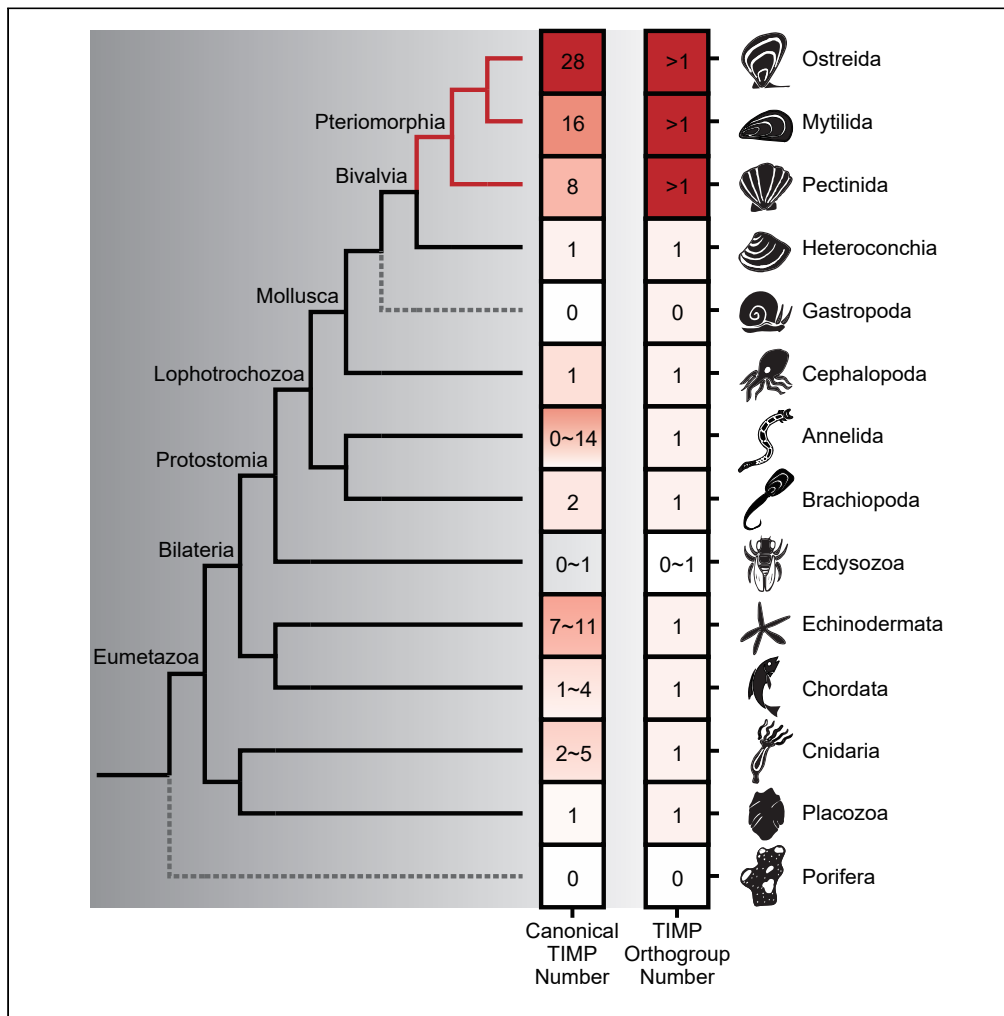


Article

Tracing the evolution of tissue inhibitor of metalloproteinases in Metazoa with the *Ptereria penguin* genome



Chao-Yi Ma, Yi Chen, Xin Zhan, Yun-Wei Dong

zhanxinuni@163.com (X.Z.)
dongyw@ouc.edu.cn (Y.-W.D.)

Highlights

The TIMP gene family originated in the common ancestor of eumetazoans

TIMP gene family sizes vary dramatically across metazoan clades

Pteriomorphia bivalves exhibit the largest and most diversified TIMP repertoires



Article

Tracing the evolution of tissue inhibitor of metalloproteinases in Metazoa with the *Ptereria penguin* genomeChao-Yi Ma,^{1,2} Yi Chen,^{3,4} Xin Zhan,^{3,4,*} and Yun-Wei Dong^{1,2,5,*}

SUMMARY

Tissue inhibitors of metalloproteinase (TIMPs) play a pivotal role in regulating extracellular matrix (ECM) dynamics and have been extensively studied in vertebrates. However, understanding their evolution across invertebrate phyla is limited. Utilizing the high-quality *Ptereria penguin* genome, we conducted phylogenomic orthology analyses across metazoans, revealing the emergence and distribution of the TIMP gene family. Our findings show that TIMP repertoires originated during eumetazoan radiation, experiencing independent duplication events in different clades, resulting in varied family sizes. Particularly, Pteriomorphia bivalves within Mollusca exhibited the most significant expansion and displayed the most diverse TIMP repertoires among metazoans. These expansions were attributed to multiple gene duplication events, potentially driven by the demands for functional diversification related to multiple adaptive traits, contributing to the adaptation of Pteriomorphia bivalves as stationary filter feeders. In this context, Pteriomorphia bivalves offer a promising model for studying invertebrate TIMP evolution.

INTRODUCTION

The evolutionary dynamics of key gene families have long been a topic of great interest in evolutionary biology.^{1,2} This includes studying the variations in gene family size and functional divergence during the evolutionary process. Analyzing the patterns of these events, while often difficult to achieve, can provide insights into the forces driving gene family evolution and their impact on the adaptation and evolutionary success of certain species.^{3,4}

The tissue inhibitor of metalloproteinase (TIMP) family is an ancient gene family that is widely distributed in the Metazoa kingdom.^{5,6} TIMPs primarily function as endogenous inhibitors of metalloproteinases, particularly matrix metalloproteinases, which are enzymes responsible for extracellular matrix (ECM) degradation.⁷ This inhibitory function is mainly mediated by the netrin-like (NTR) domain, which is conserved and corresponds to the N-terminal domain in TIMPs. The NTR domain can independently interact with the active site of MMPs, forming a tight 1:1 complex and inhibiting MMP activity.^{8,9} By regulating MMP activity, TIMPs play a crucial role in maintaining the balance between ECM synthesis and degradation, ensuring proper tissue remodeling and repair processes.^{10,11} As a key gene family instrumental in ECM modification, investigating the evolutionary trajectory of TIMPs across metazoans can provide valuable insights into the mechanisms that govern ECM dynamics in organismal development and survival.

Promising research has been conducted on the evolution and functional diversification of the TIMP gene family within specific clades in vertebrates.^{6,10} A comprehensive study revealed four TIMP members, commonly found in jawed vertebrates, originated from a common ancestor through three successive duplications during early vertebrate radiation.¹² Throughout their evolutionary divergence, different members of the vertebrate TIMP family have exhibited varying rates of evolutionary change and have undergone gain and loss of functions.¹² For instance, some members have lost inhibitory activity toward specific metalloproteinases while developing new interactions with specific targets such as integrins and receptors, while retaining their roles in regulating ECM structure.^{6,13} These findings highlight the complex evolutionary history of the TIMP gene family.

The TIMPs in invertebrates share a similar domain architecture but exhibit substantial sequence divergence compared to vertebrate TIMPs, indicating distinct evolutionary paths.^{6,10} A notable example is the study conducted on the evolutionary trajectory of TIMPs in echinoderms,¹⁴ which belong to the Deuterostomia clade along with vertebrates. Phylogenetic analysis has revealed that the TIMP gene family underwent diversification in the ancestral deuterostome. Unlike chordates, which retained only a single copy in the ancestral stage and

¹Key Laboratory of Mariculture, Ministry of Education, Fisheries College, Ocean University of China, Qingdao 266003, P.R. China

²Academy of the Future Ocean, Ocean University of China, Qingdao 266100, P.R. China

³State Key Laboratory of Marine Resources Utilization in South China Sea, Hainan University, Haikou 570228, P.R. China

⁴School of Marine Biology and Fisheries, Hainan University, Haikou 570228, P.R. China

⁵Lead contact

*Correspondence: zhanxinuni@163.com (X.Z.), dongyw@ouc.edu.cn (Y.-W.D.)

<https://doi.org/10.1016/j.isci.2023.108579>



subsequently diversified into four members in vertebrates, echinoderms retained numerous early-lineage TIMPs and continued to diversify after their divergence from chordates. Despite significant sequence differences compared to vertebrates, invertebrate TIMPs retain inhibitory activities against metalloproteinases.^{10,12} TIMPs from cnidarians exhibit a canonical TIMP fold structure, like mammalian TIMPs, with only minor structural changes, suggesting the evolutionary conservation of a common structure across metazoans.¹⁵ However, comprehensive studies focusing on the evolutionary events of invertebrate TIMPs, particularly in protostomes, are still limited despite the vast diversity within this group. A more comprehensive understanding of TIMP repertoires in invertebrates, which are currently underrepresented in evolutionary studies centered around the TIMP gene family, is essential for gaining insights into the evolution of the TIMP gene family across the entire animal kingdom and facilitating further functional studies within an appropriate evolutionary context.

Despite Mollusca being the second largest phylum of invertebrates, our understanding of the evolution of TIMP genes in this clade is still lacking. Present studies in molluscs have primarily centered around the bivalves from the Pteriomorphia clade, including *Tegillarca granosa* (Arcida),¹⁶ *Mytilus* spp. (Mytilida),^{17,18} *Magallana* spp. and *Pinctada* spp. (Ostreida),^{19–27} and *Chlamys farreri* (Pectinida).^{28–31} These studies have revealed that TIMPs found in Pteriomorphia bivalves play important roles in the immune system^{16,19} and formation of the bivalved shell and the hinge ligament,^{22,23,25} as well as the development of the bivalve byssus.^{18,25,26,31} These findings suggest that the TIMP gene family is crucial for the survival and adaptation of Pteriomorphia bivalves and may have undergone expansion in family size during their evolution. Moreover, Pteriomorphia bivalves can serve as valuable model organisms for studying the evolutionary patterns and diversification of the TIMP gene family in invertebrates. In addition to the Pteriomorphia bivalves, *Pteria penguin*, commonly known as the penguin's wing oyster, belongs to the family Pteridae, within the order Ostreida. Native to the western and central Indo-Pacific region, this species is widely cultivated for pearl production.³² While genomic resources and TIMP information for *P. penguin* are currently unavailable, it is reasonable to speculate that this species may possess a rich repertoire of TIMPs based on its phylogenetic relationship with other bivalves, as well as notable features such as a well-developed nacre layer and byssus in the adult stage. Conducting a comprehensive phylogenomic analysis with a specific focus on molluscs would significantly contribute to our understanding of the underappreciated TIMP resources in Pteriomorphia bivalves and provide insights into the broader distribution of TIMPs across metazoans.

In this study, we aim to address the gaps in our understanding of the evolution of the TIMP gene family in Mollusca and investigate the origin and diversification of canonical TIMPs across the phylogeny of Metazoa. While the size of the TIMP gene family in various molluscan clades remains inconclusive, previous studies have provided insights into the diverse functions of TIMPs in Pteriomorphia bivalves.^{19,22,28} Hence, we hypothesize that the TIMP gene family undergoes multiple duplication events and exhibits an extensive repertoire in Pteriomorphia bivalves. To address this hypothesis, we initiated the study by generating a high-quality genome assembly for *P. penguin* and conducted comprehensive phylogenomic and orthology inference analyses incorporating other metazoan genomes. Our findings offer a comprehensive overview of the distribution patterns of the TIMP gene family across the animal kingdom, suggesting the TIMP repertoires in metazoans originated from ancestral genes that emerged during the early radiation of eumetazoans. The most significant revelation from our research is that Ostreida bivalves belonging to the Pteriomorphia infraclass possess the most extensive and diverse repertoire of TIMPs in the animal kingdom. We propose that the abundance of TIMP repertoires in these species results from multiple independent duplication events driven by the various functional demands of different ECM protein-related tissues. These insights of this study could pave the way for further research into the role of TIMPs in a wide range of biological processes across metazoans by providing a crucial evolutionary context.

