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

Sarah A.M. Lucas[1], Allie M Graham[1], Jason S Presnell[1], & Nathan L Clark[1*]

[1] Department of Human Genetics, University of Utah, UT, USA

*communicating author, nclark@utah.edu

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.

Introduction

The family of paraoxonase (*PON*) genes was named based on the discovery that one member could degrade the insecticide parathion, whose active metabolite paraoxon functions as a neurotoxic cholinesterase inhibitor[1]. In 1996 it was revealed that this enzyme is part of a multigene family in humans[2], 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[3,4]. 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[5].

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[6,7]. 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)[8,9], arylesters (aromatic esters)[7,8,10], and homo-serine lactones which are key molecules for quorum sensing in bacteria[8,11,12]. Other studies have revealed PON1 has atheroprotective - protection against plaque formation - effects[13,14], antioxidant properties through the degradation of lipoperoxides[15], and an ability to co-regulate inflammation through an interaction with myeloperoxidase on HDL particles[16,17]. Meanwhile PON2 has also been shown to have atheroprotective effects through its ability to reduce superoxide release[3,4,18,19] as well as anti-apoptotic

properties[20]. PON3 has been associated with several diseases, but its exact functional role in those diseases have yet to be elucidated[21,22].

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[23]. 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*[5]; however, a more recent study challenged this finding. Through the incorporation of additional *PON* sequences, Bar-Rogovsky et. al determined that *PON3* diverged before *PON1* and *PON2*[9]. 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.

Methods

Identification of Metazoan *PON* family members

The 101 species used in this analysis (see Supplementary Table 1) 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[2,24]; thus we used *hmmsearch* in HMMR 3.0[25] to search proteins for motifs that matched PFAM profile for arylesterase (PF01731.21) – this model was created using 1047 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 > 100kb away, are designated by a different alphabet character. *PON*s on the same chromosome/scaffold and within 100kb are given the same alphabet character and a different number.

We then used PASTA (v.1.8.6)[26] 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[27]. RAxML generated the optimal phylogenetic tree from twenty random starting trees and using the protgammaauto option. Then using the Le-Gascuel (LG)[28]+G amino acid substitution model, 1000 bootstraps were performed. Trees were visualized using FigTree (version 1.4.4, http://tree.bio.ed.ac.uk/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 2) were queried against human PON1 (NP_000437.3), PON2 (NP_000296.2), and PON3 (NP_000931.1) using BLASTP[29] with a minimum query coverage of 90% and minimum percent identity of 50%. 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/)[30], and regions of high divergence or single species specific indels were trimmed manually using AliView (version 1.27)[31]. Phylogenetic analysis of this alignment was done using PhyML (version 3.3.20190909, http://www.atgc-montpellier.fr/phyml/)[32] using SPR tree improvement and 3 random starting trees with 200 bootstraps. Smart Model selection (SMS)[33] determined the Jones-Taylor-Thornton (JTT) substitution model[34] with a gamma distribution (G) (parameter = 1.29) and 0.068 proportion of sites being invariable (I) was the best model using Akaike information criterion (AIC)[35]. 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)[36]. The JTT substitution model was used. The log-likelihood scores produced by CODEML were compared to determine statistical significance using a likelihood ratio test and a chi-square distribution with one degree of freedom[37].

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 *(Trichosurus vulpecula)* is real we investigate whether RNA-seq samples could map to that region. Samples (see Supplementary Table 3) were downloaded from NCBI-SRA using sra-toolkit (v2.10.0, https://github.com/ncbi/sra-tools)[38]. Reads were trimmed using trim-galore (0.4.4, cutadapt v1.14 https://github.com/FelixKrueger/TrimGalore)[39,40], and had quality assessment done by fastqc (0.11.4), https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)[41]. Using BWA-MEM (v. 2020_03_19) [42], 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)[43] (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)[44,45]. 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)[46]. All sites which were identified by the Bayes Empirical Bayes analysis were taken to be under positive selection[47]. Additionally, another branch-site analysis was also done using BUSTED (https://www.datamonkey.org/analyses)[48] 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 PRANK[30]. Positively selected sites were then mapped onto the rabbit serum paraoxonase (protein data bank: 1V04)[49]. The protein was visualized using Chimera 1.13.1[50]. 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[51] and used that information to run random permutation simulations to statistically determine if these residues are clustered[52].

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 two 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 sequences were aligned and subjected to phylogenetic analysis.

The resulting phylogeny provides evidence of ancestral duplication among closely related species (Fig 1, Fig S1). Within echinoderms, tardigrades, and bivalves, there is evidence of a single ancestral duplications which resulted in two ancestral PON genes for each taxonomical group. Similarly, there is also evidence of two ancestral duplications in cephalochordates and mammals, resulting in three ancestral PON genes.

In addition to those instances of ancestral duplications, there are many instances of lineage specific PON expansions through tandem duplication and other gene duplication mechanisms. Two examples of tandem duplication include *PONA1* & *PONA2* for the green sea urchin (*Lytechinus variegatus*) and purple sea urchin (*Strongylocentrotus purpuratus).* Both species had an independent tandem expansion of the same PON gene as the sequences within a species cluster better together than with a *PON* from the other species and are located on the same chromosome. This contrasts with the ancestral tandem duplication observed in mammals (Fig 2A). Besides tandem duplication, there is evidence of additional gene duplication, particularly in the Great Scallop (*Pecten Maximus)*. While three of its nine *PON* genes are in a tandem array (i.e. less than 100 kb away from each other) on one of its chromosomes, the remaining six are scattered among five chromosomes and scaffolds (Supplementary Table 1) giving evidence to some alternate gene duplication mechanisms.

