

Article

Taxonomic Distribution and Molecular Evolution of Mytilectins

Marco Gerdol ^{1,*}, Daniela Eugenia Nerelli ¹, Nicola Martelossi ¹, Yukiko Ogawa ², Yuki Fujii ²,
Alberto Pallavicini ¹ and Yasuhiro Ozeki ³

¹ Department of Life Sciences, University of Trieste, Via Licio Giorgieri 5, 34127 Trieste, Italy

² Graduate School of Pharmaceutical Sciences, Nagasaki International University, 2825-7 Huis Ten Bosch, Sasebo 859-3298, Japan

³ Graduate School of NanoBio Sciences, Yokohama City University, 22-2 Seto, Kanazawa-ku, Yokohama 236-0027, Japan

* Correspondence: mgerdol@units.it

Abstract: R-type lectins are a widespread group of sugar-binding proteins found in nearly all domains of life, characterized by the presence of a carbohydrate-binding domain that adopts a β -trefoil fold. Mytilectins represent a recently described subgroup of β -trefoil lectins, which have been functionally characterized in a few mussel species (Mollusca, Bivalvia) and display attractive properties, which may fuel the development of artificial lectins with different biotechnological applications. The detection of different paralogous genes in mussels, together with the description of orthologous sequences in brachiopods, supports the formal description of mytilectins as a gene family. However, to date, an investigation of the taxonomic distribution of these lectins and their molecular diversification and evolution was still lacking. Here, we provide a comprehensive overview of the evolutionary history of mytilectins, revealing an ancient monophyletic evolutionary origin and a very broad but highly discontinuous taxonomic distribution, ranging from heteroscleromorph sponges to ophiuroid and crinoid echinoderms. Moreover, the overwhelming majority of mytilectins display a chimera-like architecture, which combines the β -trefoil carbohydrate recognition domain with a C-terminal pore-forming domain, suggesting that the simpler structure of most functionally characterized mytilectins derives from a secondary domain loss.

Keywords: marine invertebrates; innate immunity; glycan-binding; lectins; pore-forming; β -trefoil



Citation: Gerdol, M.; Nerelli, D.E.; Martelossi, N.; Ogawa, Y.; Fujii, Y.; Pallavicini, A.; Ozeki, Y. Taxonomic Distribution and Molecular Evolution of Mytilectins. *Mar. Drugs* **2023**, *21*, 614. <https://doi.org/10.3390/md21120614>

Academic Editor: Rob Keyzers

Received: 5 November 2023

Accepted: 25 November 2023

Published: 27 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Lectins are a large class of proteins nearly ubiquitously found in all living organisms, ranging from unicellular prokaryotes to complex multicellular eukaryotes, which play a pivotal role in mediating carbohydrate-protein interactions. These glycan-binding molecules display a remarkable molecular diversity, due to the presence of distinct Carbohydrate Recognition Domains (CRDs) that exhibit specificity for carbohydrate moieties. The extraordinary breadth of recognized ligands allows lectins to mediate fundamental biological processes, from cell adhesion [1], to cell-cell recognition [2], apoptotic cell clearance [3], embryogenesis [4], food particle recognition [5], and the discrimination between “self” and “nonself”. In this context, lectins have a paramount importance for pathogen recognition, which occurs thanks to the specific detection of molecular patterns associated with invading pathogenic agents, such as bacteria and viruses, the so-called Microbe Associated Molecular Patterns (MAMPs) [6]. This carbohydrate-dependent immune recognition mechanism elucidates how lectins participate in innate immune responses, triggering the activation of downstream immune effectors upon the recognition of exogenous entities. In organisms devoid of an immunoglobulin- and T-cell receptor-based adaptive immune system, lectins gain a primary role as a first barrier to prevent microbial invasion. Hence, it is not surprising that a significant number of lectin families underwent massive expansion during

evolution in different lineages, fueled by tandem gene duplication events and subsequent fast molecular diversification (and functional specialization) via positive selection [7–11].

R-type lectins (RTLs), which take their name from the plant toxin ricin, are one of the many different structural superfamilies of lectins that have been described to date. These lectins are characterized by a CRD that displays a unique β -trefoil structural organization, consisting of three homologous subdomains, which most likely derive from the duplication of an ancestral smaller glycan-binding peptide [12]. Although this CRD is often found associated with other domains in large proteins that carry out functions linked with carbohydrate biosynthesis and metabolism [13], several RTLs displaying a simple architecture, which only include a single CRD, have been previously described in a few phyla of invertebrate animals [14–17]. Among these, a group of sequences that display a highly divergent primary sequence from all other previously described RTLs have attracted significant attention due to their glycan-binding specificity, which could make them interesting targets for biotechnological applications. These molecules, which are collectively known as mytilectins from the name of MytiLec-1 [18], have been so far functionally characterized only in a small group of bivalve mollusks, all belonging to the family Mytilidae [19–23]. Due to their marked binding specificity for globotriose (Gb3), a glycan expressed at high levels by Burkitt lymphoma Raji cells, as well as due to the cytotoxic effect exerted upon binding on these and other types of cancer cells [24–26] and the ability to modulate macrophage activity in mice [27], mytilectins may find practical applications in the context of cancer diagnosis and treatment. For this reason, mytilectins have been the subject of mutagenesis studies aimed at better understanding their structure–function relationships to allow the design of molecules with improved glycan-binding properties [28,29], and have been also used as a template for the in silico design of the synthetic lectin Mitsuba [30].

While all functionally characterized mytilectins display a high primary sequence homology with each other, being characterized by the presence of a single CRD, further studies have revealed the existence of additional members of the same family in mussels. These sequences display a higher molecular weight, due to the presence of a C-terminal domain that shares a striking structural resemblance with aerolysin, a cytolytic toxin from *Aeromonas hydrophila* [25]. This observation would suggest that these mytilectins, called “chimera-type” to differentiate them from the aforementioned “proto-type” mytilectins, may be involved in the formation of pores in target membranes through the formation of oligomeric beta-barrels, as happens in other toxins that have acquired similar structural features in a convergent manner [31–33].

More recently, sequences sharing high primary sequence homology with mytilectins have been described in *Lingula anatina*, a marine invertebrate belonging to the phylum Brachiopoda, distantly related with bivalve mollusks, implying a shared ancestry for these molecules and strongly suggesting a taxonomic spread much broader than originally thought [34]. Although mytilectins have been previously referred to as members of a novel lectin family [35,36], the lack of any specific investigations concerning their evolutionary origin, taxonomic spread, and relationships has prevented, to date, a formal description of the “mytilectin family”. Here, with a comprehensive screening of available genomic and transcriptomic resources, we provide a clear overview concerning these aspects, supporting a monophyletic origin for mytilectins deeply rooted in the metazoan lineage, and reveal that these lectins are present in an unexpectedly large number of animal phyla.

2. Results and Discussion

2.1. Taxonomic Distribution of Mytilectins

The large-scale screening of available metazoan–omic resources allowed the authors to significantly expand the taxonomic range of distribution of mytilectins compared with previous reports, as schematically displayed in Figure 1. The presence of members of this lectin family in largely divergent animal phyla, ranging from Porifera to Echinodermata, suggest an ancient evolutionary origin predating the acquisition of bilateral symmetry. At

the same time, mytilectins display a highly discontinuous distribution, characterized by their absence in several major phyla. This situation could be consistent with two alternative and not mutually exclusive scenarios, namely (i) the very ancient origin of this lectin family, preceding the acquisition of bilateral symmetry, followed by multiple independent gene loss events occurring in different taxa; or (ii) a more recent origin traceable to a single phylum, followed by several horizontal gene transfer events, which greatly expanded the narrow original distribution of this lectin family. The plausibility of these two hypotheses, in light of the recent report of the presence of SaroL-1, a lectin sharing striking structural similarity with mytilectins, in the choanoflagellate *S. rosetta* (Figure 1), will be discussed after the comprehensive overview of the results of the comparative genomics analyses conducted in this study outlined in the next sections.

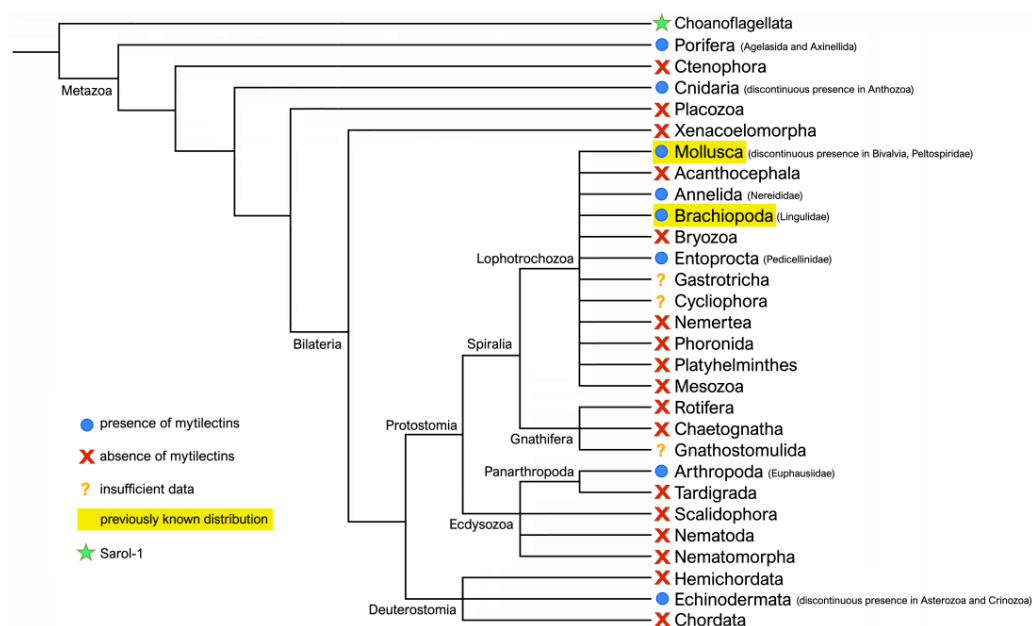


Figure 1. Schematic view of the distribution of mytilectins across the metazoan tree of life. Presence or absence is displayed at the phylum level, with further indication of their distribution at lower taxonomic ranks, whenever relevant. In this representation, presence indicates that some (but not all) the members of a given phylum had mytilectin genes, whereas absence indicates that no mytilectin sequences could be detected in any of the genomes or transcriptomes available for a given phylum. In a few cases, evidenced with a question mark, available genomic data was insufficient to draw definitive conclusions concerning the presence or absence of mytilectins. The phylogenetic placement of the phylum Choanoflagellata at the base of the tree is marked due to the detection of SaroL-1, a lectin sharing striking structural similarity with mytilectins, as discussed in Section 2.6.

2.1.1. Phylum Porifera

The most early branching metazoan phylum where mytilectins were detected was Porifera. However, mytilectin genes were only present in a single one out of the 10 sponge genomes available to date (as of November 2023), i.e., *Agelas oroides*, belonging to the order Agelasida. This finding, together with the absence of orthologous sequences in other genomes of the five species placed in the subclass Heteroscleromorpha (class Demospongiae), allowed the authors to infer the lack of mytilectins in the orders Haplosclerida, Spongillida, and Suberitida. However, further analysis of transcriptome data extended the distribution of poriferan mytilectins to the order Axinellida (in detail, mytilectin transcripts were found in *Eurypon* sp. 2 AS-2020 and *Hymenophora stellifera* [37]). No evidence supporting the presence of mytilectins could be collected in the three other extant classes of Porifera, i.e., Calcarea, Hexactinellida, and Homoscleromorpha.