RESULTS

A high-quality *P. penguin* genome

Genomic DNA was extracted from a single individual of *P. penguin* (Figure 1A) and sequenced using Illumina short-read and PacBio long-read sequencing platforms (Table S1). The *de novo* assembly process, followed by heterozygosity removal and polishing, resulted in a final genome assembly of 838.71 Mb. The assembly had a contig N50 of 5.03 Mb and a scaffold N50 of 60.66 Mb, comparable to the estimated size by k-mer analysis (812.32 Mb) (Figure S1). Additionally, leveraging 46 Gb Hi-C sequencing data, approximately 99.83% of the obtained contigs were successfully anchored into 14 pseudo chromosome linkage groups (Table S2). To assess the completeness of the genome assembly, Benchmarking Universal Single-copy Ortholog (BUSCO) analysis was performed, revealing 95.8% metazoan BUSCOs present (Table S3). These results confirm the high quality and completeness of the assembled *P. penguin* genome.

For genome annotation, repeat elements constituted 43.89% (368.14 Mb) of the assembled genome (Table S4). A total of 36,733 protein-coding genes were identified within the *P. penguin* genome using *ab initio* methods and transcripts assembly constructed from Illumina and PacBio sequencing data (Table S5). Among these genes, 99.85% (36,677) were anchored to the 14 pseudo chromosome linkage groups. BUSCO analysis further demonstrated the completeness of the protein-coding gene set in the *P. penguin* assembly, with a metazoan BUSCO score of 97.8%. Additionally, a total of 16,884 noncoding RNA genes were annotated across the *P. penguin* genome (Table S6).

To investigate the phylogenetic relationship of *P. penguin* within the molluscs, a set of 273 single-copy genes was identified among seventeen molluscan species and one brachiopod *Lingula anatina* (for detailed information, see Table S7). These genes were used to construct a phylogenomic tree, which placed *P. penguin* together with the Pacific oyster *Magallana gigas* in the Ostreida order (Figure 1B). Both species, along with the mussel *Mytilus californianus* and the scallop *Mizuhopecten yessoensis*, clustered within the Pteriomorphia infraclass. This infraclass represents the sister clade to Heteroconchia, which includes the freshwater mussel *Dreissena polymorpha*, the soft-shell clam *Mya arenaria*, the hard clam *Mercenaria mercenaria*, and the freshwater pearl mussel *Margaritifera margaritifera*.

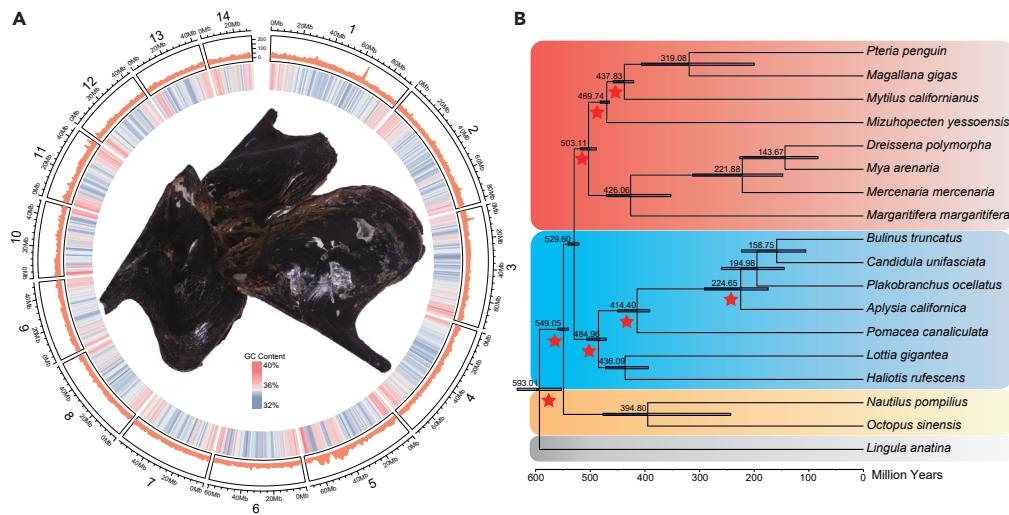


Figure 1. The penguin's wing oyster *Pteria penguin* and its phylogenetic positions

(A) The photograph of *P. penguin*. Circos plot showing the 14 pseudomolecules. The outer ring (red peaks) indicates gene density in each pseudo chromosome, and the inner ring shows GC content; the sliding window size is 1 MB, megabases.

(B) Time-calibrated phylogenomic tree showing the position of the *P. penguin* among seventeen molluscan species and one brachiopod species *Lingula anatina*. The color labeling scheme of taxa: Bivalvia: red; Gastropoda: blue; Cephalopoda: orange; other two Lophotrochozoan species: gray. The blue horizontal bars indicate the 95% confidence intervals of the divergence times. Red stars indicate the calibrated nodes based on fossil records and geographic events. The node labels are marked on the tree nodes.

Phylogenetic orthology analysis reveals the distribution of the TIMP gene family across Mollusca

Using Orthofinder software, a total of 38,849 orthogroups (OGs) were identified among eighteen species, representing 88.7% of all assigned genes. For each OG, a rooted gene tree was constructed by introducing the outgroup species *L. anatina*. Subsequently, 40,746 orthogroups at each hierarchical level (HOGs) were inferred from the OGs by analyzing the rooted gene trees. The gene copy numbers in each HOG were counted to estimate gene family expansion and contraction events across the molluscan phylogenomic tree. Among the 17,411 HOGs owned by the Ostreida clade (*P. penguin* or *M. gigas*), 82 gene families experienced significant expansion, while 13 gene families showed significant contraction compared to their parent clade (p value <0.05). Gene Ontology (GO) enrichment analysis of the aforementioned expanded HOGs in *P. penguin* genome revealed 276 highly significantly enriched GO categories (adjusted p value <0.01 ; listed in Table S8), of which fifteen were related to the TIMP gene family and TIMP activity (Figure 2A), indicating a massive expansion of the TIMP family in the species belong to the Ostreida order.

To quantify the copy numbers of TIMP genes in the eighteen species, their genomes were scanned, resulting in the identification of 144 putative TIMP sequences. These putative TIMPs had protein signature matches with "Tissue inhibitor of metalloproteinase" in the Pfam database, "METALLOPROTEASE INHIBITOR" in the PANTHER database, and "NTR domain profile" in the PROSITE database, all simultaneously. These 114 TIMP sequences were further classified into two groups based on whether they matched the conserved NTR domain in the SMART database. One group included 89 sequences, which were named "canonical TIMPs" because they had matches to SMART NTR domain. The other group consisted of 55 sequences, which were referred to as "pseudo TIMPs" since they did not have matches to the domain. Analysis of the canonical TIMP gene distribution revealed significant divergence in copy numbers among molluscan species. Cephalopods, gastropods, and some bivalves from the infraclass Heteroconchia showed either one copy or no canonical TIMP gene, except for *M. margaritifera*, which possessed three copies. In contrast, Pteriomorpha bivalves exhibited a dramatic expansion of canonical TIMP copies, ranging from eight in *M. yessoensis* to twenty-eight in *M. gigas* and *P. penguin* (Figure 2B); notably, species from the order Ostreida possessed the highest number of canonical TIMP genes in their genomes. The distribution pattern of pseudo TIMPs differed from canonical TIMPs and was mainly found in seven bivalves and one gastropod (*Pomacea canaliculata*). Notably, two Heteroconchia bivalves, *D. polymorpha* and *M. mercenaria*, possessed the highest number of pseudo TIMPs compared to other species.

The HOGs containing putative TIMP sequences were further analyzed. One HOG named HOG0001471, containing the majority of canonical TIMPs (53 out of 89 sequences) from ten out of eighteen surveyed species that encompassed canonical TIMP genes, was considered as the "core Mollusca-Brachiopoda TIMP HOG" that possessed "core Mollusca-Brachiopoda TIMPs." Additionally, eighteen HOGs containing other canonical TIMPs were considered "branch TIMP HOGs" with "branch TIMPs", while ten HOGs containing no canonical TIMPs but pseudo TIMPs were named "pseudo TIMP HOGs" (Table S9). Through inferring the gene duplication events in each HOG, the core Mollusca-Brachiopoda TIMP HOG was found to have experienced multiple duplication events across different species (Figure S2), representing the mainstream evolutionary trajectory of TIMPs from their potential common ancestor. Two major duplication events were identified in the evolution of core Mollusca-Brachiopoda TIMPs within the molluscan phylogeny (marked as "Non-Terminal" in Table S10). These events

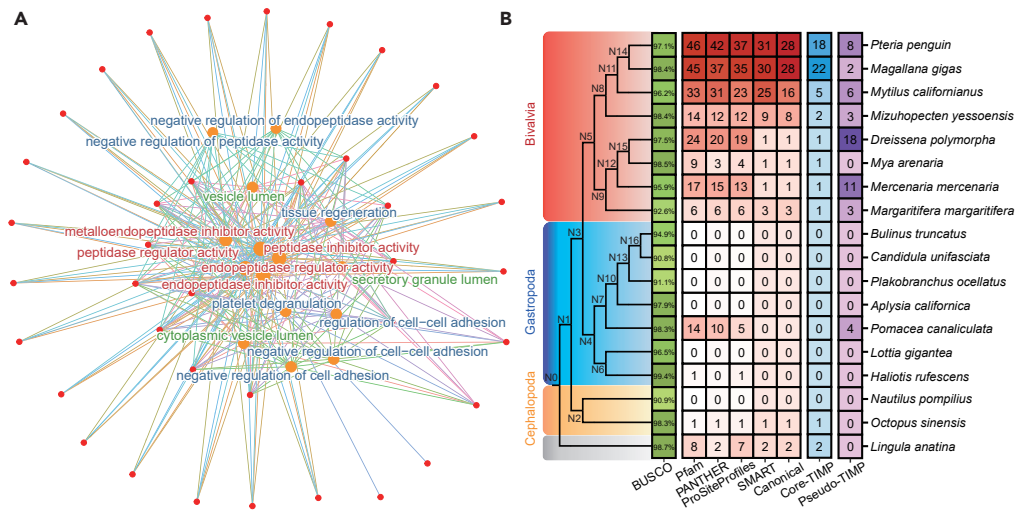


Figure 2. The TIMP family has dramatic expansion in *P. penguin* and other relative bivalves

(A) The GO pathways related to the TIMP gene family and TIMP activity, which were clustered from the expanded gene families within the order Ostreida. The molecular function (MF) GO pathways are shown in red, the biological process (BP) GO pathways are shown in blue, and the cellular component pathways are shown in green.

(B) The number of putative TIMP copies in each species and node labels were marked on the node of the phylogeny tree. With deeper color denoting the higher values, the BUSCO numbers indicating the completeness of protein sets are marked in green boxes, the number of sequences having matches in different databases as well as the finally identified canonical TIMPs are marked in red boxes, TIMP numbers in core Mollusca-Brachiopoda TIMP HOG (*HOG0001471*) are marked in blue boxes, and the numbers of pseudo TIMPs are marked in purple boxes. The cladogram tree is inherited from the results of this research.

were shared by more than one species and linked to the speciation events. The first event occurred during the separation of cephalopods from gastropods and bivalves. At the same time, the second event happened within the Pteriomorphia clade, specifically during the divergence of the Pectinida order from Mytilida and Ostreida. In contrast to the core Mollusca-Brachiopoda TIMP HOG, the branch TIMP HOG underwent species-specific or lineage-specific duplication events, giving rise to several TIMP subgroups specific to particular species. Similar to the previous two groups, the pseudo TIMP HOGs were derived from multiple duplications. Leveraging our chromosome-level *P. penguin* assembly as well as the publicly available *M. gigas* genome, the genomic location of putative TIMPs in these two species was retrieved, and the tandem duplications contributing to the TIMP expansion were detected and shown (Figure S3).