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 ancestral losses of *PON* in the ctenophores, placozoans, urochordates, arthropods, and platyhelminths (Fig 4, Fig S3). 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[53].

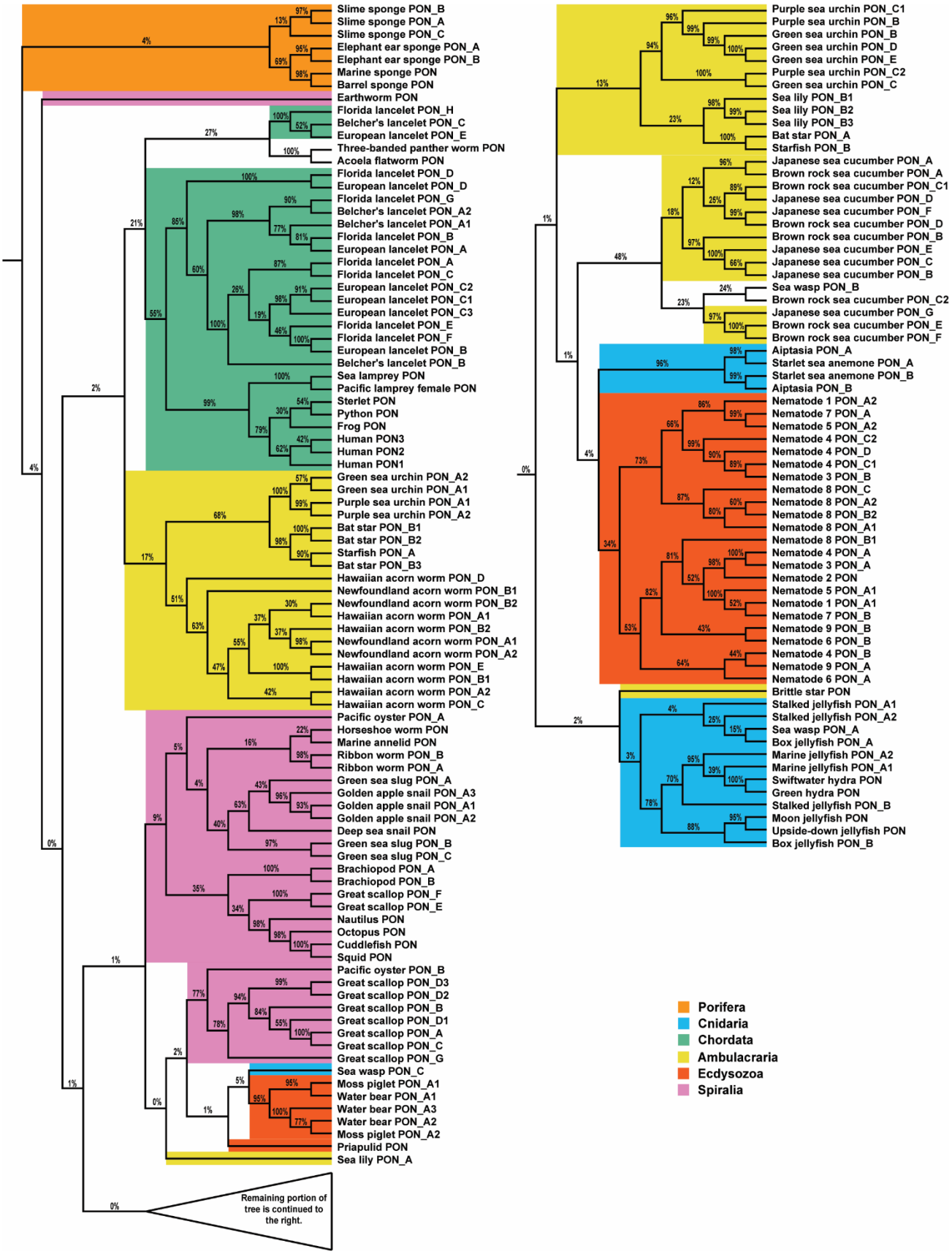

Slime sponge PON_B
Slime sponge PON_A
Slime sponge PON_C
Elephant ear sponge PON_A
Elephant ear sponge PON_B
Marine sponge PON
Barrel sponge PON
Earthworm PON
Florida lancelet PON_H
Belcher's lancelet PON_C
European lancelet PON_E
Three-banded panther worm PON
Acoela flatworm PON
Florida lancelet PON_D
European lancelet PON_D
Florida lancelet PON_G
Belcher's lancelet PON_A2
Belcher's lancelet PON_A1
Florida lancelet PON_B
European lancelet PON_A
Florida lancelet PON_A
Florida lancelet PON_C
European lancelet PON_C2
European lancelet PON_C1
European lancelet PON_C3
Florida lancelet PON_E
Florida lancelet PON_F
European lancelet PON_B
Belcher's lancelet PON_B
Sea lamprey PON
Pacific lamprey female PON
Sterlet PON
Python PON
Frog PON
Human PON3
Human PON2
Human PON1
Green sea urchin PON_A2
Green sea urchin PON_A1
Purple sea urchin PON_A1
Purple sea urchin PON_A2
Bat star PON_B1
Bat star PON_B2
Starfish PON_A
Bat star PON_B3
Hawaiian acorn worm PON_D
Newfoundland acorn worm PON_B1
Newfoundland acorn worm PON_B2
Hawaiian acorn worm PON_A1
Hawaiian acorn worm PON_B2
Newfoundland acorn worm PON_A1
Newfoundland acorn worm PON_A2
Hawaiian acorn worm PON_E
Hawaiian acorn worm PON_B1
Hawaiian acorn worm PON_A2
Hawaiian acorn worm PON_C
Pacific oyster PON_A
Horseshoe worm PON
Marine annelid PON
Ribbon worm PON_B
Ribbon worm PON_A
Green sea slug PON_A
Golden apple snail PON_A3
Golden apple snail PON_A1
Golden apple snail PON_A2
Deep sea snail PON
Green sea slug PON_B
Green sea slug PON_C
Brachiopod PON_A
Brachiopod PON_B
Great scallop PON_F
Great scallop PON_E
Nautilus PON
Octopus PON
Cuddlefish PON
Squid PON
Pacific oyster PON_B
Great scallop PON_D3
Great scallop PON_D2
Great scallop PON_B
Great scallop PON_D1
Great scallop PON_A
Great scallop PON_C
Great scallop PON_G
Sea wasp PON_C
Moss piglet PON_A1
Water bear PON_A1
Water bear PON_A3
Water bear PON_A2
Moss piglet PON_A2
Priapulid PON
Sea lily PON_A
Remaining portion of tree is continued to the right.
Purple sea urchin PON_C1
Purple sea urchin PON_B
Green sea urchin PON_B
