-Medina JP. 2019. Natural tolerance to ischemia and hypoxemia in diving mammals: a review. Front Physiol. 10:1199.

Altenhöfer S, et al. 2010. One enzyme, two functions: PON2 prevents mitochondrial superoxide formation and apoptosis independent from its lactonase activity. J Biol Chem. 285:24398.

Altenhoff AM, Glover NM, Dessimoz C. 2019. Inferring orthology and paralogy. Methods Mol Biol. 1910:149–175.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Andrews S. 2010. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. [updated 2021 Jun 30]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Anon. GitHub - ncbi/sra-tools: SRA Tools. Available from: https://github.com/ncbi/sra-tools

Aviram M, et al. 1998. A possible role for paraoxonase paraoxonase inhibits high-density lipoprotein oxidation and preserves its functions. A possible peroxidative role for paraoxonase paraoxonase • HDL • LDL • lipid peroxidation • apolipoprotein E deficient mice. J Clin Invest. 101:1581–1590.

Aviram M, et al. 2000. Human serum paraoxonases (PON1) Q and R selectively decrease lipid peroxides in human coronary and carotid atherosclerotic lesions. Circulation 101:2510–2517.

Bar-Rogovsky H, Hugenmatter A, Tawfik DS. 2013. The evolutionary origins of detoxifying enzymes the mammalian serum paraoxonases (PONS) relate to bacterial homoserine lactonases. J Biol Chem. 288:23914–23927.

Bernard FR. 1989. Uptake and elimination of coliform Bacteria by four marine Bivalve mollusks. Can. J. Fish. Aquat. Sci. 46:1592–1599.

Billecke S, et al. 2000. Human serum paraoxonase (PON1) isozymes Q and R hydrolyze lactones and cyclic carbonate esters. Drug Metab. Dispos. 28:1335–1342.

Chen M, et al. 2021. Genome warehouse: a public repository housing genome-scale data. Genomics, Proteomics Bioinforma. 19: 584–589.

Clark NL, Findlay GD, Yi X, MacCoss MJ, Swanson WJ. 2007. Duplication and selection on abalone sperm lysin in an allopatric population. Mol Biol Evol. 24:2081–2090.

Copley R, Leclere L, Houliston E. 2018. The genome of the jellyfish Clytia hemisphaerica and the evolution of the cnidarian life-cycle.

Downloaded from https://academic.oup.com/gbe/article/15/2/evad011/7013737 by University of Pittsburgh user on 23 April 2024

