

Integrative genome and transcriptome mining identifies a diverse repertoire of cyclooxygenase/prostaglandin- endoperoxide synthase-like candidates in Porifera

Sebastiano Scibelli

scibseb@libero.it

Independent researcher

Research Article

Keywords: Porifera, cyclooxygenase, prostaglandin-endoperoxide synthase, molecular evolution, comparative genomics, phylogeny

Posted Date: August 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10829400/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Cyclooxygenases/prostaglandin-endoperoxide synthases (COX/PTGS) catalyse the committed step of prostanoid biosynthesis, yet their evolutionary distribution outside Bilateria and Cnidaria remains unresolved, and previous metazoan surveys recovered no convincing Porifera homologues. We integrated translated transcriptomes, annotated proteins, 213 sponge genome assemblies and 75 public RNA-seq runs to test for COX/PTGS-like genes in Porifera. Candidates were evaluated by reciprocal curated-database annotation, exact amino-acid deduplication, domain architecture, P05979-equivalent residues, genomic structure, splice-aware RNA reconstruction and phylogenetic placement. Genome mining recovered 67 broad models; reconciliation of 65 RNA-evaluable models yielded 47 non-overlapping loci in 22 species, and 15 stringent assembly-level candidates showed complete predicted-intron reconstruction and exact transcript-derived ORF confirmation. In 18 non-redundant focal candidates, all seven monitored peroxidase/haem residues and the Tyr385-centred radical/positioning core were retained, while channel-associated substitutions separated canonical-compatible from product-positioning, channel-constriction and divergent substrate-entry profiles. A domain-restricted maximum-likelihood analysis separated all 54 primary Porifera candidates plus 12 experimentally characterised COX proteins from 33 strict non-COX animal haem-peroxidase comparators with maximal SH-aLRT/UFBoot support; the family partition persisted under an LG+C60 posterior-mean-site-frequency sensitivity analysis. These results revise the apparent absence of COX/PTGS-like homologues from Porifera and support a diverse, taxonomically structured candidate repertoire. By extending the minimum supported phylogenetic depth of the COX/PTGS-like lineage beyond the cnidarian–bilaterian node, these findings are consistent, under a vertical-inheritance interpretation, with the presence of a COX/PTGS-like homologue in the last common ancestor shared by Porifera, Cnidaria and Bilateria. Catalytic activity, product specificity and duplication history remain to be established experimentally.

1. Introduction

Cyclooxygenases (COX), also known as prostaglandin-endoperoxide synthases (PTGS) or prostaglandin G/H synthases (PGHS), are bifunctional haem enzymes that catalyse two consecutive reactions in prostanoid biosynthesis. A cyclooxygenase site converts arachidonic acid to prostaglandin G₂, and a spatially distinct peroxidase site reduces this intermediate to prostaglandin H₂, the precursor of multiple prostaglandins and thromboxanes (Simmons et al. 2004; Rouzer and Marnett 2020). Vertebrate COX-1 and COX-2 proteins are approximately 600 amino acids long and share an N-terminal signal peptide, one epidermal growth factor (EGF)-like module, a membrane-binding region formed by four amphipathic helices and a C-terminal catalytic domain containing the cyclooxygenase and peroxidase sites (Chandrasekharan and Simmons 2004; Miciaccia et al. 2021). Human PTGS1 and PTGS2 comprise 11 and 10 exons, respectively, whereas two functionally characterised octocoral COX genes each comprise nine exons; conserved exon boundaries coexist with lineage-specific intron gain and loss (Järving et al. 2004; Chandrasekharan and Simmons 2004). Within the catalytic region, Arg120, Tyr355 and Glu524 define the constriction at the entrance of the cyclooxygenase channel, Tyr385 initiates substrate oxygenation through a tyrosyl radical, Ser530 is the aspirin-acetylation site, and additional residues contribute to the haem/peroxidase environment and channel geometry (Simmons et al. 2004; Rouzer and Marnett 2020).

COX/PTGS proteins belong to the broader peroxidase-cyclooxygenase superfamily, which also includes peroxidasins, peroxinectins, myeloperoxidase, eosinophil peroxidase, lactoperoxidase and thyroid peroxidase (Zámocký et al. 2015). Consequently, a significant local similarity to a vertebrate COX query does not by itself establish COX/PTGS homology. Reliable identification requires discrimination from related peroxidases, assessment of complete protein architecture and catalytic-site context, and evolutionary placement relative to experimentally characterised or well-supported COX sequences.

The distribution of COX/PTGS homologues across Metazoa is uneven. Functional or strongly supported sequences have been described in vertebrates, octocorals, molluscs, annelids and crustaceans, whereas lineage-specific duplications and marked sequence divergence complicate orthology assignment (Koljak et al. 2001; Valmsen et al. 2001; Havird et al. 2008, 2015; Kawamura et al. 2014). Havird et al. (2015) combined public genomes and approximately 250 invertebrate transcriptomes to reconstruct COX evolution, recovered 40 previously unrecognised putative orthologues, and reported no convincing COX homologue from the poriferan resources available at that time. The subsequent expansion of sponge genomes, RNA-sequencing (RNA-seq) datasets and standardised translated transcriptomes now permits a substantially more sensitive test of that apparent absence.

Porifera are particularly relevant to this question because they occupy an early-diverging position within Metazoa and display complex lipid metabolism, symbiotic interactions, regeneration and reproductive processes in which lipid mediators may participate. Historical work in *Geodia cydonium* reported that arachidonic acid was metabolised predominantly to prostaglandin E₂ and implicated this mediator in increased DNA synthesis, although the enzyme responsible for endoperoxide formation was not molecularly identified (Gramzow et al. 1989). A combined lipidomic and transcriptomic investigation of *Phakellia ventilabrum* further documented extensive reproductive-stage lipid remodelling and provided an independent molecular context for evaluating lipid-metabolism genes in this species (Koutsouveli et al. 2022). Nevertheless, sequence similarity, lipid detection and gene expression represent different evidence types and should not be conflated with biochemical validation of a COX reaction.

The present study therefore used a staged, source-aware strategy to identify and prioritise Porifera COX/PTGS-like candidates. Translated transcriptomes, annotated protein records and unannotated genomes were screened independently. Candidate loci were modelled directly from genomic DNA, tested against curated protein databases, examined for COX/PTGS-like domain and residue architecture, and evaluated using RNA-seq, splice-aware mapping, transcript reconstruction and open reading frame (ORF) recovery. A broad animal haem peroxidase phylogeny was then used to test whether the retained sponge sequences segregated with published COX/PTGS proteins rather than with non-COX peroxidases. The resulting classifications describe putative COX/PTGS-like homologues and levels of in silico support; they do not constitute a demonstration of enzymatic activity.

2. Materials and Methods

2.1. Evidence streams and reporting levels

Candidate discovery was organised into three independent streams: translated Porifera transcriptomes, annotated/browser-derived proteins and protein-to-genome mining of unannotated assemblies (Fig. 1).

Candidate status was not assigned from a single similarity hit. Reciprocal curated-database annotation, exclusion of non-COX animal haem peroxidases, protein completeness, domain architecture and P05979-equivalent residues were assessed across streams. Genome-derived candidates were additionally evaluated for gene structure, splice-aware RNA support, transcript reconstruction and ORF agreement.

Counts were reported by analytical level rather than combined across non-independent steps. Broad genome models, coding-sequence-clean (CDS-clean) models, RNA-evaluated genome models, reconciled loci, globally unique proteins, integrated candidates and phylogenetic representatives therefore describe nested or partially overlapping evidence layers. Specifically, 67 broad genome models yielded 65 RNA-evaluable models, which were reconciled by genomic overlap to 47 non-overlapping loci and by global exact amino-acid deduplication to 38 unique protein sequences. A 45-record pre-phylogenetic matrix was used for uniform domain, residue and gene-structure comparison. The final primary phylogenetic roster was generated independently after exact amino-acid deduplication and a 500–650 amino-acid length screen and contained 54 non-redundant study candidates together with published COX/PTGS references and non-COX comparators.

2.2. Input transcriptomes, genomes and RNA-seq datasets

Translated Porifera transcriptomes were obtained from the Community bAsed Non-bilaTeriAn Transcriptome Archive (CANTATA), Porifera release version 1.0 (Vargas and Wörheide 2024; Zenodo record 8208114). Fifty-seven TransDecoder peptide sets were merged into a single FASTA file containing 4,183,296 protein sequences. Species labels, source archive names and retained peptide-file provenance are listed in Supplementary Table S1.

Two hundred and thirteen Porifera genomic assemblies were screened as nucleotide FASTA files. Assembly accessions, organism labels and source paths are listed in Supplementary Table S2. The assembly inventory included independent GenBank (GCA) and RefSeq (GCF) assembly representations when both had been downloaded; these were retained during screening and reconciled at the protein and species levels downstream. For each assembly, Supplementary Table S2 records whether at least one model met the broad tier A/B criteria, together with the number and identifiers of retained models.

RNA-seq support was evaluated using 75 unique public European Nucleotide Archive (ENA) and Sequence Read Archive (SRA) runs, represented by 78 target-to-run mappings because three runs were informative for more than one target label. Run accessions, library layout, target species, proxy status where applicable and file provenance are reported in Supplementary Table S3. Downloaded FASTQ files were checked against reported Message-Digest Algorithm 5 (MD5) checksums when checksum metadata were available.

2.3. CANTATA translated-transcriptome screening

The merged CANTATA peptide FASTA was formatted as a Basic Local Alignment Search Tool (BLAST+) protein database and queried with sheep COX-1/PTGS1 (UniProt P05979) and the octocoral *Plexaura homomalla* COX-like protein Q962I8 using BLASTp 2.14.0+ (Camacho et al. 2009). Searches used an e-value threshold of 1e-10 and retained all qualifying targets. Tabular output contained query and subject identifiers and lengths, percentage identity, alignment length, e-value, bit score, query coverage and high-scoring segment pair (HSP) coverage.

Hits were parsed at the complete-subject-header level because assembler-derived identifiers were not globally unique across transcriptomes. Priority candidates required e-value $\leq 1e-10$, query coverage $\geq 70\%$, HSP coverage $\geq 70\%$, alignment length ≥ 400 amino acids and subject length of 450–900 amino acids. Partial or domain-level candidates were retained separately when e-value was $\leq 1e-10$, bit score was ≥ 100 , query coverage was $\geq 30\%$ and alignment length was ≥ 180 amino acids, but one or more priority criteria were not satisfied. Candidate proteins were searched reciprocally against a local Swiss-Prot database as described below. All 73 source-level decisions, including non-COX classifications, were retained in Supplementary Table S4.

2.4. Annotated-protein/browser search

Human COX-2/PTGS2 (UniProt P35354; 604 amino acids) was queried against the National Center for Biotechnology Information (NCBI) nr_cluster_seq protein database in a Porifera-focused NCBI BLASTp browser search performed on 16 May 2026 (NCBI request identifier [RID] 0G48C3Z1016). The exported Extensible Markup Language (XML) file recorded BLASTP 2.17.0+, BLOSUM62, e-value 0.05, gap-open/gap-extension costs of 11/1 and low-complexity filtering disabled. The output was configured to display up to 300 descriptions; complete FASTA sequences were retrieved for all 100 unique subject accessions represented in the exported results and retained for reciprocal classification. Browser coverage, annotation string and reciprocal Swiss-Prot results were integrated for each record. Proteins were accepted as COX/PTGS-like only when the reciprocal best hit was a prostaglandin G/H synthase, cyclooxygenase, PTGS, PGH1 or PGH2 protein. Reciprocal hits to peroxidase, peroxinectin, myeloperoxidase, eosinophil peroxidase, lactoperoxidase, thyroid peroxidase, dual oxidase (DUOX) or other non-COX animal haem peroxidases were excluded as COX/PTGS-like candidates but retained for transparency and possible comparator selection. All 100 decisions are reported in Supplementary Table S5.

2.5. Protein-to-genome screening and locus definition

The 213 genome assemblies were screened using TBLASTN 2.12.0+ with a query panel containing vertebrate, octocoral and Porifera COX/PTGS-like proteins and an e-value threshold of 1e-5. HSPs were grouped by assembly, contig and strand when separated by no more than 20 kb. A locus was assigned to tier A when cumulative query coverage was $\geq 70\%$, bit score was ≥ 200 and e-value was $\leq 1e-20$, or when an individual HSP independently satisfied query coverage $\geq 70\%$ and bit score ≥ 200 . Tier B loci required cumulative query coverage $\geq 40\%$, bit score ≥ 100 and e-value $\leq 1e-10$, or at least three coherent HSPs with bit score ≥ 100 . Tier A and B loci were carried forward.

For each retained candidate genomic region, a 200-kb window was extracted to preserve intronic and flanking context. Protein-to-genome models were generated with miniprot 0.18-r281 (Li 2023) using `--gff --aln --outs = 0.35 --outc = 0.25 --outn = 5000 -N 1000`. Genomic coordinates, strand, CDS exon

coordinates, intron coordinates, splice motifs, spliced CDS and predicted protein were exported. Spliced CDS translations were compared with the miniprot proteins. The initial conservative model filter required miniprot query coverage $\geq 80\%$, region-level BLAST query coverage $\geq 80\%$, protein length ≥ 450 amino acids, a PF03098 animal haem-peroxidase-superfamily hit and no frameshift or internal stop. Broader models were retained separately if reciprocal annotation supported COX/PTGS-like identity. This produced a broad 67-record layer and a 26-record CDS-clean subset.

2.6. Reciprocal protein annotation

Candidate proteins were searched against curated protein databases with BLASTp (Camacho et al. 2009) or DIAMOND blastp (Buchfink et al. 2021). Swiss-Prot BLASTp searches used e-value $\leq 1e-10$. DIAMOND 2.1.25 genome-set searches used –sensitive, e-value $\leq 1e-5$ and up to 25 retained hits, whereas reconstructed ORF searches used DIAMOND 2.1.24, –sensitive, e-value $\leq 1e-10$ and up to 10 retained hits. For each query, the best hit was selected by descending bit score, ascending e-value, descending percentage identity and descending query coverage. COX/PTGS-like classifications required a best curated hit to prostaglandin G/H synthase, cyclooxygenase, PTGS, PGH1 or PGH2. Non-COX animal haem peroxidase best hits were classified separately.

2.7. Exact deduplication and cross-source sequence reconciliation

Exact amino-acid deduplication compared full-length protein sequences independently of FASTA headers and line wrapping. Only records with 100% complete-sequence identity were collapsed. Non-identical records were retained, including possible isoforms, alleles, assembly-specific models and paralogues. Pairwise cross-source BLASTp was used to quantify additional relationships. Exact cross-source matches required 100% identity and 100% coverage of both sequences. Very close but non-identical matches required $\geq 95\%$ identity and $\geq 90\%$ query and subject coverage. Related high-coverage matches required $\geq 80\%$ identity and $\geq 70\%$ query coverage. Partial-overlap matches were reported separately and were not collapsed. Source-specific identifiers and retained representatives were recorded so that every collapsed record remained traceable.

2.8. Domain architecture

The 45-record integrated candidate set was annotated uniformly with InterProScan 5.74–105.0 (Jones et al. 2014). Tab-separated-values (TSV), XML and General Feature Format version 3 (GFF3) outputs were generated with Gene Ontology and pathway annotations enabled. Candidate-level architecture was evaluated using strict EGF-like calls, broader EGF/laminin-like calls, Pfam PF03098 animal haem peroxidase, Protein Analysis Through Evolutionary Relationships (PANTHER) classification PTHR11903 prostaglandin G/H synthase and NCBI Conserved Domain Database (CDD) cd09816 prostaglandin-endoperoxide synthase. HMMER 3.4/Pfam-A support (Eddy 2011) used a conditional domain i-E-value threshold of $1e-3$ during genome-model evaluation.