Green sea urchin PON_D
Green sea urchin PON_E
Purple sea urchin PON_C2
Green sea urchin PON_C
Sea lily PON_B1
Sea lily PON_B2
Sea lily PON_B3
Bat star PON_A
Starfish PON_B
Japanese sea cucumber PON_A
Brown rock sea cucumber PON_A
Brown rock sea cucumber PON_C1
Japanese sea cucumber PON_D
Japanese sea cucumber PON_F
Brown rock sea cucumber PON_D
Brown rock sea cucumber PON_B
Japanese sea cucumber PON_E
Japanese sea cucumber PON_C
Japanese sea cucumber PON_B
Sea wasp PON_B
Brown rock sea cucumber PON_C2
Japanese sea cucumber PON_G
Brown rock sea cucumber PON_E
Brown rock sea cucumber PON_F
Aiptasia PON_A
Starlet sea anemone PON_A
Starlet sea anemone PON_B
Aiptasia PON_B
Nematode 1 PON_A2
Nematode 7 PON_A
Nematode 5 PON_A2
Nematode 4 PON_C2
Nematode 4 PON_D
Nematode 4 PON_C1
Nematode 3 PON_B
Nematode 8 PON_C
Nematode 8 PON_A2
Nematode 8 PON_B2
Nematode 8 PON_A1
Nematode 8 PON_B1
Nematode 4 PON_A
Nematode 3 PON_A
Nematode 2 PON
Nematode 5 PON_A1
Nematode 1 PON_A1
Nematode 7 PON_B
Nematode 9 PON_B
Nematode 6 PON_B
Nematode 4 PON_B
Nematode 9 PON_A
Nematode 6 PON_A
Brittle star PON
Stalked jellyfish PON_A1
Stalked jellyfish PON_A2
Sea wasp PON_A
Box jellyfish PON_A
Marine jellyfish PON_A2
Marine jellyfish PON_A1
Swiftwater hydra PON
Green hydra PON
Stalked jellyfish PON_B
Moon jellyfish PON
Upside-down jellyfish PON
Box jellyfish PON_B
Porifera
Cnidaria
Chordata
Ambulacraria
Ecdysozoa
Spiralia

Figure 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 1000 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.

PON expansion occurred after divergence of mammals from sauropsida

RefSeq-identified PON protein sequences from fifteen tetrapod species were collected[54]. 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 2A). Importantly, the monotreme *PON* genes clustered 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 2B). 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 (-8410.56), although it was not significantly better than the model with *PON3* diverging first (-8411.31). Thus, we cannot be certain which of those two *PON*s is ancestral to the other (p-value = 0.22); however, it can be stated with statistical significance that *PON2* did not diverge before *PON1* (-8413.77, -8410.56, p-value = 0.0113) nor did *PON2* diverge before *PON3* (-8413.77, -8411.31, p-value = 0.0267) (Fig 2B).