[updated 2022 May 19]. Available from: https://zenodo.org/record/1470436
Costa LG, et al. 1990. Serum paraoxonase and its influence on paraoxon and chlorpyrifos-oxon toxicity in rats. Toxicol Appl Pharmacol. 103:66–76.
Data. SIMRbase Organisms | SIMRbase Genomes. [updated 2022 May 19]. Available from: https://simrbase.stowers.org/
Data. OIST Marine Genomics Unit Genome Browser. [updated 2022 May 19]. Available from: https://marinegenomics.oist.jp/acornworm/gallery
Data. Aurelia_Genome - Google Drive. [updated 2022 May 19]. Available from: https://drive.google.com/drive/folders/1NC6bZ9cxWkZyofOsMPzrxlH3C7m1ySiu
Data. NeuroBase: A Comparative Neurogenomics Database. [updated 2022 May 19]. Available from: https://neurobase.rc.ufl.edu/pleurobrachia/download
Data. Dryad Data – Calcisponges have a ParaHox gene and dynamic expression of dispersed NK homeobox genes. [updated 2022 May 19]. Available from: https://datadryad.org/stash/dataset/doi:10.5061/dryad.tn0f3
Data. Resources – EphyBase. [updated 2022 May 19]. Available from: https://spaces.facsci.ualberta.ca/ephybase/resources/
Data. GigaDB Dataset - DOI 10.5524/100564 - A draft genome assembly of the acoel flatworm Praesagittifera naikaiensis. [updated 2022 May 19]. Available from: http://gigadb.org/dataset/100564
Data. GFF file, CDS and Protein Sequences.rar. [updated 2022 May 19]. Available from: https://figshare.com/articles/dataset/GFF_file_CDS_and_Protein_Sequences_rar/6142322/1
Data. MGP Portal. [updated 2022 May 19]. Available from: https://kona.nhgri.nih.gov/mnemiopsis/
Delroisse J, Mallefet J, Flammang P. 2016. De novo adult transcriptomes of two European brittle stars: spotlight on opsin-based photoreception. PLoS One 11:e0152988.
Devarajan A, et al. 2011. Paraoxonase 2 deficiency alters mitochondrial function and exacerbates the development of atherosclerosis. Antioxid Redox Signal. 14:341.
Draganov DI, et al. 2005. Human paraoxonases (PON1, PON2, and PON3) are lactonases with overlapping and distinct substrate specificities. J. Lipid Res. 46:1239.
Draganov DI, La Du BN. 2004. Pharmacogenetics of paraoxonases: a brief review. Naunyn Schmiedebergs Arch Pharmacol. 369:78–88.
Eddy SR. 1998. Profile hidden Markov models. Bioinformatics 14:755–763.
Fernández R, Gabaldón T. 2020. Gene gain and loss across the Metazoa Tree of Life. Nat Ecol Evol. 4:524.
Fraczkiewicz R, Braun W. 1998. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. J Comput Chem. 19:319–333.
Glover NM, Redestig H, Dessimoz C. 2016. Homoeologs: what are they and how do we infer them? Trends Plant Sci. 21:609.
Grevskott DH, Svanevik CS, Sunde M, Wester AL, Lunestad BT. 2017. Marine bivalve mollusks as possible indicators of multidrug-resistant *Escherichia coli* and other species of the Enterobacteriaceae family. Front Microbiol. 8:24.
Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59:307–321.
Hahn MW. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. J Hered. 100:605–617.
Harel M, et al. 2004. Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes. Nat Struct Mol Biol. 11:412–419.
Horke S, et al. 2007. Paraoxonase-2 reduces oxidative stress in vascular cells and decreases endoplasmic reticulum stress-induced caspase activation. Circulation 115:2055–2064.
Huang Y, et al. 2013. Myeloperoxidase, paraoxonase-1, and HDL form a functional ternary complex. J Clin Invest. 123:3815.
Huelsenbeck JP, Crandall KA. 1997 Phylogeny estimation and hypothesis testing using maximum likelihood. Annual Review of Ecology and Systematics 28:437–466. doi:10.1146/annurev.ecolsys.28.1.437
Jin F, et al. 2020. High-quality genome assembly of *Metaphire vulgaris*. PeerJ 8:e10313.
Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8:275–282.
Kapli P, Telford MJ. 2020. Topology-dependent asymmetry in systematic errors affects phylogenetic placement of Ctenophora and Xenacoelomorpha. Sci Adv. 6:5162–5173.
Khersonsky O, Tawfik DS. 2005. Structure-reactivity studies of serum paraoxonase PON1 suggest that its native activity is lactonase. Biochemistry 44:6371–6382.
King N, Rokas A. 2017. Embracing uncertainty in reconstructing early animal evolution. Curr Biol. 27:R1081–R1088.
Koonin EV. 2005. Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet. 39:309–338.
Krueger F. 2012. Babraham Bioinformatics - Trim Galore! [updated 2021 Jun 30]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Krüger M, Pabst AM, Al-Nawas B, Horke S, Moergel M. 2015. Paraoxonase-2 (PON2) protects oral squamous cell cancer cells against irradiation-induced apoptosis. J Cancer Res Clin Oncol. 141:1757–1766.
Kvist S, Manzano-Marín A, de Carle D, Trontelj P, Siddall ME. 2019. Supplementary data for Kvist et al. 2020 "Draft genome of the European medicinal leech *Hirudo medicinalis* (Annelida: Clitellata: Hirudiniformes) with emphasis on anticoagulants." [updated 2022 May 19]. Available from: https://zenodo.org/record/3555585
Larsson A. 2014. Aliview: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30:3276–3278.
Laumer CE, et al. 2019. Revisiting metazoan phylogeny with genomic sampling of all phyla. Proc R Soc B. 286:1–10.
Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol Biol Evol. 25:1307–1320.
Lefort V, Longueville JE, Gascuel O. 2017. SMS: smart model selection in PhyML. Mol Biol Evol. 34:2422–2424.
Li W, et al. 2018. Paraoxonase 2 prevents the development of heart failure. Free Radic Biol Med. 121:117–126.
Li Y, Shen XX, Evans B, Dunn CW, Rokas A. 2021. Rooting the animal tree of life. Mol Biol Evol. 38:4322–4333.
Liew YJ, Aranda M, Voolstra CR. 2016. Reefgenomics.org - a repository for marine genomics data. Database 2016:baw152.
Löytynoja A, Goldman N. 2010. WebPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. BMC Bioinformatics 11:579.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17:10.
Meyer WK, et al. 2018. Ancient convergent losses of paraoxonase 1 yield potential risks for modern marine mammals. Science 361:591–594.
Mirarab S, et al. 2015. PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. J Comput Biol. 22:377–386.
Murrell B, et al. 2015. Gene-wide identification of episodic selection. Mol Biol Evol. 32:1365–1371.
Muthukrishnan S, et al. 2012. Mechanistic insights into the hydrolysis of organophosphorus compounds by paraoxonase-1: exploring the limits of substrate tolerance in a promiscuous enzyme. J Phys Org Chem. 25:1247.