Phylogenetic trees were constructed using the aforementioned 144 putative TIMPs and 53 core Mollusca-Brachiopoda TIMPs, with two *L. anatina* sequences chosen as outgroups (Figures S4 and S5). Both trees divided molluscan sequences into two groups: a minor group consisting of single-copy canonical TIMPs from the four Heteroconchia bivalves and a major group comprising other sequences, primarily from Pteriomorphia bivalves. In the tree constructed from putative TIMPs, the pseudo or branch canonical TIMPs of *D. polymorpha* (eighteen pseudo TIMPs), *M. mercenaria* (eleven pseudo TIMPs), and *M. margaritifera* (three pseudo TIMPs and two branch TIMPs) were not clustered with the canonical TIMPs of these species but instead grouped within the major group. In the tree constructed from core Mollusca-Brachiopoda TIMPs, the TIMP sequence from the octopus *Octopus sinensis* clustered with the Pteriomorphia TIMPs rather than serving as an outgroup for both Heteroconchia and Pteriomorphia. Ten sequence motifs were discovered in the 144 putative TIMPs and used to compare the sequence composition of canonical and pseudo TIMPs from different phylogeny clades (Figure S4). Five motifs (*motif1* to *motif5*) were present across the majority of putative TIMPs, although absent in several sequences. Compared to other molluscan putative TIMPs, *motif7* was found only in some pseudo TIMPs of *D. polymorpha*, and *motif8* was barely present in canonical TIMPs from the two Ostreida bivalves, *P. penguin* and *M. gigas*, as well as their most closely related species, *M. californianus*. The motifs *motif9* and *motif10* were completely absent in the four sequences from *P. canaliculata*, and *motif10* was also absent in all sequences from Heteroconchia. Additionally, the arrangement order of motifs in core Mollusca-Brachiopoda TIMPs showed no reshuffling of order, but the position of *motif10* in molluscan TIMPs differed from that in *L. anatina* (Figure S5). The distribution and position of signal peptides and NTR domain in core Mollusca-Brachiopoda TIMPs were also demonstrated (Figure S5). Finally, the exon numbers in each putative TIMP gene were examined and found to range from four to six exons, with 132 out of 142 molluscan sequences encompassing five exons, similar to the exon number found in vertebrate TIMP genes.¹² This suggests that five exons are common in molluscs.

Phylogenetic orthology analysis reveals the distribution of the TIMP gene family across Metazoa

The Orthofinder software was employed to identify OGs among forty-one metazoans (Table S11). A total of 59,702 OGs were detected, encompassing 93.3% of all assigned genes. For each OG, a rooted gene tree was constructed using the sponge *Amphimedon queenslandica* as the outgroup. Subsequently, a total of 61,700 phylogenetic HOGs were inferred from the OGs by analyzing the rooted gene trees.

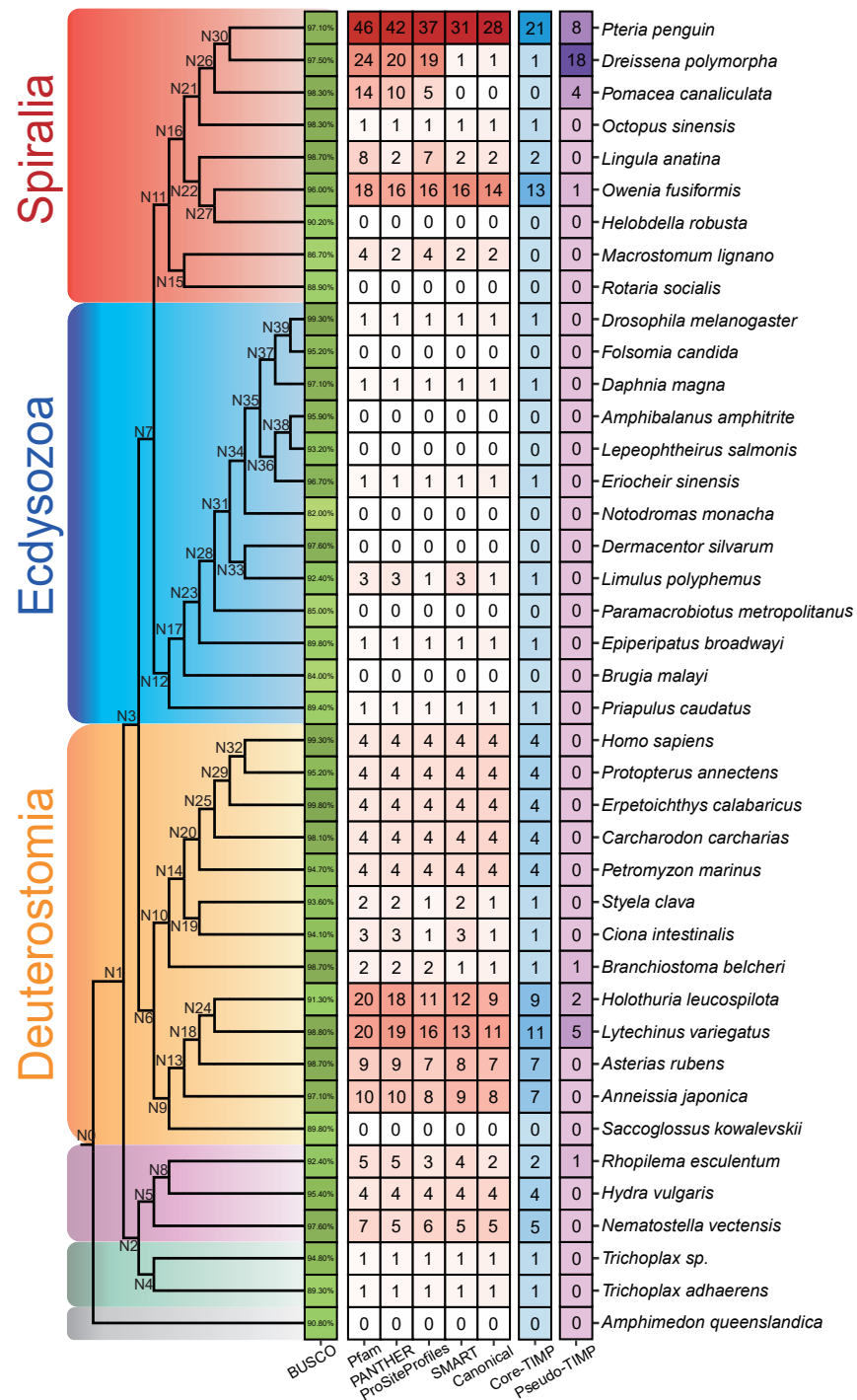


Figure 3. The distribution and phylogenetic relationship of the TIMP family from 41 metazoans

The distribution of TIMP copy numbers in each species and node labels were marked on the node of the phylogeny tree. The detailed interpretation can be referred to Figure 2B; with some differences are the core Eumetazoa TIMP HOG named HOG0001662, and the cladogram tree is constructed by referring to existing studies.³³⁻⁴⁰

Through the examination of a total of forty-one metazoan genomes, we identified 165 putative TIMPs, including 125 canonical TIMPs and 40 pseudo TIMPs (Table S12). The distribution of canonical TIMPs in the surveyed species was consistent with previous findings, showing the widespread presence of orthologs across Metazoa.¹⁰ However, the copy numbers varied among different metazoan clades (Figure 3). In the

poriferan (*Amphimedon queenslandica*), TIMP gene was absent, while in the clade containing placozoans (*Trichoplax adhaerens* and *T. sp.*) and cnidarians, the copy numbers ranged from one to multiple copies (Cnidaria, ranging from two to five). In Ambulacraria within Deuterostomia, Echinodermata species exhibited multiple TIMPs in their genomes, ranging from seven (*Anneissia japonica*, *Asterias rubens*) to eleven (*Lytechinus variegatus*). In Chordata of Deuterostomia clade, basal species such as cephalochordates (*Branchiostoma belcheri*) and tunicates (*Styela clava*, *Ciona intestinalis*) possessed one canonical TIMP, while the vertebrates included in this study, along with the jawless fish *Petromyzon marinus*, were found to have four canonical TIMPs. In Ecdysozoa of Protostome, only one or no copy of canonical TIMPs was detected. In Spiralia of Protostome, canonical TIMP copy numbers varied significantly. For instance, in Platyzoa, *Macrostomum lignano* had two copies, while *Rotaria socialis* had none. Among Trochozoa, which includes Annelida, Brachiopoda, and Mollusca, the TIMP copy numbers varied not only between sister clades but also within a single clade. Annelida exhibited a range from no copies in the leech *Helobdella robusta* to up to fourteen copies in the spindle worm *Owenia fusiformis*. And, as previously mentioned, the range of TIMP repertoire sizes in molluscs varied from zero copy to a remarkable twenty-eight copies, which were found in *P. penguin*, making it the largest repertoire of TIMP copies among all animals employed in this study. The distribution of pseudo TIMPs was also investigated and found to be present in species from Cnidaria (*Rhopilema esculentum*), Echinodermata (*Holothuria leucospilota*, *L. variegatus*), Chordate (*B. belcheri*), Annelida (*O. fusiformis*), and Mollusca (*D. polymorpha*, *P. canaliculata*, and *P. penguin*). Notably, the bivalves *D. polymorpha* had the highest number of pseudo TIMPs compared to other species.

Among the HOGs derived from the forty-one surveyed metazoans, the “core Eumetazoa TIMP HOG”, named *HOG0001662*, comprised the majority of canonical TIMPs (114 out of 125 sequences) from twenty-nine out of forty-one surveyed animals that encompassed canonical TIMP genes and all belong to Eumetazoa clade, the sister group of the Porifera (sponges). The number of core Eumetazoa TIMPs from each species followed the pattern observed in canonical TIMPs, with *P. penguin* possessing the highest number. Additionally, eight branch and pseudo TIMP HOGs were identified. Multiple gene duplication events were detected within each HOG containing putative TIMPs (Figure S6). Two major duplication events were identified in the evolution of core Eumetazoa TIMPs within the metazoan phylogeny; one occurred during the divergence of bilaterians from cnidarians and placozoans, while the other took place during the separation of protostomes from deuterostomes (Table S13).

Phylogenetic trees were constructed using the aforementioned 165 putative TIMPs and 114 core Eumetazoa TIMPs, with the two sequences of Placozoa species chosen as the outgroups (Figures S7 and S8). Among the ten sequence motifs discovered in the 165 putative TIMPs, *motif1* and *motif3* were present in the majority of putative TIMPs, with only a few sequences lacking them (Figure S7). The *motif2* and *motif4* were commonly found in canonical TIMPs, but one or both were absent in the pseudo TIMPs. Consistent with the previous findings, *motif10* was exclusively observed in some of the pseudo TIMPs from *D. polymorpha*. Four motifs (*motif1* to *motif4*) and seven motifs (*motif1* to *motif5*, *motif7*, *motif8*) were concurrently present in certain canonical TIMPs from eleven phyla representing the four major metazoan clades. These clades included Placozoa and Cnidaria (which were sister groups), Echinodermata, Chordata (both belonging to Deuterostomia), Arthropoda, Onychophora, Priapulida (belonging to Ecdysozoa of Protostomia), Annelida, Brachiopoda, Mollusca, and Platyhelminthes (belonging to Spiralia of Protostomia). No reshuffling of motif order was observed in the 114 core Eumetazoa TIMPs (Figure S8). The distribution and position of signal peptides and NTR domain in core Eumetazoa TIMPs were also demonstrated (Figure S8). Lastly, the exon number in each gene of the canonical and pseudo TIMPs was calculated. Out of the 165 sequences, 127 encompassed four to six exons, with 60 sequences having five exons. In Echinodermata, 25 out of 41 sequences had four exons, while in Placozoa and Cnidaria 10 out of 13 sequences encompassed two exons, indicating the common exon numbers in these two clades.