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

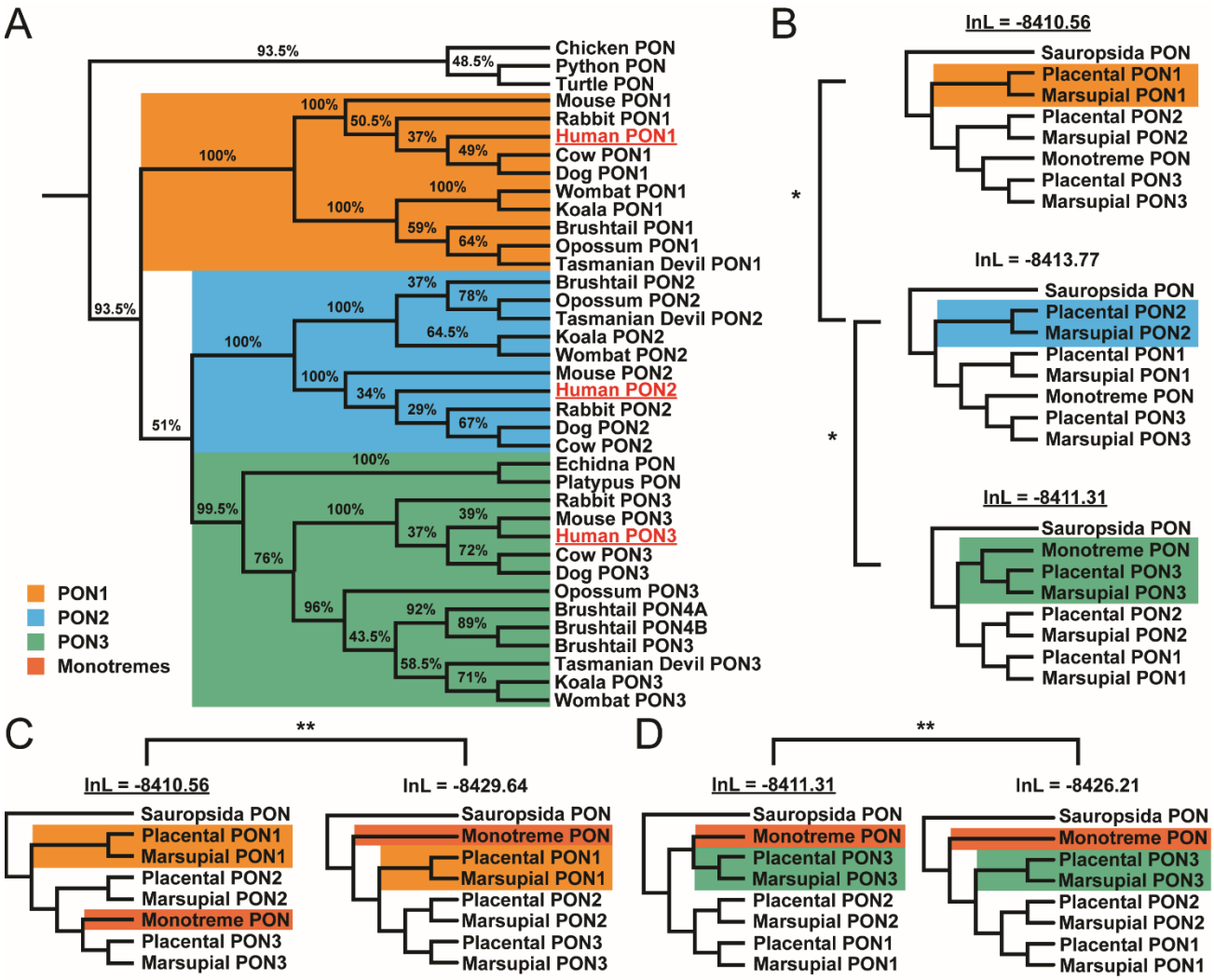


Figure 2 Evolution of PON in tetrapods. (A) Phylogenetic tree of PON family proteins in tetrapods determined by PhyML based on multiple sequence alignment. Bootstrap support values are shown as percentages out of 200. Brushtail PON4 was split into two separate genes based on RNA-seq evidence. Human sequences were underlined to help orient the reader. (B) Phylogenetic tree models represent the scenario in which each of the PONs is ancestral compared to the other two. * indicates p-value was less than 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). ** indicates p-value < 1e-8. The underlined log likelihood indicates the significantly better model. Rooted and unrooted trees produced the same log likelihood score.

Case Study: Recent *PON* expansion under positive selection in brushtail possum.

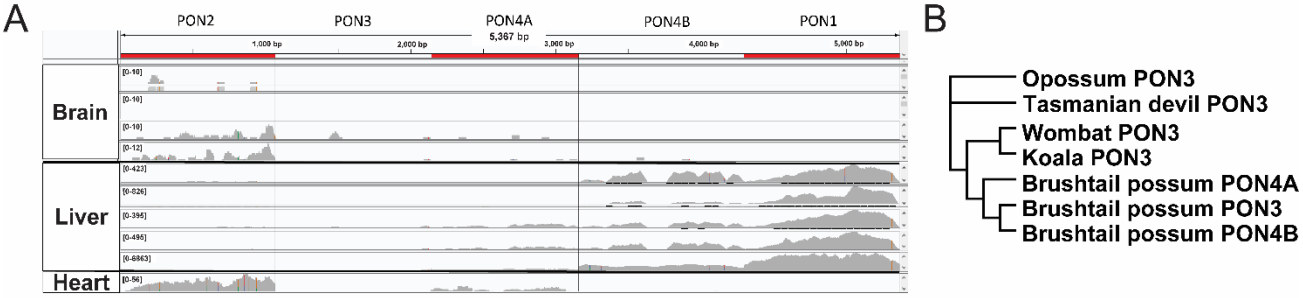
In addition to the three *PON*s that were expected to be found in the brushtail possum, a marsupial mammal, BLAST identified another locus (XP_036615431.1) in between the brushtail *PON3* and *PON1*. Upon closer examination this single locus in brushtail contained what could be two new *PON* genes. We 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 3A). PON4A is comprised of annotated residues 1-337 of XP_036615431.1, and PON4B consists of residues 338-717, although this study suggests the 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 2A).

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 3A). 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. There was no *PON3* expression in the liver as was expected given 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 surprisingly, there was also low expression of *PON4B*.