Downloaded from https://academic.oup.com/gbe/article/15/2/evad011/7013737 by University of Pittsburgh user on 23 April 2024

Ng CJ, et al. 2001. Paraoxonase-2 is a ubiquitously expressed protein with antioxidant properties and is capable of preventing cell-mediated oxidative modification of low density lipoprotein. J Biol Chem. 276:44444–44449.
Nosenko T, et al. 2013. Deep metazoan phylogeny: when different genes tell different stories. Mol Phylogenet Evol. 67:223–233.
Olanrewaju TO, McCarron M, Dooley JSG, Arnscheidt J. 2019. Transfer of antibiotic resistance genes between *Enterococcus faecalis* strains in filter feeding zooplankton *Daphnia magna* and *Daphnia pulex*. Sci Total Environ 659:1168–1175.
Pandey A, Braun EL. 2020. Phylogenetic analyses of sites in different protein structural environments result in distinct placements of the metazoan root. Biology 9:64.
Pearson WR. 2013. An introduction to sequence similarity ("homology") searching. Curr Protoc Bioinformatics. 3:3.1.1–3.1.8.
Pettersen EF, et al. 2004. UCSF Chimera - a visualization system for exploratory research and analysis. J Comput Chem. 25:1605–1612.
Philippe H, et al. 2019. Mitigating anticipated effects of systematic errors supports sister-group relationship between *Xenacoelomorpha* and *Ambulacraria*. Curr Biol. 29:1818–1826.e6.
Primo-Parmo SL, Sorenson RC, Teiber J, La Du BN. 1996. The human serum paraoxonase/arylesterase gene (PON1) is one member of a multigene family. Genomics 33:498–507.
Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 33:D501.
Qingxiang G. 2022. A myxozoan genome reveals mosaic evolution in a parasitic cnidarian. BMC Biol. 20(1):51.
Redmond AK, McLysaght A. 2021. Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding. Nat Commun. 12:1–14.
Rodrigo L, et al. 1997. Purification and characterization of paraoxon hydrolase from rat liver. Biochem J. 321:595–601.
Rokas A, King N, Finnerty J, Carroll SB. 2003. Conflicting phylogenetic signals at the base of the metazoan tree. Evol Dev. 5:346–359.
Rozanski A, et al. 2019. Planmine 3.0 - improvements to a mineable resource of flatworm biology and biodiversity. Nucleic Acids Res. 47:D812–D820.
Rull A, et al. 2012. Serum paraoxonase-3 concentration is associated with insulin sensitivity in peripheral artery disease and with inflammation in coronary artery disease. Atherosclerosis 220:545–551.
Ryan J. Ohdera_et_al_2018/AA_Files at master josephryan/Ohdera_et_al_2018 · GitHub. [updated 2022 May 19]. Available from: https://github.com/josephryan/Ohdera_et_al_2018/tree/master/AA_Files
Ryan J. Ryan Lab: Genome data. [updated 2022 May 19]. Available from: http://ryanlab.whitney.ufl.edu/genomes/
Sagane Y, et al. 2010. Functional specialization of cellulose synthase genes of prokaryotic origin in chordate larvaceans. Development 137:1483–1492.
Schultz T, Francis W. 2020. Conchoecia/hormiphora: Annotation of the genome - Hcv1.av93 - for zenodo. [updated 2022 May 19]. Available from: https://zenodo.org/record/4074309
Shih DM, et al. 1998. Mice lacking serum paraoxonase are susceptible to organophosphate toxicity and atherosclerosis. Nature 394:284–287.
Shih DM, et al. 2007. Decreased obesity and atherosclerosis in human paraoxonase 3 transgenic mice. Circ Res. 100:1200.
Smith RS, Iglewski BH. 2003. *P. aeruginosa* quorum-sensing systems and virulence. Curr Opin Microbiol. 6:56–60.
Steenwyk JL, Buida TJ, Li Y, Shen XX, Rokas A. 2020. ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. 18:e3001007.
Stoltz DA, et al. 2007. Paraoxonase-2 deficiency enhances *Pseudomonas aeruginosa* quorum sensing in murine tracheal epithelia. Am J Physiol - Lung Cell Mol Physiol. 292:852–860.
Teiber JF, et al. 2008. Dominant role of paraoxonases in inactivation of the *Pseudomonas aeruginosa* quorum-sensing signal N-(3-oxododecanoyl)-l-homoserine lactone. Infect Immun. 76:2512.
Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14:178–192.
Tward A, et al. 2002. Decreased atherosclerotic lesion formation in human serum paraoxonase transgenic mice. Circulation 106:484–490.
Van der Auwera GA, et al. 2013. From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinforma. 11:11.10.1–11.10.33.
Variji A, et al. 2019. The combined utility of myeloperoxidase (MPO) and paraoxonase 1 (PON1) as two important HDL-associated enzymes in coronary artery disease: which has a stronger predictive role? Atherosclerosis 280:7–13.
Venkat A, Hahn MW, Thornton JW. 2018. Multinucleotide mutations cause false inferences of lineage-specific positive selection. Nat Ecol Evol. 2:1280.
Wolf YI, Koonin EV. 2013. Genome reduction as the dominant mode of evolution. Bioessays 35:829.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Yang Z, Swanson WJ. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol Biol Evol. 19:49–57.
Yang Z, Wong WSW, Nielsen R. 2005. Bayes Empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. 22:1107–1118.
Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 22:2472–2479.
Zverkov OA, et al. 2019. Dicyemida and orthonectida: two stories of body plan simplification. Front Genet. 10:443.

Associate editor: Federico Hoffmann

Downloaded from https://academic.oup.com/gbe/article/15/2/evad011/7013737 by University of Pittsburgh user on 23 April 2024