2.1.2. Phylum Cnidaria

In the phylum Cnidaria, the presence of mytilectins was restricted to the class Anthozoa, where this gene family was represented across a broad range of orders belonging to two out of three subclasses (Hexacorallia and Octocorallia; no genomic data is available for Ceriantharia), with significant gaps across phylogeny, suggesting a complex evolutionary history characterized by multiple independent losses. Namely, mytilectin genes were detected in 7 out of 85 available genomes of Hexacorallia (8% of the total), i.e., *Orbicella faveolata*, *Montipora capitata* [38], *Montipora* sp. Colony 1 RG-2022, *Stylophora pistillata* [39], *Palythoa heliodiscus* [40], *Palythoa grandis* [40], and *Ricordea florida*. A few additional species could be added to this list (*Acropora tenuis* [41], *Montipora digitata* [42], *Alveopora japonica* [43], *Parachrysogorgia stellata* [44], and *Fimbriaphyllia ancora* [45]) thanks to transcriptomic evidence. Most of these species belong to different families within the order Scleractinia (i.e. Acroporidae, Chrysogorgiidae, Euphylliidae, Faviidae, Merulinidae, and Pocilloporidae), but matches were also found in Corallimorpharia and Zoantharia.

The prevalence of mytilectins in available octocorallian genomes (2 out of 11) was similar to that outlined above in Hexacorallia, with positive hits in *Paramuricea clavata* [46] and *Eunicella verucosa* [47], both classified within the order Malacalcyonacea. Transcriptome data, besides further matches in *Eleutherobia rubra* [48], *Clavularia* sp. cla_tr77125 [49], and *Scleronephthya gracillima*, expanded the range of distribution of cnidarian mytilectins to Scleralcyonacea (i.e., *Heliopora coerulea* [50]).

No evidence supporting of the presence of mytilectins could be collected in the other cnidarian classes (i.e., Cubozoa, Hydrozoa, Myxozoa, Scyphozoa, and Staurozoa), neither at the genomic nor at the transcriptomic level.

2.1.3. Phylum Mollusca, Class Bivalvia

The class Bivalvia was the taxonomic group with the highest number of mytilectin sequences identified, in part due to its high species richness and abundance of genomic and transcriptomic resources, in part due to the fact that several species displayed multiple paralogous gene copies.

Fully sequenced genomes are still lacking for the subclass Protobranchia, an early offshoot of the Bivalve lineage. However, transcriptomic evidence supports the presence of mytilectins in *Solemya velum* [51] and *Ennucula tenuis* [52], which belong to two different protobranch orders, i.e., Nuculida and Solemyida. This finding indicates the likely presence of a mytilectin gene in the latest common ancestor of all bivalves. As far as the second bivalve subclass (i.e., Autobranchia), is concerned, mytilectin genes were identified both in the infraclass Pteriomorphia and in the infraclass Heteroconchia. Nevertheless, the distribution of mytilectins was sparse with significant gaps, mirroring the general situation outlined at higher taxonomic ranks in Figure 1.

According to the current WoRMS classification, Pteriomorphia includes five orders: Arcida, Limida, Mytilida, Ostreida, and Pectinida. Mytilectins were clearly missing in Ostreida, as no significant homology could be detected in any of the 14 fully sequenced genomes available to date. Although no genome data was available for Limida, a similar conclusion could be drawn for this order based on the analysis of transcriptome data [53]. On the other hand, mytilectins were present in one out of the two genomes of Arcida, i.e. *Tegillarca granulosa* [54]. Mytilectins were present in the genomes of 6 out of 13 species belonging to the order Mytilida. Five of these (*Mytilus californianus*, *Mytilus chilensis*, *Mytilus coruscus*, *Mytilus edulis*, and *Mytilus galloprovincialis*) [55–59] are congeneric, confirming the previous reports of MytiLec-1, CGL, and MTL in *Mytilus* and *Crenomytilus* spp. [20,26,60]. The sixth species was *Perna viridis* [61], which also belongs to the subfamily Mytilinae. Mytilectin-encoding transcripts were also detected in the congeneric species *Perna perna* [62] and *Perna canaliculus*. The transcriptome of the ribbed mussel *Geukensia demissa* [63] allowed the authors to expand the range of distribution of mytilectins to a second mytilid family, i.e., Brachidontinae. Mytilectins were widespread in Pectinida, as evidenced by their presence in all the six species with a fully sequenced genome available (*Pecten maximus*, *Mizuhopecten yessoensis*, *Argopecten*

irradians, *Argopecten purpuratus*, and *Mimachlamys varia*) [64–68] and by their detection in the transcriptomes of *Adamussium colbecki* and *Nodipecten subnodosus* [69,70].

In the infraclass Heteroconchia, mytilectins were clearly absent in the genomes of all Palaeoheterodonta, which include the large majority of freshwater mussels and clams, and likely absent also in Archiheterodonta. Despite being absent in the superorder Anomalodesmata, mytilectins were found in 5 out of 24 available fully sequenced genomes available for the superorder Imparidentia, part of Euheterodonta, the largest group of heteroconch bivalves. Three of these were members of the order Myida (*Congeria kusceri*, *Dreissena polymorpha*, and *Mya arenaria*) [71,72] and the other two (*Mercenaria mercenaria* and *Saximodus purpurata*) [73] belonged to the order Venerida. Nevertheless, several other species classified as venerid clams lacked mytilectin genes. The abundance of available genomes for the order Cardiida allowed ruling out the presence of mytilectins in this taxa. Although -omic resources are still limited for other minor Imparidentia orders, neither genomic, nor transcriptomic data supported the presence of mytilectins, with the lone exception of a partial transcript sequence detected in *Hiatella arctica* (order Adapedonta) [53].

2.1.4. Other Mollusca

Despite their frequent occurrence in Bivalvia, mytilectins were absent in the overwhelming majority of other mollusks. The lone exception was represented by the three species of the family Peltospiridae (class Gastropoda) with a sequenced genome available, i.e., *Chrysomallon squamiferum* [74], *Dracogyra subfusca*, and *Gigantopelta aegis* [75], which are all deep-sea hydrothermal vent snails. The lack of genomic resources for sister taxa did not allow the authors to ascertain whether mytilectins were restricted to Peltospiridae or more broadly distributed in all the members of the order Neomphalida. Although over 50 other gastropod genomes have been sequenced and assembled to date, none of these presented mytilectin-encoding genes, indicating a highly reduced representation of this gene family in the most species-rich molluscan class.

Albeit significant taxonomic gaps are still present for the minor molluscan classes (e.g., Monoplacophora, Scaphopoda, and Caudofoveata), no evidence supporting the existence of mytilectins could be found in Solenogastres, Cephalopoda, and Polyplachophora.

2.1.5. Other Lophotrochozoa

The presence of mytilectins was previously reported in *Lingula anatina*, a member of the phylum Brachiopoda [34]. Although this species remains to date the only brachiopod with a sequenced genome, current evidence suggests that mytilectin distribution in this phylum might be very narrow, due to the lack of orthologous sequences in the transcriptomes of several other brachiopod species.

The only other lophotrochozoan phylum with compelling evidence supporting the presence of mytilectins was Annelida. In detail, matches were identified, either at a genomic or a transcriptomic level, in the family Nereididae, consisting of polychaete worms (subclass Errantia, order Phyllodocida). Namely, mytilectin sequences were found in *Alitta virens* [76], *Perinereis aibuhitensis* [77], and *Platynereis dumerilii* [78]. The absence of mytilectins in several genomes belonging to the same order, but different families, points to a distribution restricted to nereidid polychaete worms.

The only other instance of detection of partial sequences with a clear homology with mytilectins was *Pedicellina cernua* (Entoprocta, Pedicellinidae) [79]. However, the lack of -omic resources for this neglected minor lophotrochozoan phylum impeded further confirmation of this finding at the genome level and the exploration in other species belonging to closely related taxa.

2.1.6. Ecdysozoa

Mytilectins were largely absent in Ecdysozoa, which include the most species-rich group of animals, i.e., arthropods. In fact, only a single out of the nearly 2700 arthropod genomes available to date (most of which are from Insecta) carried mytilectin genes. This

surprising finding was made in the Antarctic krill *Euphausia superba*, whose recently sequenced genome is one of the largest ever reported in the animal kingdom [80]. Further transcriptomic investigations confirmed this finding and allowed the authors to detect mytillectins in the transcriptome of a second species belonging to the family Euphausiidae, i.e., *Meganyctiphanes norvegica* [81]. The absence of sequences bearing detectable homology in the genomes of about 90 other crustacean species clearly marked the restricted presence of mytillectins in krill as a unique case in Ecdysozoa.

2.1.7. Phylum Echinodermata

Within Deuterostomata, Echinodermata was the only phylum where mytillectin sequences could be unambiguously identified and confirmed both at a genomic and at a transcriptomic level. Nevertheless, like other previously described cases, mytillectins were detected just in a restricted number of taxa.

In detail, mytillectin genes were found only in two out of the three extant echinoderm subphyla, i.e. Asterozoa and Crinozoa, thereby pointing out their absence in both sea urchins and sea cucumbers. The presence of mytillectins could be only confirmed in a subgroup of asterozoans, i.e., those belonging to the class Ophiuroidea. In detail, complete or partial sequences were recovered in *Amphiura filiformis* [82], *Ophiothrix exigua* [83], *Ophiothrix spiculata* [84], and *Ophioderma brevispina* [85]. Albeit these four species belong to the subclass Myophiuroidea, the lack of genomic data for the other ophiuroid species prevented a more precise assessment of taxonomic delimitation in this case. Mytillectin genes were also detected in a single crinoid species, i.e., *Nesometra sesokonis* [86], and further investigations failed to identify orthologous sequences in the transcriptomes of several other crinoids, suggesting a narrow taxonomic distribution within this subphylum.

2.1.8. Other Deuterostomes

No significant matches supporting the presence of mytillectins were found in the genomes of deuterostome phyla other than Echinodermata. These comprised all representatives of the phyla Hemichordata and Chordata, including cephalochordates and tunicates. This was strongly supported by transcriptomic evidence, even though a few assembled transcripts linked to *Sardinops melanostictus* (Actinopterygii) matching mytillectins are present in the TSA repository. However, upon further inspection of their phylogenetic placement, these sequences were found to bear close similarity with bivalve mytillectins. Taking into account that bivalve larvae are usual components of the diet of sardines [87], we could safely classify this finding as a false positive linked to the ingestion of an unidentified bivalve species by the fish.

2.2. Structural Features of Mytillectins

We had previously described the presence of two distinct types of mytillectins [88], defined as “proto-type” and “chimera-type”, respectively, characterized by markedly different length and domain architecture. Proto-type mytillectins, which include all functionally characterized mytillectins to date (i.e. CGL, MTL and MytiLec-1), display a single β -trefoil CRD with no accessory domains, being approximately 150 aa long. On the other hand, other mytillectins identified in *M. galloprovincialis* and *L. anatina* were significantly longer due to the presence of a C-terminal extension encoding an additional 140 aa-long domain, whose three-dimensional structure was predicted to resemble that of *Aeromomas hydrophyla* aerolysin and other pore-forming toxins [34,89]. The presence of these two distinct domain architectures leaves an open question concerning the ancestral structural configuration of mytillectins, due to the existence of two alternative evolutionary scenarios: indeed, the CRD/pore-forming domain combination observed in chimera-type mytillectins could be the result of a gene fusion event involving an ancestral proto-type mytillectin gene, or, alternatively, proto-type mytillectins could be the result of the secondary loss of the pore-forming domain present in the ancestral chimera-type gene. To provide an answer to this question, we collected over one hundred mytillectin sequences from multiple phyla,

significantly expanding the repertoire of both mytilectin types. The multiple sequence alignment displayed in Figure 2 reports a representative group of the mytilectins recovered in this study.