To eliminate source-dependent missing values from the phylogenetic metadata, all 157 proteins used in the length-sensitivity analysis were subsequently rescanned under one protocol. EGF-like domains were called with the PROSITE PS50026 profile at level 0 using ps_scan and PROSITE release 2026_02 (Sigrist et al. 2013). Supporting HMMER searches used Pfam PF00008.34 (EGF-like) and PF03098.21 (animal haem peroxidase) with model-specific gathering thresholds. For proteins without a PS50026 hit, evaluability of the EGF-bearing N-terminal region was assessed from the untrimmed MAFFT alignment (Katoh and Standley 2013) across columns corresponding to sheep COX-1 P05979 residues 32–70. Proteins shorter than 500 amino acids that represented less than 50% of this interval were classified as EGF-non-evaluable partial sequences rather than EGF-negative. All other no-hit proteins were reported explicitly as profile-not-detected. Domain coordinates, evaluability calls and partial-protein inventories are provided in Supplementary Table S6. EGF status was interpreted together with the catalytic core, reciprocal annotation and residue profile, not as an independent proof of COX/PTGS identity.

2.9. P05979-equivalent residue analysis

Candidate proteins were aligned with reference COX/PTGS proteins using MAFFT L-INS-i (Katoh and Standley 2013). Sheep COX-1/PTGS1 UniProt P05979 was used as the numbering anchor. The P05979-equivalent positions R120, Q203, H207, Y355, Y385, H386, H388, E524, S530 and G533 were mapped by scanning the aligned P05979 sequence and incrementing the residue number only at non-gap positions. Residues occupying the corresponding alignment columns were extracted for every candidate. Scores were reported for all ten positions, the six COX-channel/active-site positions R120, Y355, Y385, E524, S530 and G533, and the four peroxidase/haem-context positions Q203, H207, H386 and H388. The complete sequence-by-position matrix is provided in Supplementary Table S7. Residue conservation was interpreted as compatibility with a COX/PTGS-like architecture, not as proof of catalytic activity.

2.10. Mutagenesis-informed functional-site assessment

A focused mechanistic comparison was performed on the 15 strict genome/RNA-supported candidates together with five focal curated or lineage-divergent candidates. Exact amino-acid duplicates were removed, yielding 18 non-redundant proteins from 11 species-level labels. Twelve experimentally characterised animal COX proteins were included as positive references: COX-1/COX-2 pairs from *Ovis aries*, *Homo sapiens*, *Mus musculus* and *Danio rerio*, the two *Gersemia fruticosa* isoforms and the 15R/15S *Plexaura homomalla* enzymes (Koljak et al. 2001; Grosser et al. 2002; Järving et al. 2004; Valmsen et al. 2001, 2004). Proteins were aligned with MAFFT 7.526 L-INS-i using –localpair –maxiterate 1000 (Katoh and Standley 2013). Eighteen positions were mapped by reference-projected numbering from full-length sheep COX-1 P05979: R120, Q203, H204, T206, H207, H309, Y348, Y349, Y355, Y385, H386, W387, H388, Y504, E524, S530, G533 and L534. These positions represent substrate entry and channel constriction, the peroxidase (POX)/haem environment, the Tyr385-centred radical and positioning network, and upper-channel/product-geometry determinants.

Primary mutagenesis and functional studies were curated for each mapped site (DeWitt et al. 1990; Shimokawa et al. 1990; Shimokawa and Smith 1991, 1992; Bhattacharyya et al. 1996; Thuresson et al. 2000, 2001; Seibold et al. 2003; Valmsen et al. 2004; Rogge et al. 2009). Evidence was distinguished as: an exact amino-acid state present in an experimentally active COX; an exact substitution tested directly; a different substitution tested at the same site; or site-level mechanistic evidence without an exact-state experiment. No additive activity score was calculated. Candidates were assigned descriptive sequence profiles only: canonical-channel-compatible, product-positioning variant, channel-constriction/radical-network variant or divergent substrate-entry. These labels prioritise biochemical hypotheses and do not establish COX or POX activity.

2.11. RNA-seq screening, splice support and transcript reconstruction

RNA-seq validation was restricted to genome-derived models. Bowtie2 2.5.2 (Langmead and Salzberg 2012) was first used to map reads to predicted spliced CDS sequences and identify candidate-informative runs. Candidate-associated reads were then mapped to the corresponding genomic spans using STAR 2.7.11b (Dobin et al. 2013) with `--outFilterMultimapNmax 50, --alignIntronMin 20, --alignIntronMax 250000` and coordinate-sorted binary alignment/map (BAM) output. Exonic breadth was calculated with bedtools 2.31 coverage-split (Quinlan and Hall 2010); BAM sorting, indexing and summaries used samtools 1.19.2 (Danecek et al. 2021).

Strong splice support required exon breadth $\geq 80\%$, at least 10 mapped reads and at least two-read support for one or more exact predicted junctions, with two exact junctions required when two or more introns were available. Records with exon breadth $\geq 80\%$ and at least 10 reads but limited junction recovery were reported as strong exonic/low-junction support. Partial support required exon breadth $\geq 30\%$ or at least one exact junction. Species-matched and same-genus proxy RNA data were distinguished in the output.

Targeted transcript reconstruction was performed with StringTie 2.2.1 (Pertea et al. 2015) using `-c 1 -j 1 -m 100 -f 0.01`. Reconstructed transcripts were compared with predicted exon-intron models. The longest ORF in each selected transcript was identified by six-frame translation, compared with the genome-derived candidate and searched against UniProtKB using DIAMOND. Exact ORF confirmation required the reconstructed protein to match the expected genome-derived protein across the complete comparable sequence. Candidate-associated transcription, splice confirmation and exact protein-model confirmation were reported as separate evidence states.

2.12. Host-attribution and genomic-context audit

All 67 broad genome-derived models were screened for assembly identity and level, contig length, gene-versus-assembly guanine-cytosine (GC) content deviation, exon-intron structure, and available RNA support. The 26 CDS-clean models were mapped to 23 distinct 200-kb candidate regions for translated-neighbourhood analysis. The candidate span and 10 kb on each side were excluded, and the remaining sequence was divided into 20-kb bins. Bins were searched against a local Swiss-Prot database with DIAMOND 2.1.25 blastx using `--sensitive, e-value  $\leq 1e-5$` and up to 20 targets per bin. Swiss-Prot organism-taxonomy (OX) identifiers were classified with the NCBI Taxonomy rankedlineage.dmp snapshot dated 1 August 2026 (Schoch et al. 2020). The primary screen retained alignments with bit score ≥ 50 and aligned length ≥ 30 amino acids and evaluated taxa within 80% of the best bit score in each bin. Sensitivity analyses used thresholds of 60/40 and 80/50 for bit score/aligned length. Final host-attribution calls integrated genomic context with same-species or proxy RNA evidence, multiexon structure and assembly metadata; translated-neighbourhood similarity was not treated as independent proof of host origin. Candidate-level results are reported in Supplementary Table S8.

2.13. Evidence classes

Evidence classes were assigned before phylogenetic analysis. Class A denoted the same or nearly identical protein recovered independently from annotated or transcriptomic evidence and genome mining, together with RNA-seq/StringTie support, reconstructed ORF and coherent exon-intron structure. Class B denoted a new genome-derived candidate with strong RNA-seq/StringTie, ORF, domain and residue support but no equivalent record in the pre-existing positive source set. Class C denoted reciprocal/domain/residue-supported COX/PTGS-like candidates without complete cross-stream genome/RNA confirmation. C1 was used for priority transcriptome candidates and C2 for full annotated/browser candidates. Class D denoted partial or secondary evidence, subdivided as D1 for transcriptome fragments and D2 for the partial browser record. Class E denoted unresolved or excluded records. Definitions are provided in Supplementary Table S9.

2.14. Phylogenetic inference and sensitivity analysis

The evolutionary analysis combined non-redundant study candidates, the COX and animal haem-peroxidase-superfamily sequence set of Havird et al. (2015), additional curated COX references and the documented strict non-COX comparator panel: 27 peroxidasin (PXDN)-like and six myeloperoxidase (MPO)-like proteins. These comparators sampled the non-COX lineages represented in the analysed panel rather than serving as arbitrary distant outgroups (Zámocký et al. 2015). The primary dataset retained COX/PTGS-like proteins of 500–650 amino acids and contained 128 taxa: 95 candidate or reference COX/PTGS-like proteins, including 54 study candidates, and 33 strict non-COX comparators. A length-sensitivity dataset retained partial and extended candidate/reference records and contained 157 taxa: 124 candidate or reference roster proteins and the same 33 non-COX comparators. Taxon metadata, source, family call, evidence tier, sequence length and publication label are provided in Supplementary Tables S10 and S11.

Protein sequences were aligned with MAFFT 7.526 (Katoh and Standley 2013) using `--auto` and trimmed with trimAl 1.5.0 (Capella-Gutiérrez et al. 2009) using `-gt 0.1`. Maximum-likelihood trees were inferred with IQ-TREE 3.1.3 (Wong et al. 2026) using ModelFinder Plus (`-m MFP`; Kalyaanamoorthy et al. 2017), `-alrt 1000, -B 1000, -bnni, -T 16` and `-seed 20260719`. Shimodaira-Hasegawa-like approximate likelihood-ratio test (SH-aLRT) and ultrafast-bootstrap (UFBoot) support (Hoang et al. 2018) were therefore estimated from 1,000 replicates each. Raw IQ-TREE .treefile and .cintree files were retained unmodified and unrooted. Publication display copies of the UFBoot consensus trees were rooted at the unique edge that separated the a priori COX/PTGS-like set from all 33 strict non-COX comparators. This rooting was used only to orient the displayed family split; no support value was assigned to the root edge when one was not encoded by IQ-TREE.

Primary and sensitivity topologies were compared after pruning the sensitivity tree to the 128 common taxa. Shared internal splits and Robinson-Foulds distance were calculated. Rogue-taxon sensitivity was evaluated with the R package Rogue 2.2.0 (Smith 2021), using its implementation of the RogueNaRok algorithm (Aberer et al. 2013), `C_RogueNaRok`, computed support values, `dropsetSize = 1`, `threshold 50`, `label penalty 0` and `majority-rule-extended` optimisation disabled. Complete model, topology and rogue-taxon summaries are reported in Supplementary Table S12.

A separate family-discrimination phylogeny was constructed to test the central annotation alternative directly. The dataset comprised 54 non-redundant Porifera candidates of 500–650 amino acids, 33 strict non-COX comparators (27 PXDN-like and six MPO-like proteins) and 12 COX proteins supported by biochemical, heterologous-expression or structural evidence. The reference panel comprised COX-1/COX-2 pairs from *Ovis aries*, *Homo sapiens*, *Mus*

musculus and *Danio rerio*, two functional *Gersemia fruticosa* COX isoforms and the 15R/15S enzymes of *Plexaura homomalla*; the experimental basis and associated PubMed identifiers (PMIDs), digital object identifiers (DOIs) and Protein Data Bank (PDB) identifiers are listed in Supplementary Table S13. To avoid artefacts from lineage-specific accessory domains, phylogenetic characters were restricted to the homologous PROSITE PS50292 animal haem-peroxidase-superfamily interval. EGF-like status, full-protein length, gene structure, RNA evidence and P05979-equivalent residues were retained only as external annotations.

The 99 extracted PS50292 regions were aligned with MAFFT 7.526 (Katoh and Standley 2013) using L-INS-i (--localpair --maxiterate 1000) and trimmed with trimAl 1.5.0 (Capella-Gutiérrez et al. 2009) using -gt 0.1. Maximum-likelihood (ML) inference used IQ-TREE 3.1.3 (Wong et al. 2026), with ModelFinder (Kalyaanamoorthy et al. 2017), 5,000 ultrafast-bootstrap version 2 (UFBoot2) replicates (Hoang et al. 2018) and 1,000 SH-aLRT replicates, -bnni and seed 20260803. Sensitivity to across-site compositional heterogeneity was assessed using the posterior mean site frequency (PMSF) procedure (Wang et al. 2018): site frequencies were estimated under LG+C60 + F + G from the final guide topology (-ft, -n 0), then supplied with -fs for an independent IQ-TREE search with 1,000 UFBoot2 and 1,000 SH-aLRT replicates, -bnni and seed 20260804. Raw ML and consensus trees remained unrooted. The publication display was oriented at the exact edge separating the candidate/reference and strict-comparator sets; this graphical root was not used to infer evolutionary direction. Complete taxon metadata, residue states, experimental-reference evidence and model-sensitivity results are provided in Supplementary Tables S14–S16.

2.15. Expanded reanalysis of the Havird et al. sequence set

To place the Porifera discovery set within the metazoan framework of Havird et al. (2015), the 148 protein records represented in their maximum-likelihood analysis were recovered from current NCBI or UniProt records or from accession-traceable sequence archives and assigned stable identifiers H15_001–H15_148. These records were initially combined with the 67 exact-amino-acid-nonredundant Porifera candidates identified in the present study (P67_001–P67_067). One 115-amino-acid *Ephydatia muelleri* transcript fragment (P67_035) was excluded before phylogenetic inference because 114 of its 115 residues were identical to human PTGS1 over the aligned interval and the originating transcript contig had very low sequence coverage, a pattern considered more compatible with contamination or taxonomic misassignment than with a Porifera locus. The resulting curated full roster comprised 148 historical records and 66 Porifera candidates (214 proteins; Supplementary Tables S17 and S18).

Because the 66-candidate roster included closely related records recovered from alternate assemblies or from genome and transcriptome representations of the same species, a second, conservative locus/assembly-representative roster was defined for the primary analysis. Candidate records were consolidated only when they belonged to the same species, had equal protein length, shared at least 98% direct amino-acid identity and had concordant query, architecture and source context. Within each group, the genome/RNA-supported or otherwise best-supported model was retained. This removed eight near-identical source or assembly replicas (98.599–99.669% identity; 2–8 amino-acid differences) while preserving distinct non-overlapping loci from the same assembly. The primary roster therefore comprised 148 historical records and 58 Porifera representatives (206 proteins; Supplementary Table S19). This procedure was intended to reduce source and assembly replication and was not used to infer one-to-one orthology.

Phylogenetic characters were restricted to the homologous PROSITE PS50292 animal haem-peroxidase-superfamily interval. Full-protein EGF-like status, protein length, exon count, RNA support, evidence tier and P05979-equivalent residue profiles were retained as external annotations and did not contribute characters to tree inference. The 206- and 214-protein datasets were aligned independently with MAFFT 7.526 using L-INS-i (--localpair --maxiterate 1000) and trimmed with trimAl 1.5.0 using -gt 0.1. The primary alignment contained 1,308 columns before and 514 columns after trimming; the full sensitivity alignment contained 513 trimmed columns. ModelFinder selected LG + R7 by the Bayesian information criterion (BIC) for both datasets. Maximum-likelihood trees were inferred with IQ-TREE 3.1.3 using 1,000 UFBoot2 replicates, 1,000 SH-aLRT replicates, bootstrap nearest-neighbour-interchange (NNI) correction and fixed random seeds (Supplementary Table S20). Raw .treefile and .cintree outputs were retained unmodified and unrooted. For display only, trees were oriented at the recovered bipartition containing the 13 historical MPO/PXDN references and H15_134; this orientation was not interpreted as evidence for the direction of COX evolution.

The primary and curated-full topologies were compared after projection to their 206 common taxa using shared internal splits, split Jaccard similarity and normalised Robinson-Foulds distance. Rogue-taxon sensitivity was evaluated on the 1,000 primary bootstrap trees with the R package Rogue 2.2.0 (Smith 2021), using its implementation of C_RogueNaRok (Aberer et al. 2013), computed support values, dropsetSize = 1, threshold 50, label penalty 0 and majority-rule-extended optimisation disabled. A second 205-protein tree was inferred after removal of the single taxon identified by this analysis. Across-site compositional heterogeneity was evaluated by the two-stage posterior mean site frequency procedure (Wang et al. 2018) under LG+C60 + F + G using the primary LG + R7 topology as the guide tree, followed by 1,000 UFBoot2 and 1,000 SH-aLRT replicates with bootstrap NNI correction. Prespecified family- and class-enriched bipartitions were tested as exact splits after projection to each retained taxon set (Supplementary Tables S21–S23).

