




Article

Divergences of the *RLR* Gene Families across Lophotrochozoans: Domain Grafting, Exon–Intron Structure, Expression, and Positive Selection

Shanshan Yao ^{1,2,3}, Jiulin Chan ^{2,3}, Yue Xu ^{2,3}, Shimei Wu ¹ and Linlin Zhang ^{2,3,4,*} 

¹ College of Life Sciences, Qingdao University, Qingdao 266071, China; 17805421231@163.com (S.Y.); shimeiwu2016@126.com (S.W.)

² CAS and Shandong Province Key Laboratory of Experimental Marine Biology & Center of Deep Sea Research, Center for Ocean Mega-Science, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China; chanjiulin@qdio.ac.cn (J.C.); xuyue@qdio.ac.cn (Y.X.)

³ Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071, China

⁴ College of Marine Science, University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: linlinzhang@qdio.ac.cn

Abstract: Invertebrates do not possess adaptive immunity but have evolved a variety of unique repertoires of innate immune sensors. In this study, we explored the immune diversity and specificity of invertebrates based on the lophotrochozoan *RLRs*, a major component in antiviral immune recognition. By annotating *RLRs* in the genomes of 58 representative species across metazoan evolution, we explored the gene expansion of *RLRs* in Lophotrochozoa. Of note, the N-terminal domains of lophotrochozoan *RLRs* showed the most striking diversity which evolved independently by domain grafting. Exon–intron structures were revealed to be prevalent in the domain grafting of lophotrochozoan *RLRs* based on an analysis of sibling paralogs and orthologs. In more than half of the cases, the mechanism of ‘exonization/pseudoexonization’ led to the generation of non-canonical N-terminal domains. Transcriptomic studies revealed that many non-canonical *RLRs* display immune-related expression patterns. Two of these *RLRs* showed obvious evidence of positive selection, which may be the result of host defense selection pressure. Overall, our study suggests that the complex and unique domain arrangement of lophotrochozoan *RLRs* might result from domain grafting, exon–intron divergence, expression diversification, and positive selection, which may have led to functionally distinct lophotrochozoan *RLRs*.

Keywords: *RIG-I*-like receptor; lophotrochozoan; innate immune; molecular evolution; gene expression; domain grafting; exon–intron structure; positive selection



Citation: Yao, S.; Chan, J.; Xu, Y.; Wu, S.; Zhang, L. Divergences of the *RLR* Gene Families across

Lophotrochozoans: Domain Grafting, Exon–Intron Structure, Expression, and Positive Selection. *Int. J. Mol. Sci.* **2022**, *23*, 3415. <https://doi.org/10.3390/ijms23073415>

Academic Editor: Manickam Sugumaran

Received: 29 January 2022

Accepted: 10 March 2022

Published: 22 March 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

The immune system has long been known for its remarkably rapid evolution due to strong selective drivers, such as fast-evolving pathogens [1,2]. Lophotrochozoa is one of the most species-rich superphyla, including mollusks, segmented worms, and other invertebrates. Despite lacking adaptive immunity, these animals exhibit a high level of biodiversity and a wide range of ecological adaptations, including adaptation to a complex pathogenic environment [3]. Innate immunity is not only the main defense mechanism of crown rotors against pathogens, but also the most conservative defense mechanism of multicellular organisms against pathogens [4,5]. Innate immunity is characterized by a rapid, nonspecific response to infection and injury. Invertebrates lacking immunoglobulin-mediated adaptive immunity have evolved a variety of broad, unique, and complex repertoires of innate immune sensors [6,7]. An alternative strategy for invertebrates is the large-scale expansion and diversification of multigene families encoding genome-encoded pattern recognition receptors (PRRs) [8–10].

Animals distinguish self from non-self by using a variety of PRRs [11,12]. PRRs interact with pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs) [13–15]. PRR-binding events typically trigger signaling cascades that lead to the transcription of immune effector genes encoding products such as antibacterial and antiviral proteins, leading to the production of proinflammatory cytokines and the death of host cells [16,17]. PRRs are divided into the following major categories: cytoplasmic NOD-like receptors (*NLRs*), membrane-bound Toll-like receptors (*TLRs*), C-type lectin receptors (*CLRs*), scavenger receptors, AIM2-like receptors (*ALRs*), and RIG-I-like receptors (*RLRs*) [14].