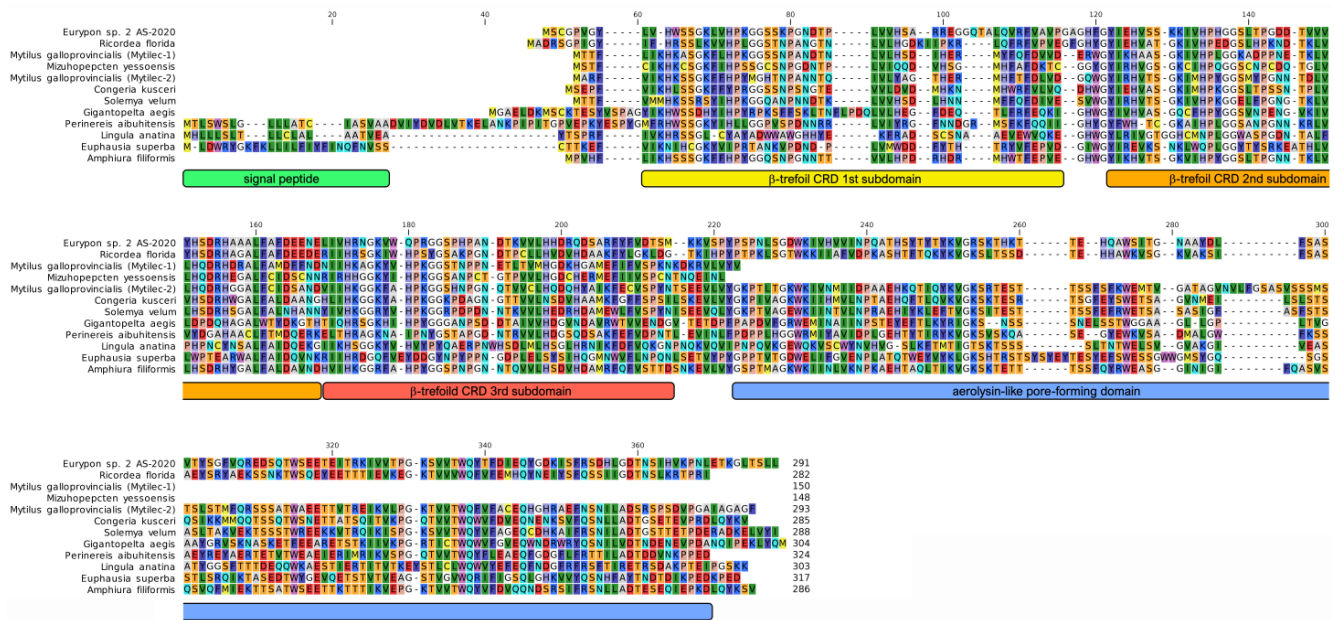


Figure 2. Multiple sequence alignment of representative mytilectins from different phyla. The most significant structural features discussed in the text (i.e., the presence/absence of a signal peptide, the delimitation of the CRD, and pore-forming domain) are highlighted. Note that the *M. galloprovincialis* MytilLec-1, and the *M. yessoensis* sequences reported in the MSA are proto-type mytilectins.

The data we collected strongly support the second out of the two aforementioned evolutionary scenarios, due to the presence of chimera-type mytilectins in all early branching metazoan phyla, such as Porifera and Cnidaria. Moreover, chimera-type mytilectins were far more widespread than proto-type mytilectins, as the latter were only detected in *Mytilus* spp., in the brachiopod *L. anatina* and in all pectinid bivalves. Nevertheless, while both domain architectures were simultaneously present in mussels and brachiopods, scallops were the only species to uniquely display proto-type sequences. Moreover, within bivalves, proto-type mytilectins were absent in Arcida and Heteroconchia, further supporting a secondary domain loss in the proto-type mytilectins of mussels and scallops.

Another interesting structural feature which has been likely acquired in a convergent manner by distinct phyla concerns the N-terminal region of mytilectins. As previously reported [88], mussel mytilectins lack a detectable signal peptide for secretion, thereby being targeted to the extracellular environment through a non-canonical secretion pathway. This peculiarity is shared by the overwhelming majority of the mytilectins detected in this study and clearly represents the ancestral status of the mytilectin precursor gene, as suggested by the lack of signal peptides in the sequences of sponges and anthozoans (Figure 2). Nevertheless, three notable exceptions were identified: brachiopods (as previously reported in [34]), nereidid polychaetes, and krill. Indeed, these three taxa displayed well-supported signal peptides, which would support the secretion of these mytilectins following the canonical route mediated by signal recognition particles.

2.3. Gene Architecture Strongly Supports a Monophyletic Origin for all Mytilectins

As briefly discussed above, the highly discontinuous taxonomic distribution of mytilectins summarized in Figure 1 could be consistent with two alternative scenarios, the likelihood of which will be here evaluated using phylogenetic inference and gene architecture information. The conservation of exon/intron boundaries and splicing sites among genes from distantly related phyla is generally considered as strong evidence of orthology [90–92], to the point that

this factor is often incorporated in gene prediction algorithms [93]. On the other hand, convergent independent intron gain in the same position, with the very same phase, is considered unlikely [94] and the presence of a reduced number of introns, or their entire lack, has been previously reported as a typical feature of eukaryotic genes recently acquired by horizontal gene transfer [95]. Hence, the detection of shared intron positions by mytilectin genes of distantly related phyla would provide strong evidence supporting an ancient monophyletic origin for mytilectins, which would be consequently inferred to be already present in the latest common ancestor of all metazoans. We will here report the intron/exon architecture of mytilectins with regard to the coding sequence only, even though some mytilectin genes most certainly include an exon in the 5' end, before the ATG start codon [88]. This is necessary due to the focus placed by standard genome annotation pipelines on coding sequences, which may therefore often entirely miss or mis-annotate 5' and 3' UTR regions. At the same time, we will here mostly focus on the gene architecture of proto-type mytilectins, as the genes encoding proto-type mytilectins are significantly shorter due to the lack of the pore-forming domain (see Section 2.2).

The complete analysis of the available genomic data revealed that the full ORF of mytilectins was contained within a single exon in poriferan and cnidarian genes, as well as in those from Peltospiridae gastropods. Due to the basal placement of sponges and cnidarians in the animal tree of life (see Figure 1), this data would suggest that the ancestral metazoan mytilectin gene was intronless. However, mytilectin genes acquired introns in several other taxa during evolution, leading to slightly more complex splicing patterns, which may split the ORF between either two or three coding exons. Interestingly, despite the presence of lineage-specific losses and acquisitions, the placement of splicing sites was often conserved across largely divergent phyla, thereby strongly supporting a shared monophyletic evolutionary origin for animal mytilectins (Figure 3).

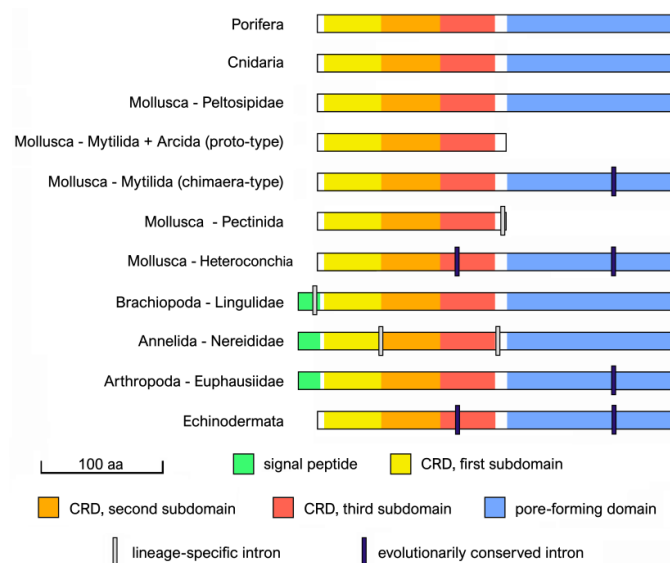


Figure 3. Schematic view of the architecture of mytilectin genes in the major animal taxa where this lectin family was identified. Colored bars indicate the full-length sequence of mytilectin precursor proteins, with indication of the position of introns relative to the N-terminal CRD (subdivided among the three subdomains of the β -trefoil fold) and to the C-terminal pore-forming domain. Please note that only the mytilectins of Brachiopoda, Annelida, and Arthropoda display a signal peptide, which determines a slight extension of the protein sequence at the N-terminal end. As discussed in the text, introns located in the 5' and 3' UTR were disregarded due to the frequent lack of accurate annotations of these regions in available genome assemblies.

In detail, the ORF of chimera-type mytilectins from echinoderms (including both ophiurids and crinoids), bivalves, and krill was interrupted by the presence of an intron in the very same position, i.e., roughly breaking in two equal parts the C-terminal pore-

forming domain. Echinoderms and heteroconch bivalves also shared an additional intron in the N-terminal region, within the third subdomain of the β -trefoil CRD. This exon was missing in krill, mussels, and ark shells, but also in scallops (which only have shorter proto-type mytilectin genes), which nevertheless displayed a single intron in a different position, i.e., after the end of the CRD.

Additional independent lineage-specific intron gain events could be inferred in nereidid polychaetes, which displayed two introns, placed roughly between the first and the second CRD subdomains, and in the region connecting the CRD to the pore-forming domain, respectively, and in Brachiopoda, where the ORF was interrupted close to the N-terminal end. Interestingly, the placement of this intron may explain the acquisition of a signal peptide by brachiopod mytilectins, perhaps due to exon shuffling phenomena (see Section 2.2).

2.4. Phylogeny of Mytilectins

We investigated the evolutionary relationships among the mytilectins identified in this study through maximum likelihood phylogenetic inference, revealing a complex picture which did not fully mirror the well-established taxonomic placement of the taxa where mytilectin genes were present (Figure 4).

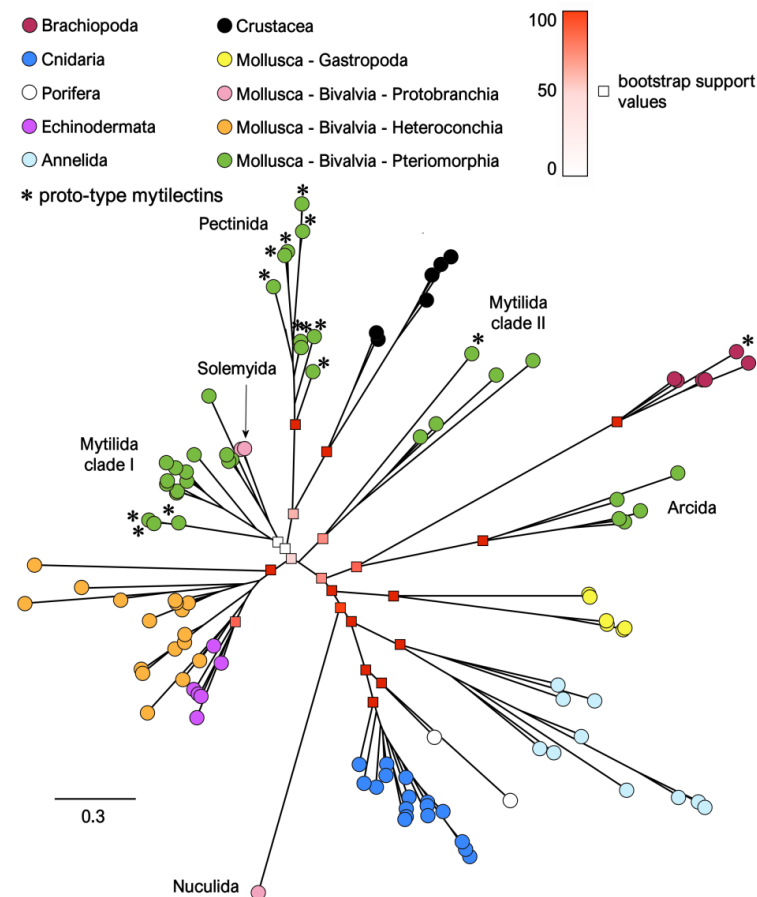


Figure 4. Maximum likelihood phylogeny of mytilectins, represented as an unrooted tree. Individual sequences are displayed as circles, colored based on the major taxonomic groups discussed in the main text (i.e. Brachiopoda, Cnidaria, Porifera, Echinodermata, Crustacea, and Mollusca, further subdivided among Gastropoda and three taxonomic groups of Bivalvia, i.e., Protobranchia, Heteroconchia, and Pteriomorphia). Further lower-rank classifications discussed in the main text are also indicated close to the relevant clades. For the sake of simplicity, statistical support for major basal nodes is reported with squares, whose color indicates bootstrap values. Proto-type mytilectins are marked with an asterisk.