The lineage summary of Havird et al. (2015, their Fig. 6) was updated as an evidence-distribution diagram rather than an ancestral-state reconstruction. Lineages with published COX sequence evidence in the historical study retained their original evidence status, whereas Porifera was added as containing candidate COX/PTGS-like homologues supported by the sequence, domain, gene-model and/or RNA analyses reported here. The diagram does not encode enzymatic validation, ancestral presence, duplication timing or lineage-specific loss.

2.16. Genomic-locus multiplicity and candidate paralogy

Genome-derived models were reconciled at locus level before copy-number interpretation. Models on the same contig and strand were assigned to one locus when their genomic intervals overlapped by at least 50% of the shorter interval. Exact predicted proteins were identified by Secure Hash Algorithm 256-bit (SHA-256) sequence hashes. Multiple predictions at one locus were treated as redundant or alternative models, and homologous loci found only in different assemblies of the same species were treated as cross-assembly replication rather than copy-number evidence. Two or more non-overlapping COX/PTGS-like loci in one assembly were classified as candidate paralogous loci. This designation required spatially distinct gene models and was evaluated using

complete-protein identity, exon-intron architecture, P05979-equivalent residues, RNA/splice evidence and reconstructed-ORF status; it does not establish the timing of duplication or exclude assembly duplication at very high sequence identity.

One representative per non-overlapping locus was retained for an exploratory locus-level phylogeny. The dataset comprised 47 genomic loci and 12 COX/PTGS references. Proteins were aligned with MAFFT, trimmed with trimAl -gt 0.1 to 608 columns and analysed with IQ-TREE 3.1.3. ModelFinder selected LG + R4 by BIC, and support was estimated with 1,000 SH-aLRT and 1,000 ultrafast-bootstrap replicates. The complete locus inventory, within-assembly comparisons and tree context are provided in Supplementary Tables S24–S26 and Supplementary Fig. S1. Supplementary Fig. S1 is midpoint-rooted for visualisation only; the unmodified unrooted IQ-TREE outputs are retained in the data package. Formal orthology or duplication timing was not inferred because chromosome-scale synteny and gene-tree/species-tree reconciliation were unavailable.

3. Results

3.1. Independent searches produced a staged, non-additive evidence funnel

The search and validation stages produced distinct but overlapping candidate layers (Table 1; Fig. 1; Supplementary Tables S27 and S28).

Table 1 Staged discovery and validation funnel. Counts describe successive, overlapping evidence layers and are not additive. Complete source-specific decisions and integrated evidence are provided in Supplementary Tables S4, S5 and S28–S36.				
Evidence stream	Stage	Records	Species/labels	Outcome
CANTATA translated transcriptomes	merged peptide search space	4,183,296	not counted at this stage	local translated-transcriptome database screened with COX/PTGS queries
CANTATA translated transcriptomes	deduplicated reciprocal-decision table	73	31	16 priority COX/PTGS-like; 9 partial/secondary; 48 non-COX/outgroup-like
CANTATA translated transcriptomes	positive COX/PTGS-like set	25	18	16 priority; 9 partial/secondary
NCBI/browser annotated proteins	full sequences recovered from browser BLAST	100	not all retained as COX/PTGS-like	98 had a Swiss-Prot hit at the selected threshold; 2 had no strict Swiss-Prot hit
NCBI/browser annotated proteins	reciprocal Swiss-Prot decision table	100	100 protein records	4 high-confidence COX/PTGS-like; 1 partial/secondary; 95 non-COX/no-hit/other
Unannotated genome mining	broad genome-derived reciprocal COX/PTGS-like records	67	23	all 67 broad records had Swiss-Prot best-hit behaviour interpreted as supporting COX-like annotation
Unannotated genome mining	CDS-clean subset	26	11	subset with cleaner CDS/protein recovery after locus extraction and model filtering
RNA-seq/StringTie support for genome models	genome-derived candidates evaluated with RNA-seq	65	22	20 full CDS/all introns; 9 all introns length-consistent; 30 partial; 4 divergent junctions; 2 no transcript
RNA-seq/StringTie support for genome models	ORF-level validation of reconstructed transcripts	65	22	34 ORFs confirmed the original genome protein; 28 ORFs were COX-like but not the expected top self-match; 2 no ORF; 1 unresolved
Genome-model reconciliation	overlap and exact-protein reconciliation	47 loci; 38 proteins	22	65 RNA-evaluated models reconciled to 47 non-overlapping loci and 38 globally unique amino-acid sequences
Integrated pre-phylogenetic set	selected records with uniform domain/residue/gene-structure annotation	45	25	25 CANTATA-derived, 5 annotated/browser-derived, 15 genome/RNA-supported records
Integrated genome/RNA strict subset	strongest direct genome/RNA-supported candidates	15	9	6 class A and 9 class B records in the integrated table

The translated-transcriptome stream reduced 4,183,296 CANTATA peptides to 73 reciprocal-decision records and 25 positive COX/PTGS-like candidates. The annotated-protein stream reduced 100 browser-recovered proteins to five positive candidates. Genome mining yielded 67 broad reciprocal COX/PTGS-like models from 23 species-level labels, of which 26 models from 11 labels passed the conservative CDS-clean filter. Sixty-five genome-derived candidates from 22 labels were evaluated using RNA-seq, and 15 candidates from nine labels met the integrated class A or B criteria. The integrated 45-record matrix comprised 25 CANTATA, five annotated/browser and 15 genome/RNA-supported candidates. Because these layers are nested or overlap across assemblies and sources, they are not additive estimates of gene number.

Across the positive source tables, evidence was detected in 37 species or species-level source labels, and four labels were represented in more than one positive stream (Supplementary Table S29). Source labels were retained as deposited and were not silently taxonomically reconciled; in particular, *Cymbastela concentrica* and *Cymbastella concentrica* are reported as distinct source-level labels. The integrated 45-record matrix comprised six class A, nine class B, 16 class C1, four class C2, nine class D1 and one class D2 records. Rejected and unresolved sequences were preserved in the complete source-specific decision tables but were not counted as positive candidates.

3.2. CANTATA transcriptome mining recovered full-length and partial candidates from 18 species-level labels

The reciprocal CANTATA decision table contained 73 non-redundant records from 31 species-level labels. Sixteen records met all priority criteria, nine were retained as partial or secondary candidates, and 48 were rejected as COX/PTGS-like because their reciprocal annotations supported non-COX haem peroxidases or other protein families (Supplementary Table S4). The 25 positive records represented 18 species-level labels (Supplementary Table S30), while the complete input inventory is provided in Supplementary Table S1.

The 16 priority candidates are reported individually in Table 2.

Table 2 Sixteen priority COX/PTGS-like candidates recovered from the CANTATA translated-transcriptome database. aa, amino-acid length; qcov, query coverage; recid, reciprocal identity; rqcov, reciprocal query coverage. The complete 73-decision set and all 25 retained candidates are provided in Supplementary Tables S30 and S31.								
Species	Candidate	aa	Initial query	Initial identity (%)	Initial qcov (%)	Reciprocal Swiss-Prot hit	Reciprocal identity (%)	Reciprocal qcov (%)
<i>Dysidea avara</i>	SOAP.k25.C3288810.p1	590	Q962I8	63.7	99.0	P27607 (PGH2_CHICK)	52.1/96.0	99.0
<i>Leucetta chagosensis</i> PHIL	SOAP.k53.C1004013.p1	670	Q962I8	25.5	70.0	P35355 (PGH2_RAT)	26.6/61.0	70.0
<i>Plakina jani</i>	SPADES.k25.NODE_16377_length_2181_cov_8.873377_g6138_i1.p1	606	Q962I8	39.1	96.0	O02768 (PGH2_RABIT)	42.0/93.0	96.0
<i>Farrea similis</i>	SPADES.k25.NODE_249_length_1893_cov_47.593683_g176_i0.p1	594	Q962I8	47.4	97.0	O62698 (PGH2_BOVIN)	48.1/93.0	97.0
<i>Spongia officinalis</i>	SPADES.k25.NODE_33591_length_2112_cov_18.596550_g9718_i0.p1	603	Q962I8	49.0	96.0	O62698 (PGH2_BOVIN)	51.7/93.0	96.0
<i>Axinella polypoides</i>	SPADES.k25.NODE_8540_length_2094_cov_88.678589_g4356_i0.p1	593	Q962I8	58.6	99.0	Q05769 (PGH2_MOUSE)	50.1/98.0	99.0
<i>Corticium candelabrum</i>	SPADES.k41.NODE_19503_length_2196_cov_34.353596_g6791_i2.p1	634	Q962I8	40.5	93.0	P27607 (PGH2_CHICK)	41.7/89.0	93.0
<i>Farrea occa</i>	SPADES.k41.NODE_2941_length_1864_cov_32.177729_g2024_i0.p1	594	Q962I8	47.4	97.0	O62698 (PGH2_BOVIN)	48.3/93.0	97.0
<i>Halichondria panicea</i>	SPADES.k41.NODE_78661_length_1853_cov_3.693157_g28047_i0.p1	592	Q962I8	46.7	97.0	O62698 (PGH2_BOVIN)	51.3/94.0	97.0
<i>Oscarella pearsei</i>	SPADES.k53.NODE_10982_length_2139_cov_52.570470_g5343_i0.p1	628	Q962I8	40.6	95.0	P27607 (PGH2_CHICK)	44.3/88.0	95.0
<i>Halisarca dujardini</i>	TransABYSS.k25.R3805392.p1	612	Q962I8	60.8	99.0	P27607 (PGH2_CHICK)	53.2/89.0	99.0
<i>Chondrosia reniformis</i> ASG	TransABYSS.k25.R426490.p1	451	Q962I8	58.3	76.0	P70682 (PGH2_CAVPO)	47.8/99.0	76.0
<i>Cliona varians</i>	TransABYSS.k25.R851381.p1	596	Q962I8	73.5	99.0	O62698 (PGH2_BOVIN)	53.3/92.0	99.0
<i>Cliona varians</i>	TransABYSS.k25.S851946.p1	604	Q962I8	65.6	98.0	P79208 (PGH2_SHEEP)	53.3/92.0	98.0
<i>Oscarella lobularis</i>	TransABYSS.k25.S94931.p1	596	Q962I8	40.3	97.0	P27607 (PGH2_CHICK)	43.3/93.0	97.0
<i>Tethya wilhelma</i>	TransABYSS.k53.R278387.p1	829	Q962I8	60.9	95.0	O62698 (PGH2_BOVIN)	53.4/67.0	95.0

Initial identities to the best COX query ranged from 25.5% to 73.5%, and initial query coverage ranged from 70% to 99%. Reciprocal Swiss-Prot query coverage ranged from 61% to 99%. All 16 retained the P05979-equivalent radical-site Tyr385, and complete monitored-position scores ranged from 6/10 to 10/10. Most priority records had a complete or near-complete EGF-like plus COX-core architecture. *Dysidea avara* SOAP.k25.C3288810.p1 was 590 amino acids long, aligned to octocoral Q962I8 with 63.7% identity across 99% of the query and recovered chicken PGH2 as the reciprocal best hit with 52.1% identity and 96% query coverage. Its P05979-equivalent profile conserved 9/10 monitored positions.

Two distinct *Cliona varians* proteins were recovered. TransABYSS.k25.R851381.p1 and TransABYSS.k25.S851946.p1 were 596 and 604 amino acids long, respectively, aligned to Q962I8 with 73.5% and 65.6% identity at 99% and 98% query coverage, and both conserved 9/10 monitored positions. Additional near-

full candidates with 9/10 residue profiles were detected in *Farrea occa*, *Farrea similis*, *Halichondria panicea* and *Spongia officinalis*. *Halisarca dujardinii* TransABYSS.k25.R3805392.p1 conserved all 10 monitored positions. *Tethya wilhelma* TransABYSS.k53.R278387.p1 also conserved 10/10 positions and showed 60.9% identity to Q96218 across 95% of the query, but its 829-amino-acid length and extended architecture warranted exclusion from the length-screened primary phylogeny and retention in the sensitivity analysis. By contrast, the 451-amino-acid *Chondrosia reniformis* ASG sequence and the nine partial records were treated as incomplete evidence despite reciprocal COX/PTGS support.

Strict cross-source reconciliation confirmed that the expanded CANTATA search did not simply reproduce a small annotated set. Five positive CANTATA proteins were exact full-sequence amino-acid matches to candidates recovered from another stream, four were very close but non-identical matches, and 16 remained distinct at the applied thresholds. Exact duplicates were collapsed only for downstream non-redundant analyses; non-identical candidates were retained.

3.3. Reciprocal filtering reduced 100 annotated/browser proteins to four full and one partial candidate

The human COX-2 browser search recovered 100 complete protein records. Ninety-eight produced a Swiss-Prot hit at e-value $\leq 1e-10$, whereas two had no hit at this threshold. Reciprocal best-hit classification retained only five COX/PTGS-like records and excluded 95 records as non-COX, other or unresolved (Table 3; Supplementary Tables S31 and S5).

Table 3 Annotated/browser-derived COX/PTGS-like candidates retained after reciprocal Swiss-Prot classification. aa, amino-acid length; qcov, query coverage. The partial XP_062518513.1 record was retained as secondary evidence only; all 100 reciprocal decisions are provided in Supplementary Table S5.									
Species	Protein accession	aa	Browser qcov (%)	Reciprocal Swiss-Prot hit	Reciprocal identity/qcov (%)	Decision	Architecture	P05979 residues	Y385
<i>Dysidea avara</i>	XP_065916514.1	590	94.2	P27607 (PGH2_CHICK)	52.1/96.0	accept high confidence COX like	EGF-like + COX core	9/10	conserved
<i>Oscarella lobularis</i>	XP_065833991.1	596	92.1	P27607 (PGH2_CHICK)	43.3/93.0	accept high confidence COX like	EGF-like + COX core	6/10	conserved
<i>Oscarella lobularis</i>	XP_065827972.1	597	94.5	O19183 (PGH2_HORSE)	42.0/93.0	accept high confidence COX like	EGF-like + COX core	6/10	conserved
<i>Corticium candelabrum</i>	XP_062518712.1	604	93.7	P27607 (PGH2_CHICK)	41.9/93.0	accept high confidence COX like	EGF-like + COX core	7/10	conserved
<i>Corticium candelabrum</i>	XP_062518513.1	324	50.5	P79208 (PGH2_SHEEP)	40.6/90.0	keep partial or secondary COX like	short fragment COX like signature	4/10	substituted_N

This result illustrates the limited specificity of a one-directional COX query within the animal haem peroxidase superfamily.

The four full-length candidates were *Dysidea avara* XP_065916514.1, *Oscarella lobularis* XP_065833991.1 and XP_065827972.1, and *Corticium candelabrum* XP_062518712.1 (Table 3). Their lengths were 590–604 amino acids, browser query coverage was 92.1–94.5%, reciprocal Swiss-Prot query coverage was 93–96%, and Tyr385 was conserved in all four. The *Dysidea* protein conserved 9/10 monitored residues and was identical to a transcriptome-derived candidate. The two *Oscarella* proteins conserved 6/10 residues, whereas the full *Corticium* protein conserved 7/10, indicating conservation of the core COX/peroxidase framework with divergence at several channel-associated positions. *Corticium candelabrum* XP_062518513.1 was only 324 amino acids long, covered 50.5% of human COX-2 and showed Tyr385-to-Asn at the mapped position; it was retained only as a partial D2 record.