DISCUSSION

The primary objective of this study is to investigate the evolutionary dynamics and distribution of the TIMP gene family within the animal kingdom, with a particular focus on molluscs. To achieve this, we performed comprehensive analyses of the TIMP gene family across the metazoan phylogeny. By assembling a high-quality genome assembly of *P. penguin* and comparing it with other animals, we identified TIMP orthologs and examined their distribution in various animal clades. While the canonical TIMP gene is absent in the genome of the poriferan *A. queenslandica*, its orthologs are widely distributed among the eumetazoans, spanning from simple placozoans to bilaterians with complex tissue structures. It is worth mentioning that the phylogeny orthology analyses reveal a shared TIMP orthology group among the analyzed animals with canonical TIMPs, indicating a common ancestral origin in the evolution of the TIMP gene family. These findings also suggest the TIMP gene family likely emerged in the common ancestral lineage of eumetazoans after diverging from poriferans but before the split between the Placozoa + Cnidaria clade and Bilateria. Within the TIMP repertoires of various animal clades, a remarkable expansion was observed in the bivalves of the infraclass Pteriomorpha. From the common ancestor of eumetazoans to Pteriomorpha bivalves, four major duplication events of TIMP family during speciation events have occurred, along with multiple lineage- or species-specific duplication events after speciation events, potentially driven by adaptive demands. Accompanying the TIMP duplication events are the secondary loss events of the TIMP members in some lineage or species along the metazoan phylogeny. The duplication and loss events shape the distribution pattern of TIMPs in extant animals (Figure 4).

Our study provides comprehensive genomic analyses of TIMP gene copy number distribution across the Metazoa kingdom. Regarding the protostomes, ecdysozoans either completely lose TIMPs or retain a single-copy TIMP within their genomes. In contrast, the Spiralia displays two distinct expansions. One occurs in Annelida, where we identified fourteen canonical TIMPs in *O. fusiformis* from the Polychaeta class. Conversely, no TIMP gene was found in *H. robusta* from the Clitellata class. The variation in the size of the TIMP gene family within Annelida necessitates further elucidation through future studies incorporating additional genomic resources.

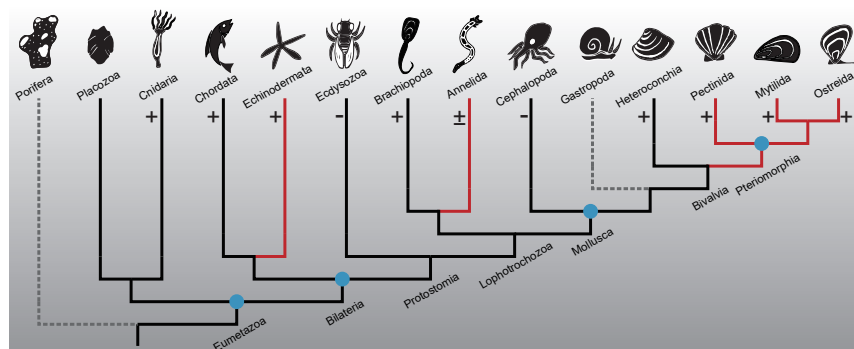


Figure 4. Schematic of features in the evolution of TIMP gene family in Metazoa kingdom

The phylogeny tree represents relationships among major metazoan clades employed in this study. Branch lengths are not indicative for evolutionary distances. Solid branches signify clades retaining TIMP gene copies, while dashed branches denote complete TIMP gene loss within a clade. Red branches highlight three remarkable TIMP expansions across metazoans: in echinoderms, the Polychaeta annelid *O. fusiformis*, and Pteriomorphia bivalves. Circles on tree nodes indicate four major duplication events of TIMP family along metazoan phylogeny, including the divergence of bilaterians from cnidarians and placozoans, protostomes from deuterostomes, bivalves and gastropods from cephalopods, and the separation of Mytilida and Ostreida order from Pectinida order. Plus signs (+) along the branches indicate lineage- or species-specific duplication events of TIMP genes within these clades, while minus signs (-) indicate complete TIMP gene loss in at least one species.

Another noteworthy expansion was observed in the Mollusca phylum, which exhibited considerable variation in the size of the TIMP gene family across its different clades. Some cephalopods have one TIMP gene, while gastropods have lost the canonical TIMP family during evolution. This observation highlights the need for further studies to investigate the evolutionary mechanisms behind the loss of TIMP members in gastropods. Compared with other molluscs, the Pteriomorphia bivalves exhibit a general expansion of the TIMP gene family. Within this clade, the copy number of the TIMP family has a linear relationship along the evolution phylogeny and reaches its peak in the Ostreidae bivalves with the most recent origins.⁴¹ The Ostreidae bivalves employed in this study, *P. penguin* and *M. gigas*, possess the highest number of canonical TIMP genes within their genomes, surpassing all other animals included in our research. The distribution of TIMP genes in the genomes of *P. penguin* and *M. gigas* provides compelling evidence supporting frequent gene tandem duplication as the primary mechanism underlying the notable expansion of the TIMP gene family. This mechanism is also implicated in the expansion of other gene families to unparalleled sizes in bivalves.^{42,43} Shifting the focus to the non-protostomes, our findings generally align with previous reports on TIMP copy numbers in cnidarians,¹⁵ echinoderms,¹⁴ and vertebrates,¹² with a few exceptions. In the jawless fish *Petromyzon marinus*, we observed four TIMP genes, which deviates from the previously reported number of three.¹² The four TIMPs in jawed vertebrates most likely originated from three successive duplications before the split of Chondrichthyes (cartilaginous fishes) and Euteleostomes (ray-fins and lobe-fins). In contrast, similar duplications are considered to occur only once or twice in Agnatha (jawless fishes). These suggest an additional Agnatha-specific duplication for the TIMP gene after their separation from the vertebrate stem. Another exception pertains to echinoderms, where the size of the TIMP repertoire identified in our study is relatively smaller than that previously reported.¹⁴ This discrepancy may arise from alternative splicing events, wherein a single TIMP gene can generate different transcript isoforms.^{44,45} Previous transcriptome-based methods may have consequently overestimated the distribution of TIMPs in echinoderms to some extent. In brief, our phylogenomic analysis offers a comprehensive investigation of the size variations in the TIMP gene family across the Metazoa kingdom. It highlights an unparalleled expansion of the TIMP gene family in the Pteriomorphia bivalves, emphasizing a unique evolutionary process.

High diversity is also a prominent feature of the TIMP repertoires of Pteriomorphia bivalves. In contrast to the echinoderms and *O. fusiformis*, which also exhibit an expansion of the TIMP gene family in the genome, all their TIMP orthologs are grouped within the Eumetazoa orthology group shared by other species containing TIMP genes, namely the core Eumetazoa HOG identified in our results. This indicates the limited diversity of the TIMP family in echinoderms and *O. fusiformis*, suggesting a relatively recent origin from a single expansion event.⁴² The canonical TIMPs from *P. penguin* display a high diversity and are clustered into five different orthology groups, including the core Eumetazoa orthology group and the other four restricted to *P. penguin*. Indeed, 25% of the canonical TIMPs (seven sequences) from *P. penguin* were not included in the core Eumetazoa orthology group but dispersed in other branch orthology groups, suggesting the evolution novelty of these sequences.⁴² In our molluscan orthology analysis, the diversity of the TIMP gene family in Pteriomorphia bivalves is particularly pronounced. Along with the core Mollusca-Brachiopoda orthology group, the canonical TIMPs of Pteriomorphia bivalves disperse into twelve different orthology groups. Seven of these groups are shared among different species, while the remaining five orthologs are specific to a single species. This suggests that the large-scale TIMP expansion in Pteriomorphia bivalve results from multiple, independent duplication events occurring at distinct stages along the evolutionary phylogeny. The initial expansion likely occurred in the latest common ancestor of Pteriomorphia bivalves, contributing to their larger repertoire compared with their sister clade, the Heteroconchia. Subsequent independent expansions occurred during the radiation of Pteriomorphia bivalves, with some leading to recent species-specific evolutionary innovations, resulting in observed phylogenetic TIMP orthology branches exclusive to Pteriomorphia bivalves.

The remarkable expansion and diversity of the TIMP family in Pteriomorpha bivalves can be attributed to their adaptation to habitats. This adaptation-driven expansion of the TIMP family has been indicated in echinoderms, as these animals have high demands for tissue remodeling as adaptive mechanisms.¹⁴ For Pteriomorpha bivalves, these marine bivalves primarily inhabit intertidal zones or shallow waters as stationary filter feeders, which expose them to constant threats, including aquatic pathogens and pervasive predators, as well as the fluctuating physicochemical conditions of the surroundings.⁴⁶ Faced with these multiple stresses, Pteriomorpha bivalves require robust adaptive mechanisms, which can be facilitated by the TIMP family and may further drive the subsequent duplication and neofunctionalization events of this family in Pteriomorpha bivalves. Firstly, these bivalves have developed an effective innate immunity system involving multiple adaptive evolutionary processes.⁴⁷ A crucial aspect of this adaptation is the expansion of specific immune-related gene families, such as the Toll-like receptor (TLR) and inhibitors of apoptosis (IAPs) families.^{42,43} These genetic changes are closely associated with the immune needs of bivalves in response to the stress from their living environments. In Pteriomorpha bivalves, certain TIMP members specifically expressed in hemocytes have been confirmed to play a pivotal role in immune mechanisms, including wound healing and the defense against microbial pathogens.^{16,19,21} The demand for immune processes via TIMP members may drive the adaptive duplication of a subset of TIMP repertoires related to the immune system. Secondly, the survival of Pteriomorpha bivalves relies on their bivalved shells, mainly serving as a defensive barrier against external predators. In the case of true oysters like those in the Ostreidae family, the shell firmly cements these oysters to substrates. The components within the shell of Pteriomorpha bivalves contain various matrix proteins contributing to shell formation and growth.^{48,49} Among these shell matrix proteins of Pteriomorpha bivalves, the TIMP family has consistently been identified as one component, further substantiating its role in facilitating shell formation and growth. Some TIMP members in the Iwagaki oyster *M. nippona* are specifically expressed in the mantle,²⁷ the tissue responsible for shell development. Moreover, studies conducted through RNA interference experiments in the pearl oyster *Pinctada fucata* have provided evidence of TIMP proteins participating in the formation of shell and ligament structure.^{23,25} Additionally, some TIMP proteins or TIMP-related domains have been identified as integral components of the shell structure,^{17,22,24,27,49} possibly serving to protect the shell protein components from degradation by external proteinases. These studies indicate the involvement of the TIMP family in shell development can be a key driver for its general expansion and diversity in Pteriomorpha bivalves. This expansion aligns with the observed duplication events in other genes encoding shell matrix proteins, regarded as fundamental mechanisms in the evolution of these families.^{27,50} Thirdly, byssus, an adhesion structure made of protein filaments and secreted by the foot, serves as an adaptive trait in Pteriomorpha bivalves. Byssus is crucial for the larvae's settlement and metamorphosis. In adults, many Pteriomorpha bivalves, such as mussels in the Mytiloidea family and pearl oysters in the Pteriidae family, retain byssus for attachment to hard substrates.⁵¹ Specific TIMP members in Pteriomorpha bivalves not only play roles in byssus fabrication, as indicated by multiple studies based on foot transcriptomes,^{18,28,29,31} but also have been shown to be integrated into the protein components of byssus.^{18,26,28} Some of these structural TIMPs even have evolved beyond their inhibitory function, transforming into cross-linkers that may facilitate the self-assembly of byssus.^{30,31} These byssus-related TIMPs undergo duplication and subsequent neofunctionalization events as adaptive mechanisms in Pteriomorpha bivalves to meet their dependence on byssus. Lastly, some other factors contributing to the expansion of the TIMP family in Pteriomorpha bivalves may be attributed to neutral or non-adaptive processes, such as neutral gene duplication events,^{52,53} and contribute to the initial duplication event in the ancestor of Pteriomorpha bivalves. Nevertheless, the adaptive force is still recognized as the main driving force behind the multiple duplication events of this family, as supported by the close relationship between the functional diversity exhibited by multiple TIMP members and the adaptive traits in Pteriomorpha bivalves. In summary, the expansion of the TIMP family can be driven by the need for multiple adaptive traits in Pteriomorpha bivalves and plays a crucial role in the survival and adaptation of these bivalves in living environments. Further investigation is needed to explore more precise reasons behind the expansion of this family.