For example, even though sponge and cnidarian sequences were placed in the same monophyletic clade with maximum statistical support (bootstrap value = 100), thereby reflecting the early branching position of these phyla in the animal tree of life, neither protostome nor deuterostome mytillectins were grouped in monophyletic clades, as would have been expected in the case mytillectin evolution closely following the evolutionary relationships among species. Nevertheless, while considering lower taxonomic ranks, a high number of sequence groups matched highly supported monophyletic clades: this was the case with Pectinida (bootstrap support = 100), Arcida (bootstrap support = 100), Crustacea (bootstrap support = 100), Brachiopoda (bootstrap support = 100), Annelida (bootstrap support = 100), and Gastropoda (bootstrap support = 100).

On the other hand, this was not the case for the sequences of Mytilida, which, as reported in Section 2.2, are highly diversified from a structural point of view. These sequences were indeed grouped in two clades, also highlighted in Figure 4: the poorly supported clade I (bootstrap support = 22), included a mixture of mytilid proto-type and chimera-type sequences, together with the two sequences from Solemyida (Protobranchia). On the contrary, the better supported clade II (bootstrap support = 61) exclusively included mytilid sequences (both proto- and chimera-type). A second large clade of sequences characterized by high statistical support (bootstrap support = 100) included a subgroup of bivalve sequences (i.e., those from Heteroconchia) and all echinoderm mytillectins. Although the bootstrap support for the internal nodes of this branch of the tree were low, hindering further investigation into the detailed relationships between the mytillectins of these two phylogenetically distant taxa, their high primary sequence similarity, strongly supported by ML inference, would point to an interesting case of convergent evolution.

The scattered position of proto-type mytillectins in the phylogenetic tree (marked by asterisks in Figure 4) further supports the previously hypothesized independent origin for these modified mytillectins in Mytilida, Pectinida, and Brachiopoda, as outlined in Section 2.2. Similarly, the mytillectins that display a signal peptide (i.e., those from Branchiopoda, Crustacea, and Annelida) would create a polyphyletic group, thereby strongly supporting the independent acquisition of a signal for canonical secretion in these three taxa.

Overall, the molecular phylogeny of mytillectins opens several interesting evolutionary questions, which can only be partially solved at the moment due to the current lack of information concerning the functional specialization (if any) of proto-type and chimera-type mytillectins. The significant discrepancies between gene and species phylogeny in this case may have several different explanations. Undoubtedly, mytillectins are relatively short proteins (i.e., the total size of the multiple sequence alignment analyzed with phylogenetic inference was 268 amino acids), which often display limited primary sequence homology (with p-distances in the range of 0.7–0.8 in inter-phyla pairwise comparisons), and the low number of phylogenetically informative sites may have led to the incorrect or poorly supported relative placement of some sequence groups, fundamentally altering the ordering of some nodes of the tree. Nevertheless, as highlighted above, mytillectins display several convergent features, which include the independent loss of the pore-forming domain, as well as the independent acquisition of a signal peptide in different phyla. For similar reasons, one might expect to observe the independent occurrence of other convergent sequence features, whose weight on such a short MSA may have understandably led to unexpected branching patterns. Moreover, the lack of non-metazoan sequences showing high homology with mytillectins to be used for rooting purposes, together with the low number of available mytillectins from the most basal animal group (i.e. Porifera, with just two sequences available) represented another limitation for phylogenetic inference. Finally, it needs to be considered that the unusual and highly discontinuous taxonomic distribution of mytillectins in extant species is most likely the product of massive gene loss events, which resulted in the availability of a particularly low number of sequences in a few key taxa.

2.5. Optimization of Mytilectin-Specific Hidden Markov Models

The β -trefoil and pore-forming domains of mytilectins share a significant structural homology with several R-type lectins and aerolysin-like toxins from multiple organisms, thereby allowing the identification of these conserved domains using Hidden Markov Models (HMM). For example, the presence of the mytilectin β -trefoil domain would allow the recognition of mytilectins as members of the Ricin B-like lectins homologous superfamily (IPR035992) in InterPro [96], as well as members of the CATH superfamily 2.80.10.50 [97]. Nevertheless, such classifications are extremely generic, as to date over 93 and 184 thousand protein sequences fit within these large superfamilies. Similarly, the pore-forming domain of chimera-type mytilectins matches the CATH superfamily G3DSA:2.170.15.10, which includes over 7000 different sequences.

Taking this into account, we investigated the possibility of generating HMM profiles that could be used to specifically identify mytilectins by separately detecting the two structural units found in chimera-type sequences. The performance of the two resulting profile HMMs, obtained as described in Section 3, and named “mytilectin N-terminal domain” and “mytilectin C-terminal domain”, respectively, was tested against the full set of mytilectins described in this manuscript and all the protein sequences deposited in UniProt.

These tests confirmed the high specificity of both HMMs. In detail, the N-terminal HMM could positively identify all mytilectins, with e-values ranging from $6.7e^{-70}$ to $3.6e^{-31}$, with the worst results usually obtained for the detection for mytilectins from polychaete worms, crustaceans, and ark shells. The C-terminal HMM was also able to identify all mytilectins, with e-values ranging from $1.7e^{-54}$ to $1.3e^{-28}$. In this case, the worst results were observed for the mytilectins of brachiopods, crustaceans, and gastropods. These good performances were accompanied by a lack of false-positive detections in UniProt: as expected, the non-target hits achieving the best scores for the two HMMs (i.e. the mucoricin RLT1_RHIO9 and the monalysin MONAL_PSEE4, respectively) belonged to the same structural superfamilies. However, their e-values (2.9 and 0.097, respectively) remained far below the detection threshold.

2.6. On the Occurrence of the β -Trefoil/Aerolysin-like Pore-Forming Domain Combination in Non-Metazoan Phyla: Convergent Evolution or Shared Ancestry?

Although our recursive homology search strategy (see Section 3.1) was limited to Metazoa, we noticed the presence of statistically significant similarities between a few animal mytilectins and a single sequence recently described in the choanoflagellate *S. rosetta*, belonging to a sister group of Metazoa within Opisthokonta. This sequence was Sarol-1, characterized as a pore-forming protein from this unicellular marine eukaryote in 2022 [98]. As previously noted by Notova and colleagues, the N-terminal region of Sarol-1, corresponding to the β -trefoil domain, displayed a significant homology, both at the primary sequence and at the structural level, with MytiLec-1, CGL, and MTL. Most importantly, unlike functionally characterized mytilectins, Sarol-1 also displayed a long C-terminal extension which structurally resembled aerolysin, thereby completely matching the domain architecture of chimera-type mytilectins (see Section 2.2). In light of this observation and of our novel identification of mytilectins in the phylum Porifera (see Section 2.1), one might wonder whether the evolutionary origins of mytilectins might be pushed further back deeply into the Opisthokonta lineage, at the very least to the latest common ancestor of the Metazoa and Choanoflagellata lineages.

While the remarkable primary sequence similarity (i.e. up to 30–35%) between the β -trefoil domains of Sarol-1 and some mytilectins may support this view, a much higher divergence was observed between the aerolysin domains of Sarol-1 and chimera-type mytilectins. This was fully consistent with the results obtained with the analysis of the Sarol-1 sequence with mytilectin-specific profile HMMs described in Section 2.5. Indeed, the β -trefoil domain of Sarol-1 could be recognized with a highly significant e-value (i.e., $1e^{-30}$), whereas the aerolysin-like domain could not be recognized at all. Hence, although Sarol-1 could be reasonably considered as a mytilectin-related sequence that most likely shares ancestry with

metazoan mytillectins, its primary sequence does not fully conform with that of chimera-type mytillectins. Based on the available sequence data, it is therefore unclear whether the pore-forming domain was acquired independently in the choanoflagellate and metazoan lineages, or whether the significant divergence observed between Sarol-1 and mytillectins in this domain is due to rapid evolution that made the presence of homology unrecognizable.

As a side note, it needs to be remarked that structural similarity-based detection methods, such as HHPRED [99], allowed the identification of other proteins sharing the very same domain combination of chimera-type mytillectins and Sarol-1 in other phyla, even though, in these cases, primary sequence similarity was too low to allow the detection of homology through BLAST. As highlighted by Figure 5, this was the case with LSL, a toxin produced by the mushroom *Laetiporus sulphureus* [31], and with the insecticidal toxins Tpp80Aa1 and BinAB from *Bacillus thuringiensis* and *Lysinibacillus sphaericus* [100,101]. Since the possibility of detecting such structural convergence is currently restricted to protein sequences whose three-dimensional structure has previously been experimentally determined, the number of proteins found in nature that have acquired the very same domain combination in a convergent manner is likely much higher.

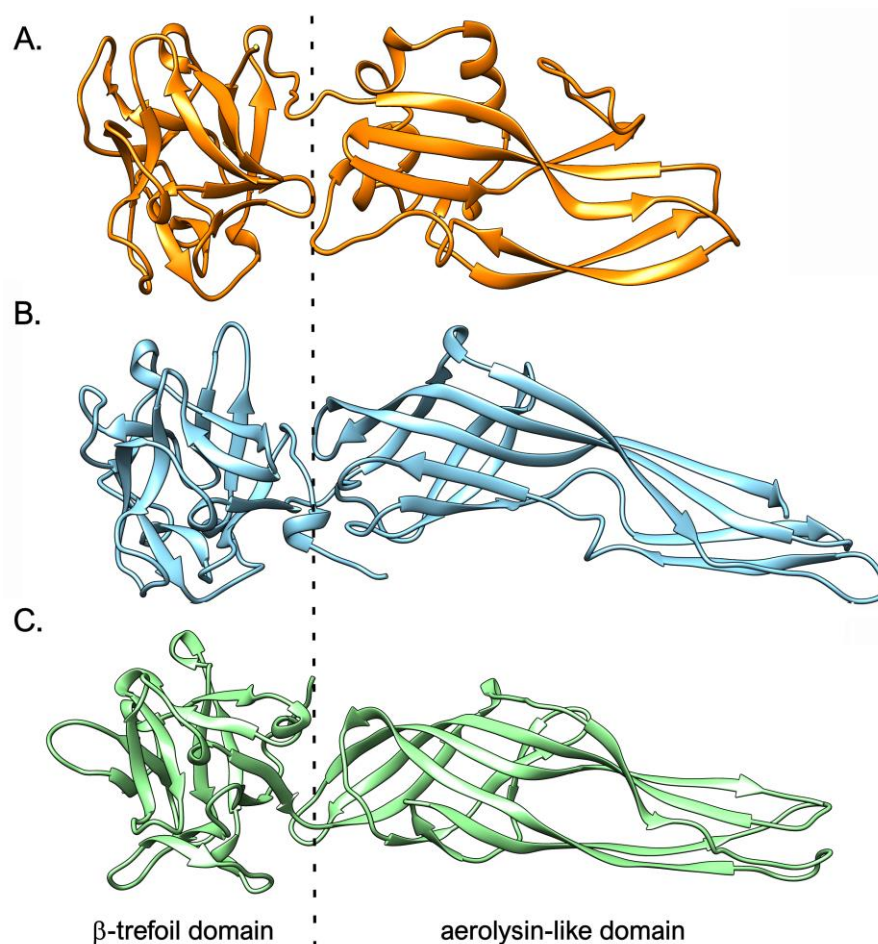


Figure 5. Comparison between the three-dimensional structures of *M. galloprovincialis* mytillectin-3, selected as a representative member of metazoan chimera-type mytillectins (panel (A)), *S. rosetta* Sarol-1 (panel (B), PDB ID: 7QE4), and *L. sulphureus* LSL (panel (C), PDB ID: 1W3F).