3.4. Genome mining recovered 67 broad models and 15 stringent genome/RNA-supported records

TBLASTN screening and miniprot modelling recovered 67 broad genome-derived proteins from 23 species-level labels (Supplementary Tables S32 and S33). Reciprocal Swiss-Prot results supported a COX/PTGS-like interpretation for all 67, but the broad layer included replicated assemblies, sequence-identical models, incomplete proteins and models without complete RNA confirmation. The conservative CDS-clean subset contained 26 records from 11 labels (Supplementary Table S34). These two layers document the breadth of the genomic signal and should not be interpreted as 67 distinct genes.

Broad candidates were recovered from 40 of the 213 screened assemblies; 173 assemblies yielded no model meeting the broad tier A/B criteria (Supplementary Table S2). These negative screening outcomes were retained explicitly but were not interpreted as biological gene absence, because assembly completeness, locus fragmentation and sequence divergence affect detectability.

The strict class A/B set contained 15 assembly-level models from nine species-level labels (Table 4).

Table 4

Class A and B genome-derived assembly-level models with RNA-seq, complete predicted-intron reconstruction and exact StringTie ORF confirmation. aa, amino acid; Reciprocal Swiss-Prot hit, Reciprocal Swiss-Prot hit; Identity/qcov, Identity/Query Coverage; Exons/introns, Exons/Introns; STAR, STAR; Reads/breadth/junctions, Reads/breadth/junctions; St, Status.

Class	Species	Assembly	Model	aa	Reciprocal Swiss-Prot hit	Identity/qcov (%)	Exons/introns	STAR	Reads; breadth; junctions	St
A	<i>Cliona varians</i>	GCA_978019885.1	MPROT000324	596	O62698 (PGH2_BOVIN)	53.5/92.0	9/8	strong splice support	998; 100.0%; 8	full + a int
A	<i>Cliona varians</i>	GCA_978019885.1	MPROT000327	596	O62698 (PGH2_BOVIN)	53.3/92.0	9/8	strong splice support	4428; 100.0%; 8	full + a int
A	<i>Cliona varians</i>	GCA_978019885.1	MPROT000343	604	P79208 (PGH2_SHEEP)	53.3/92.0	9/8	strong splice support	6792; 100.0%; 8	full + a int
A	<i>Cliona varians</i>	GCA_978019905.1	MPROT000326	596	O62698 (PGH2_BOVIN)	52.8/92.0	9/8	strong splice support	2750; 100.0%; 8	full + a int
A	<i>Cliona varians</i>	GCA_978019905.1	MPROT000344	604	P79208 (PGH2_SHEEP)	52.9/92.0	9/8	strong splice support	4922; 100.0%; 8	full + a int
A	<i>Dysidea avara</i>	GCA_963966705.1	MPROT000353	590	P27607 (PGH2_CHICK)	52.1/96.0	9/8	strong splice support	874; 100.0%; 8	full + a int
B	<i>Biemna sp. UH-2024</i>	GCA_965654295.1	MPROT000153	570	O62698 (PGH2_BOVIN)	51.1/97.0	9/8	strong splice support	2556; 100.0%; 8	full + a int
B	<i>Cliona cf. celata</i>	GCA_982375335.1	MPROT000333	594	O62698 (PGH2_BOVIN)	51.3/92.0	8/7	strong splice support	686; 100.0%; 7	full + a int
B	<i>Cliona cf. celata</i>	GCA_982375485.1	MPROT000234	552	P27607 (PGH2_CHICK)	51.3/99.0	9/8	strong splice support	1578; 100.0%; 8	full + a int
B	<i>Cliona cf. orientalis</i>	GCA_963930775.1	MPROT000328	596	O62698 (PGH2_BOVIN)	53.0/92.0	9/8	strong splice support	350; 100.0%; 8	full + a int
B	<i>Cliona cf. orientalis</i>	GCA_963930775.1	MPROT000348	600	P27607 (PGH2_CHICK)	51.8/96.0	9/8	strong splice support	312; 100.0%; 8	full + a int
B	<i>Dysidea cf. avara</i>	GCA_963678995.2	MPROT000356	590	P27607 (PGH2_CHICK)	52.4/96.0	9/8	strong splice support	516; 100.0%; 8	full + a int
B	<i>Phakellia ventilabrum</i>	GCA_963921505.1	MPROT000122	552	O02768 (PGH2_RABIT)	52.3/99.0	9/8	strong splice support	3332; 99.6%; 8	full + a int
B	<i>Pione sp. GHC0000220</i>	GCA_982552475.1	MPROT000337	571	P79208 (PGH2_SHEEP)	51.5/100.0	8/7	strong splice support	2470; 100.0%; 7	full + a int
B	<i>Spheciospongia papillosa</i>	GCA_982375275.1	MPROT000345	598	P79208 (PGH2_SHEEP)	53.0/92.0	9/8	strong splice support	1890; 100.0%; 8	full + a int

Six class A records comprised five *Cliona varians* models and one *Dysidea avara* model that were recovered independently across evidence streams. The nine class B records were new genome/RNA-supported models from *Biemna sp. UH-2024*, *Cliona cf. celata*, *Cliona cf. orientalis*, *Dysidea cf. avara*, *Phakellia ventilabrum*, *Pione sp. GHC0000220* and *Spheciospongia papillosa*. All 15 had reciprocal PGH1/PGH2 best hits with 51.1–53.5% identity and 92–100% query coverage.

Each of the five *Cliona varians* models encoded a 596–604-amino-acid protein with nine exons and eight introns. The proteins formed two length groups and were retained as distinct records when their complete amino-acid sequences were not identical. *Dysidea avara* MPROT000353 encoded a 590-amino-acid protein with nine exons and eight introns and matched the independently recovered transcriptome/annotated protein. Among class B candidates, *Phakellia ventilabrum* MPROT000122 encoded a 552-amino-acid protein, had rabbit PGH2 as the reciprocal best hit at 52.3% identity and 99% query coverage, and conserved Tyr385 together with 8/10 monitored positions. This locus is especially relevant because it provides a direct genomic candidate in the species examined independently for reproductive lipid metabolism by Koutsouveli et al. (2022). *Pione sp. GHC0000220* MPROT000337 encoded a 571-amino-acid 8-exon/7-intron model with sheep PGH2 as the best hit at 51.5% identity and 100% query coverage. The *Biemna*, *Spheciospongia* and both *Cliona cf. orientalis* proteins were 570–600 amino acids long and conserved Tyr385.

3.5. Splice-aware RNA reconstruction distinguished transcription from exact model confirmation

RNA-seq was evaluated for 65 genome-derived candidates rather than only for the 15 strict models. Twenty candidates had a reconstructed transcript containing the full predicted CDS and all predicted introns; nine additional candidates had all predicted introns in a transcript whose length was consistent with the genome model. Thirty candidates had partial predicted-intron reconstruction, four produced spliced transcripts with junctions that differed from the miniprot model, and two had no reconstructed transcript (Supplementary Table S35).

ORF comparison provided a stricter test than local read coverage. Thirty-four reconstructed StringTie ORFs exactly confirmed the corresponding genome-derived COX/PTGS-like protein. Twenty-eight produced a COX/PTGS-like ORF but did not recover the expected genome protein as the top self-match, two produced no ORF and one remained unresolved. The latter categories support local COX-like transcription but not exact validation of the proposed coding model.

All 15 class A/B models had strong splice support, near-complete or complete exonic breadth, exact recovery of every predicted intron and exact confirmation of the expected protein by the StringTie ORF. Mapped-read counts ranged from 312 to 6,792. Fourteen models had 100% exonic breadth, and the *Phakellia ventrilabrum* model had 99.6% breadth. Thirteen models comprised nine CDS exons and eight introns, while *Cliona cf. celata* MPROT000333 and *Pione* sp. GHC0000220 MPROT000337 comprised eight exons and seven introns. All exported introns in the strict direct-model set had canonical GT-AG splice motifs in transcript orientation.

3.6. Genomic context supported host attribution and exposed one assembly-level taxonomic conflict

The assembly/RNA/composition screen covered all 67 broad genome models, and the translated-neighbourhood audit covered the 26 CDS-clean models mapped to 23 distinct genomic regions (Supplementary Table S8). Fifteen CDS-clean models occurred on chromosome-level assemblies and 11 on scaffold-level assemblies. The absolute gene-versus-assembly GC difference had a median of 0.594 percentage points and a maximum of 6.591 percentage points.

After excluding each candidate plus 10-kb flanks, 7.06 Mb of neighbourhood sequence were represented by 365 bins. The primary DIAMOND/taxonomy screen retained 4,246 alignments and classified 229 bins as metazoan-supported, 16 as other-eukaryote, 21 as mixed metazoan/prokaryote and three as prokaryote-only. The three permissive prokaryote-only calls were isolated within regions containing multiple metazoan-supported bins; only one persisted at the 60-bit/40-aa threshold and none persisted at 80 bits/50 aa.

Twenty of the 26 models received strong host-attribution calls, two *Cymbastela concentrica* models were supported by partial/junction RNA evidence and eukaryotic context, and two *Dysidea avara* models were supported by eukaryotic context but lacked same-species RNA in the selected validation set. The remaining two loci, APLY_MPROT000014 and APLY_MPROT000016, were supported as multiexonic metazoan host loci but not assigned to a species or Porifera class because the current NCBI organism label for GCA_977007035.1 (*Aplysilla rubra*) conflicts with the assembly name, BioProject, BioSample attributes and curation comments identifying *Oscarella rubra*. All 26 final calls were unchanged under both stricter taxonomic-hit filters, and none received a contamination-concern classification.

3.7. Domain and residue profiles separated canonical-like, channel-divergent and partial candidates

InterProScan detected a PANTHER prostaglandin G/H synthase signature in all 45 integrated candidates and Pfam PF03098 animal haem peroxidase in 44/45. Thirty-two records had complete or near-complete EGF-like plus COX-core organisation. Nine were short COX-like fragments, and the remaining records included partial, EGF-undetected or extended architectures (Supplementary Table S36). The partial records were retained because they preserve traceable COX-like sequence evidence but were not used alone to support full-length protein claims.

The subsequent 157-protein uniform profile audit resolved all EGF metadata used in the phylogenetic figures (Supplementary Tables S10, S11 and S6). PS50026 detected an EGF-like domain in all 54 primary study candidates of 500–650 amino acids. Across the complete 67-candidate study set, 55 proteins were EGF-positive, ten proteins shorter than 500 amino acids lacked sufficient representation of the P05979-equivalent N-terminal interval and were therefore classified as EGF-non-evaluable partials, and two evaluable extended proteins, *Halichondria panicea* SPADES.k25.NODE_11647_length_3389_cov_6.044590_g4708_i0.p1 (1,056 amino acids) and *Leucetta chagosensis* C10040 (670 amino acids), lacked a PS50026 hit. All ten short candidates retained an animal haem peroxidase PS50292 profile; nine also passed the Pfam PF03098 gathering threshold, whereas the 119-amino-acid *Farrea occa* fragment was detected only by PS50292. Among the 33 strict non-COX comparators, no PS50026 hit was detected: 21 represented evaluable non-detections, whereas 12 lacked sufficient N-terminal representation for assessment. All 33 retained both animal haem peroxidase profiles. Thus, the primary candidates consistently combined the EGF-like module with the shared catalytic superfamily core, while the fragment calls explicitly distinguish missing sequence from a detected domain absence.

Across the 135 Havird et al. (2015) COX/PTGS records in the curated 214-protein roster, 104 produced a PS50026 hit. Of the 31 no-hit records, 26 were shorter than 500 amino acids and the remaining five were 501–559 amino acids long. Because the basis of the latter non-detections was unresolved, they were not interpreted as biological losses of the EGF-like module (Supplementary Table S17).

The complete P05979-equivalent residue matrix is reported in Supplementary Table S7. Twenty-two integrated records showed a strong COX-like profile, nine retained the peroxidase/haem core but varied at channel-associated positions, four were divergent COX-like, one was weak/problematic and nine could not

be scored completely because of short or extended sequence architecture. Tyr385 was conserved in 35/45 records, unassessed in nine and replaced by Asn in the partial *Corticium candelabrum* XP_062518513.1 record.

The residue matrix identified a biologically informative contrast among full-length candidates. Demospongiae such as *Cliona varians*, *Dysidea avara*, *Halisarca dujardini*, *Tethya wilhelma*, *Farrea*, *Halichondria* and *Spongia* generally conserved Tyr385 and the complete Q203/H207/H386/H388 peroxidase/haem context together with most channel residues. In *Dysidea avara*, E524 was replaced by Ala, whereas the remaining monitored positions were conserved. The *Oscarella lobularis* and *Corticium candelabrum* full-length proteins retained Tyr385 and the Q203/H207/H386/H388 core but showed more substitutions at R120-, Y355-, E524- or S530-equivalent positions. These sequences were therefore interpreted as channel-divergent COX/PTGS-like candidates rather than canonical vertebrate-like enzymes.

3.8. Mutagenesis-informed profiles separated canonical-compatible and channel-divergent candidates

All 18 focused candidates retained the seven monitored POX/haem-environment residues Q203, H204, H207, H309, H386, W387 and H388 and the three-residue COX radical/positioning core Y348, Y385 and G533 (Fig. 2; Supplementary Fig. S2; Supplementary Tables S37–S39).

None carried a Y385F-equivalent change, which abolishes COX while retaining POX activity in ovine PGHS-1, or substitutions at H207, H309 or H388 equivalent to the experimentally disruptive histidine replacements (Shimokawa et al. 1990; Shimokawa and Smith 1991). This uniform retention supports sequence compatibility with the monitored catalytic framework but does not establish coupled COX/POX chemistry or measure catalytic turnover.

Nine candidates had a canonical-channel-compatible profile: one each from *Cliona cf. celata*, *Cliona cf. orientalis*, *Dysidea avara*, *Dysidea cf. avara* and *Pione sp. GH0000220*, together with four *Cliona varians* proteins. Their monitored states were either identical to P05979 or represented by E524A, an exact state present in the experimentally active octocoral enzymes of *Gersemia fruticosa* and *Plexaura homomalla* (Koljak et al. 2001; Valmsen et al. 2001, 2004). E524A was therefore not interpreted as a loss-of-function lesion.

Four proteins, from *Cliona cf. orientalis*, *Cliona varians* and *Spheciospongia papillosa*, combined V349I, E524T and L534I and were classified as product-positioning variants. V349I occurs in the functionally characterised 15R-specific *P. homomalla* enzyme, whereas the exact E524T and L534I states have not been tested. Because changes at V349 and L534 alter arachidonate positioning and oxygenated-product distributions in experimental COX systems, these candidates generate a testable product-specificity hypothesis rather than evidence of reduced activity (Thuresson et al. 2000, 2001; Valmsen et al. 2004).

Three candidates, from *Biemna sp. UH-2024*, *C. cf. celata* and *Phakellia ventilabrum*, carried S530Y and were classified as channel-constriction/radical-network variants; the *Biemna* protein additionally carried Y504V, whereas the other two carried Y504F. S530 and Y504 influence channel geometry and radical distribution, but the observed combined states have not been tested directly (Shimokawa and Smith 1992; Rogge et al. 2009). The *Corticium candelabrum* and *Oscarella lobularis* proteins retained the complete monitored POX/haem and Tyr385-centred cores but carried R120G or R120S together with Y355G. Because R120 and Y355 form the canonical substrate-entry constriction, these Homoscleromorpha proteins were designated divergent substrate-entry candidates, not inactive enzymes (Bhattacharyya et al. 1996; Thuresson et al. 2001).

3.9. Phylogenetic analyses separated candidate COX/PTGS-like proteins from non-COX animal haem peroxidases

The dedicated family-discrimination alignment contained 99 proteins and 676 retained amino-acid sites. ModelFinder selected Q.PFAM + I+R4. All 54 Porifera candidates and all 12 experimentally characterised COX references occurred on one side of a single edge, whereas all 33 strict PXDN-like or MPO-like comparators occurred on the other; this exact 66-versus-33 bipartition received SH-aLRT/UFBoot support of 100/100 in the maximum-likelihood tree and UFBoot 100 in the bootstrap consensus (Fig. 3; Supplementary Tables S14 and S16).