RLRs are a family of three DExD/H box-containing RNA helicases [18]. In vertebrates, *RLRs* include retinoic acid-inducible gene I (*RIG-I*), melanoma differentiation-associated gene 5 (*MDA5*), and laboratory of genetics and physiology 2 (*LGP2*) [19]. *MDA5* and *RIG-I* contain two N-terminal caspase (cysteine-dependent aspartate-specific proteases) recruitment domains (CARDs), followed by an intermediate DEAD/DEAH box helicase domain (DEXDc + HELICc domains) and a C-terminal RIG-I_C-RD domain [20]. Helicases use the energy provided by the hydrolysis of ATP to catalyze the unwinding of nucleic acid duplexes. CARD is a structural domain composed of amphipathic α -helices and is predicted to function through protein–protein interactions with apoptotic and antiapoptotic proteins. In contrast, *LGP2* lacks CARDs and retains the DEAD/DEAH box helicase domain and RIG-I_C-RD domain. The simultaneous presence of CARD and RNA helicase domains in a single molecule, as observed in *MDA5* and *RIG-I*, is unique [21]. According to previous studies, this unique mechanism for preventing viral infection is caused by the shuffling of domains [21]. Recent studies have shown that the domain organization of *RLRs* in early animals differs from that in vertebrates. Four types of structural compositions of *RLR* proteins are found in the brachiopod *Lingula anatina* [22]. The N-terminus of these *LanRLRs* presents different domains, including a death effector domain (DED) and caspase catalytic (CASC) domains, in addition to CARD. Only one CARD has been found at the N-terminus of the *RLRs* in cnidarians [23]. However, the evolutionary lineage of *RLRs* across all metazoans is unknown, especially the evolutionary pattern of their domain architecture.

As one of the main cytoplasmic PRRs, *RLRs* are responsible for the intracellular dsRNA sensor induced by type I IFN [20,24]. An analysis of *MDA5* or *RIG-I* knockout mice revealed that this TLR-independent pathway is central to innate immunity against viral infection [25–27]. Moreover, as both *MDA5* and *RIG-I* are IFN-stimulated genes, a positive feedback loop that generates a potent antiviral state is established [28,29]. The prototypical *RLRs* contain three types of functional domains: the two CARD domains at the N-terminus that are responsible for downstream signaling transduction, the DEXDc domain and the HELICc domain in the center that are responsible for RNA recognition, and the C-terminal RIG-I_C-RD domain that assists in pathogen recognition by binding to specific viral RNAs [30]. *MDA5* and *RIG-I* interact with the CARD domain of the mitochondrial protein IFN- β promoter stimulator-1 (*IPS-1*, also known as *MAVS*, *VISA*, and *CARDIF*). Thereafter, TNF-receptor-associated factor-3 (*TRAF-3*) is recruited, and *TRAF* family member-associated NF- κ B-activator binding kinase-1 (*TBK1*) and inducible I κ B kinase (*IKK*) are activated [31–33]. These kinases phosphorylate *IRF-3* and *IRF-7* and activate NF- κ B, which translocates to the nucleus to induce type I IFN expression [34]. *MAVS*-dependent *RLR* signaling has been shown to be involved in viral immunity in mollusks [35]. Further, as *LGP2* lacks CARDs, it does not trigger immune responses; however, it can regulate *RIG-I* and *MDA5* signaling [36,37]. The regulatory function of *LGP2* is attributed to its retained helicase domain and RD [38,39].