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 vs M2 and M7 vs M8, to test if there was positive selection within the entire marsupial PON3 clade (Fig 3B, 3C), 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 to 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-value = 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 3D) and we determined that these sites under positive selection are clustered together more closely than would be expected by chance (permutation p-value < 1e-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 = -3726.8) with the brushtail *PON3*, *PON4A*, and *PON4B* as the foreground sequences fit the data better than the null model (logL = -3728.1), it did not reject the null at an alpha of 0.05 (p-value 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.



C

		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	

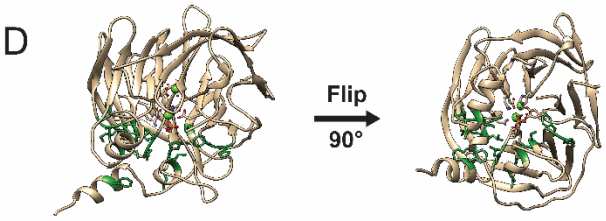


Figure 3 Brushtail PON3 and recent PON4A/B under positive selection. (A) Screenshot from IGV showing a mapped reads to a concatenated transcriptome for all brushtail possum *PON* genes. Tissue samples include four brain, five liver, and one heart sample. The order of the concatenated PON genes is listed 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. (C) Log likelihood values were determined by PAML. The marsupial tree in figure 3B was used in the marsupial analysis (D) 3D protein image of rabbit PON1 (PDB:1V04). Corresponding sites under positive selection in brushtail are colored green and show their atomic structure. Residues which are part of the active site also show their atomic structure.

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 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 initial study was perhaps limited by the use of four placental mammals[5]. In a more recent study, which used six placental mammals, *PON3* was identified as likely being ancestral to *PON1* and *PON2*[9]. 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 monotremes had *PON1* and *PON2* or their ancestor but then lost them.

Throughout the metazoan tree, there are multiple examples of *PON* expansion and several independent suggestions of PON loss (Fig 4, Fig S3). 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 selection 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[55]. Oxidative stress management is different in aquatic environments compared to living on land because diving mammals need to tolerate repeated diving-induced ischemia and reperfusion[56]. 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[23]. 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[57]. A common signaling molecule bacteria use is homoserine lactone (HSL) which PON2 has shown evidence of being able to degrade[9,57]. 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 the biofilm does not inhibit their ability to extract nutrients and sustenance from the water. Another possible explanation for the large *PON* expansion observed in these water filtering species is they are the frequent recipients of multiple horizontal transfer from bacteria given their close contact with bacteria.

While two selective pressures have been offered as explanation for *PON* expansions, neither yet sufficiently explains the brushtail possum specific duplication of *PON3*. 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 to 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, more RNA-seq samples from different tissue types are needed to determine where each of the *PON* genes are being expressed. Brushtail *PON4B* and *PON1* have been shown to be 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 the results of these studies, 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 theories. Certainly, the expansion of *PON3* within the brushtail possum hints that there are still other explanations waiting to be discovered.

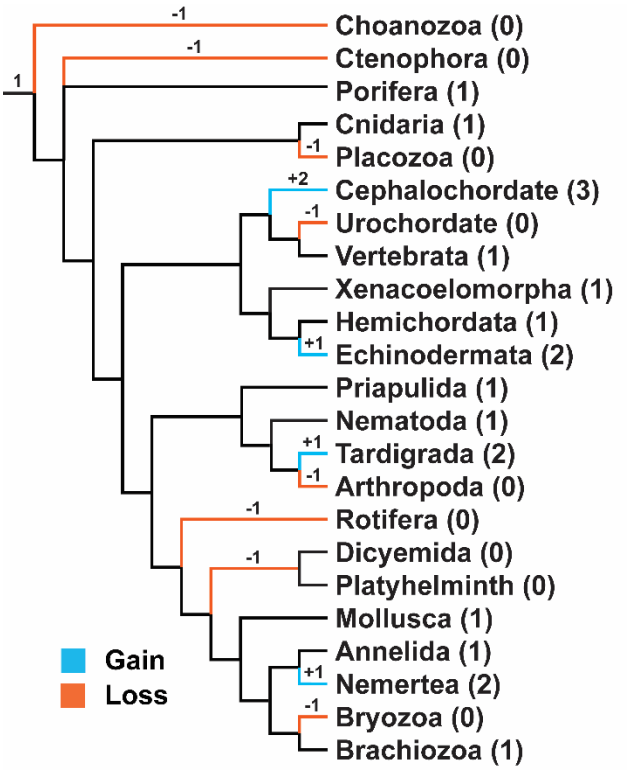


Figure 4 Overview of observed changes in the number of PON genes throughout metazoa. Phylogenetic tree represent the approximate time when changes in the number of PON genes occurred in metazoan broadly[58–60]. More detailed phylogenetic tree is available as supplemental figure 3. The relative timing 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[58,61,62]. Placement of Dicyemida is not well resolved at the time of this publication[63].

References

1 Costa LG, McDonald BE, Murphy SD, Omenn GS, Richter RJ, Motulsky AG, *et al.* Serum paraoxonase and its influence on paraoxon and chlorpyrifos-oxon toxicity in rats. *Toxicol Appl Pharmacol* 1990;**103**:66–76. https://doi.org/10.1016/0041-008X(90)90263-T.

2 Primo-Parmo SL, Sorenson RC, Teiber J, La Du BN. The human serum paraoxonase/arylesterase gene (PON1) is one member of a multigene family. *Genomics* 1996;**33**:498–507. https://doi.org/10.1006/geno.1996.0225.