The candidate set represented 31 species-level labels and comprised 43 Demospongiae, six Homoscleromorpha and three Hexactinellida proteins, together with two models from the taxonomically unresolved GCA_977007035.1 assembly.

The molecular annotation associated with the tree was consistent with, but not used to impose, the family partition. All 99 proteins contained the PS50292/PF03098 animal haem-peroxidase-superfamily core. EGF-like PS50026 was detected in all 54 Porifera candidates, whereas no comparator produced a PS50026 hit; 21 comparator sequences were evaluable profile non-detections and 12 lacked sufficient representation of the corresponding N-terminal interval. Fifty-one of 54 candidates retained all four monitored POX/haem-context residues, all 54 retained Tyr385, and 33 retained at least five of the six monitored COX-site/channel positions. The shared peroxidase-superfamily core was therefore not treated as COX-specific; discrimination rested on the combined topology, EGF-like architecture, COX-site context and independent annotation evidence (Fig. 3; Supplementary Tables S14 and S15).

The three 3/4 POX/haem profiles were all Hexactinellida: *Aphrocallistes beatrix* C001, *Farrea occa* C029 and *Farrea similis* C030 each carried H386R while retaining Tyr385 and all six monitored COX-site/channel positions. The shared substitution was therefore treated as a class-associated POX-context variant rather than an isolated loss-of-function call.

The LG+C60 posterior-mean-site-frequency sensitivity tree retained the exact family bipartition with UFBoot 100 in its consensus. The ModelFinder and site-heterogeneous maximum-likelihood trees shared 90 of 96 internal bipartitions (unweighted Robinson-Foulds distance 12; normalised distance 0.0625). Thus, limited internal rearrangement under the more composition-aware model did not alter the separation of any Porifera candidate from the strict non-COX comparator panel (Supplementary Table S16).

The broader primary alignment contained 128 proteins and 1,018 trimmed amino-acid sites. ModelFinder selected Q.PFAM + I+R6. The length-sensitivity alignment contained 157 proteins and 965 sites, and Q.PFAM+R7 was selected (Table 5; Supplementary Table S12).

Table 5

Family-discrimination, broader primary, length-sensitivity and expanded Havird-Porifera phylogenetic analyses, including source-inclusive, site-heterogeneity and current-rogue sensitivities. Support values are SH-aLRT/UFBoot percentages. SH-aLRT, Shimodaira-Hasegawa-like approximate likelihood-ratio test; UFBoot, ultrafast bootstrap; RF, Robinson-Foulds distance; PMSF, posterior mean site frequency.

Analysis	Taxa	Candidate/reference roster	Non-COX comparators	Sites	Model	Result
Family discrimination	99	66 (54 candidates + 12 experimental references)	33	676	Q.PFAM + I+R4	Exact 66/33 split; SH-aLRT/UFBoot 100/100; PMSF consensus UFBoot 100
Primary	128	95	33	1,018	Q.PFAM + I+R6	Exact COX/non-COX bipartition; SH-aLRT/UFBoot 100/100
Primary candidate-rich grouping	68	68	0	1,018	Q.PFAM + I+R6	46/54 study candidates + 7 octocoral COX; SH-aLRT/UFBoot 86.8/52
Length sensitivity	157	124	33	965	Q.PFAM+R7	Exact COX/non-COX bipartition
Common-taxon comparison	128	95	33	not applicable	not applicable	109/125 shared internal splits; normalised RF 0.128
RogueNaRok sensitivity	157	124	33	not applicable	not applicable	0 taxa flagged
Expanded primary roster	206	192 (134 historical + 58 candidates)	14 historical family-side records	514	LG+R7	189-protein COX-rich split (134 + 55), 96.4/97; 37 Demospongiae + 7 octocoral split, 99.9/100
Expanded source-inclusive sensitivity	214	200 (134 historical + 66 candidates)	14 historical family-side records	513	LG+R7	197-protein COX-rich split (134 + 63), 92.8/95; 45 Demospongiae + 7 octocoral split, 100/100
Expanded PMSF sensitivity	206	192	14	514	LG+C60 + F + G PMSF	COX-rich split, 90.1/94; Demospongiae + octocoral split, 91/74
Expanded rogue-pruned sensitivity	205	191	14	513	LG+R7	P67_046 removed; COX-rich split, 96.8/89; projected normalised RF 0.040

In both analyses, every declared COX/PTGS-like protein was separated from all 33 strict non-COX comparators by one edge. The raw trees were inferred without a root; rooting at this family-separating edge was applied only to the complete publication display copies (Supplementary Figs. S3 and S4). These broader trees were analysed separately from the independently inferred 99-protein family-discrimination topology shown in Fig. 3.

The primary tree contained 54 non-redundant study candidates. Forty-six occurred in a 68-taxon non-vertebrate grouping that also contained seven published octocoral COX sequences. Support for this split was moderate by SH-aLRT but limited by UFBoot (86.8/52), so it was treated as a candidate-rich evolutionary grouping rather than a fully resolved deep clade. Based on patristic proximity within the primary tree, 25 study candidates were closest to an octocoral COX reference group, 21 to other metazoan COX references and eight to vertebrate COX references. The eight candidates outside the candidate-rich octocoral-containing grouping comprised the two taxonomically unresolved GCA_977007035.1 models, two *Corticium candelabrum* proteins, two *Oscarella lobularis* proteins, one *Oscarella pearsei* transcript and one *Plakina jani* transcript. Their placement does not by itself reject COX/PTGS homology because all remained on the COX/PTGS-like side of the exact family bipartition, but it identifies a divergent subset that should be treated separately. The candidate-focused annotation view further showed that all three Hexactinellida representatives remained together, whereas six of eight candidates in the Homoscleromorpha-enriched terminal group carried the shared channel-divergent profile (Fig. 3). The two unresolved GCA_977007035.1 models were excluded from class-specific interpretation.

The length-sensitivity tree retained the exact family bipartition after adding partial and extended proteins. After pruning to the 128 common taxa, 109 of 125 primary internal splits were shared, corresponding to 87.2%, and the normalised Robinson-Foulds distance was 0.128. RogueNaRok did not flag any taxon under the declared single-taxon settings. Thus, partial and extended sequences caused local rearrangements but did not drive the separation between COX/PTGS-like and non-COX families or the major candidate-rich clade.

3.10. Expanded Havird-Porifera phylogeny recovered a broad COX core and a Demospongiae-octocoral association

The historical 148-protein dataset was reanalysed together with 58 conservative Porifera locus/assembly representatives in the primary tree and with all 66 curated Porifera candidates in a source-inclusive sensitivity tree (Fig. 4a; Table 5; Supplementary Tables S17–S21).

The two topologies were similar after projection to their 206 shared taxa: 190 of 203 internal splits were shared (split Jaccard 0.880; normalised Robinson-Foulds distance 0.064; Supplementary Table S22). Thus, removing eight near-identical source or assembly replicas reduced tip redundancy without materially changing the overall topology.

In the 206-protein LG + R7 analysis, the 14-protein bipartition containing the 13 historical MPO/PXDN references and the short historical *Amphimedon queenslandica* record H15_134 received SH-aLRT/UFBoot support of 100/100. A broad 189-protein COX-rich bipartition received 96.4/97 support and contained 134 historical COX records and 55 of the 58 Porifera representatives. The three Porifera representatives outside this bipartition were the previously flagged 390-amino-acid *Dysidea avara* fragment P67_033, the 1,056-amino-acid extended *Halichondria panicea* model P67_040 and the 670-amino-acid *Leucetta chagosensis* model P67_045. In the 214-protein source-inclusive analysis, the corresponding 197-protein bipartition comprised the same 134 historical records and 63 of the 66 Porifera candidates and received 92.8/95 support.

Within the primary broad COX-rich region, a 44-protein bipartition comprised 37 Demospongiae representatives and all seven historical octocoral COX sequences. This split received 99.9/100 support, and the octocoral subclade received 100/100 (Fig. 4b; Supplementary Table S21). The source-inclusive analysis recovered the corresponding 52-protein bipartition, containing 45 Porifera candidates and the same seven octocoral sequences, with 100/100 support. All five Hexactinellida representatives formed an exact clade with 100/100 support. Seven Homoscleromorpha representatives and two taxonomically unresolved models from GCA_977007035.1 formed a separate nine-protein clade with 99.9/100 support. These class-enriched arrangements describe gene-tree groupings and were not interpreted as species-tree relationships or as evidence of one-to-one orthology.

The independent RogueNaRok analysis identified one current rogue taxon, P67_046, a 135-amino-acid *Mycale cecilia* fragment for which the EGF-bearing N-terminal interval was not evaluable (Supplementary Table S23). Removing P67_046 retained the broad COX-rich bipartition (96.8/89), the Demospongiae-octocoral bipartition (99.9/100), the Hexactinellida clade (100/100) and the Homoscleromorpha-plus-unresolved clade (100/100); relative to the primary topology projected to 205 taxa, split Jaccard similarity was 0.924 and normalised Robinson-Foulds distance was 0.040. Under the LG+C60 + F + G PMSF model, the broad COX-rich bipartition remained supported at 90.1/94, whereas support for the Demospongiae-octocoral bipartition decreased to 91/74. The Hexactinellida and Homoscleromorpha-plus-unresolved clades remained supported at 97.8/90 and 99.9/100, respectively. The seven-sequence octocoral subclade was recovered but weakly supported under PMSF (61.5/44). Relative to the primary tree, the PMSF topology shared 177 of 203 internal splits (split Jaccard 0.773; normalised Robinson-Foulds distance 0.128). These sensitivity analyses distinguish robust family- or class-level placement from model-sensitive deep and within-octocoral relationships. Complete uncollapsed primary and source-inclusive trees are provided in Supplementary Figs. S5–S7, with analysis definitions and split-level support in Supplementary Tables S17–S23.

The lineage-level update changes one central feature of the 2015 summary: Porifera can no longer be represented simply as a lineage without recovered COX homologues (Fig. 5).

The primary roster contained 58 locus/assembly representatives across 36 species labels: 43 representatives from 27 Demospongiae species, seven from four Homoscleromorpha species, five from three Hexactinellida species, one Calcarea representative and two candidates from one taxonomically unresolved Porifera assembly label. This extension records candidate COX/PTGS-like homology in Porifera; it does not demonstrate prostaglandin-endoperoxide synthase activity or establish the ancestral state of the phylum.

3.11. Six species contained multiple candidate COX/PTGS-like loci in the same assembly

Reconciliation of the 65 genome-mined models produced 47 non-overlapping genomic loci, 38 unique predicted protein sequences, 39 represented assemblies and 22 species (Supplementary Tables S24 and S26). Most species had no same-assembly copy-number evidence. Six species contained at least two non-overlapping COX/PTGS-like loci within one assembly: *Axinella dissimilis*, *Cliona cf. celata*, *Cliona cf. orientalis*, *Cliona varians*, *Phakellia ventilabrum* and *Spheciospongia papillosa* (Table 6).

Table 6

Species with two or more non-overlapping COX/PTGS-like candidate loci in the same genome assembly. RNA-confirmed loci are those with strong splice support and exact StringTie ORF confirmation under the available sampling. Candidate paralogy is supported but formal duplication timing and orthology are unresolved.

Species	Assembly	Candidate loci in assembly	Representative models	Protein identity (%)	CDS exon/intron architectures	RNA- and ORF-confirmed loci	Interpretation
<i>Axinella dissimilis</i>	GCA_978159415.1	2	MPROT000131; MPROT000198	63.8	9/8; 10/9	1/2	Two divergent candidate loci; one lacked RNA support under available sampling
<i>Cliona cf. celata</i>	GCA_982375335.1	2	MPROT000071; MPROT000333	59.5	10/9; 8/7	2/2	Two divergent RNA-confirmed candidate loci
<i>Cliona cf. orientalis</i>	GCA_963930775.1	2	MPROT000348; MPROT000328	72.2	9/8; 9/8	2/2	Two divergent RNA-confirmed candidate loci
<i>Cliona varians</i>	GCA_978019885.1	3	MPROT000324; MPROT000343; MPROT000327	73.3– 98.7	9/8	3/3	At least two divergent lineages; 98.7%-identical pair may be a recent duplicate or unpurged haplotype
<i>Phakellia ventilabrum</i>	GCA_963921505.1	2	MPROT000122; MPROT000172	64.7	9/8; 8/7	2/2	Two divergent RNA-confirmed candidate loci
<i>Spheciospongia papillosa</i>	GCA_982375275.1	2	MPROT000345; MPROT000321	84.7	9/8; 8/7	1/2	Two candidate loci; one lacked RNA support under available sampling

Five supported a minimum of two candidate loci, whereas one *C. varians* assembly contained three.

The compared same-assembly proteins were not exact duplicates. Divergent pairs shared 59.5–84.7% amino-acid identity across at least 93.8% mutual aligned overlap (Supplementary Table S25). A *C. varians* pair shared 98.7% identity and remains compatible with either a recent duplication or an unpurged haplotype; the same assembly also contained a third, more divergent locus. Both compared loci had strong splice support and an exactly confirming reconstructed ORF in *C. cf. celata*, *C. cf. orientalis*, *C. varians* and *P. ventilabrum*. Only one of two loci was RNA-supported under the available sampling in *A. dissimilis* and *S. papillosa*. All compared proteins retained strong COX/PTGS-like residue profiles.

Locus architectures comprised 28 nine-exon/eight-intron models, 13 eight-exon/seven-intron models and six ten-exon/nine-intron models. Different architectures co-occurred within single assemblies in *A. dissimilis* (9/8 and 10/9), *C. cf. celata* (10/9 and 8/7), *P. ventilabrum* (9/8 and 8/7) and *S. papillosa* (9/8 and 8/7). The locus-level tree placed most divergent copies in broader COX/PTGS-like lineages rather than as exclusive two-tip sister pairs (Supplementary Fig. S1). These observations support multiple candidate COX/PTGS-like loci and are compatible with paralogy, but they do not resolve duplication ages or one-to-one orthology. Integrated candidate multiplicity, exon-intron architecture and residue profiles are summarised in Supplementary Fig. S8.

4. Discussion

4.1. Porifera contain a broader COX/PTGS-like candidate repertoire than previously recognised

The combined evidence is inconsistent with a simple absence of COX/PTGS homologues from Porifera. Positive candidates were recovered independently from translated transcriptomes, annotated proteins and unannotated genomes, and the strictest loci were confirmed by exon-spanning RNA-seq and exact reconstruction of the predicted ORF. This conclusion differs from the absence reported by Havird et al. (2015), but the difference can be explained without invoking a biological contradiction. Porifera genome and transcriptome resources have expanded substantially since that survey (Havird et al. 2015; Vargas and Wörheide 2024), and the present analysis combined protein-to-genome modelling, reciprocal family discrimination and targeted RNA validation at a scale not previously available for the phylum.

The staged counts are essential to this interpretation. Sixty-seven broad genome records do not represent 67 independently validated genes. Replicate assemblies, identical proteins, partial models and candidates with incomplete transcript support are included in that layer. Conversely, restricting the result to the 15 class A/B assembly-level models would discard genuine transcriptome and annotated-protein evidence from additional species. The broad and strict layers answer different questions: the broad layer describes the distribution of recoverable COX-like genomic signal, whereas the strict layer identifies the most defensible direct gene models. The genomic-context audit adds a separate safeguard: it found no neighbourhood-level contamination concern among the 26 CDS-clean models, but it also demonstrated that host-locus support cannot repair conflicting source taxonomy. The two GCA_977007035.1 models are therefore retained as metazoan host loci with unresolved species and class rather than being used to support an *Aplysilla*- or *Oscarella*-specific claim.