The emergence of similar adaptive traits through distinct genetic mechanisms is a common phenomenon in nature.^{54–57} This process contributes to a better understanding of the lineage-specific expansion of the TIMP family in Pteriomorpha bivalves. Beyond Pteriomorpha bivalves, this study also investigated the distribution of TIMP families in other animals such as stationary filter-feeding animals in aquatic environments, but none of them exhibited the remarkable TIMP expansion seen in Pteriomorpha bivalves. Among these animals, Heteroconchia bivalves, which are the sister group to Pteriomorpha bivalves and share similar adaptive traits,⁵⁸ including effective innate immunity, bivalved shells, and adhesive byssus, have much smaller TIMP repertoires in comparison to Pteriomorpha bivalves. Considering the more than 500 million years of divergence between these two groups, despite the deep conservation in fundamental biological processes, specific divergences remain in the genetic mechanisms underlying some similar traits. In the case of gene families regulating innate immunity in bivalves, although some families have ancestrally conserved duplications across bivalve lineages, the expansion of specific families is identified as lineage or species specific.⁴⁶ And regarding the shell and byssus of bivalves, despite the conservation in the morphology, the molecular components of these structures have diverged across bivalve lineages, indicating independent evolution in multiple gene families related to shell formation and byssus fabrication.^{51,59,60} Therefore, despite their limited TIMP family, it is speculated that Heteroconchia bivalves may have evolved specific alternative genes to achieve functions related to adaptive traits similar to those carried out by Pteriomorpha TIMPs. For other stationary filter-feeding animals beyond bivalves, the reasons for not experiencing TIMP expansion may follow a similar pattern. One example comes from the barnacles, which are sessile arthropods and have undergone convergent evolution with bivalves. A comparative genomics analysis in barnacles and Pteriomorpha bivalves revealed genome-wide substantial convergent molecular evolution between these two groups. However, there are still multiple divergences in the genetic mechanisms related to convergent adaptive traits, including shell formation and substrate settlement.⁵⁷ The results of this study indicate that barnacles have much less TIMP members than Pteriomorpha bivalves. These suggest that, in the convergent evolution of barnacles and Pteriomorpha bivalves, the TIMP families in these two groups have

divergent evolutionary processes. In brief, diverse functions of large TIMP families contribute to some adaptive traits in Pteriomorpha bivalves, while other animals may achieve similar functions through distinct genetic mechanisms. This divergence is hypothesized to be caused by the initial duplication of the TIMP family in Pteriomorpha bivalves, specific to the latest common ancestor of this bivalve clade. The resulting several TIMP gene copies in the ancestral Pteriomorpha bivalves may have undergone neofunctionalization^{61,62} and further contribute to the multiple adaptive duplication events of TIMP families in their descendants. More studies are required to compare further the evolution dynamics of the TIMP family in Pteriomorpha bivalves and other animals.

Moreover, our findings reveal the widespread presence of pseudo TIMPs in metazoans, which lack conserved NTR domains compared to canonical TIMPs. These pseudo TIMPs exhibit the most extensive repertoire and diversity in bivalves, reaching unprecedented sizes in two Heteroconchia bivalves (18 copies in *D. polymorpha*, 11 copies in *M. mercenaria*). Among the pseudo TIMPs in molluscs, the majority of pseudo TIMPs in Pteriomorpha bivalves were found grouped within the same orthology groups as canonical TIMPs, suggesting a coeval origin with the multiple independent duplications that contribute to the extensive TIMP repertoire in these bivalves, as discussed earlier. In these cases, the pseudo TIMPs may have emerged through gene duplication events of TIMP, resulting in excess copies, which subsequently underwent N-terminal structural changes to balance the overall functionality of TIMPs within the organism. Conversely, most pseudo TIMPs in Heteroconchia bivalves and the four pseudo TIMPs found in the gastropod *P. canaliculata* belong to the orthology groups that do not contain any canonical TIMP, indicating the existence of specific and independent duplications. The expansion of these pseudo TIMPs may have played specific functional roles within the organism, conferring advantages for the survival of certain species and leading to positive natural selection. One hypothesis is that these multiple pseudo TIMPs partially compensate for the functional deficiencies and losses of canonical TIMPs in certain species, such as Heteroconchia bivalves and the gastropod *P. canaliculata*. However, the precise roles of these pseudo TIMPs in these molluscs require further investigation in future studies to be fully understood.

In conclusion, our study utilized a high-quality genome of *P. penguin* to investigate the evolutionary dynamics of the TIMP gene family within the Metazoa kingdom. Through comparative genomic analyses, we investigated the distribution of the TIMP family across the animal phylogeny and suggested this family originated in the common ancestor of eumetazoans. Our investigation revealed variations in TIMP gene copy numbers across animal lineages, with a massive expansion occurring in the Pteriomorpha infraclass and peaking in the Ostreoida order. These bivalves also exhibit highly diverse TIMP repertoires, which contribute to their multiple adaptive traits, including an effective immune system, and the formation of bivalved shell and byssus. Moreover, our study unveils the presence of abundant pseudo TIMPs, primarily concentrated in Heteroconchia bivalves, highlighting potential future research directions to uncover the exact functions of these genes. Overall, our study provides valuable insights into the evolution of the TIMP gene family in molluscs, bridging the gap in our understanding of the TIMP gene family evolution across the entire Metazoa kingdom. Further investigations into the diverse functions and evolutionary trajectory within the TIMP gene family in Pteriomorpha bivalves are warranted to deepen our understanding.

Limitations of the study

This study investigates the distribution of TIMP numbers across various metazoan clades using comparative genomic analyses. The reliability of our findings relies on the quality of the investigated genomes, their annotations, and the quantity of genomes studied in each metazoan clade. We gathered metazoan genomes from sources like NCBI (<https://www.ncbi.nlm.nih.gov/>) and relevant literature and then assessed genome annotation completeness using the BUSCO score as an index. Despite these efforts, a scarcity of publicly available genomes with highly comprehensive annotations, especially in some clades like the Heteroconchia infraclass within bivalves, potentially limited our research scope. A more thorough exploration of TIMP family distribution across animals awaits the release of future high-quality genomes. Additionally, previous research has established the involvement of TIMP family members in various organismal processes within Pteriomorpha bivalves, like immune responses and shell formation. This study refrains from conducting extra systematic experiments to further validate the functions of multiple and diverse TIMP orthologs in *P. penguin* or other Pteriomorpha bivalves. Finally, the sex of *P. penguin* individuals utilized for genome and transcriptome sequencing was undetermined, as species within the Pteriidae family are usually protandrous hermaphrodites. The gender does not have an impact on the quality of our genome assembly.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
 - The penguin's wing oyster
- METHOD DETAILS

- Samples preparation and sequencing
- Genome survey, assembly, scaffolding and assessment
- Genome annotation
- Phylogenomic analyses in Mollusca
- Orthologous gene family size change analysis in Mollusca
- Phylogenetic orthology inference in Metazoa
- TIMP gene family analyses with Mollusca and Metazoa
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.108579>.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China grant 31860727, grant 32160860, and grant 42025604 and the Fundamental Research Funds for the Central Universities.

AUTHOR CONTRIBUTIONS

C.-Y.M.: investigation, software, visualization, writing-original draft; Y.C.: animal work and experiments; Y.-W.D. and X.Z.: conceptualization, funding acquisition, project administration, manuscript revision, and editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 11, 2023

Revised: September 13, 2023

Accepted: November 23, 2023

Published: November 25, 2023

REFERENCES

1. Demuth, J.P., and Hahn, M.W. (2009). The life and death of gene families. *Bioessays* 31, 29–39.
2. Fernández, R., and Gabaldón, T. (2020). Gene gain and loss across the metazoan tree of life. *Nat. Ecol. Evol.* 4, 524–533.
3. Richter, D.J., Fozouni, P., Eisen, M.B., and King, N. (2018). Gene family innovation, conservation and loss on the animal stem lineage. *Elife* 7, e34226.
4. Guijarro-Clarke, C., Holland, P.W.H., and Paps, J. (2020). Widespread patterns of gene loss in the evolution of the animal kingdom. *Nat. Ecol. Evol.* 4, 519–523.
5. Brew, K., Dinakarandian, D., and Nagase, H. (2000). Tissue inhibitors of metalloproteinases: evolution, structure and function. *Biochim. Biophys. Acta* 1477, 267–283.
6. Brew, K., and Nagase, H. (2010). The tissue inhibitors of metalloproteinases (TIMPs): an ancient family with structural and functional diversity. *Biochim. Biophys. Acta* 1803, 55–71.
7. Brinckerhoff, C.E., and Matrisian, L.M. (2002). Matrix metalloproteinases: a tail of a frog that became a prince. *Nat. Rev. Mol. Cell Biol.* 3, 207–214.
8. Gomis-Rüth, F.X., Maskos, K., Betz, M., Bergner, A., Huber, R., Suzuki, K., Yoshida, N., Nagase, H., Brew, K., Bourenkov, G.P., et al. (1997). Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. *Nature* 389, 77–81.
9. Fernandez-Catalan, C., Bode, W., Huber, R., Turk, D., Calvete, J.J., Lichte, A., Tschesche, H., and Maskos, K. (1998). Crystal structure of the complex formed by the membrane type 1-matrix metalloproteinase with the tissue inhibitor of metalloproteinases-2, the soluble progelatinase A receptor. *EMBO J.* 17, 5238–5248.
10. Murphy, G. (2011). Tissue inhibitors of metalloproteinases. *Genome Biol.* 12, 233.
11. Arpino, V., Brock, M., and Gill, S.E. (2015). The role of TIMPs in regulation of extracellular matrix proteolysis. *Matrix Biol.* 44–46, 247–254.
12. Brew, K. (2019). Reflections on the evolution of the vertebrate tissue inhibitors of metalloproteinases. *Faseb. J.* 33, 71–87.
13. Stetler-Stevenson, W.G. (2008). Tissue inhibitors of metalloproteinases in cell signaling: metalloproteinase-independent biological activities. *Sci. Signal.* 1, re6.
14. Clouse, R.M., Linchangco, G.V., Jr., Kerr, A.M., Reid, R.W., and Janies, D.A. (2015). Phylotranscriptomic analysis uncovers a wealth of tissue inhibitor of metalloproteinases variants in echinoderms. *R. Soc. Open Sci.* 2, 150377.
15. Nicosia, A., Maggio, T., Costa, S., Salamone, M., Tagliavia, M., Mazzola, S., Gianguzza, F., and Cuttitta, A. (2016). Maintenance of a protein structure in the dynamic evolution of TIMPs over 600 million years. *Genome Biol. Evol.* 8, 1056–1071.
16. Wang, Q., Bao, Y., Huo, L., Gu, H., and Lin, Z. (2012). A novel tissue inhibitor of metalloproteinase in blood clam *Tegillarca granosa*: molecular cloning, tissue distribution and expression analysis. *Fish Shellfish Immunol.* 33, 645–651.
17. Liao, Z., Bao, L.F., Fan, M.H., Gao, P., Wang, X.X., Qin, C.L., and Li, X.M. (2015). In-depth proteomic analysis of nacre, prism, and myostracum of *Mytilus* shell. *J. Proteomics* 122, 26–40.
18. Qin, C.L., Pan, Q.D., Qi, Q., Fan, M.H., Sun, J.J., Li, N.N., and Liao, Z. (2016). In-depth proteomic analysis of the byssus from marine mussel *Mytilus coruscus*. *J. Proteomics* 144, 87–98.
19. Montagnani, C., Le Roux, F., Berthe, F., and Escoubas, J.M. (2001). Cg-TIMP, an inducible tissue inhibitor of metalloproteinase from the Pacific oyster *Crassostrea gigas* with a potential role in wound healing and defense mechanisms. *FEBS Lett.* 500, 64–70.
20. Montagnani, C., Tirape, A., Boulo, V., and Escoubas, J.M. (2005). The two Cg-timp mRNAs expressed in oyster hemocytes are generated by two gene families and differentially expressed during ontogenesis. *Dev. Comp. Immunol.* 29, 831–839.
21. Montagnani, C., Avarre, J.C., de Lorgeril, J., Quiquand, M., Boulo, V., and Escoubas, J.M. (2007). First evidence of the activation of Cg-timp, an immune response component of Pacific oysters, through a damage-associated molecular pattern pathway. *Dev. Comp. Immunol.* 31, 1–11.
22. Marie, B., Joubert, C., Tayalé, A., Zanella-Cléon, I., Belliard, C., Piquemal, D., Cochenec-Laureau, N., Marin, F., Gueguen, Y., and Montagnani, C. (2012).

- Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. *Proc. Natl. Acad. Sci. USA* 109, 20986–20991.
23. Yan, F., Jiao, Y., Deng, Y., Du, X., Huang, R., Wang, Q., and Chen, W. (2014). Tissue inhibitor of metalloproteinase gene from pearl oyster *Pinctada martensii* participates in nacre formation. *Biochem. Biophys. Res. Commun.* 450, 300–305.
 24. Liu, C., Li, S., Kong, J., Liu, Y., Wang, T., Xie, L., and Zhang, R. (2015). In-depth proteomic analysis of shell matrix proteins of *Pinctada fucata*. *Sci. Rep.* 5, 17269.
 25. Kubota, K., Tsuchihashi, Y., Kogure, T., Maeyama, K., Hattori, F., Kinoshita, S., Sakuda, S., Nagasawa, H., Yoshimura, E., and Suzuki, M. (2017). Structural and functional analyses of a TIMP and MMP in the ligament of *Pinctada fucata*. *J. Struct. Biol.* 199, 216–224.
 26. Liu, C., and Zhang, R. (2021). Identification of novel adhesive proteins in pearl oyster by proteomic and bioinformatic analysis. *Biofouling* 37, 299–308.
 27. Bai, Y., Liu, S., Hu, Y., Yu, H., Kong, L., Xu, C., and Li, Q. (2023). Multi-omic insights into the formation and evolution of a novel shell microstructure in oysters. *BMC Biol.* 21, 204.
 28. Miao, Y., Zhang, L., Sun, Y., Jiao, W., Li, Y., Sun, J., Wang, Y., Wang, S., Bao, Z., and Liu, W. (2015). Integration of transcriptomic and proteomic approaches provides a core set of genes for understanding of scallop attachment. *Mar. Biotechnol.* 17, 523–532.
 29. Li, Y., Sun, X., Hu, X., Xun, X., Zhang, J., Guo, X., Jiao, W., Zhang, L., Liu, W., Wang, J., et al. (2017). Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. *Nat. Commun.* 8, 1721.
 30. Zhang, X., Dai, X., Wang, L., Miao, Y., Xu, P., Liang, P., Dong, B., Bao, Z., Wang, S., Lyu, Q., and Liu, W. (2018). Characterization of an atypical metalloproteinase inhibitors like protein (Sbp8-1) From scallop byssus. *Front. Physiol.* 9, 597.
 31. Dai, X., Zhu, X., Bao, L., Chen, X., Miao, Y., Li, Y., Li, Y., Lv, J., Zhang, L., Huang, X., et al. (2022). Decoding the byssus fabrication by spatiotemporal secretome analysis of scallop foot. *Comput. Struct. Biotechnol. J.* 20, 2713–2722.
 32. Southgate, P.C., Strack, E., Hart, A., Wada, K.T., Monteforte, M., Cariño, M., Langy, S., Lo, C., Acosta-Salmón, H., and Wang, A. (2008). Chapter 9 - Exploitation and Culture of Major Commercial Species. In *The Pearl Oyster*, P.C. Southgate and J.S. Lucas, eds. (Elsevier), pp. 303–355.
 33. Telford, M.J., Lowe, C.J., Cameron, C.B., Ortega-Martinez, O., Aronowicz, J., Oliveri, P., and Copley, R.R. (2014). Phylogenomic analysis of echinoderm class relationships supports Asterozoa. *Proc. Biol. Sci.* 281, 20140479.
 34. Delsuc, F., Philippe, H., Tsagkogeorga, G., Simon, P., Tilak, M.K., Turon, X., López-Legentil, S., Piette, J., Lemaire, P., and Douzery, E.J.P. (2018). A phylogenomic framework and timescale for comparative studies of tunicates. *BMC Biol.* 16, 39.
 35. Laumer, C.E., Fernández, R., Lemer, S., Combsch, D., Kocot, K.M., Riesgo, A., Andrade, S.C.S., Sterrer, W., Sørensen, M.V., and Giribet, G. (2019). Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proc. Biol. Sci.* 286, 20190831.
 36. Marletaz, F., Peijnenburg, K., Goto, T., Satoh, N., and Rokhsar, D.S. (2019). A new spiralian phylogeny places the enigmatic arrow worms among gnathiferans. *Curr. Biol.* 29, 312–318.e3.
 37. Kapli, P., Natsidis, P., Leite, D.J., Fursman, M., Jeffrie, N., Rahman, I.A., Philippe, H., Copley, R.R., and Telford, M.J. (2021). Lack of support for Deuterostomia prompts reinterpretation of the first Bilateria. *Sci. Adv.* 7, eabe2741.
 38. Drábková, M., Kocot, K.M., Halanych, K.M., Oakley, T.H., Moroz, L.L., Cannon, J.T., Kuris, A., Garcia-Vedrenne, A.E., Pankey, M.S., Ellis, E.A., et al. (2022). Different phylogenomic methods support monophyly of enigmatic ‘Mesozoa’ (Dicyemida + Orthonectida, Lophotrochozoa). *Proc. Biol. Sci.* 289, 20220683.
 39. Howard, R.J., Giacomelli, M., Lozano-Fernandez, J., Edgecombe, G.D., Fleming, J.F., Kristensen, R.M., Ma, X., Olesen, J., Sørensen, M.V., Thomsen, P.F., et al. (2022). The Ediacaran origin of Ecdysozoa: integrating fossil and phylogenomic data. *J. Geol. Soc. London.* 179.
 40. Schultz, D.T., Haddock, S.H.D., Bredeson, J.V., Green, R.E., Simakov, O., and Rokhsar, D.S. (2023). Ancient gene linkages support ctenophores as sister to other animals. *Nature* 618, 110–117.
 41. Lemer, S., González, V.L., Bieler, R., and Giribet, G. (2016). Cementing mussels to oysters in the pteriomorphian tree: a phylogenomic approach. *Proc. Biol. Sci.* 283, 20160857.
 42. Saco, A., Novoa, B., Greco, S., Gerdol, M., and Figueras, A. (2023). Bivalves present the largest and most diversified repertoire of toll-like receptors in the animal kingdom, suggesting broad-spectrum pathogen recognition in marine waters. *Mol. Biol. Evol.* 40, msad133.
 43. Song, H., Guo, X., Sun, L., Wang, Q., Han, F., Wang, H., Wray, G.A., Davidson, P., Wang, Q., Hu, Z., et al. (2021). The hard clam genome reveals massive expansion and diversification of inhibitors of apoptosis in *Bivalvia*. *BMC Biol.* 19, 15.
 44. Obro, N.F., Lademann, U., Birkenkamp-Demtroder, K., Holten-Andersen, L., Brünner, N., and Offenberg, H. (2008). A TIMP-1 splice variant transcript: possible role in regulation of TIMP-1 expression. *Cancer Lett.* 262, 64–70.
 45. Usher, P.A., Sieuwerts, A.M., Bartels, A., Lademann, U., Nielsen, H.J., Holten-Andersen, L., Foekens, J.A., Brünner, N., and Offenberg, H. (2007). Identification of alternatively spliced TIMP-1 mRNA in cancer cell lines and colon cancer tissue. *Mol. Oncol.* 1, 205–215.
 46. Regan, T., Stevens, L., Peñaloza, C., Houston, R.D., Robledo, D., and Bean, T.P. (2021). Ancestral physical stress and later immune gene family expansions shaped bivalve mollusc evolution. *Genome Biol. Evol.* 13, evab177.
 47. Canesi, L., and Pruzzo, C. (2016). Specificity of innate immunity in bivalves. In *Lessons in Immunity* (Elsevier), pp. 79–91.
 48. Addadi, L., Joester, D., Nudelman, F., and Weiner, S. (2006). Mollusk Shell Formation: A Source of New Concepts for Understanding Biomineralization Processes. *Chem. Eur J.* 12, 980–987.
 49. Yarra, T., Blaxter, M., and Clark, M.S. (2021). A bivalve biomineralization toolbox. *Mol. Biol. Evol.* 38, 4043–4055.
 50. Shimizu, K., Takeuchi, T., Negishi, L., Kurumizaka, H., Kuriyama, I., Endo, K., and Suzuki, M. (2022). Evolution of Epidermal Growth Factor (EGF)-like and Zona Pellucida Domains Containing Shell Matrix Proteins in Mollusks. *Mol. Biol. Evol.* 39, msac148.
 51. McCartney, M.A. (2021). Structure, function and parallel evolution of the bivalve byssus, with insights from proteomes and the zebra mussel genome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376, 20200155.
 52. Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. USA* 104, 8597–8604.
 53. Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108.
 54. Qi, X., Liu, Y., Vigueira, C.C., Young, N.D., Caicedo, A.L., Jia, Y., Gealy, D.R., and Olsen, K.M. (2015). More than one way to evolve a weed: parallel evolution of US weedy rice through independent genetic mechanisms. *Mol. Ecol.* 24, 3329–3344.
 55. Thurber, C.S., Jia, M.H., Jia, Y., and Caicedo, A.L. (2013). Similar traits, different genes? Examining convergent evolution in related weedy rice populations. *Mol. Ecol.* 22, 685–698.
 56. Qiu, J., Jia, L., Wu, D., Weng, X., Chen, L., Sun, J., Chen, M., Mao, L., Jiang, B., Ye, C., et al. (2020). Diverse genetic mechanisms underlie worldwide convergent rice fertilization. *Genome Biol.* 21, 70.
 57. Yuan, J., Zhang, X., Li, S., Liu, C., Yu, Y., Zhang, X., Xiang, J., and Li, F. (2022). Convergent evolution of barnacles and molluscs sheds lights in origin and diversification of calcareous shell and sessile lifestyle. *Proc. Biol. Sci.* 289, 20221535.
 58. Lemer, S., Bieler, R., and Giribet, G. (2019). Resolving the relationships of clams and cockles: dense transcriptome sampling drastically improves the bivalve tree of life. *Proc. Biol. Sci.* 286, 20182684.
 59. Rüdiger, B., Mikkelsen, P.M., Collins, T.M., Glover, E.A., González, V.L., Graf, D.L., Harper, E.M., Healy, J., Kawachi, G.Y., Sharma, P.P., et al. (2014). Investigating the Bivalve Tree of Life – an exemplar-based approach combining molecular and novel morphological characters. *Invertebr. Systemat.* 28, 32–115.
 60. Arivalagan, J., Yarra, T., Marie, B., Sleight, V.A., Duvernois-Berthet, E., Clark, M.S., Marie, A., and Berland, S. (2017). Insights from the Shell Proteome: Biomineralization to Adaptation. *Mol. Biol. Evol.* 34, 66–77.
 61. Assis, R., and Bachtrog, D. (2013). Neofunctionalization of young duplicate genes in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 110, 17409–17414.
 62. Rastogi, S., and Liberles, D.A. (2005). Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol. Biol.* 5, 28.
 63. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770.
 64. Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of