3 Horke S, Witte I, Wilgenbus P, Krüger M, Strand D, Förstermann U. Paraoxonase-2 reduces oxidative stress in vascular cells and decreases endoplasmic reticulum stress-induced caspase activation. *Circulation* 2007;**115**:2055–64. https://doi.org/10.1161/CIRCULATIONAHA.106.681700.

4 Ng CJ, Wadleigh DJ, Gangopadhyay A, Hama S, Grijalva VR, Navab M, *et al.* 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* 2001;**276**:44444–9. https://doi.org/10.1074/JBC.M105660200.

5 Draganov DI, La Du BN. Pharmacogenetics of paraoxonases: A brief review. *Naunyn Schmiedebergs Arch Pharmacol* 2004;**369**:78–88. https://doi.org/10.1007/S00210-003-0833-1.

6 Muthukrishnan S, Shete VS, Sanan TT, Vyas S, Oottikkal S, Porter LM, *et al.* 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* 2012;**25**:1247. https://doi.org/10.1002/POC.3002.

7 Billecke S, Draganov D, Counsell R, Stetson P, Watson C, Hsu C, *et al.* HUMAN SERUM PARAOXONASE (PON1) ISOZYMES Q AND R HYDROLYZE LACTONES AND CYCLIC CARBONATE ESTERS. *Drug Metab Dispos* 2000;**28**:1335–42.

8 Draganov DI, Teiber JF, Speelman A, Osawa Y, Sunahara R, Du BN La. Human paraoxonases (PON1, PON2, and PON3) are lactonases with overlapping and distinct substrate specificities. *J Lipid Res* 2005;**46**:1239. https://doi.org/10.1194/jlr.M400511-JLR200.

9 Bar-Rogovsky H, Hugenmatter A, Tawfik DS. The Evolutionary Origins of Detoxifying Enzymes THE MAMMALIAN SERUM PARAOXONASES (PONs) RELATE TO BACTERIAL HOMOSERINE LACTONASES *. *J Biol Chem* 2013;**288**:23914–27. https://doi.org/10.1074/jbc.M112.427922.

10 Khersonsky O, Tawfik DS. Structure-Reactivity Studies of Serum Paraoxonase PON1 Suggest that Its Native Activity Is Lactonase. *Biochemistry* 2005;**44**:6371–82. https://doi.org/10.1021/bi047440d.

11 Stoltz DA, Ozer EA, Ng CJ, Yu JM, Reddy ST, Lusis AJ, *et al.* Paraoxonase-2 deficiency enhances Pseudomonas aeruginosa quorum sensing in murine tracheal epithelia. *Am J Physiol - Lung Cell Mol Physiol* 2007;**292**:852–60. https://doi.org/10.1152/AJPLUNG.00370.2006/ASSET/IMAGES/LARGE/ZH50030748090008.JPEG.

12 Teiber JF, Horke S, Haines DC, Chowdhary PK, Xiao J, Kramer GL, *et al.* Dominant Role of Paraoxonases in Inactivation of the Pseudomonas aeruginosa Quorum-Sensing Signal N-(3-Oxododecanoyl)-l-Homoserine Lactone. *Infect Immun* 2008;**76**:2512. https://doi.org/10.1128/IAI.01606-07.

13 Shih DM, Gu L, Xia YR, Navab M, Li WF, Hama S, *et al.* Mice lacking serum paraoxonase are susceptible to organophosphate toxicity and atherosclerosis. *Nat 1998 3946690* 1998;**394**:284–7. https://doi.org/10.1038/28406.

14 Tward A, Xia YR, Wang XP, Shi YS, Park C, Castellani LW, *et al.* Decreased atherosclerotic lesion formation in human serum paraoxonase transgenic mice. *Circulation* 2002;**106**:484–90. https://doi.org/10.1161/01.CIR.0000023623.87083.4F.

15 Aviram M, Rosenblat M, Bisgaier CL, Newton RS, Primo-Parmo SL, Du BN La. 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* 1998;**101**:1581–90.

16 Huang Y, Wu Z, Riwanto M, Gao S, Levison BS, Gu X, *et al.* Myeloperoxidase, paraoxonase-1, and

HDL form a functional ternary complex. *J Clin Invest* 2013;**123**:3815. https://doi.org/10.1172/JCI67478DS1.

17 Variji A, Shokri Y, Fallahpour S, Zargari M, Bagheri B, Abediankenari S, *et al.* 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* 2019;**280**:7–13. https://doi.org/10.1016/J.ATHEROSCLEROSIS.2018.11.004.

18 Altenhöfer S, Witte I, Teiber JF, Wilgenbus P, Pautz A, Li H, *et al.* One Enzyme, Two Functions: PON2 PREVENTS MITOCHONDRIAL SUPEROXIDE FORMATION AND APOPTOSIS INDEPENDENT FROM ITS LACTONASE ACTIVITY*. *J Biol Chem* 2010;**285**:24398. https://doi.org/10.1074/JBC.M110.118604.

19 Devarajan A, Bourquard N, Hama S, Navab M, Grijalva VR, Morvardi S, *et al.* Paraoxonase 2 Deficiency Alters Mitochondrial Function and Exacerbates the Development of Atherosclerosis. *Antioxid Redox Signal* 2011;**14**:341. https://doi.org/10.1089/ARS.2010.3430.

20 Krüger M, Pabst AM, Al-Nawas B, Horke S, Moergel M. Paraoxonase-2 (PON2) protects oral squamous cell cancer cells against irradiation-induced apoptosis. *J Cancer Res Clin Oncol 2015 14110* 2015;**141**:1757–66. https://doi.org/10.1007/S00432-015-1941-2.