3. Materials and Methods

3.1. Retrieval of Mytillectin Sequences from Public Databases

Mytillectin sequences were retrieved through a recursive homology search approach using BLASTp [102], which used the previously described sequences of *M. galloprovincialis* (MytiLec-1, -2, and -3) [18,88] and *L. anatina* [34] as initial queries. In the first round

of analysis, homology searches were carried out against the NCBI nr database and all hits with e-values lower than 0.05 were extracted and manually inspected to verify the reliability of protein sequences, with particular attention to those deriving from automated gene annotation pipelines. The sequences displaying unusual features (e.g., missing N- and C-terminal ends or with significant internal gaps) were marked as suspicious and subjected to manual confirmation as briefly described below. All newly recovered sequences were clustered by pairwise similarity with CD-HIT [103] based on an arbitrary similarity threshold of 0.5. The representative sequences of each cluster were then selected as queries for a second round of sequence homology searches. The process was performed recursively until no novel hits could be recovered.

A similar strategy was used to recover mytilectin sequences from the de novo assembled transcriptomes deposited in the NCBI TSA database, using tBLASTn [102]. In this case, positive matches were translated in silico using the ExPasy translate tool [104]. Incomplete sequences (i.e., those lacking the initial ATG codon or the stop codon), as well as those characterized by obvious mis-assembly or encoding identical proteins (e.g., those encoded by redundant assembled contigs) were removed.

Finally, the homology search approach was extended to all assembled genomes deposited in the NCBI Whole Genome Shotgun database, with particular attention to genomes devoid of gene annotation. Here, mytilectin genes were manually annotated, by combining the identification of High Scoring Segment Pairs (HSPs) via tBLASTn and the prediction of canonical donor and acceptor splicing sites via Genie [105]. All the amino acid sequences of the mytilectins described in this manuscript are reported in Supplementary Material S1.

3.2. Protein Sequence Analysis

All the protein sequences obtained as described in the previous section were subjected to signal peptide prediction with SignalP v.6.0 [106] and analyzed with InterProScan v.5 [107] to verify the identification of a complete profile HMM attributable to the Ricin B-like lectins homologous superfamily (IPR035992). The sequences that only displayed this structurally recognizable domain were classified as proto-type mytilectins, according to previously suggested nomenclature [88], whereas those displaying an additional match in the C-terminal region, consistent with the CATH superfamily 2.80.10.50 [64], were classified as chimera-type mytilectins. The reliability of all proto-type mytilectin protein sequences derived from automated gene prediction pipelines was evaluated through the comparison of these sequences with assembled RNA-sequencing data, whenever available.

The presence of structural similarities with proteins with experimentally determined three-dimensional structures deposited in the Protein DataBank (PDB) database was investigated with HHPRED [99]. The three-dimensional structure of *M. galloprovincialis* mytilectin-3 was predicted with AlphaFold v2.3.0 [108], using CASP14-like settings.

3.3. Phylogenetic Analysis

All mytilectin protein sequences were aligned with MUSCLE [109], obtaining a multiple sequence alignment (MSA) file that was used as an input for subsequent analyses. To reduce background noise and only keep phylogenetically informative sites, poorly alignable positions were removed. In detail, the MSA was trimmed by removing the signal peptide region (whenever present) and all residues located at the N-terminal side of the CRD. Similarly, all residues located at the C-terminal side of the pore-forming domain were also removed. Moreover, all alignment positions characterized by missing data (i.e., gaps) in >50% sequences were deleted. The resulting clean MSA, including 274 positions, was analyzed with ModelFinder [110] to detect the best-fitting model of molecular evolution for this dataset, which was determined to be a WAG + R5 model [111], according to the Bayesian Information Criterion [112]. A Maximum Likelihood (ML) phylogenetic inference analysis was subsequently run with IqTree [113]. The reliability of the generated tree was tested with 1000 ultrafast bootstrap replicates. Due to the lack of suitable outgroups, the tree was graphically represented as an unrooted tree.

3.4. Creation and Validation of Mytilectin-Specific Profile HMMs

The MSA of all mytilectins was modified to remove highly similar sequences (those displaying pairwise homology > 90%) to reduce the bias linked with the over-representation of mytilectin sequences from phylogenetically close taxa. The MSA was split in two parts, reflecting the position of the N-terminal β -trefoil and C-terminal pore-forming domains, respectively, as previously defined by other studies [88]. N-terminal and C-terminal extensions were removed, whenever present. The two profile HMMs were built with the HMMer hmmbuild module [114], and the hmmsearch module was used for testing their performance against two distinct datasets. Dataset (a) included all the mytilectin sequences previously described in the literature [20,21,26,88,115], and dataset (b) included the complete Uniprot sequence database [116]. Positive matches were detected using default e-value thresholds.

The “mytilectin N-terminal domain” and “mytilectin C-terminal domain” HMMs are available as Supplementary Materials S2 and S3, respectively.

4. Conclusions

Although previous studies have already highlighted the potential biotechnological use of mytilectins [26,30,115], so far, all available functional data concern a small number of sequences isolated from bivalve mollusks, whose biological role in the context of innate immunity still remains rather elusive. We reported here an overview of the distribution of this type of lectins in the animal kingdom, formally defining a widespread gene family. We noted a very ancient evolutionary origin and a complex distribution pattern characterized by a great number of gene loss events that occurred independently of each other in different phyla, also identifying the chimera-type architecture as the most ancestral one, from which proto-type mytilectins likely derived following the loss of the pore-forming domain. Although unusual, this patchy taxonomic distribution mirrors that previously described for other effectors of innate immunity in invertebrate organisms [117]. Understanding which evolutionary factors underlie the maintenance and loss of mytilectins in different taxa is a prerequisite for better understanding their biological role and improving their functional study, which will be made easier with the availability of profile HMMs specifically developed for their identification in large sequence databases.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/md21120614/s1>, File S1: amino acid sequences of all the mytilectins reported in this study (note that “G” and “T” indicate sequences identified from genomes and transcriptomes, respectively); File S2: profile HMM for the mytilectin N-terminal domain; File S3: profile HMM for the mytilectin C-terminal domain.

Author Contributions: Conceptualization, M.G., Y.F. and Y.O. (Yasuhiro Ozeki); methodology, M.G. and A.P.; validation, M.G., D.E.N. and N.M.; formal analysis, M.G., D.E.N. and N.M.; data curation, M.G. and N.M.; investigation, M.G., D.E.N., Y.O. (Yukiko Ogawa); writing—original draft preparation, M.G.; writing—review and editing, M.G., D.E.N., N.M., Y.O. (Yukiko Ogawa), Y.F., A.P. and Y.O. (Yasuhiro Ozeki); visualization, M.G.; supervision, M.G.; All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Assisted Joint Research Program (Exploration type) of the J-GlycoNet cooperative network, accredited by the Ministry of Education, Culture, Sports, Science, and Technology, MEXT, Japan, as a Joint Usage/Research Center. Yuki Fujii, Yukiko Ogawa, and Yasuhiro Ozeki are supported by a research grant (19K06239, 23K10950, 23K06190) from JSPS. This work was also funded by the European Union under NextGenerationEU. Ecosystem INEST Interconnected Nord-Est Innovation; PNRR Mission 4, Component 2, Investment 1.5, Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or The European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: All mytilectin protein sequences are available in Supplementary Material S1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kaltner, H.; Stierstorfer, B. Animal Lectins as Cell Adhesion Molecules. *Acta Anat.* **1998**, *161*, 162–179. [[CrossRef](#)] [[PubMed](#)]
2. Sharon, N.; Lis, H. Lectins as Cell Recognition Molecules. *Science* **1989**, *246*, 227–234. [[CrossRef](#)] [[PubMed](#)]
3. Nauta, A.J.; Castellano, G.; Xu, W.; Woltman, A.M.; Borrias, M.C.; Daha, M.R.; van Kooten, C.; Roos, A. Opsonization with C1q and Mannose-Binding Lectin Targets Apoptotic Cells to Dendritic Cells. *J. Immunol.* **2004**, *173*, 3044–3050. [[CrossRef](#)]
4. Maki, J.S.; Mitchell, R. Involvement of Lectins in the Settlement and Metamorphosis of Marine Invertebrate Larvae. *Bull. Mar. Sci.* **1985**, *37*, 675–683.
5. Emmanuelle, P.E.; Mickael, P.; Evan, W.; Shumway, S.E.; Bassem, A. Lectins Associated with the Feeding Organs of the Oyster *Crassostrea virginica* Can Mediate Particle Selection. *Biol. Bull.* **2009**, *217*, 130–141.
6. Stuart, L.M.; Paquette, N.; Boyer, L. Effector-Triggered versus Pattern-Triggered Immunity: How Animals Sense Pathogens. *Nat. Rev. Immunol.* **2013**, *13*, 199–206. [[CrossRef](#)]
7. Gerdol, M.; Venier, P.; Pallavicini, A. The Genome of the Pacific Oyster *Crassostrea gigas* Brings New Insights on the Massive Expansion of the C1q Gene Family in Bivalvia. *Dev. Comp. Immunol.* **2015**, *49*, 59–71. [[CrossRef](#)]
8. Jiang, S.-Y.; Ma, Z.; Ramachandran, S. Evolutionary History and Stress Regulation of the Lectin Superfamily in Higher Plants. *BMC Evol. Biol.* **2010**, *10*, 79. [[CrossRef](#)]
9. Hofberger, J.A.; Nsibo, D.L.; Govers, F.; Bouwmeester, K.; Schranz, M.E. A Complex Interplay of Tandem- and Whole-Genome Duplication Drives Expansion of the L-Type Lectin Receptor Kinase Gene Family in the Brassicaceae. *Genome Biol. Evol.* **2015**, *7*, 720–734. [[CrossRef](#)]
10. Saco, A.; Suárez, H.; Novoa, B.; Figueras, A. A Genomic and Transcriptomic Analysis of the C-Type Lectin Gene Family Reveals Highly Expanded and Diversified Repertoires in Bivalves. *Mar. Drugs* **2023**, *21*, 254. [[CrossRef](#)]
11. Yang, Y.; Labbé, J.; Muchero, W.; Yang, X.; Jawdy, S.S.; Kennedy, M.; Johnson, J.; Sreedasyam, A.; Schmutz, J.; Tuskan, G.A.; et al. Genome-Wide Analysis of Lectin Receptor-like Kinases in *Populus*. *BMC Genom.* **2016**, *17*, 699. [[CrossRef](#)]
12. Rutenber, E.; Ready, M.; Robertus, J.D. Structure and Evolution of Ricin B Chain. *Nature* **1987**, *326*, 624–626. [[CrossRef](#)]
13. Cummings, R.D.; Schnaar, R.; Ozeki, Y. R-Type Lectins. In *Essentials of Glycobiology*, 4th ed.; Cold Spring Harbor Laboratory Press: New York, NY, USA, 2022; Chapter 31.
14. Kawsar, S.M.A.; Takeuchi, T.; Kasai, K.; Fujii, Y.; Matsumoto, R.; Yasumitsu, H.; Ozeki, Y. Glycan-Binding Profile of a D-Galactose Binding Lectin Purified from the Annelid, *Perinereis nuntia* ver. *vallata*. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* **2009**, *152*, 382–389. [[CrossRef](#)]
15. Hirabayashi, J.; Dutta, S.K.; Kasai, K. Novel Galactose-Binding Proteins in Annelida. Characterization of 29-kDa Tandem Repeat-Type Lectins from the Earthworm *Lumbricus Terrestris*. *J. Biol. Chem.* **1998**, *273*, 14450–14460. [[CrossRef](#)] [[PubMed](#)]
16. Fujii, Y.; Gerdol, M.; Kawsar, S.M.A.; Hasan, I.; Spazzali, F.; Yoshida, T.; Ogawa, Y.; Rajia, S.; Kamata, K.; Koide, Y.; et al. A GM1b/Asialo-GM1 Oligosaccharide-Binding R-Type Lectin from Purplish Bifurcate Mussels *Mytilisepta virgata* and Its Effect on MAP Kinases. *FEBS J.* **2020**, *287*, 2612–2630. [[CrossRef](#)] [[PubMed](#)]
17. Gerdol, M. First Insights into the Repertoire of Secretory Lectins in Rotifers. *Mar. Drugs* **2022**, *20*, 130. [[CrossRef](#)] [[PubMed](#)]
18. Fujii, Y.; Dohmae, N.; Takio, K.; Kawsar, S.M.A.; Matsumoto, R.; Hasan, I.; Koide, Y.; Kanaly, R.A.; Yasumitsu, H.; Ogawa, Y.; et al. A Lectin from the Mussel *Mytilus galloprovincialis* Has a Highly Novel Primary Structure and Induces Glycan-Mediated Cytotoxicity of Globotriaosylceramide-Expressing Lymphoma Cells. *J. Biol. Chem.* **2012**, *287*, 44772–44783. [[CrossRef](#)] [[PubMed](#)]
19. Belogortseva, N.I.; Molchanova, V.I.; Kurika, A.V.; Skobun, A.S.; Glazkova, V.E. Isolation and Characterization of New GalNAc/Gal-Specific Lectin from the Sea Mussel *Crenomytilus grayanus*. *Comp. Biochem. Physiol. Part C Pharmacol. Toxicol. Endocrinol.* **1998**, *119*, 45–50. [[CrossRef](#)] [[PubMed](#)]
20. Chikalovets, I.V.; Kovalchuk, S.N.; Litovchenko, A.P.; Molchanova, V.I.; Pivkin, M.V.; Chernikov, O.V. A New Gal/GalNAc-Specific Lectin from the Mussel *Mytilus trossulus*: Structure, Tissue Specificity, Antimicrobial and Antifungal Activity. *Fish. Shellfish. Immunol.* **2016**, *50*, 27–33. [[CrossRef](#)]
21. García-Maldonado, E.; Cano-Sánchez, P.; Hernández-Santoyo, A. Molecular and Functional Characterization of a Glycosylated Galactose-Binding Lectin from *Mytilus californianus*. *Fish. Shellfish. Immunol.* **2017**, *66*, 564–574. [[CrossRef](#)]
22. Golotin, V.A.; Filshtein, A.P.; Chikalovets, I.V.; Yu, K.N.; Molchanova, V.I.; Chernikov, O.V. Expression and Purification of a New Lectin from Mussel *Mytilus trossulus*. *Protein Expr. Purif.* **2019**, *154*, 62–65. [[CrossRef](#)]
23. Kovalchuk, S.N.; Chikalovets, I.V.; Chernikov, O.V.; Molchanova, V.I.; Li, W.; Rasskazov, V.A.; Lukyanov, P.A. CDNA Cloning and Structural Characterization of a Lectin from the Mussel *Crenomytilus grayanus* with a Unique Amino Acid Sequence and Antibacterial Activity. *Fish. Shellfish. Immunol.* **2013**, *35*, 1320–1324. [[CrossRef](#)] [[PubMed](#)]
24. Liao, J.-H.; Chien, C.-T.H.; Wu, H.-Y.; Huang, K.-F.; Wang, I.; Ho, M.-R.; Tu, I.-F.; Lee, I.-M.; Li, W.; Shih, Y.-L.; et al. A Multivalent Marine Lectin from *Crenomytilus grayanus* Possesses Anti-Cancer Activity through Recognizing Globotriose Gb3. *J. Am. Chem. Soc.* **2016**, *138*, 4787–4795. [[CrossRef](#)] [[PubMed](#)]
25. Hasan, I.; Asaduzzaman, A.K.M.; Swarna, R.R.; Fujii, Y.; Ozeki, Y.; Uddin, M.B.; Kabir, S.R. MytiLec-1 Shows Glycan-Dependent Toxicity against Brine Shrimp *Artemia* and Induces Apoptotic Death of Ehrlich Ascites Carcinoma Cells In Vivo. *Mar. Drugs* **2019**, *17*, 502. [[CrossRef](#)] [[PubMed](#)]