4.2. The strongest candidates combine independent sequence, locus and expression evidence

The class A candidates in *Cliona varians* and *Dysidea avara* provide the strongest pre-functional evidence because related or identical proteins were recovered from independent sources and then confirmed at the genomic locus by splice-aware RNA-seq and transcript-derived ORFs. The exact correspondence between *Dysidea avara* SOAP.k25.C3288810.p1, the annotated XP_065916514.1 protein and genome/RNA-supported models is particularly compelling. These proteins also have approximately canonical COX length, EGF-like plus catalytic-core architecture, curated PGH2 best hits and conservation of Tyr385 and most monitored residues.

The class B loci extend the evidence beyond species with existing annotated COX-like proteins. *Phakellia ventilabrum*, *Pione* sp. GHC0000220, *Biemna* sp. UH-2024 and *Spheciospongia papillosa* each yielded a coherent genomic model, canonical splice structure, strong RNA support, exact ORF reconstruction and a full-length COX/PTGS-like protein. These records are therefore not products of joining unrelated genomic HSPs into an unsupported protein. The *Phakellia ventilabrum* model is a high-priority target for biochemical work because its genomic and transcriptomic evidence can be evaluated alongside independent reproductive lipidomic data from the same species (Koutsouveli et al. 2022).

4.3. Conserved and divergent residue profiles may reflect functional diversification

The focused residue analysis narrows the functional question without converting sequence conservation into an activity claim. All 18 candidates retain Tyr385 and the seven monitored POX/haem-environment residues. Mutagenesis of the P05979-equivalent Tyr385, H207, H309 and H388 positions can disrupt one or both PGHS activities, whereas none of the sponge proteins carries an equivalent experimentally established switch-off state (Shimokawa et al. 1990; Shimokawa and Smith 1991; Seibold et al. 2003). The nine canonical-channel-compatible proteins are therefore the most defensible initial targets for assays of coupled arachidonate oxygenation and peroxide reduction, not proteins already shown to catalyse those reactions.

Several non-reference states are compatible with functional diversification rather than simple loss. E524A occurs in active *Gersemia* and *Plexaura* COX enzymes, and V349I is present in the 15R-specific *P. homomalla* enzyme. The product-positioning variants that combine V349I with E524T and L534I may therefore alter substrate pose or product stereochemistry. This interpretation is consistent with mutagenesis showing that alternative arachidonate arrangements within the COX channel generate different oxygenated products, but the exact sponge combinations require direct product profiling (Thuresson et al. 2000, 2001; Valmsen et al. 2004).

The S530Y-containing proteins define a different experimental problem. S530 lies in the upper channel and is the classical aspirin-acetylation position; a Tyr replacement introduces a substantially bulkier side chain. Variation at Y504 may additionally alter the radical network (Shimokawa and Smith 1992; Rogge et al. 2009). These proteins are therefore predicted to be mechanistically informative channel variants, but neither COX inactivity nor isolated POX activity can be assigned from the available substitutions.

The two Homoscleromorpha proteins are the clearest substrate-entry outliers. Their R120G/S and Y355G states remove side chains that anchor the carboxylate and constrict the canonical arachidonate channel in mammalian COX enzymes (Bhattacharyya et al. 1996; Thuresson et al. 2001). Nevertheless, both retain the monitored POX/haem core, Tyr385, EGF-like architecture and COX/PTGS-like phylogenetic placement. They should therefore be tested for altered fatty-acid affinity, substrate range or product spectrum rather than discarded by a vertebrate motif template. The partial *Corticium* XP_062518513.1 record remains a separate secondary candidate because incomplete sequence architecture, rather than a coherent full-length variant profile, limits its interpretation.

4.4. Phylogenetic placement links most sponge candidates to non-vertebrate COX diversity

The strongest family-assignment evidence remains the dedicated domain-restricted analysis, in which every primary Porifera candidate and every experimentally characterised COX reference was separated from all strict PXDN-like and MPO-like comparators with maximal SH-aLRT/UFBoot support. The exact partition was unchanged under the LG+C60 posterior-mean-site-frequency sensitivity model. This conclusion does not rest on the shared animal haem-peroxidase-superfamily core: all comparators possessed that core, whereas the candidate set additionally combined EGF-like architecture, conserved Tyr385 and substantially greater retention of the monitored COX-site profile.

Within this restricted comparison of metazoan proteins sharing the animal haem-peroxidase core, EGF-like detection provides positive architectural support for a COX/PTGS-like assignment, but the module is neither specific to COX proteins nor sufficient to establish orthology or catalytic activity. Its detection among the expanded historical COX roster was frequent rather than universal, and the basis of the five profile non-detections in 501–559-amino-acid records remains unresolved. Two evaluable but atypical Porifera records, *Leucetta chagosensis* C10040 and the extended *Halichondria panicea* sequence, also lacked PS50026 and occurred outside the broad COX-rich grouping. Their exclusion from the strict family-discrimination roster was based on length rather than EGF status, avoiding circularity between domain annotation and phylogenetic interpretation; nevertheless, their atypical architectures preclude treating non-detection as evidence of evolutionary domain loss. The experimentally characterised prostaglandin H synthase of the red alga *Gracilaria vermiculophylla* lacks an EGF-like domain while retaining the capacity to convert arachidonic acid to PGG2 and PGH2, demonstrating that the module is not required for cyclooxygenase catalysis (Varvas et al. 2013; Gupta and Selinsky 2015). EGF-like detection should therefore be interpreted together with phylogenetic placement, reciprocal annotation and catalytic-site conservation, whereas non-detection alone does not exclude COX/PTGS homology.

Reanalysis of the complete recoverable 148-sequence Havird et al. framework with a conservatively dereplicated Porifera roster provides a macroevolutionary test independent of the targeted family-discrimination tree. Fifty-five of 58 primary Porifera representatives entered a supported COX-rich bipartition, whereas all three exceptions had been flagged independently as partial or extended models. The association of 37 Demospongiae

representatives with seven published octocoral COX sequences was strongly supported under LG + R7 and was also recovered in the source-inclusive dataset. Its retention with reduced support under PMSF (91/74) indicates that the grouping is reproducible but that confidence in this deep relationship is model-dependent; the weak PMSF support for the internal octocoral subclade provides an additional reason not to interpret a resolved sponge-octocoral sister relationship. By contrast, the Hexactinellida and Homoscleromorpha-enriched groupings retained high or maximal support under both LG + R7 and PMSF. These patterns strengthen assignment to a COX/PTGS-like evolutionary radiation and reveal class-structured sequence diversification, but they do not establish the root of COX evolution, one-to-one orthology, the number or timing of duplications, or catalytic function.

The Porifera candidates were not distributed randomly within the COX-rich region. Taxonomically coherent local groupings included representatives of Axinellida (*Axinella*, *Phakellia* and *Cymbastela*), Clionaida (*Cliona*, *Spheciospongia* and *Pione*), Hexactinellida and Homoscleromorpha, although the gene tree was not perfectly congruent with the species tree. The records originated from independent assemblies, transcriptomes, projects and geographic sampling, and the strongest genomic loci were multiexonic, had canonical splice junctions and were supported by intron-spanning RNA reads. Together with the genomic-neighbourhood audit, this distribution makes systematic contamination by cnidarian sequences an unlikely explanation for the overall pattern. It does not exclude isolated contamination, source misidentification or assembly error at individual loci, particularly on short or weakly annotated scaffolds; the GCA_977007035.1 taxonomic conflict illustrates that remaining limitation.

The Demospongiae-octocoral grouping has several plausible evolutionary explanations. An ancient duplication predating the Porifera–Cnidaria split, followed by differential retention or loss, could place sponge and octocoral homologues in one paralogous lineage while other metazoan homologues occupy another (Järving et al. 2004; Havird et al. 2008, 2015). Alternatively, lineage-specific rates and residual site-compositional heterogeneity could generate or strengthen the association, consistent with its reduced PMSF support. Because *Plexaura homomalla* Q962I8 was included in the discovery query panel, its high pairwise identity to some sponge candidates is partly conditioned by ascertainment and was not treated as independent evidence for ancient paralogy or a resolved sponge-octocoral sister relationship. Distinguishing these alternatives will require species-tree-aware reconciliation, chromosome-scale synteny and wider taxon sampling; ancient paralogy is therefore a testable hypothesis rather than a resolved historical inference.

Havird et al. (2015) inferred that a COX homologue was probably present in the common ancestor of Cnidaria and Bilateria, while reporting no convincing homologues in the poriferan resources then available. The recovery of phylogenetically supported COX/PTGS-like candidates across multiple Porifera classes therefore extends the minimum supported phylogenetic depth of this protein lineage beyond the cnidarian–bilaterian node. Under a vertical-inheritance model, the observed distribution is most parsimoniously consistent with the presence of a COX/PTGS-like homologue in the last common ancestor shared by Porifera, Cnidaria and Bilateria, followed by lineage-specific retention, diversification, or loss. This inference neither identifies the branch on which the family originated nor establishes its ancestral distribution in Ctenophora or Placozoa. Its precise interpretation remains conditional on the debated branching order among the earliest-diverging metazoan lineages (Redmond and McLysaght 2021; Schultz et al. 2023).

Molecular-clock analyses place the early diversification of extant animal lineages deep in the Neoproterozoic, with some analyses inferring pre-Sturtian divergences for the major non-bilaterian lineages (Dohrmann and Wörheide 2017). However, broad credibility intervals, fossil-calibration uncertainty, and sensitivity to deep metazoan topology prevent these external estimates from being translated into a defensible number of millions of years by which the COX/PTGS-like lineage is extended (dos Reis et al. 2015). The present results therefore increase the minimum supported phylogenetic depth of the lineage without assigning an absolute age to its origin. They concern the evolutionary history of COX/PTGS-like homologues and do not establish that the ancestral protein possessed the substrate specificity or coupled catalytic activities of experimentally characterised cyclooxygenases.

The primary tree should not be interpreted as a resolved species tree or as proof that all candidate duplications are ancient. Multiple candidates within a species may represent paralogues, alleles, isoforms or assembly-specific models. Exact amino-acid duplicates were removed, but non-identical proteins were deliberately retained so that biologically meaningful copies were not collapsed. More detailed orthology inference will require chromosome-scale assemblies, synteny and species-tree-aware reconciliation. The present tree instead supports family-level placement and identifies candidate groups for focused evolutionary and structural comparison.

4.5. Exon-intron architecture supports coherent COX-like gene models

Exon-intron architecture provides evidence independent of protein similarity and phylogenetic placement. Among the 47 reconciled genomic loci, 28 contained nine coding exons and eight introns, 13 contained eight exons and seven introns, and six contained 10 exons and nine introns. The predominant 9/8 organisation matches each of the two functionally characterised COX genes of the octocoral *Gersemia fruticosa*, which also retain exon boundaries broadly conserved with vertebrate COX genes (Järving et al. 2004). By comparison, human PTGS1 and PTGS2 contain 11 and 10 exons, respectively, and ascidian COX genes can contain 12, illustrating that a conserved domain architecture can coexist with lineage-specific intron gain or loss (Järving et al. 2004; Chandrasekharan and Simmons 2004).

The correspondence is supportive rather than diagnostic. A nine-exon model does not by itself establish COX activity or one-to-one orthology, and the 8/7 and 10/9 structures should not be treated as evidence against COX-like homology. Their canonical splice junctions, coherent full-length translations and, for the strongest loci, RNA reads spanning the predicted exon boundaries instead argue that the recovered proteins are internally consistent gene models rather than products assembled from distant genomic matches.

The co-occurrence of different architectures within single assemblies in *Axinella dissimilis*, *Cliona cf. celata*, *Phakellia ventilabrum* and *Spheciospongia papillosa* further supports the presence of distinct loci, but it does not determine whether exon-count differences preceded or followed duplication. Independent COX duplications are known in octocoral, ascidian and vertebrate lineages (Järving et al. 2004; Havird et al. 2008, 2015), making paralogy plausible for the sponge copy sets; chromosome-scale synteny and species-tree-aware reconciliation remain necessary to resolve duplication timing and orthology.

4.6. Multiple loci indicate candidate lineage-specific expansion

Same-assembly locus multiplicity strengthens the interpretation that at least part of the observed sequence diversity is biological rather than a consequence of comparing alternative predictions or independent assemblies. The strongest cases are *C. cf. celata*, *C. cf. orientalis*, *C. varians* and *P. ventilabrum*, in which spatially distinct loci are accompanied by splice-aware RNA and exact ORF support for each compared copy. Differences in protein sequence and, in several species, exon-intron architecture further argue against simple duplicate reporting of one gene model. The coexistence of two RNA-supported *P. ventilabrum* loci is noteworthy in view of the reproductive lipid remodelling described for this species, although neither locus can be assigned a physiological prostanoid function from sequence evidence (Koutsouveli et al. 2022). This pattern is compatible with the recurrent, lineage-specific COX duplications documented elsewhere in Metazoa, but does not identify the age or direction of the sponge duplications (Järving et al. 2004; Havird et al. 2008, 2015).

The copy relationships nevertheless remain provisional. A same-assembly locus pair can reflect true paralogy, uncollapsed haplotypes or local assembly duplication, particularly for the 98.7%-identical *C. varians* pair. Conversely, the more divergent pairs are distributed across the COX/PTGS-like tree and need not represent recent species-specific duplications. The appropriate conclusion is therefore that six species contain multiple candidate COX/PTGS-like loci, with strong transcriptional support for both copies in four species. Establishing orthology, duplication timing and lineage-specific retention will require chromosome-scale flanking synteny and species-tree-aware gene-tree reconciliation.

4.7. Limitations and next tests

Several limitations remain. First, curated best hits and phylogenetic placement support homology but do not demonstrate prostaglandin-endoperoxide synthase activity. Second, RNA-seq evidence is condition-dependent: absence of reconstruction cannot distinguish gene absence from low or restricted expression. Targeted reconstruction confirms transcript compatibility with predicted loci but does not necessarily provide locus-unique expression evidence for nearly identical copies; this limitation is particularly relevant to the 98.7%-identical *C. varians* pair. Third, same-genus proxy RNA was used for a small subset and is explicitly distinguished from same-species support. Fourth, partial and extended sequences remain sensitive to assembly and ORF-boundary errors, which is why the 500-650-amino-acid primary tree was accompanied by a full length-sensitivity analysis. Fifth, the inferred gene models were targeted annotations of candidate loci rather than whole-genome annotations. Sixth, translated neighbourhoods provide an affinity screen rather than a formal contamination model based on raw-read coverage, k-mer composition or chromosome-scale synteny; the audit therefore supports, but does not alone establish, organismal provenance.

The next experimental priorities are therefore biochemical. Canonical-channel-compatible representatives should be expressed recombinantly, assessed for haem incorporation and tested in paired cyclooxygenase and peroxidase assays. Product-positioning variants require liquid chromatography-tandem mass spectrometry (LC-MS/MS) identification and stereochemical analysis of products formed from arachidonic acid and alternative polyunsaturated fatty acids. Channel-constriction and divergent substrate-entry variants should be examined for substrate range, kinetic efficiency and inhibitor sensitivity. These experiments are future work: the present sequence, gene-model, RNA and phylogenetic evidence establishes testable COX/PTGS-like homologues, not functional cyclooxygenases or prostaglandin biosynthesis *in vivo*.

5. Conclusions

Independent transcriptome, annotated-protein, genome and RNA analyses recovered a staged set of Porifera COX/PTGS-like candidates. The strongest 15 class A/B assembly-level models combine full-length COX-like proteins, curated COX/PTGS best hits, EGF-like architecture plus an animal haem-peroxidase-superfamily profile, coherent exon-intron structures, splice-aware RNA support and exact transcript-derived ORF confirmation. A focused 18-candidate comparison retained the complete monitored POX/haem and Tyr385-centred COX radical framework in every protein while separating nine canonical-channel-compatible candidates from product-positioning, channel-constriction and divergent substrate-entry variants. After exclusion of the short transcript-derived record P67_035 from the conservative primary roster because of its anomalous length and taxonomically discordant placement, and conservative consolidation of near-identical source or assembly replicas, 58 representatives across 36 species-level labels were retained for the primary expanded phylogeny. Fifty-five entered a supported COX-rich bipartition, and class-enriched Porifera groupings were recovered across model and rogue-taxon sensitivity analyses. The recovery of COX/PTGS-like homologues in Porifera extends the minimum supported phylogenetic depth of this protein lineage beyond the cnidarian–bilaterian node and, under a vertical-inheritance interpretation, is consistent with its presence in the last common ancestor shared by Porifera, Cnidaria and Bilateria. These data support a broader and taxonomically structured Porifera COX/PTGS-like repertoire than previously recognised; enzyme activity, substrate and product specificity, and the timing of candidate gene duplications remain to be established experimentally.