21 Rull A, García R, Fernández-Sender L, García-Heredia A, Aragonès G, Beltrán-Debón R, *et al.* Serum paraoxonase-3 concentration is associated with insulin sensitivity in peripheral artery disease and with inflammation in coronary artery disease. *Atherosclerosis* 2012;**220**:545–51. https://doi.org/10.1016/J.ATHEROSCLEROSIS.2011.11.021.

22 Shih DM, Xia YR, Wang XP, Wang SS, Bourquard N, Fogelman AM, *et al.* Decreased Obesity and Atherosclerosis in Human Paraoxonase 3 Transgenic Mice. *Circ Res* 2007;**100**:1200. https://doi.org/10.1161/01.RES.0000264499.48737.69.

23 Meyer WK, Jamison J, Richter R, Woods SE, Partha R, Kowalczyk A, *et al.* Ancient convergent losses of Paraoxonase 1 yield potential risks for modern marine mammals. *Science (80-)* 2018;**361**:591–4. https://doi.org/10.1126/science.aap7714.

24 Rodrigo L, Gil F, Hernandez AF, Marina A, Vazquez J, Pla A. Purification and characterization of paraoxon hydrolase from rat liver. *Biochem J* 1997;**321**:595–601. https://doi.org/10.1042/bj3210595.

25 Eddy SR. Profile hidden Markov models. *Bioinformatics* 1998;**14**:755–63. https://doi.org/10.1093/bioinformatics/14.9.755.

26 Mirarab S, Nguyen N, Guo S, Wang LS, Kim J, Warnow T. PASTA: Ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. *J Comput Biol* 2015;**22**:377–86. https://doi.org/10.1089/cmb.2014.0156.

27 Steenwyk JL, Buida TJ, Li Y, Shen XX, Rokas A. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLOS Biol* 2020;**18**:e3001007. https://doi.org/10.1371/JOURNAL.PBIO.3001007.

28 Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol* 2008;**25**:1307–20. https://doi.org/10.1093/molbev/msn067.

29 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*

1990;**215**:403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

30 Löytynoja A, Goldman N. WebPRANK: A phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics* 2010;**11**:579. https://doi.org/10.1186/1471-2105-11-579.

31 Larsson A. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 2014;**30**:3276–8. https://doi.org/10.1093/bioinformatics/btu531.

32 Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O, *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol* 2010;**59**:307–21. https://doi.org/10.1093/sysbio/syq010.

33 Lefort V, Longueville JE, Gascuel O. SMS: Smart Model Selection in PhyML. *Mol Biol Evol* 2017;**34**:2422–4. https://doi.org/10.1093/molbev/msx149.

34 Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 1992;**8**:275–82. https://doi.org/10.1093/bioinformatics/8.3.275.

35 Akaike H. A New Look at the Statistical Model Identification. *IEEE Trans Automat Contr* 1974;**19**:716–23. https://doi.org/10.1109/TAC.1974.1100705.

36 Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol* 2007;**24**:1586–91. https://doi.org/10.1093/molbev/msm088.

37 Huelsenbeck JP, Crandall KA. *PHYLOGENY ESTIMATION AND HYPOTHESIS TESTING USING MAXIMUM LIKELIHOOD*. vol. 28. 1997.

38 *GitHub - ncbi/sra-tools: SRA Tools*. n.d. URL: https://github.com/ncbi/sra-tools (Accessed 30 June 2021).

39 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJournal* 2011;**17**:10. https://doi.org/10.14806/ej.17.1.200.

40 Krueger F. *Babraham Bioinformatics - Trim Galore!*. 2012. URL: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (Accessed 30 June 2021).

41 Andrews S. *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*. 2010. URL: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

42 Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, *et al.* From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr Protoc Bioinforma* 2013;**11**:11.10.1. https://doi.org/10.1002/0471250953.bi1110s43.

43 Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 2013;**14**:178–92. https://doi.org/10.1093/bib/bbs017.

44 Yang Z, Swanson WJ. Codon-Substitution Models to Detect Adaptive Evolution that Account for Heterogeneous Selective Pressures Among Site Classes. *Mol Biol Evol* 2002;**19**:49–57. https://doi.org/10.1093/oxfordjournals.molbev.a003981.

45 Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007;**24**:1586–91. https://doi.org/10.1093/molbev/msm088.

46 Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 2005;**22**:2472–9. https://doi.org/10.1093/molbev/msi237.

47 Yang Z, Wong WSW, Nielsen R. Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Mol Biol Evol* 2005;**22**:1107–18. https://doi.org/10.1093/MOLBEV/MSI097.

48 Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, *et al.* Gene-wide identification of episodic selection. *Mol Biol Evol* 2015;**32**:1365–71. https://doi.org/10.1093/molbev/msv035.

49 Harel M, Aharoni A, Gaidukov L, Brumshtein B, Khersonsky O, Meged R, *et al.* Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes. *Nat Struct Mol Biol* 2004;**11**:412–9. https://doi.org/10.1038/nsmb767.

50 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 2004;**25**:1605–12. https://doi.org/10.1002/jcc.20084.

51 Fraczkiewicz R, Braun W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comput Chem* 1998;**19**:319–33.

52 Clark NL, Findlay GD, Yi X, MacCoss MJ, Swanson WJ. Duplication and selection on abalone sperm lysin in an allopatric population. *Mol Biol Evol* 2007;**24**:2081–90. https://doi.org/10.1093/molbev/msm137.