Evolutionary studies have painted a complex picture of how *RLRs* emerge, and their functional diversity. Previous studies suggested that full-length *RLRs* are a vertebrate-specific evolutionary novelty, although the building blocks of *RLRs* may have been present in closely related prevertebrate animals [21,40]. These studies differ in the ordering of gene duplication events leading to the three *RLRs* present in mammals; however, the studies ultimately concluded that *RLR* evolution was driven by a complex series of domain grafting

and gene fusion events in CARD domains. Based on recent studies, the *RLR*-based immune system is not vertebrate-specific but originates in the earliest multicellular animals [41]. *RLRs* functionally diversified through a series of gene duplication events followed by protein coding changes, which modulate the RNA-binding properties of different *RLRs* by altering key contact residues within the C-terminal RD [41]. There is strong evidence that *RLRs* are involved in a long-term evolutionary arms race with viral RNA molecules, suggesting that the structure of viral RNA may have shaped the evolution of animal innate immunity [41]. Although there is no consensus on the exact evolutionary history of *RLRs*, these receptors (and/or their building blocks) have been reported to originate early in metazoan evolution [41,42] and, most importantly, are evolutionary hot spots [42]; this is illustrated by the lineage-specific loss of *RLR* genes in many species. For example, most mammals possess *RIG-I*; however, *RIG-I* is lost in at least one mammalian species, the Chinese tree shrew [43]. Interestingly, with the loss of *RIG-I*, both *MDA5* and *LGP2* have undergone strong positive selection in Chinese tree shrews, and positively selected sites in *MDA5* endowed the substitute function for the lost *RIG-I* [44]. Adaptive evolution analysis of the *RLR* gene family in reptiles, birds, and other chordates also revealed that purification selection was dominant [45,46]. However, an overview of large-scale metazoan macroevolution is lacking.

Here, we explored the immune diversity and specificity of the *RLR* gene family in lophotrochozoans. Briefly, we annotated 227 *RLR* genes from 58 species across the metazoan phylogeny, with emphasis on the molecular evolutionary dynamics of *RLRs* in lophotrochozoans. Further, we explored their phylogeny and domain composition, which revealed that lophotrochozoan *RLRs* exhibited highly diverse and complex N-terminal domain integration. We also established an exon–intron structure to investigate the molecular mechanism underlying domain grafting. Evolutionary selection signals and tissue and infective gene expression levels were also calculated. Overall, a comprehensive molecular evolution analysis of the *RLR* gene family, which is not only associated with rapid domain grafting but also with the potential for an immune response by positive selection, was carried out.

2. Results

2.1. Identification of the *RIG-I*-like Receptor Repertoires

To investigate the composition difference in the *RLR* gene family across all metazoans, we first annotated 227 *RLRs* in the genomes of 58 representative species with different evolutionary positions (Supplementary Tables S1 and S2). Two *RLR* genes were identified in the sponge of the phylum Porifera (*Amphimedon queenslandica*), suggesting potential origins of *RLRs* in early metazoans (Figure 1). The *RLR* gene was completely lost in arthropods of the 58 species investigated (Figure 1), which is consistent with the results of previous studies [41]. Two or three *RLRs* in most (88%) of the chordates were annotated, while seven *RLRs* were annotated in *Branchiostoma floridae*. Compared with chordate animals, *RLRs* were extensively expanded in most (60%) lophotrochozoans. Further, 13 *RLRs* in the genome of *Crassostrea gigas*, 12 in *Bathymodiolus platifrons*, and 19 in *Mytilus coruscus* were predicted. The expansion of *RLRs* was also identified in the echinoderm *Strongylocentrotus purpuratus*.

Further genomic distribution exploration found that the expansions of the lophotrochozoan *RLRs* can be attributed to multiple local tandem duplication events (Figure 2 and Supplementary Table S3). The tandem duplication phenomenon is the most noticeable in the bivalve species with extensive expansion of *RLRs*. Specifically, 8 of 13, 6 of 12, and 6 of 19 *RLRs* were found to be linked in tandem arrays in the bivalves *C. gigas*, *B. platifrons*, and *M. coruscus*.