26. Hasan, I.; Sugawara, S.; Fujii, Y.; Koide, Y.; Terada, D.; Iimura, N.; Fujiwara, T.; Takahashi, K.G.; Kojima, N.; Rajia, S.; et al. MytilLec, a Mussel R-Type Lectin, Interacts with Surface Glycan Gb3 on Burkitt's Lymphoma Cells to Trigger Apoptosis through Multiple Pathways. *Mar. Drugs* **2015**, *13*, 7377–7389. [[CrossRef](#)] [[PubMed](#)]
27. Chernikov, O.V.; Wong, W.-T.; Li, L.-H.; Chikalovets, I.V.; Molchanova, V.I.; Wu, S.-H.; Liao, J.-H.; Hua, K.-F. A GalNAc/Gal-Specific Lectin from the Sea Mussel *Crenomytilus grayanus* Modulates Immune Response in Macrophages and in Mice. *Sci. Rep.* **2017**, *7*, 6315. [[CrossRef](#)] [[PubMed](#)]
28. Kovalchuk, S.N.; Buinovskaya, N.S.; Likhatskaya, G.N.; Rasskazov, V.A.; Son, O.M.; Tekutyeva, L.A.; Balabanova, L.A. Mutagenesis Studies and Structure-Function Relationships for GalNAc/Gal-Specific Lectin from the Sea Mussel *Crenomytilus grayanus*. *Mar. Drugs* **2018**, *16*, 471. [[CrossRef](#)]
29. Kovalchuk, S.N.; Golotin, V.A.; Balabanova, L.A.; Buinovskaya, N.S.; Likhatskaya, G.N.; Rasskazov, V.A. Carbohydrate-Binding Motifs in a Novel Type Lectin from the Sea Mussel *Crenomytilus grayanus*: Homology Modeling Study and Site-Specific Mutagenesis. *Fish. Shellfish. Immunol.* **2015**, *47*, 565–571. [[CrossRef](#)]
30. Terada, D.; Voet, A.R.D.; Noguchi, H.; Kamata, K.; Ohki, M.; Addy, C.; Fujii, Y.; Yamamoto, D.; Ozeki, Y.; Tame, J.R.H.; et al. Computational Design of a Symmetrical β -Trefoil Lectin with Cancer Cell Binding Activity. *Sci. Rep.* **2017**, *7*, 5943. [[CrossRef](#)]
31. Mancheño, J.M.; Tateno, H.; Sher, D.; Goldstein, I.J. Laetiporus Sulphureus Lectin and Aerolysin Protein Family. *Adv. Exp. Med. Biol.* **2010**, *677*, 67–80. [[CrossRef](#)]
32. Jia, N.; Liu, N.; Cheng, W.; Jiang, Y.-L.; Sun, H.; Chen, L.-L.; Peng, J.; Zhang, Y.; Ding, Y.-H.; Zhang, Z.-H.; et al. Structural Basis for Receptor Recognition and Pore Formation of a Zebrafish Aerolysin-like Protein. *EMBO Rep.* **2016**, *17*, 235–248. [[CrossRef](#)] [[PubMed](#)]
33. Dang, L.; Rougé, P.; Van Damme, E.J.M. Amaranthin-like Proteins with Aerolysin Domains in Plants. *Front. Plant Sci.* **2017**, *8*, 1368. [[CrossRef](#)] [[PubMed](#)]
34. Gerdol, M.; Luo, Y.-J.; Satoh, N.; Pallavicini, A. Genetic and Molecular Basis of the Immune System in the Brachiopod *Lingula anatina*. *Dev. Comp. Immunol.* **2018**, *82*, 7–30. [[CrossRef](#)] [[PubMed](#)]
35. Fujii, Y.; Gerdol, M.; Hasan, I.; Koide, Y.; Matsuzaki, R.; Ikeda, M.; Rajia, S.; Ogawa, Y.; Kawsar, S.M.A.; Ozeki, Y. Phylogeny and Properties of a Novel Lectin Family with β -Trefoil Folding in Mussels. *Trends Glycosci. Glycotechnol.* **2018**, *30*, E195–E208. [[CrossRef](#)]
36. Chikalovets, I.; Filshtein, A.; Molchanova, V.; Mizgina, T.; Lukyanov, P.; Nedashkovskaya, O.; Hua, K.-F.; Chernikov, O. Activity Dependence of a Novel Lectin Family on Structure and Carbohydrate-Binding Properties. *Molecules* **2020**, *25*, 150. [[CrossRef](#)]
37. Strehlow, B.W.; Schuster, A.; Francis, W.R.; Eckford-Soper, L.; Kraft, B.; McAllen, R.; Nielsen, R.; Mandrup, S.; Canfield, D.E. Transcriptomic Responses of Sponge Holobionts to in Situ, Seasonal Anoxia and Hypoxia. *bioRxiv* **2023**.
38. Helmkampf, M.; Bellinger, M.R.; Geib, S.M.; Sim, S.B.; Takabayashi, M. Draft Genome of the Rice Coral *Montipora capitata* Obtained from Linked-Read Sequencing. *Genome Biol. Evol.* **2019**, *11*, 2045–2054. [[CrossRef](#)]
39. Voolstra, C.R.; Li, Y.; Liew, Y.J.; Baumgarten, S.; Zoccola, D.; Flot, J.-F.; Tambutté, S.; Allemand, D.; Aranda, M. Comparative Analysis of the Genomes of *Stylophora pistillata* and *Acropora digitifera* Provides Evidence for Extensive Differences between Species of Corals. *Sci. Rep.* **2017**, *7*, 17583. [[CrossRef](#)]
40. Santos, M.E.A.; Kise, H.; Julie Loïs Fourreau, C.; Polisenio, A.; Pirro, S.; Reimer, J.D. The Complete Genome Sequences of 13 Species of Brachycnemina (Cnidaria, Hexacorallia, Anthozoa, Zoantharia). *Biodivers. J.* **2023**, *2023*. [[CrossRef](#)]
41. Takekata, H.; Hamazato, H.; Suan, T.E.; Izumi, R.; Yaguchi, H.; Matsunami, M.; Isomura, N.; Takemura, A. Transcriptome Analysis in a Scleractinian Coral, *Acropora tenuis*, during the Spawning Season With Reference to the Gonadal Condition. *Zool. Sci.* **2022**, *39*, 570–580. [[CrossRef](#)]
42. Liu, Y.; Liao, X.; Han, T.; Su, A.; Guo, Z.; Lu, N.; He, C.; Lu, Z. Full-Length Transcriptome Sequencing of the Scleractinian Coral *Montipora foliosa* Reveals the Gene Expression Profile of Coral–Zooxanthellae Holobiont. *Biology* **2021**, *10*, 1274. [[CrossRef](#)] [[PubMed](#)]
43. Ryu, T.; Cho, W.; Yum, S.; Woo, S. Holobiont Transcriptome of Colonial Scleractinian Coral *Alveopora japonica*. *Mar. Genom.* **2019**, *43*, 68–71. [[CrossRef](#)]
44. Ryu, T.; Hwang, S.-J.; Woo, S. Transcriptome Assemblies of Two Deep-Sea Octocorals *Calyptrophora lyra* and *Chrysogorgia stellata* from West Pacific Seamount, Godin Guyot. *Mar. Genom.* **2023**, *67*, 101006. [[CrossRef](#)] [[PubMed](#)]
45. Chiu, Y.-L.; Shikina, S.; Yoshioka, Y.; Shinzato, C.; Chang, C.-F. De Novo Transcriptome Assembly from the Gonads of a Scleractinian Coral, *Euphyllia ancora*: Molecular Mechanisms Underlying Scleractinian Gametogenesis. *BMC Genom.* **2020**, *21*, 732. [[CrossRef](#)] [[PubMed](#)]
46. Ledoux, J.-B.; Cruz, F.; Gómez-Garrido, J.; Antoni, R.; Blanc, J.; Gómez-Gras, D.; Kipson, S.; López-Sendino, P.; Antunes, A.; Linares, C.; et al. The Genome Sequence of the Octocoral *Paramuricea clavata*—A Key Resource To Study the Impact of Climate Change in the Mediterranean. *G3 (Bethesda)* **2020**, *10*, 2941–2952. [[CrossRef](#)]
47. Macleod, K.L.; Paris, J.R.; Jenkins, T.L.; Stevens, J.R. The First Genome of the Cold-Water Octocoral, the Pink Sea Fan, *Eunicella verrucosa*. *Genome Biol. Evol.* **2023**, *15*, evad083. [[CrossRef](#)]
48. Ryu, T.; Cho, I.-Y.; Hwang, S.-J.; Yum, S.; Kim, M.-S.; Woo, S. First Transcriptome Assembly of the Temperate Azooxanthellate Octocoral *Eleutherobia rubra*. *Mar. Genom.* **2019**, *48*, 100682. [[CrossRef](#)]