- polyploid genomes. *Nat. Commun.* 11, 1432.
65. Hu, J., Wang, Z., Sun, Z., Hu, B., Ayoola, A.O., Liang, F., Li, J., José, R.S., Cooper, D.N., Ye, K., et al. (2023). An efficient error correction and accurate assembly tool for noisy long reads. Preprint at bioRxiv. <https://doi.org/10.1101/2023.03.09.531669>.
 66. Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36, 2896–2898.
 67. Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* 36, 2253–2255.
 68. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890.
 69. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3, 95–98.
 70. Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95.
 71. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at bioRxiv. <https://doi.org/10.48550/ARXIV.1303.3997>.
 72. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10.
 73. Manni, M., Berkeley, M.R., Seppely, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38, 4647–4654.
 74. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* 117, 9451–9457.
 75. Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics Chapter 4*, 4.10.11–14.10.14.
 76. Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935.
 77. Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* 3, lqaa108.
 78. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652.
 79. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915.
 80. Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31, 5654–5666.
 81. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, R7.
 82. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10, 421.
 83. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240.
 84. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829.
 85. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238.
 86. Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24, 1586–1591.
 87. Mendes, F.K., Vanderpool, D., Fulton, B., and Hahn, M.W. (2021). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518.
 88. R Core Team (2022). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
 89. Carlson, M., and Pagès, H. (2023). AnnotationForge: Tools for Building SQLite-Based Annotation Data Packages. R package version 1.44.0. <https://bioconductor.org/packages/AnnotationForge>.
 90. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* 2, 100141.
 91. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534.
 92. Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gislason, M.H., Pihl, S.I., Tsirigos, K.D., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* 40, 1023–1025.
 93. Adzigiqli, L., Wang, Z., Lai, Z., Li, J., and Deng, Y. (2019). Sex determination in pearl oyster: A mini review. *Aquac. Rep.* 15, 100214.
 94. Gordon, S.P., Tseng, E., Salamov, A., Zhang, J., Meng, X., Zhao, Z., Kang, D., Underwood, J., Grigoriev, I.V., Figueroa, M., et al. (2015). Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLoS One* 10, e0132628.
 95. Wu, Y.-C., Rasmussen, M.D., Bansal, M.S., and Kellis, M. (2014). Most parsimonious reconciliation in the presence of gene duplication, loss, and deep coalescence using labeled coalescent trees. *Genome Res.* 24, 475–486.
 96. Huerta-Cepas, J., Dopazo, H., Dopazo, J., and Gabaldón, T. (2007). The human phylome. *Genome Biol.* 8, R109.
 97. dos Reis, M., and Yang, Z. (2011). Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol. Biol. Evol.* 28, 2161–2172.
 98. Benton, M.J., Donoghue, P.C.J., and Asher, R.J. (2009). Calibrating and constraining molecular clocks. In *The Timetree of Life* (Oxford University Press), pp. 35–86.
 99. Jörgen, K.M., Stöger, I., Kano, Y., Fukuda, H., Knebelberger, T., and Schrödl, M. (2010). On the origin of Acochlidia and other enigmatic euthyneuran gastropods, with implications for the systematics of Heterobranchia. *BMC Evol. Biol.* 10, 323.
 100. Berry, W.B.N., and Boucot, A.J. (1973). Correlation of the African Silurian Rocks. In *Correlation of the African Silurian Rocks*, W.B.N. Berry and A.J. Boucot, eds. (Geological Society of America).
 101. Berry, W.B.N., and Boucot, A.J. (1973). Correlation of the African Silurian Rocks. 147 : Correlation of the African Silurian Rocks (Geological Society of America), pp. 1–74.
 102. Stöger, I., Sigwart, J.D., Kano, Y., Knebelberger, T., Marshall, B.A., Schwabe, E., and Schrödl, M. (2013). The continuing debate on deep molluscan phylogeny: evidence for Serialia (Mollusca, Monoplacophora + Polyplacophora). *BioMed Res. Int.* 2013, 407072.
 103. Pojeta, J., Jr., Runnegar, B., and Kriz, J. (1973). *Fordilla troyensis* Barrande: The Oldest Known Pelecypod. *Science* 180, 866–868.
 104. Benton, M.J., Donoghue, P.C.J., Vinther, J., Asher, R., Friedman, M., and Near, T. (2015). Constraints on the timescale of animal evolutionary history. *Palaeontol. Electron.* 18.
 105. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419.
 106. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., et al. (2023). InterPro in 2022. *Nucleic Acids Res.* 51, D418–D427.
 107. Sigrist, C.J.A., de Castro, E., Cerutti, L., Cucho, B.A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Res.* 41, D344–D347.
 108. Letunic, I., Khedkar, S., and Bork, P. (2021). SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* 49, D458–D460.

109. Williamson, R.A., Marston, F.A., Angal, S., Koklitis, P., Panico, M., Morris, H.R., Carne, A.F., Smith, B.J., Harris, T.J., and Freedman, R.B. (1990). Disulphide bond assignment in human tissue inhibitor of metalloproteinases (TIMP). *Biochem. J.* 268, 267–274.
110. Morgunova, E., Tuuttila, A., Bergmann, U., and Tryggvason, K. (2002). Structural insight into the complex formation of latent matrix metalloproteinase 2 with tissue inhibitor of metalloproteinase 2. *Proc. Natl. Acad. Sci. USA* 99, 7414–7419.
111. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. *Nucleic Acids Res.* 43, W39–W49.
112. Katoh, K., Misawa, K., Kuma, K.i., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066.
113. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321.
114. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522.
115. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
<i>Pteria penguin</i> genome	This study	PRJNA782667
Software and algorithms		
Illumina Novaseq	Illumina	https://www.illumina.com/systems/sequencing-platforms/novaseq.html
PacBio Sequel system	Pacbio	https://www.pacb.com/technology/hifi-sequencing/sequel-system/
Jellyfish version 2.2.10	Marcais and Kingsford ⁶³	https://github.com/gmarcais/Jellyfish
Genomescope2 version 2.0	Ranallo-Benavidez et al. ⁶⁴	http://qb.cshl.edu/genomescope/genomescope2.0/
NextDenovo version 2.4.0	Hu et al. ⁶⁵	https://github.com/Nextomics/NextDenovo
purge_dups version 1.2.5	Guan et al. ⁶⁶	https://github.com/dfguan/purge_dups
NextPolish version 1.4.0	Hu et al. ⁶⁷	https://github.com/Nextomics/NextPolish
fastp version 0.23.2	Chen et al. ⁶⁸	https://github.com/OpenGene/fastp
Juicer version 1.6	Durand et al. ⁶⁹	https://github.com/aidenlab/juicer
3D de novo assembly pipeline version 190716	Dudchenko et al. ⁷⁰	https://github.com/aidenlab/3d-dna
Juicebox version 1.11.08	Durand et al. ⁶⁹	https://github.com/aidenlab/Juicebox
Burrows-Wheeler Aligner (BWA) version 0.7.17	Li ⁷¹	https://github.com/lh3/bwa
bcftools version 1.15.1	Danecek et al. ⁷²	https://github.com/samtools/bcftools
BUSCOs version 5.3.2	Manni et al. ⁷³	https://busco.ezlab.org/
RepeatModeler version 2.0.3	Flynn et al. ⁷⁴	https://www.repeatmasker.org/RepeatModeler/
RepeatMasker version 4.1.2	Tarailo-Graovac and Chen ⁷⁵	https://www.repeatmasker.org/RepeatMasker/
Inferral version 1.1.4	Nawrocki et al. ⁷⁶	http://eddylab.org/inferral/
BRAKER2 version 2.1.6	Bruna et al. ⁷⁷	https://github.com/Gaius-Augustus/BRAKER
Trinity version 2.14.0	Haas et al. ⁷⁸	https://github.com/trinityrnaseq/trinityrnaseq/wiki
HISAT2 version 2.2.1	Kim et al. ⁷⁹	http://daehwankimlab.github.io/hisat2/
Isoseq3 version 3.7.0	PacificBiosciences	https://github.com/PacificBiosciences/pbbioconda
cDNA_Cupcake version 28.0.0	Magdoll	https://github.com/Magdoll/cDNA_Cupcake
PASApipeline version 2.5.2	Haas et al. ⁸⁰	https://github.com/PASApipeline/PASApipeline
EVidenceModeler version 1.1.1	Haas et al. ⁸¹	https://github.com/EvidenceModeler/EvidenceModeler
BLAST version 2.13.0	Camacho et al. ⁸²	https://blast.ncbi.nlm.nih.gov/Blast.cgi
InterProScan version 5.52	Jones et al. ⁸³	https://github.com/ebi-pf-team/interproscan
eggNOG-mapper website	Cantalapiedra et al. ⁸⁴	http://eggnog-mapper.embl.de/
Orthofinder version 2.5.4	Emms and Kelly ⁸⁵	https://github.com/davidemms/OrthoFinder
Phylogenetic Analysis by Maximum Likelihood (PAML) version 4.9	Yang ⁸⁶	http://abacus.gene.ucl.ac.uk/software/paml.html

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Computational Analysis of gene Family Evolution (CAFE) pipeline version 5.1	Mendes et al. ⁸⁷	https://github.com/hahnlab/CAFE5
R software version 4.2.1	R Core Team ⁸⁸	https://www.r-project.org/
AnnotationForge version 1.42.2	Carlson et al. ⁸⁹	https://www.bioconductor.org/packages/AnnotationForge
clusterProfiler 4.0	Wu et al. ⁹⁰	https://bioconductor.org/packages/clusterProfiler/
HMMER software version 3.2.1	European Bioinformatics Institute	http://hmmer.org/
IQ-Tree version 2.2.2.6	Minh et al. ⁹¹	http://www.iqtree.org/
SignalP 6.0 website	Teufel et al. ⁹²	https://services.healthtech.dtu.dk/services/SignalP-6.0/
Adobe Photoshop and Illustrator	Adobe	https://www.adobe.com/
Bioinformatic code for this study	This study	https://github.com/YunweiDongLab/iScience_Chao-yi_Ma

RESOURCE AVAILABILITY**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yun-Wei Dong (dongyw@ouc.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The raw genomic sequencing data and *P. penguin* genome generated from this study have been deposited in GenBank under the BioProject number PRJNA782667, and are publicly available as of the date of publication. The accession number has been listed in the [key resources table](#).
- All original code has been deposited in the GitHub repository and is publicly available as of the date of publication. The URL link is https://github.com/YunweiDongLab/iScience_Chao-yi_Ma, and has been listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**The penguin's wing oyster**

A total of three adult penguin's wing oysters *P. penguin* (Pteria, Pteriidae) were collected from Wuzhizhou Island, Sanya City, Hainan Province, China, for molecular extraction and sequencing library construction. However, the precise gender of the *P. penguin* individuals used in this study was indeterminate, as oysters within the Pteriidae family are usually protandrous hermaphrodites, undergoing successive sex reversals throughout their lifespan in response to various influencing factors.⁹³

METHOD DETAILS**Samples preparation and sequencing**

Genomic DNA was extracted from the adductor muscle of one adult *P. penguin* using the QIAGEN DNeasy Kit (QIAGEN, China), following the manufacturer's instructions. The genome sequencing was conducted with a combination of long- and short-read technologies. For conducting the genome survey, a paired-end Illumina sequence library with insert size of 350 bp library was prepared and sequenced on an Illumina NovaSeq system (Novogene Co. Ltd., China). The genomic DNA from the same individual was also used for PacBio sequencing library construction, which was then sequenced on a PacBio Sequel II system (Novogene Co. Ltd., China) after the size-selection for enriching DNA fragments longer than 10 kb. The adductor muscle of the same individual was used for Hi-C library construction, in which the tissue was treated with polyformaldehyde (PFA) and the obtained genomic DNA was prepared following a procedure that included cross-linking, digestion by the restriction enzyme DpnII, repairing ends as well as labeling by biotinylated residues. After the genomic DNA was purified and sheared into 350 bp fragments, an Illumina NovaSeq platform (Novogene Co. Ltd., China) was used for subsequent sequencing.

Seven tissues (including adductor muscle, foot, root of foot, gill, labial palps, liver, and mantle) were collected from two adult *P. penguin*. Total mRNA was extracted with TRIzol reagent (Invitrogen, USA) based on the manufacturer's protocol and the assessment of RNA integrity was conducted using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, USA). To construct sequencing libraries, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Then, a total of fourteen Illumina sequencing libraries derived from seven tissues of two adults were sequenced on an Illumina Novaseq system (Novogene Co. Ltd., China) and one Pacbio sequencing library containing all tissues was sequenced on a PacBio Sequel II system (Novogene Co. Ltd., China). Raw reads were filtered by removing low-quality reads and sequencing-adaptor-contaminated reads also were removed; the output data were used for the following genome annotation.