Declarations

Competing interests

The author has no relevant financial or non-financial interests to disclose.

Ethics approval

Not applicable. This study analysed publicly available genomic, transcriptomic and protein sequence data and did not involve new collection or experimental use of animals.

Funding

No funding was received for conducting this study.

Author Contribution

S.S. conceived and designed the study, acquired and curated the data, developed and executed the analyses, interpreted the results, prepared the figures and supplementary materials, and wrote and revised the manuscript.

Data Availability

The complete data and reproducibility package supporting this study is publicly available from Zenodo (Scibelli 2026; <https://doi.org/10.5281/zenodo.22041914>). All accession-level genome, transcriptome and RNA-seq inputs used in the analysis are listed in Supplementary Tables S1–S3. Complete CANTATA and annotated-protein decisions, uniform domain and residue calls, host-attribution results, evidence definitions, phylogenetic metadata and sensitivity analyses are reported in Supplementary Tables S4–S23. Locus multiplicity, gene architecture, the discovery funnel, source-specific candidate sets, RNA evidence and the integrated evidence matrix are reported in Supplementary Tables S24–S39. The analysis-to-file registry and software specifications are provided in Supplementary Tables S40 and S41. The Zenodo record includes FASTA files, alignments, raw unrooted IQ-TREE outputs, mutagenesis-evidence tables, figure source data and versioned scripts.

Code availability

Versioned scripts used for parsing, candidate classification, gene-model extraction, RNA support, residue mapping, uniform domain analysis, table construction, phylogenetic roster curation, alignment, tree inference, rogue-taxon analysis, topology comparison and figure production are included in the Zenodo reproducibility package (<https://doi.org/10.5281/zenodo.22041914>). The final expanded-phylogeny specifications and support tests are reported in Supplementary Tables S17–S23, the analysis-to-file registry in Supplementary Table S40 and software versions and parameters in Supplementary Table S41.

References

1. Aberer AJ, Krompass D, Stamatakis A (2013) Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Syst Biol* 62:162–166. <https://doi.org/10.1093/sysbio/sys078>
2. Bhattacharyya DK, Lecomte M, Rieke CJ, Garavito RM, Smith WL (1996) Involvement of arginine 120, glutamate 524, and tyrosine 355 in the binding of arachidonate and 2-phenylpropionic acid inhibitors to the cyclooxygenase active site of ovine prostaglandin endoperoxide H synthase-1. *J Biol Chem* 271:2179–2184. <https://doi.org/10.1074/jbc.271.4.2179>
3. Buchfink B, Reuter K, Drost HG (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 18:366–368. <https://doi.org/10.1038/s41592-021-01101-x>
4. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>
5. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
6. Chandrasekharan NV, Simmons DL (2004) The cyclooxygenases. *Genome Biol* 5:241. <https://doi.org/10.1186/gb-2004-5-9-241>
7. Danecek P, Bonfield JK, Liddle J et al (2021) Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>
8. DeWitt DL, el-Harith EA, Kraemer SA, Andrews MJ, Yao EF, Armstrong RL, Smith WL (1990) The aspirin and heme-binding sites of ovine and murine prostaglandin endoperoxide synthases. *J Biol Chem* 265:5192–5198. [https://doi.org/10.1016/S0021-9258\(19\)34105-5](https://doi.org/10.1016/S0021-9258(19)34105-5)
9. Dobin A, Davis CA, Schlesinger F et al (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
10. Dohrmann M, Wörheide G (2017) Dating early animal evolution using phylogenomic data. *Sci Rep* 7:3599. <https://doi.org/10.1038/s41598-017-03791-w>
11. dos Reis M, Thawornwattana Y, Angelis K, Telford MJ, Donoghue PCJ, Yang Z (2015) Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr Biol* 25:2939–2950. <https://doi.org/10.1016/j.cub.2015.09.066>
12. Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
13. Gramzow M, Schröder HC, Fritsche U, Kurelec B, Robitzki A, Zimmermann H, Friese K, Kreuter MH, Müller WEG (1989) Role of phospholipase A2 in the stimulation of sponge cell proliferation by homologous lectin. *Cell* 59:939–948. [https://doi.org/10.1016/0092-8674\(89\)90616-8](https://doi.org/10.1016/0092-8674(89)90616-8)
14. Grosser T, Yusuff S, Cheskis E, Pack MA, FitzGerald GA (2002) Developmental expression of functional cyclooxygenases in zebrafish. *Proc Natl Acad Sci USA* 99:8418–8423. <https://doi.org/10.1073/pnas.112217799>
15. Gupta K, Selinsky BS (2015) Bacterial and algal orthologs of prostaglandin H2 synthase: novel insights into the evolution of an integral membrane protein. *Biochim Biophys Acta Biomembr* 1848:83–94. <https://doi.org/10.1016/j.bbamem.2014.09.011>
16. Havird JC, Miyamoto MM, Choe KP, Evans DH (2008) Gene duplications and losses within the cyclooxygenase family of teleosts and other chordates. *Mol Biol Evol* 25:2349–2359. <https://doi.org/10.1093/molbev/msn183>

17. Havird JC, Kocot KM, Brannock PM, Cannon JT, Waits DS, Weese DA, Santos SR, Halanych KM (2015) Reconstruction of cyclooxygenase evolution in animals suggests variable, lineage-specific duplications, and homologs with low sequence identity. *J Mol Evol* 80:193–208. <https://doi.org/10.1007/s00239-015-9670-3>
18. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35:518–522. <https://doi.org/10.1093/molbev/msx281>
19. Järving R, Järving I, Kurg R, Brash AR, Samel N (2004) On the evolutionary origin of cyclooxygenase isozymes: characterization of marine invertebrate COX genes points to independent duplication events in vertebrate and invertebrate lineages. *J Biol Chem* 279:13624–13633. <https://doi.org/10.1074/jbc.M313258200>
20. Jones P, Binns D, Chang HY et al (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
21. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
22. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
23. Kawamura M, Inaoka H, Obata S, Harada Y (2014) Why do a wide variety of animals retain multiple isoforms of cyclooxygenase? Prostaglandins Other Lipid Mediators 109–111:14–22. <https://doi.org/10.1016/j.prostaglandins.2014.03.002>
24. Koljak R, Järving I, Kurg R et al (2001) The basis of prostaglandin synthesis in coral: molecular cloning and expression of a cyclooxygenase from the Arctic soft coral *Gersemia fruticosa*. *J Biol Chem* 276:7033–7040. <https://doi.org/10.1074/jbc.M009803200>
25. Koutsouveli V, Balgoma D, Checa A, Hedeland M, Riesgo A, Cárdenas P (2022) Oogenesis and lipid metabolism in the deep-sea sponge *Phakellia ventilabrum* (Linnaeus, 1767). *Sci Rep* 12:6317. <https://doi.org/10.1038/s41598-022-10058-6>
26. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
27. Li H (2023) Protein-to-genome alignment with miniprot. *Bioinformatics* 39:btad014. <https://doi.org/10.1093/bioinformatics/btad014>
28. Miciaccia M, Belviso BD, Iaselli M et al (2021) Three-dimensional structure of human cyclooxygenase (hCOX)-1. *Sci Rep* 11:4312. <https://doi.org/10.1038/s41598-021-83438-z>
29. Perteu M, Perteu GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33:290–295. <https://doi.org/10.1038/nbt.3122>
30. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>
31. Redmond AK, McLysaght A (2021) Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding. *Nat Commun* 12:1783. <https://doi.org/10.1038/s41467-021-22074-7>
32. Rogge CE, Liu W, Kulmacz RJ, Tsai AL (2009) Peroxide-induced radical formation at Tyr385 and Tyr504 in human PGHS-1. *J Inorg Biochem* 103:912–922. <https://doi.org/10.1016/j.jinorgbio.2009.04.002>
33. Rouzer CA, Marnett LJ (2020) Structural and chemical biology of the interaction of cyclooxygenase with substrates and non-steroidal anti-inflammatory drugs. *Chem Rev* 120:7592–7641. <https://doi.org/10.1021/acs.chemrev.0c00215>
34. Schoch CL, Ciufo S, Domrachev M et al (2020) NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* 2020:baaa062. <https://doi.org/10.1093/database/baaa062>
35. Schultz DT, Haddock SHD, Bredeson JV, Green RE, Simakov O, Rokhsar DS (2023) Ancient gene linkages support ctenophores as sister to other animals. *Nature* 618:110–117. <https://doi.org/10.1038/s41586-023-05936-6>
36. Scibelli S (2026) Porifera COX/PTGS-like candidates: data and reproducibility package. <https://doi.org/10.5281/zenodo.22041914>. Version 1.0.0. Zenodo
37. Seibold SA, Ball T, Hsi LC, Mills DA, Abeysinghe RD, Micielli R, Rieke CJ, Cukier RI, Smith WL (2003) Histidine 386 and its role in cyclooxygenase and peroxidase catalysis by prostaglandin-endoperoxide H synthases. *J Biol Chem* 278:46163–46170. <https://doi.org/10.1074/jbc.M306319200>
38. Shimokawa T, Kulmacz RJ, DeWitt DL, Smith WL (1990) Tyrosine 385 of prostaglandin endoperoxide synthase is required for cyclooxygenase catalysis. *J Biol Chem* 265:20073–20076. [https://doi.org/10.1016/S0021-9258\(17\)30468-4](https://doi.org/10.1016/S0021-9258(17)30468-4)
39. Shimokawa T, Smith WL (1991) Essential histidines of prostaglandin endoperoxide synthase. His-309 is involved in heme binding. *J Biol Chem* 266:6168–6173. [https://doi.org/10.1016/S0021-9258\(18\)38099-2](https://doi.org/10.1016/S0021-9258(18)38099-2)
40. Shimokawa T, Smith WL (1992) Prostaglandin endoperoxide synthase. The aspirin acetylation region. *J Biol Chem* 267:12387–12392. [https://doi.org/10.1016/S0021-9258\(19\)49852-9](https://doi.org/10.1016/S0021-9258(19)49852-9)
41. Sigrist CJA, de Castro E, Cerutti L et al (2013) New and continuing developments at PROSITE. *Nucleic Acids Res* 41:D344–D347. <https://doi.org/10.1093/nar/gks1067>
42. Simmons DL, Botting RM, Hla T (2004) Cyclooxygenase isozymes: the biology of prostaglandin synthesis and inhibition. *Pharmacol Rev* 56:387–437. <https://doi.org/10.1124/pr.56.3.3>
43. Smith MR (2021) Rogue: identify rogue taxa in sets of phylogenetic trees. <https://doi.org/10.5281/zenodo.5037327>. Zenodo
44. Thuresson ED, Lakkides KM, Smith WL (2000) Different catalytically competent arrangements of arachidonic acid within the cyclooxygenase active site of prostaglandin endoperoxide H synthase-1 lead to the formation of different oxygenated products. *J Biol Chem* 275:8501–8507. <https://doi.org/10.1074/jbc.275.12.8501>

45. Thuresson ED, Lakkides KM, Rieke CJ, Sun Y, Wingerd BA, Micielli R, Mulichak AM, Malkowski MG, Garavito RM, Smith WL (2001) Prostaglandin endoperoxide H synthase-1: the functions of cyclooxygenase active site residues in the binding, positioning, and oxygenation of arachidonic acid. *J Biol Chem* 276:10347–10357. <https://doi.org/10.1074/jbc.M009377200>
46. Valmsen K, Järving I, Boeglin WE et al (2001) The origin of 15R-prostaglandins in the Caribbean coral *Plexaura homomalla*: molecular cloning and expression of a novel cyclooxygenase. *Proc Natl Acad Sci USA* 98:7700–7705. <https://doi.org/10.1073/pnas.131022398>
47. Valmsen K, Boeglin WE, Järving I, Schneider C, Varvas K, Brash AR, Samel N (2004) Structural and functional comparison of 15S- and 15R-specific cyclooxygenases from the coral *Plexaura homomalla*. *Eur J Biochem* 271:3533–3538. <https://doi.org/10.1111/j.0014-2956.2004.04289.x>
48. Vargas S, Wörheide G (2024) Phylum Porifera CANTATA Transcriptomes, version 1.0. Zenodo. <https://doi.org/10.5281/zenodo.8208114>
49. Varvas K, Kasvandik S, Hansen K, Järving I, Morell I, Samel N (2013) Structural and catalytic insights into the algal prostaglandin H synthase reveal atypical features of the first non-animal cyclooxygenase. *Biochim Biophys Acta Mol Cell Biol Lipids* 1831:863–871. <https://doi.org/10.1016/j.bbalip.2012.11.010>
50. Wang H-C, Minh BQ, Susko E, Roger AJ (2018) Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst Biol* 67:216–235. <https://doi.org/10.1093/sysbio/syx068>
51. Wong TKF, Ly-Trong N, Ren H et al (2026) IQ-TREE 3: phylogenomic inference software using complex evolutionary models. *Mol Biol Evol* 43:msag117. <https://doi.org/10.1093/molbev/msag117>
52. Zámocký M, Hofbauer S, Schaffner I, Gasselhuber B, Nicolussi A, Soudi M, Pirker KF, Furtmüller PG, Obinger C (2015) Independent evolution of four heme peroxidase superfamilies. *Arch Biochem Biophys* 574:108–119. <https://doi.org/10.1016/j.abb.2014.12.025>

Figures

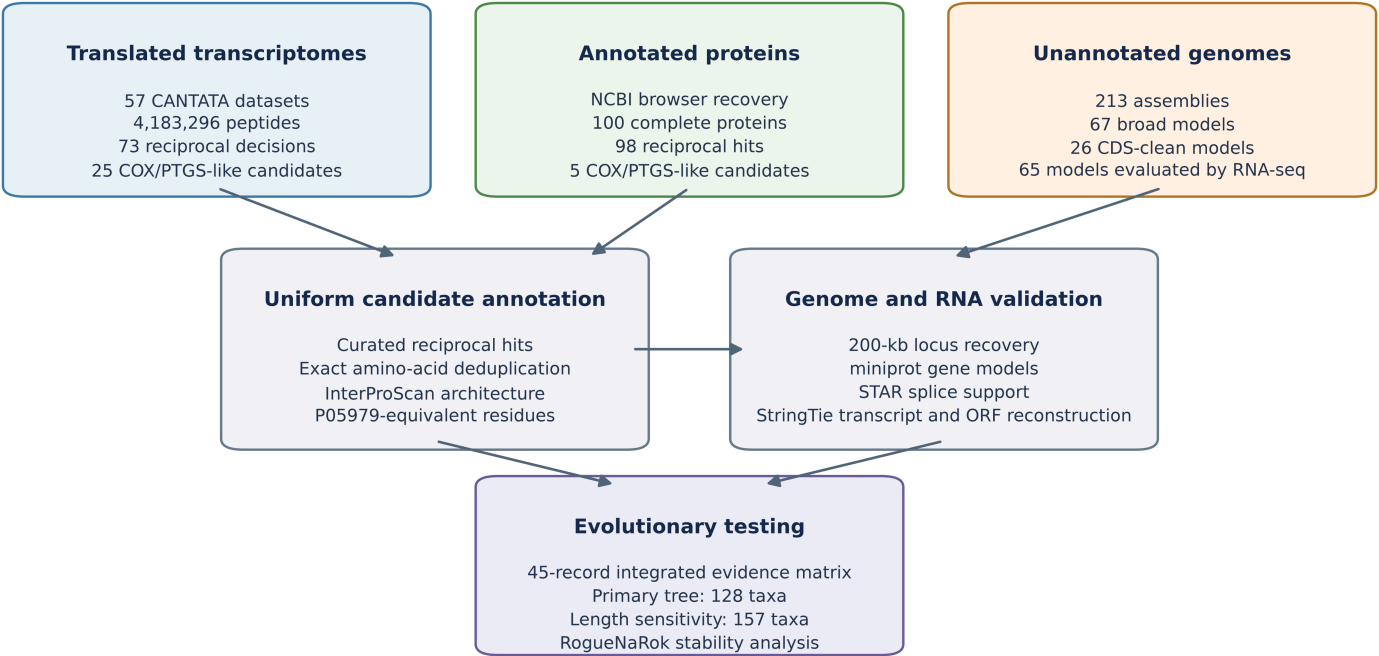


Figure 1

Integrative discovery and validation workflow. Transcriptome, annotated-protein and unannotated-genome evidence streams were evaluated independently before uniform reciprocal, architecture and residue analysis. Genome-derived candidates were then tested using region-scale gene modelling, splice-aware RNA-seq and transcript/ORF reconstruction. The 65 RNA-evaluable genome models were reconciled to 47 non-overlapping genomic loci and 38 globally unique amino-acid sequences. Counts represent successive evidence layers and are not additive