49. Khalturin, K.; Shinzato, C.; Khalturina, M.; Hamada, M.; Fujie, M.; Koyanagi, R.; Kanda, M.; Goto, H.; Anton-Erxleben, F.; Toyokawa, M.; et al. Medusozoan Genomes Inform the Evolution of the Jellyfish Body Plan. *Nat. Ecol. Evol.* **2019**, *3*, 811–822. [[CrossRef](#)]
50. Guzman, C.; Shinzato, C.; Lu, T.-M.; Conaco, C. Transcriptome Analysis of the Reef-Building Octocoral, *Heliopora coerulea*. *Sci. Rep.* **2018**, *8*, 8397. [[CrossRef](#)]
51. Stewart, F.; Dmytrenko, O.; DeLong, E.; Cavanaugh, C. Metatranscriptomic Analysis of Sulfur Oxidation Genes in the Endosymbiont of *Solemya velum*. *Front. Microbiol.* **2011**, *2*, 134. [[CrossRef](#)]
52. Smith, S.A.; Wilson, N.G.; Goetz, F.E.; Feehery, C.; Andrade, S.C.S.; Rouse, G.W.; Giribet, G.; Dunn, C.W. Resolving the Evolutionary Relationships of Molluscs with Phylogenomic Tools. *Nature* **2011**, *480*, 364–367. [[CrossRef](#)] [[PubMed](#)]
53. Lemer, S.; Bieler, R.; Giribet, G. Resolving the Relationships of Clams and Cockles: Dense Transcriptome Sampling Drastically Improves the Bivalve Tree of Life. *Proc. R. Soc. B Biol. Sci.* **2019**, *286*, 20182684. [[CrossRef](#)]
54. Bao, Y.; Zeng, Q.; Wang, J.; Zhang, Z.; Zhang, Y.; Wang, S.; Wong, N.-K.; Yuan, W.; Huang, Y.; Zhang, W.; et al. Genomic Insights into the Origin and Evolution of Molluscan Red-Bloodedness in the Blood Clam *Tegillarca granosa*. *Mol. Biol. Evol.* **2021**, *38*, 2351–2365. [[CrossRef](#)] [[PubMed](#)]
55. Gerdol, M.; Moreira, R.; Cruz, F.; Gómez-Garrido, J.; Vlasova, A.; Rosani, U.; Venier, P.; Naranjo-Ortiz, M.A.; Murgarella, M.; Greco, S.; et al. Massive Gene Presence-Absence Variation Shapes an Open Pan-Genome in the Mediterranean Mussel. *Genome Biol.* **2020**, *21*, 275. [[CrossRef](#)] [[PubMed](#)]
56. Gallardo-Escárate, C.; Valenzuela-Muñoz, V.; Nuñez-Acuña, G.; Valenzuela-Miranda, D.; Tapia, F.J.; Yévenes, M.; Gajardo, G.; Toro, J.E.; Oyarzún, P.A.; Arriagada, G.; et al. Chromosome-Level Genome Assembly of the Blue Mussel *Mytilus chilensis* Reveals Molecular Signatures Facing the Marine Environment. *Genes* **2023**, *14*, 876. [[CrossRef](#)]
57. Gualandi, N.; Fracrossi, D.; Riommi, D.; Sollitto, M.; Greco, S.; Mardirossian, M.; Pacor, S.; Hori, T.; Pallavicini, A.; Gerdol, M. Unveiling the Impact of Gene Presence/Absence Variation in Driving Inter-Individual Sequence Diversity within the CRP-I Gene Family in *Mytilus* spp. *Genes* **2023**, *14*, 787. [[CrossRef](#)]
58. Yang, J.-L.; Feng, D.-D.; Liu, J.; Xu, J.-K.; Chen, K.; Li, Y.-F.; Zhu, Y.-T.; Liang, X.; Lu, Y. Chromosome-Level Genome Assembly of the Hard-Shell Mussel *Mytilus coruscus*, a Widely Distributed Species from the Temperate Areas of East Asia. *Gigascience* **2021**, *10*, giab024. [[CrossRef](#)]
59. Paggeot, L.X.; DeBiasse, M.B.; Escalona, M.; Fairbairn, C.; Marimuthu, M.P.A.; Nguyen, O.; Sahasrabudhe, R.; Dawson, M.N. Reference Genome for the California Ribbed Mussel, *Mytilus californianus*, an Ecosystem Engineer. *J. Hered.* **2022**, *113*, 681–688. [[CrossRef](#)]
60. Jakób, M.; Lubkowski, J.; O’Keefe, B.R.; Wlodawer, A. Structure of a Lectin from the Sea Mussel *Crenomytilus grayanus* (CGL). *Acta Crystallogr. F Struct. Biol. Commun.* **2015**, *71*, 1429–1436. [[CrossRef](#)]
61. Inoue, K.; Yoshioka, Y.; Tanaka, H.; Kinjo, A.; Sassa, M.; Ueda, I.; Shinzato, C.; Toyoda, A.; Itoh, T. Genomics and Transcriptomics of the Green Mussel Explain the Durability of Its Byssus. *Sci. Rep.* **2021**, *11*, 5992. [[CrossRef](#)]
62. Nicastro, K.R.; Pearson, G.A.; Ramos, X.; Pearson, V.; McQuaid, C.D.; Zardi, G.I. Transcriptome Wide Analyses Reveal Intraspecific Diversity in Thermal Stress Responses of a Dominant Habitat-forming Species. *Sci. Rep.* **2023**, *13*, 5645. [[CrossRef](#)]
63. Erlenbach, T.R.; Wares, J.P. Latitudinal Variation and Plasticity in Response to Temperature in *Geukensia demissa*. *Ecol. Evol.* **2023**, *13*, e9856. [[CrossRef](#)]
64. Wang, S.; Zhang, J.; Jiao, W.; Li, J.; Xun, X.; Sun, Y.; Guo, X.; Huan, P.; Dong, B.; Zhang, L.; et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat. Ecol. Evol.* **2017**, *1*, 0120. [[CrossRef](#)] [[PubMed](#)]
65. Kenny, N.J.; McCarthy, S.A.; Dudchenko, O.; James, K.; Betteridge, E.; Corton, C.; Doluacan, J.; Mead, D.; Oliver, K.; Omer, A.D.; et al. The Gene-Rich Genome of the Scallop *Pecten maximus*. *Gigascience* **2020**, *9*, gaaa037. [[CrossRef](#)]
66. Li, C.; Liu, X.; Liu, B.; Ma, B.; Liu, F.; Liu, G.; Shi, Q.; Wang, C. Draft Genome of the Peruvian Scallop *Argopecten purpuratus*. *Gigascience* **2018**, *7*, giy031. [[CrossRef](#)]
67. Liu, X.; Li, C.; Chen, M.; Liu, B.; Yan, X.; Ning, J.; Ma, B.; Liu, G.; Zhong, Z.; Jia, Y.; et al. Draft Genomes of Two Atlantic Bay Scallop Subspecies *Argopecten irradians irradians* and *A. i. concentricus*. *Sci. Data* **2020**, *7*, 99. [[CrossRef](#)]
68. The Genome Sequence of the Variegated Scallop, *Mimachlamys varia*. Available online: <https://wellcomeopenresearch.org/articles/8-307> (accessed on 25 August 2023).
69. Greco, S.; Voltarel, G.; Gaetano, A.S.; Manfrin, C.; Pallavicini, A.; Giulianini, P.G.; Gerdol, M. Comparative Transcriptomic Analysis Reveals Adaptive Traits in Antarctic Scallop *Adamussium colbecki*. *Fishes* **2023**, *8*, 276. [[CrossRef](#)]
70. Galindo-Torres, P.; Abreu-Goodger, C.; Llera-Herrera, R.; Escobedo-Fregoso, C.; García-Gasca, A.; Ibarra, A.M. Triploid-Induced Complete Sterility in the Scallop *Nodipecten subnodosus* Might Be Triggered by an Early and Sustained DNA Damage Response. *Aquaculture* **2022**, *559*, 738422. [[CrossRef](#)]
71. Plachetzki, D.C.; Pankey, M.S.; MacManes, M.D.; Lesser, M.P.; Walker, C.W. The Genome of the Softshell Clam *Mya arenaria* and the Evolution of Apoptosis. *Genome Biol. Evol.* **2020**, *12*, 1681–1693. [[CrossRef](#)] [[PubMed](#)]
72. McCartney, M.A.; Auch, B.; Kono, T.; Mallez, S.; Zhang, Y.; Obille, A.; Becker, A.; Abrahante, J.E.; Garbe, J.; Badalamenti, J.P.; et al. The Genome of the Zebra Mussel, *Dreissena polymorpha*: A Resource for Comparative Genomics, Invasion Genetics, and Biocontrol. *G3 Genes | Genomes | Genet.* **2022**, *12*, jkab423. [[CrossRef](#)]