Genome survey, assembly, scaffolding and assessment

Genome size, heterozygosity, and repeat content of *P. penguin* were estimated by using k-mer analysis. Illumina short reads were first trimmed to remove adaptors and reads with >10% ambiguous or >20% low-quality bases. Then clean reads were used for calculating the distribution of 21-mer frequency with Jellyfish version 2.2.10⁶³ and estimating the genome size with Genomescope2 version 2.0.⁶⁴

Contig assembly of *P. penguin* using PacBio long reads was conducted via NextDenovo version 2.4.0,⁶⁵ with the parameter "genome_size = 850m". Both haplotigs and contig overlaps in the primary denovo assembly were identified and removed using purge_dups version 1.2.5⁶⁶ with cutoff values set as "-l 10 -m 110 -u 400". The non-redundant assembly was subsequently polished by NextPolish version 1.4.0⁶⁷ with Illumina short reads, which had been cleaned via fastp version 0.23.2.⁶⁸

Hi-C sequencing data were used for chromosome-level scaffolding. Polished contigs were first automatically scaffolded using clean Hi-C sequencing reads following the pipeline (<https://github.com/esrice/hic-pipeline>). The assembly at a scaffold-level was input into Juicer version 1.6⁶⁹ for filtering and deduplicating Hi-C reads. Then, genomic scaffolding was conducted with the 3D *de novo* assembly pipeline version 190716⁷⁰ using the default diploid mode. Several manual corrections were done based on the Hi-C contact maps in Juicebox version 1.11.08⁶⁹ to ensure the scaffolds within the same pseudo chromosomal linkage groups met the Hi-C linkage characteristics. Lastly, 346 contigs were scaffolded into 14 pseudo chromosomal linkage groups, and seven contig debris were produced during the manual correcting process. These debris were not anchored due to insufficient Hi-C linkage found on them. All raw data have been deposited in the database of the National Center for Biotechnology Information (NCBI).

The integrity and accuracy of the genome assembly were evaluated at the single-base level. Illumina short-insert library reads were mapped onto the contigs by using Burrows-Wheeler Aligner (BWA) version 0.7.17⁷¹ and genetic variants were called out by bcftools version 1.15.1.⁷² Genome completeness was also assessed by using BUSCOs version 5.3.2⁷³ and analyzed by searching metazoan gene sets (metazoa_odb10) with genomic mode.

Genome annotation

Repetitive sequences in *P. penguin* genome assembly were identified and soft-masked prior to gene structure annotation. RepeatModeler version 2.0.3⁷⁴ with option "-LTRStruct" were used to construct a *de novo* canonical database of repetitive elements. Repeat families of Bivalvia were extracted from Repbase (derived RepeatMasker libraries) and Dfam database version 3.2 to build a homology canonical database. These two databases were combined together for identifying and classifying repeats in the *P. penguin* genome using RepeatMasker version 4.1.2.⁷⁵ Screens for the genome-wide distribution of noncoding RNA genes were conducted using cmscan program within Infernal software version 1.1.4⁷⁶ with default parameters.

Protein-coding genes were annotated by incorporating ab initio prediction, and transcriptome-assisted methods. Ab initio gene prediction pipeline BRAKER2 version 2.1.6⁷⁷ was used to predict genes with repeat-masked genome sequences by default settings. Transcriptome-assisted annotation was run with Illumina RNA-Seq and Pacbio Iso-Seq data combined. Illumina RNA-Seq data from seven tissues were cleaned via fastp and used for building a comprehensive transcriptome using both *de novo* and genome-based assembly via Trinity version 2.14.0.⁷⁸ For building the genome-based assembly, HISAT2 version 2.2.1⁷⁹ was used to align the clean RNA-Seq data to the *P. penguin* genome assembly. Pacbio Iso-seq data were first preprocessed, clustered and filtered to build transcripts structures via Isoseq3 version 3.7.0 (<https://github.com/PacificBiosciences/pbbioconda>) and cDNA_Cupcake version 28.0.0 (https://github.com/Magdoll/cDNA_Cupcake) following ToFU pipeline.⁹⁴ The *de novo* and genome-guided RNA-Seq assemblies and transcript structures based on Iso-Seq data were inputted to PASA version 2.5.2⁸⁰ following its comprehensive transcriptome database-generating pipeline. Finally, all ab initio gene predictions and transcript alignments were combined into weighted consensus gene structures using EvidenceModeler version 1.1.1.⁸¹

Functional annotation was performed by homology comparisons of the predicted protein sequences against reviewed SwissProt databases (<https://www.uniprot.org/>, last modified May 25, 2022) using BLASTP command within BLAST version 2.13.0⁸² with an E-value threshold of 1e-5. Besides, InterProScan version 5.52⁸³ and eggNOG-mapper website⁸⁴ was used for gene function and pathway annotation by searching through the Pfam and Gene Ontology (GO) database.

Phylogenomic analyses in Mollusca

Full protein sets of eighteen genomes public for use, including seventeen Mollusca species and one Brachiopoda species (*Lingula anatina*), were downloaded and used for phylogenomic analysis as well as orthology inference with *P. penguin*. The quality of completeness of full protein sets were evaluated using BUSCOs version 5.3.2 and analyzed by searching metazoan gene sets (metazoa_odb10) with protein mode. And when a gene in each set possessed multiple isoforms, only the longest protein sequence was selected as the representative that was

used. Proteomes from all eighteen species were used as the input to Orthofinder version 2.5.4,⁸⁵ which conducted the inference of orthogroups and orthologs, constructed the complete set of gene trees for all orthogroups and the rooted maximum-likelihood species tree. Based on the phylogeny relationship among the species analyzed, the Orthofinder rooted the gene tree for each orthogroup identified, and further conducted a hybrid duplication-loss-coalescence (DLC) algorithm of the rooted gene trees to identify gene duplication events. The hybrid method integrates the advantages of with DLCpar (full) method the highest accuracy,⁹⁵ and the species-overlap method with high efficiency.⁹⁶ MCMCtree,⁹⁷ part of Phylogenetic Analysis by Maximum Likelihood (PAML) version 4.9,⁸⁶ was used to yield the time-calibrated tree, with nodes constrained by fossil records and geographic events. The time frames were used to constrain the nodes in the MCMC tree: minimum = 168.6 Ma and soft maximum = 473.4 Ma for Panpulmonata and Tectipleura;⁹⁸ hard minimum bound = 390 Ma for Caenogastropoda and Heterobranchia;⁹⁹ minimum = 470.2 Ma and soft maximum = 531.5 Ma for *A. californica* and *L. gigantea*;⁹⁸ hard minimum bound = 419.2 Ma for the first appearance of Mytilidae;^{100,101} hard minimum = 465.0 Ma for the first appearance of Pteriomorpha;¹⁰² hard minimum bound = 520 Ma for the first appearance of bivalves;¹⁰³ minimum = 532 Ma and soft maximum = 549 Ma for the first appearance of molluscs;¹⁰⁴ and minimum = 550.25 Ma and soft maximum = 636.1 Ma for the first appearance of Lophotrochozoan.¹⁰⁴ The rooted tree from Orthofinder results was used to provide a reference tree topology, and the burn-in, sample frequency, and number of samples were set as 1000000, 10, and 500000, respectively.

Orthologous gene family size change analysis in Mollusca

The Orthofinder inferred HOGs, orthogroups at each hierarchical level by analyzing the rooted gene trees for each orthogroups. The dataset containing the species gene number in each HOG was used for the gene family expansion and contraction analysis by Computational Analysis of gene Family Evolution (CAFE) pipeline version 5.1,⁸⁷ and the rooted ultrametric tree from the MCMCtree results were used to provide time-calibrated evolution information. In the calculating process with CAFE 5, the root equilibrium frequency was assumed to be not a uniform distribution and the -p flag is given without a parameter. Gamma model was used to run calculations as if each gene family could belong to a different evolutionary rate category, and $K = 3-4$ gamma rate categories were tested in two independent analyses.⁸⁷ The highest likelihood was found using $K = 3$, with predicted $\lambda = 0.00226$ and $\alpha = 0.781194$, because of its maximum Model Gamma Final Likelihood (-lnL) (254497). The cutoff of the p -value was set to 0.05 to determine which families underwent a significant expansion/contraction.

The Gene Ontology (GO) Enrichment Analysis of gene sets in *P. penguin* significantly expanded gene families was conducted on R software version 4.2.2.⁸⁸ The R package AnnotationForge⁸⁹ was first used for building a comprehensive annotation data package (org.Ppenguin.eg.db) by incorporating the gene pathway annotation results from InterProScan and eggNOG-mapper website; then the R package clusterProfiler⁹⁰ was used for enrichment analysis.

Phylogenetic orthology inference in Metazoa

Full protein sets of forty-one publicly available metazoan genomes (from sixteen Phylum, thirty-nine Class) public for using were downloaded, evaluated using BUSCOs version 5.3.2 analyzed by searching metazoan gene sets (metazoa_odb10) with protein mode, and used for HOGs inference together with *P. penguin* by Orthofinder version 2.5.4. The longest protein sequence was selected as the representative that was used when a gene in each set possessed multiple isoforms. To ensure the accuracy of the orthogroups found, the metazoan phylogeny was clarified and a rooted species tree containing the phylogenetic relationship between or within each clade was constructed by referring to multiple studies.³³⁻⁴⁰

TIMP gene family analyses with Mollusca and Metazoa

In two independent analyses, the TIMP gene family was surveyed across the Molluscan and then Metazoan to determine the genomic distribution and calculate the copy numbers. This gene family was first searched from the full protein sets of eighteen molluscs as well as forty-one metazoans mentioned above first using hmmsearch program within HMMER software version 3.2.1 (<http://hmmer.org/>) with an E-value threshold of $1e^{-5}$ against a TIMP hmm file (PF00965) from Pfam-A database;¹⁰⁵ in this procedure, the identical sequences were removed and isoforms coded by one gene (like two sequences in *M. gigas*: NP_001292268.1 and XP_011421049.2, coded by the same gene with Gene symbol as LOC105323677) were checked to only keep one. The TIMP sequences were then identified from the sequences obtained using InterProScan;¹⁰⁶ which searches for protein signature matches, including "METALLOPROTEASE INHIBITOR" in the PANTHER version 17.0, "NTR domain profile" in PROSITE version 2022_05,¹⁰⁷ and "NTR_2 domain" in SMART version 9.0¹⁰⁸ databases. Sequences failing to have matches in both the former two protein structure databases were excluded, the remaining were named "putative TIMP" and divided into two groups according to whether matching the conserved NTR domain in the SMART database^{109,110}: one having matches were considered as canonical TIMP, while the other failing to have matches were considered as "pseudo TIMPs". All of the putative TIMPs were used as inputs for MEME website (<https://meme-suite.org/meme/tools/meme>) for protein motif discovery;¹¹¹ and in each run, ten motifs were set and expected to be found. According to the accession ID of each TIMP protein, the gene structure information was derived from the genome annotation files in Gff3 format; the whole length and number of exons in each TIMP gene were calculated.

For the phylogenetic analyses of molluscan and metazoan TIMP gene family, the putative TIMP sequences (144 sequences in molluscan TIMP gene family and 114 sequences in the metazoan) and the canonical TIMP sequences from core Mollusca-Brachiopoda TIMP HOG (53 sequences) and core Eumetazoa TIMP HOG (114 sequences) were aligned in Mafft version 7.505¹¹² using strategy "L-INS-i", both the original Mafft alignment of putative sequences and sequences from the two core HOGs were put into IQ-Tree version 2.2.2.6⁹¹ for Maximum Likelihood (ML) tree construction. In each run, branch supports were evaluated using both 1000 SH-aLRT bootstrap replicates¹¹³ and 10000 ultrafast

bootstrap replicates;¹¹⁴ and also optimized using a hill-climbing nearest neighbor interchange (NNI) search by adding “-bnni” flag. For selecting the best model, the ModelFinder contained with IQ-Tree was used,¹¹⁵ which identified the best-fitting model as “VT+R6” and “VT+R4” for ML trees constructed from the alignments of putative and core Mollusca-Brachiopoda TIMPs, as “LG+R7” and “LG+R6” for trees constructed from the alignments of putative and core Eumetazoa TIMPs, according to their BIC values. Signal peptides in these two groups of core TIMPs were predicted using SignalP 6.0 program.⁹²

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical tests were performed with R (version 4.2.2).