53 Wolf YI, Koonin E V. Genome reduction as the dominant mode of evolution. *Bioessays* 2013;**35**:829. https://doi.org/10.1002/BIES.201300037.

54 Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2005;**33**:D501. https://doi.org/10.1093/NAR/GKI025.

55 Aviram M, Hardak E, Vaya J, Mahmood S, Milo S, Hoffman A, *et al.* Human Serum Paraoxonases (PON1) Q and R Selectively Decrease Lipid Peroxides in Human Coronary and Carotid Atherosclerotic Lesions. *Circulation* 2000;**101**:2510–7. https://doi.org/10.1161/01.CIR.101.21.2510.

56 Allen KN, Vázquez-Medina JP. Natural Tolerance to Ischemia and Hypoxemia in Diving Mammals: A Review. *Front Physiol* 2019:1199. https://doi.org/10.3389/fphys.2019.01199.

57 Smith RS, Iglewski BH. P. aeruginosa quorum-sensing systems and virulence. *Curr Opin Microbiol* 2003;**6**:56–60. https://doi.org/10.1016/S1369-5274(03)00008-0.

58 Laumer CE, Fernández R, Lemer S, Combosch D, Kocot KM, Riesgo A, *et al.* Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proc R Soc B* 2019;**286**:. https://doi.org/10.1098/RSPB.2019.0831.

59 Philippe H, Poustka AJ, Chiodin M, Hoff KJ, Dessimoz C, Tomiczek B, *et al.* Mitigating Anticipated Effects of Systematic Errors Supports Sister-Group Relationship between Xenacoelomorpha and Ambulacraria. *Curr Biol* 2019;**29**:1818-1826.e6. https://doi.org/10.1016/J.CUB.2019.04.009.

60 Kapli P, Telford MJ. Topology-dependent asymmetry in systematic errors affects phylogenetic

placement of Ctenophora and Xenacoelomorpha. *Sci Adv* 2020;**6**:5162–73. https://doi.org/10.1126/SCIADV.ABC5162/SUPPL_FILE/ABC5162_SM.PDF.

61 Li Y, Shen XX, Evans B, Dunn CW, Rokas A. Rooting the Animal Tree of Life. *Mol Biol Evol* 2021;**38**:4322–33. https://doi.org/10.1093/MOLBEV/MSAB170.

62 Redmond AK, McLysaght A. Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding. *Nat Commun 2021 121* 2021;**12**:1–14. https://doi.org/10.1038/s41467-021-22074-7.

63 Zverkov OA, Mikhailov K V., Isaev S V., Rusin LY, Popova O V., Logacheva MD, *et al.* Dicyemida and Orthonectida: Two stories of body plan simplification. *Front Genet* 2019;**10**:443. https://doi.org/10.3389/FGENE.2019.00443/BIBTEX.

64 Kayal E, Bentlage B, Sabrina Pankey M, Ohdera AH, Medina M, Plachetzki DC, *et al.* Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. *BMC Evol Biol* 2018;**18**:1–18. https://doi.org/10.1186/S12862-018-1142-0/FIGURES/7.

65 Chen MY, Liang D, Zhang P. Selecting Question-Specific Genes to Reduce Incongruence in Phylogenomics: A Case Study of Jawed Vertebrate Backbone Phylogeny. *Syst Biol* 2015;**64**:1104–20. https://doi.org/10.1093/SYSBIO/SYV059.

66 Mongiardino Koch N, Thompson JR, Hiley AS, McCowin MF, Armstrong AF, Coppard SE, *et al.* Phylogenomic analyses of echinoid diversification prompt a re-evaluation of their fossil record. *Elife* 2022;**11**:. https://doi.org/10.7554/ELIFE.72460.

67 Giribet G, Edgecombe GD. The Phylogeny and Evolutionary History of Arthropods. *Curr Biol* 2019;**29**:R592–602. https://doi.org/10.1016/J.CUB.2019.04.057.

68 Egger B, Lapraz F, Tomiczek B, Müller S, Dessimoz C, Girstmair J, *et al.* A Transcriptomic-Phylogenomic Analysis of the Evolutionary Relationships of Flatworms. *Curr Biol* 2015;**25**:1347. https://doi.org/10.1016/J.CUB.2015.03.034.

69 Kocot KM, Poustka AJ, Stöger I, Halanych KM, Schrödl M. New data from Monoplacophora and a carefully-curated dataset resolve molluscan relationships. *Sci Reports 2020 101* 2020;**10**:1–8. https://doi.org/10.1038/s41598-019-56728-w.

70 Magalhães WF, Hutchings P, Oceguera-Figueroa A, Martin P, Schmelz RM, Wetzel MJ, *et al.* Segmented worms (Phylum Annelida): a celebration of twenty years of progress through Zootaxa and call for action on the taxonomic work that remains. *Zootaxa* 2021;**4979**:190–211–190–211. https://doi.org/10.11646/ZOOTAXA.4979.1.18.

71 Wörheide G, Dohrmann M, Erpenbeck D, Larroux C, Maldonado M, Voigt O, *et al.* Deep Phylogeny and Evolution of Sponges (Phylum Porifera). *Adv Mar Biol* 2012;**61**:1–78. https://doi.org/10.1016/B978-0-12-387787-1.00007-6.

72 Hodda M. Phylum Nematoda: a classification, catalogue and index of valid genera, with a census of valid species. *Zootaxa* 2022;**5114**:001–289.