73. Farhat, S.; Bonnivard, E.; Pales Espinosa, E.; Tanguy, A.; Boutet, I.; Guiglielmoni, N.; Flot, J.-F.; Allam, B. Comparative Analysis of the *Mercenaria mercenaria* Genome Provides Insights into the Diversity of Transposable Elements and Immune Molecules in Bivalve Mollusks. *BMC Genom.* **2022**, *23*, 192. [[CrossRef](#)] [[PubMed](#)]
74. Sun, J.; Chen, C.; Miyamoto, N.; Li, R.; Sigwart, J.D.; Xu, T.; Sun, Y.; Wong, W.C.; Ip, J.C.H.; Zhang, W.; et al. The Scaly-Foot Snail Genome and Implications for the Origins of Biomineralised Armour. *Nat. Commun.* **2020**, *11*, 1657. [[CrossRef](#)] [[PubMed](#)]
75. Lan, Y.; Sun, J.; Chen, C.; Sun, Y.; Zhou, Y.; Yang, Y.; Zhang, W.; Li, R.; Zhou, K.; Wong, W.C.; et al. Hologenome Analysis Reveals Dual Symbiosis in the Deep-Sea Hydrothermal Vent Snail *Gigantopelta aegis*. *Nat. Commun.* **2021**, *12*, 1165. [[CrossRef](#)] [[PubMed](#)]
76. Fletcher, C.; Pereira da Conceicao, L.; Natural History Museum Genome Acquisition Lab; Darwin Tree of Life Barcoding Collective; Wellcome Sanger Institute Tree of Life Programme; Wellcome Sanger Institute Scientific Operations: DNA Pipelines Collective; Tree of Life Core Informatics Collective; Darwin Tree of Life Consortium. The Genome Sequence of the King Ragworm, *Alitta virens* (Sars, 1835). *Wellcome Open Res.* **2023**, *8*, 297. [[CrossRef](#)] [[PubMed](#)]
77. Lv, F.; Wang, T.; Liu, F.; Yu, Y.; Qiao, G.; Lv, L.; Wang, Z.; Qi, Z. *De Novo* Assembly and Characterization of Transcriptome in Somatic Muscles of the Polychaete *Perinereis aibuhitensis*. *J. Coast. Res.* **2017**, *33*, 931–937. [[CrossRef](#)]
78. Chou, H.-C.; Acevedo-Luna, N.; Kuhlman, J.A.; Schneider, S.Q. PdmBase: A Transcriptome Database and Research Tool for *Platynereis dumerilii* and Early Development of Other Metazoans. *BMC Genom.* **2018**, *19*, 618. [[CrossRef](#)] [[PubMed](#)]
79. Khalturin, K.; Shunatova, N.; Shchenkov, S.; Sasakura, Y.; Kawamitsu, M.; Satoh, N. Polyzoa Is Back: The Effect of Complete Gene Sets on the Placement of Ectoprocta and Entoprocta. *Sci. Adv.* **2022**, *8*, eabo4400. [[CrossRef](#)] [[PubMed](#)]
80. Shao, C.; Sun, S.; Liu, K.; Wang, J.; Li, S.; Liu, Q.; Deagle, B.E.; Seim, I.; Biscontin, A.; Wang, Q.; et al. The Enormous Repetitive Antarctic Krill Genome Reveals Environmental Adaptations and Population Insights. *Cell* **2023**, *186*, 1279–1294.e19. [[CrossRef](#)]
81. Blanco-Bercial, L.; Maas, A.E. A Transcriptomic Resource for the Northern Krill *Meganyctiphanes norvegica* Based on a Short-Term Temperature Exposure Experiment. *Mar. Genom.* **2018**, *38*, 25–32. [[CrossRef](#)]
82. Delroisse, J.; Ullrich-Lüter, E.; Ortega-Martinez, O.; Dupont, S.; Arnone, M.-I.; Mallefet, J.; Flammang, P. High Opsin Diversity in a Non-Visual Infaunal Brittle Star. *BMC Genom.* **2014**, *15*, 1035. [[CrossRef](#)]
83. Fuad, M.T.I.; Shi, W.; Liao, X.; Li, Y.; Sharifuzzaman, S.M.; Zhang, X.; Liu, X.; Xu, Q. Transcriptomic Response of Intertidal Brittle Star *Ophiothrix exigua* to Seasonal Variation. *Mar. Genom.* **2022**, *64*, 100957. [[CrossRef](#)] [[PubMed](#)]
84. Kudtarkar, P.; Cameron, R.A. Echinobase: An Expanding Resource for Echinoderm Genomic Information. *Database* **2017**, *2017*, bax074. [[CrossRef](#)] [[PubMed](#)]
85. Mashanov, V.; Machado, D.J.; Reid, R.; Brouwer, C.; Kofsky, J.; Janies, D.A. Twinkle Twinkle Brittle Star: The Draft Genome of *Ophioderma brevispinum* (Echinodermata: Ophiuroidea) as a Resource for Regeneration Research. *BMC Genom.* **2022**, *23*, 574. [[CrossRef](#)] [[PubMed](#)]
86. Virgili, R.; Polisenio, A.; Pirro, S.; Reimer, J.D. A Draft Genome Sequence of *Dorometra sesokonis* (Antedonidae, Comatulida, Crinoidea, Echinodermata). *Biodivers. Genomes* **2023**, *2023*. [[CrossRef](#)]
87. Borme, D.; Legovini, S.; de Olazabal, A.; Tirelli, V. Diet of Adult Sardine *Sardina pilchardus* in the Gulf of Trieste, Northern Adriatic Sea. *J. Mar. Sci. Eng.* **2022**, *10*, 1012. [[CrossRef](#)]
88. Hasan, I.; Gerdol, M.; Fujii, Y.; Rajia, S.; Koide, Y.; Yamamoto, D.; Kawsar, S.M.A.; Ozeki, Y. cDNA and Gene Structure of MytiLec-1, A Bacteriostatic R-Type Lectin from the Mediterranean Mussel (*Mytilus galloprovincialis*). *Mar. Drugs* **2016**, *14*, 92. [[CrossRef](#)]
89. Wilmsen, H.U.; Pattus, F.; Buckley, J.T. Aerolysin, a Hemolysin from *Aeromonas hydrophila*, Forms Voltage-Gated Channels in Planar Lipid Bilayers. *J. Membr. Biol.* **1990**, *115*, 71–81. [[CrossRef](#)]
90. Henricson, A.; Forslund, K.; Sonnhammer, E.L. Orthology Confers Intron Position Conservation. *BMC Genom.* **2010**, *11*, 412. [[CrossRef](#)]
91. Fedorov, A.; Merican, A.F.; Gilbert, W. Large-Scale Comparison of Intron Positions among Animal, Plant, and Fungal Genes. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16128–16133. [[CrossRef](#)]
92. Rogozin, I.B.; Wolf, Y.I.; Sorokin, A.V.; Mirkin, B.G.; Koonin, E.V. Remarkable Interkingdom Conservation of Intron Positions and Massive, Lineage-Specific Intron Loss and Gain in Eukaryotic Evolution. *Curr. Biol.* **2003**, *13*, 1512–1517. [[CrossRef](#)]
93. Keilwagen, J.; Wenk, M.; Erickson, J.L.; Schattat, M.H.; Grau, J.; Hartung, F. Using Intron Position Conservation for Homology-Based Gene Prediction. *Nucleic Acids Res.* **2016**, *44*, e89. [[CrossRef](#)] [[PubMed](#)]
94. Kondrashov, F.A.; Koonin, E.V.; Morgunov, I.G.; Finogenova, T.V.; Kondrashova, M.N. Evolution of Glyoxylate Cycle Enzymes in Metazoa: Evidence of Multiple Horizontal Transfer Events and Pseudogene Formation. *Biol. Direct* **2006**, *1*, 31. [[CrossRef](#)] [[PubMed](#)]
95. Gerdol, M.; Sollitto, M.; Pallavicini, A.; Castellano, I. The Complex Evolutionary History of Sulfoxide Synthase in Ovothiol Biosynthesis. *Proc. R. Soc. B Biol. Sci.* **2019**, *286*, 20191812. [[CrossRef](#)] [[PubMed](#)]
96. Hunter, S.; Apweiler, R.; Attwood, T.K.; Bairoch, A.; Bateman, A.; Binns, D.; Bork, P.; Das, U.; Daugherty, L.; Duquenne, L.; et al. InterPro: The Integrative Protein Signature Database. *Nucleic Acids Res.* **2009**, *37*, D211–D215. [[CrossRef](#)] [[PubMed](#)]
97. Dawson, N.L.; Lewis, T.E.; Das, S.; Lees, J.G.; Lee, D.; Ashford, P.; Orengo, C.A.; Sillitoe, I. CATH: An Expanded Resource to Predict Protein Function through Structure and Sequence. *Nucleic Acids Res.* **2017**, *45*, D289–D295. [[CrossRef](#)] [[PubMed](#)]
98. Notova, S.; Bonnardel, F.; Rosato, F.; Siukstaite, L.; Schwaiger, J.; Lim, J.H.; Bovin, N.; Varrot, A.; Ogawa, Y.; Römer, W.; et al. The Choanoflagellate Pore-Forming Lectin SaroL-1 Punches Holes in Cancer Cells by Targeting the Tumor-Related Glycosphingolipid Gb3. *Commun. Biol.* **2022**, *5*, 954. [[CrossRef](#)]

99. Söding, J.; Biegert, A.; Lupas, A.N. The HHpred Interactive Server for Protein Homology Detection and Structure Prediction. *Nucleic Acids Res.* **2005**, *33*, W244–W248. [[CrossRef](#)]
100. Best, H.L.; Williamson, L.J.; Lipka-Lloyd, M.; Waller-Evans, H.; Lloyd-Evans, E.; Rizkallah, P.J.; Berry, C. The Crystal Structure of *Bacillus thuringiensis* Tpp80Aa1 and Its Interaction with Galactose-Containing Glycolipids. *Toxins* **2022**, *14*, 863. [[CrossRef](#)]
101. Colletier, J.-P.; Sawaya, M.R.; Gingery, M.; Rodriguez, J.A.; Cascio, D.; Brewster, A.S.; Michels-Clark, T.; Hice, R.H.; Coquelle, N.; Boutet, S.; et al. Mosquito Larvicide BinAB Revealed by de Novo Phasing with an X-Ray Laser. *Nature* **2016**, *539*, 43–47. [[CrossRef](#)]
102. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
103. Li, W.; Godzik, A. Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [[CrossRef](#)] [[PubMed](#)]
104. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. [[CrossRef](#)] [[PubMed](#)]
105. Reese, M.G.; Eeckman, F.H.; Kulp, D.; Haussler, D. Improved Splice Site Detection in Genie. *J. Comput. Biol.* **1997**, *4*, 311–323. [[CrossRef](#)] [[PubMed](#)]
106. Teufel, F.; Almagro Armenteros, J.J.; Johansen, A.R.; Gíslason, M.H.; Pihl, S.I.; Tsirigos, K.D.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 6.0 Predicts All Five Types of Signal Peptides Using Protein Language Models. *Nat. Biotechnol.* **2022**, *40*, 1023–1025. [[CrossRef](#)] [[PubMed](#)]
107. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-Scale Protein Function Classification. *Bioinformatics* **2014**, *30*, 1236–1240. [[CrossRef](#)]
108. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [[CrossRef](#)]
109. Edgar, R.C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)] [[PubMed](#)]
110. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermini, L.S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* **2017**, *14*, 587–589. [[CrossRef](#)]
111. Whelan, S.; Goldman, N. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Mol. Biol. Evol.* **2001**, *18*, 691–699. [[CrossRef](#)]
112. Neath, A.A.; Cavanaugh, J.E. The Bayesian Information Criterion: Background, Derivation, and Applications. *WIREs Comput. Stat.* **2012**, *4*, 199–203. [[CrossRef](#)]
113. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [[CrossRef](#)]
114. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER Web Server: Interactive Sequence Similarity Searching. *Nucl. Acids Res.* **2011**, *39*, W29–W37. [[CrossRef](#)]
115. Chernikov, O.; Kuzmich, A.; Chikalovets, I.; Molchanova, V.; Hua, K.-F. Lectin CGL from the Sea Mussel *Crenomytilus grayanus* Induces Burkitt’s Lymphoma Cells Death via Interaction with Surface Glycan. *Int. J. Biol. Macromol.* **2017**, *104*, 508–514. [[CrossRef](#)] [[PubMed](#)]
116. UniProt Consortium. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [[CrossRef](#)] [[PubMed](#)]
117. Gerdol, M.; Schmitt, P.; Venier, P.; Rocha, G.; Rosa, R.D.; Destoumieux-Garzón, D. Functional Insights From the Evolutionary Diversification of Big Defensins. *Front. Immunol.* **2020**, *11*, 758. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.