	AA anchor																POX / haem																COX positioning				radical / haem interface				upper COX channel				Residue profile	Evidence support	RNA	Exons
	R120	Q203	H204	T206	H207	H209	Y246	V249	Y255	Y265	H266	W267	H268	Y269	Y274	E274	S280	G283	L284																													
Cliona cf celata C010 GCA_982375335	R	Q	H	T	H	H	Y	V	Y	Y	H	W	H	Y	A	S	G	L	Canonical	Genome/RNA	Strong	8																										
Cliona cf orientalis C014 GCA_963930775	R	Q	H	T	H	H	Y	V	Y	Y	H	W	H	Y	A	S	G	L	Canonical	Genome/RNA	Strong	9																										
Cliona varians C017 GCA_978019885	R	Q	H	T	H	H	Y	V	Y	Y	H	W	H	Y	A	S	G	L	Canonical	Cross-validated	Strong	9																										
Cliona varians C018 GCA_978019885	R	Q	H	T	H	H	Y	V	Y	Y	H	W	H	Y	A	S	G	L	Canonical	Cross-validated	Strong	9																										
Cliona varians C020 GCA_978019905	R	Q	H	T	H	H	Y	V	Y	Y	H	W	H	Y	A	S	G	L	Canonical	Cross-validated	Strong	9																										
Cliona varians C021 TransAllyS3.K25.R051301.p1	R	Q	H	T	H	H	Y	V	Y	Y	H	W	H	Y	A	S	G	L	Canonical	Curated protein	Transcript	Transcript only																										
Dysidea avara C023 GCA_963966705	R	Q	H	T	H	H	Y	V	Y	Y	H	W	H	Y	A	S	G	L	Canonical	Cross-validated	Strong	9																										
Dysidea cf avara C028 GCA_963678995	R	Q	H	T	H	H	Y	V	Y	Y	H	W	H	Y	A	S	G	L	Canonical	Genome/RNA	Strong	9																										
Pione sp GNC0000220 C046 GCA_982532475	R	Q	H	T	H	H	Y	V	Y	Y	H	W	H	Y	A	S	G	L	Canonical	Genome/RNA	Strong	8																										
Cliona cf orientalis C053 GCA_983930775	R	Q	H	T	H	H	Y	I	Y	Y	H	W	H	Y	T	S	G	I	Product variant	Genome/RNA	Strong	9																										
Cliona varians C016 GCA_978019885	R	Q	H	T	H	H	Y	I	Y	Y	H	W	H	Y	T	S	G	I	Product variant	Cross-validated	Strong	9																										
Cliona varians C019 GCA_978019905	R	Q	H	T	H	H	Y	I	Y	Y	H	W	H	Y	T	S	G	I	Product variant	Cross-validated	Strong	9																										
Sphaeclopongia papillosa C048 GCA_982375275	R	Q	H	T	H	H	Y	I	Y	Y	H	W	H	Y	T	S	G	I	Product variant	Genome/RNA	Strong	9																										
Biemna sp UH 2024 C008 GCA_965654295	R	Q	H	T	H	H	Y	V	Y	Y	H	W	H	V	D	Y	G	L	Channel variant	Genome/RNA	Strong	9																										
Cliona cf celata C012 GCA_982375485	R	Q	H	T	H	H	Y	V	Y	Y	H	W	H	F	D	Y	G	L	Channel variant	Genome/RNA	Strong	9																										
Phakella ventralis C042 MPR07000122	R	Q	H	T	H	H	Y	V	Y	Y	H	W	H	F	D	Y	G	L	Channel variant	Genome/RNA	Strong	9																										
Corticium candelabrum C024 XP_062518712.1	G	Q	H	S	H	H	Y	V	G	Y	H	W	H	F	E	A	G	I	Entry-divergent	Curated protein	Not tested	2																										
Oscarella lobularis C036 XP_065827972.1	S	Q	H	S	H	H	Y	L	G	Y	H	W	H	F	H	A	G	L	Entry-divergent	Curated protein	Not tested	4																										

PO5979-identical

Exact state in an experimentally active COX

Exact substitution tested directly

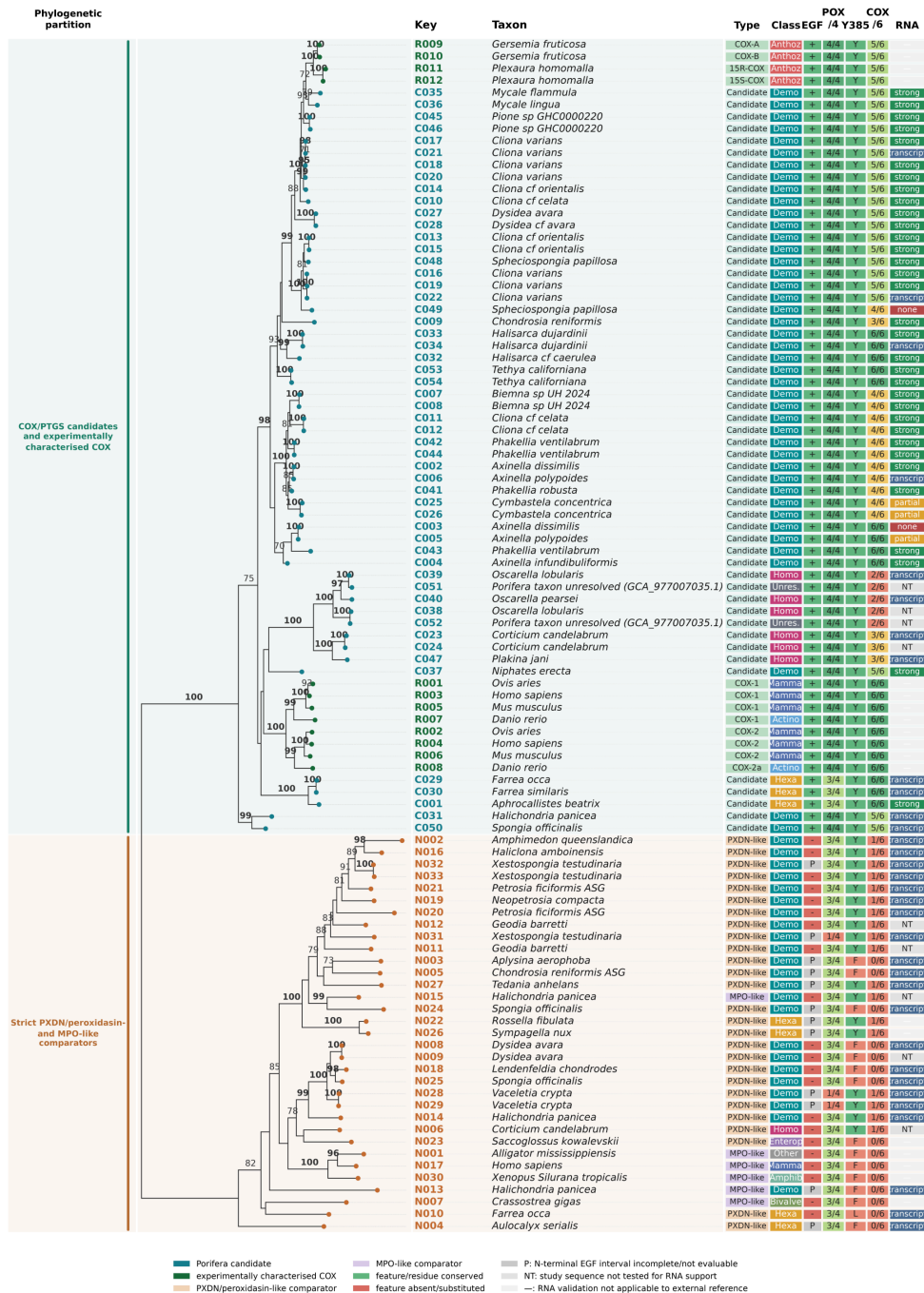
Same-site evidence / moderate uncertainty

Mechanistic concern; exact activity untested

P05979-identical
Exact substitution tested directly
Mechanistic concern; exact activity untested
Exact state in an experimentally active COX
Same-site evidence / moderate uncertainty

Figure 2

Mutagenesis-informed functional-site profiles of 18 non-redundant Porifera COX/PTGS-like candidates. Residues were mapped to full-length sheep COX-1 P05979 numbering after MAFFT L-INS-i alignment. Rows are grouped into canonical-channel-compatible, product-positioning variant, channel-constriction/radical-network variant and divergent substrate-entry profiles. Colours distinguish P05979-identical states, exact states present in experimentally active COX enzymes, exact substitutions tested by mutagenesis, same-site evidence and mechanistic concerns for which the observed substitution remains untested. Right-hand columns report descriptive evidence support, RNA evidence and CDS-exon count. The profiles describe sequence-level compatibility and do not predict catalytic rates or establish cyclooxygenase (COX) or peroxidase (POX) activity. Complete reference comparison and source evidence are provided in Supplementary Fig. S2 and Supplementary Tables S37–S39



POX and COX scores count exact P05979-equivalent residues (POX: Q203/H207/H386/H388; COX: R120/Y355/Y385/E524/S530/G533). COX-1/COX-2 labels are used only for experimentally characterised orthologues; Porifera sequences remain COX/PTGS-like candidates. Raw IQ-TREE topology is unrooted; rooting is for display on the exact 33-comparator edge.

Figure 3

Family-level discrimination of 54 Porifera COX/PTGS-like candidates from 12 experimentally characterised COX proteins and 33 strict non-COX animal haem-peroxidase comparators. The topology is the UFBoot consensus from the 99-protein PS50292-domain alignment (MAFFT L-INS-i; trimAl -gt 0.1; 676 sites; ModelFinder-selected Q.PFAM+H+R4; 5,000 UFBoot2 replicates). The raw IQ-TREE topology is unrooted; the display is oriented at the unique edge separating the 66-sequence candidate/reference side from the 33 peroxidasin (PXDN)-like or myeloperoxidase (MPO)-like comparators, without inferring evolutionary direction. Node labels show UFBoot values ≥ 70 . Columns report candidate/reference/comparator status, Porifera class or reference lineage, EGF-like status, exact P05979-equivalent POX/haem score, Tyr385, exact COX-site score and RNA support. P indicates that the N-terminal EGF-equivalent interval was incomplete or not evaluable. Complete accession, length, source, evidence-tier and gene-structure metadata, together with the posterior-mean-site-frequency sensitivity analysis, are provided in Supplementary Tables S14–S16

a Metazoan lineage-level COX evidence matrix

	2015 evidence (Havird et al.)	Current evidence summary
Mollusca	COX sequence	COX sequence
Nemertea	COX sequence	COX sequence
Brachiopoda	mixed / incomplete	mixed / incomplete
Annelida and Sipuncula	COX sequence	COX sequence
Platyhelminthes	not reported	not reported
Nematoda	not reported	not reported
Chelicerata	not reported	not reported
Myriapoda	COX sequence	COX sequence
Crustacea	COX sequence	COX sequence
Insecta	not reported	not reported
Echinodermata	not reported	not reported
Hemichordata	not reported	not reported
Teleostei	COX sequence	COX sequence
Tetrapoda	COX sequence	COX sequence
Tunicata	COX sequence	COX sequence
Cephalochordata	COX sequence	COX sequence
Cnidaria	mixed / incomplete	mixed / incomplete
Ctenophora	not reported	not reported
Placozoa	not reported	not reported
Porifera	not reported	candidate COX-like
Choanoflagellata	not reported	not reported

The current column retains the historical lineage summary and adds Porifera candidates recovered here; lineages other than Porifera were not systematically rescreened in this study.

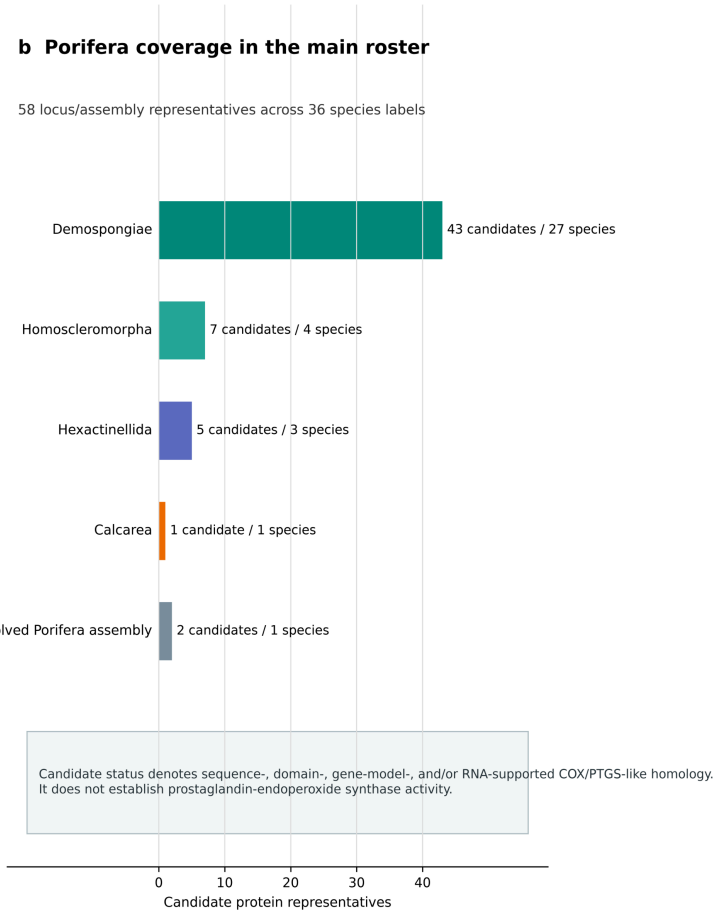


Figure 5

Updated distribution of COX sequence evidence across metazoan lineages. (a) Evidence matrix comparing the lineage-level COX sequence status summarised by Havird et al. (2015) with the present study. Historical categories distinguish published COX sequence evidence, mixed or incomplete evidence and lineages for which no COX was reported in that survey. The current column retains those historical states and adds Porifera candidate COX/PTGS-like homologues recovered here. The matrix is an evidence summary and not an ancestral-state reconstruction. (b) Taxonomic composition of the 58 Porifera locus/assembly representatives in the primary expanded phylogeny. Candidate status denotes sequence-, domain-, gene-model- and/or RNA-supported homology and does not establish enzyme activity

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESM1.xlsx](#)
- [ESM2.pdf](#)