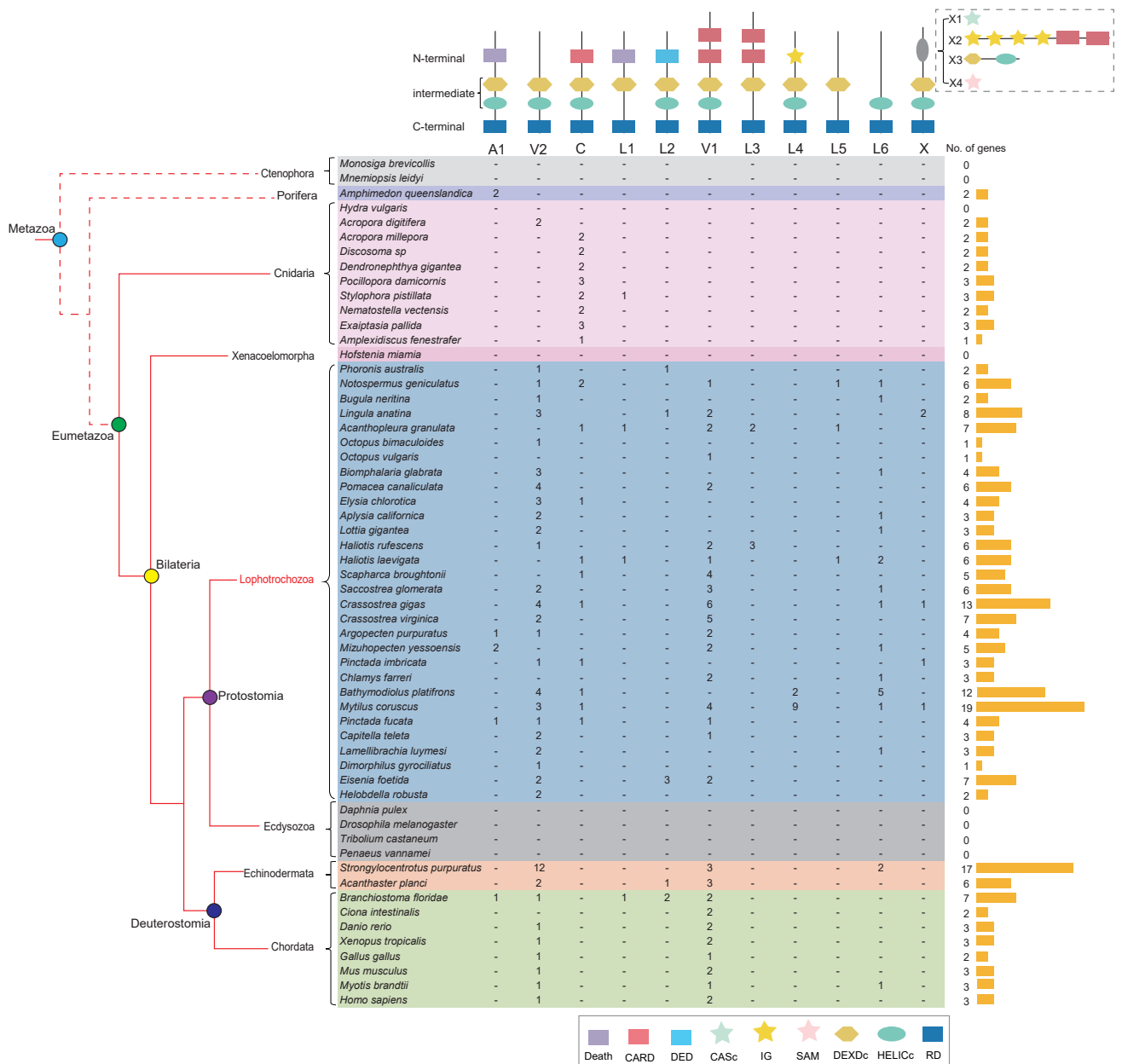


Figure 1. Comparison of gene families encoding RLR immune receptors in representative animals across metazoans. Domain architecture in the lophotrochozoans is more complex than that in other clades. Species colors represent different phyla. Short lines indicate none of the RLRs in the subtype were annotated in this species. Phylogenetic relations among species are indicated by the red cladogram on the left of the table, and dashed lines represent unresolved phylogenetic positions for ctenophores and sponges. The column on the right counts the total number of RLR genes in each species and draws a yellow column chart. RLR diagrams show death family domain in purple, CARD domain in red, DED domain in light blue, CASc domain in light green, IG domain in light yellow, SAM domain in pink, DEXDc domain in yellow, HELICc domain in green, and RD domain in blue. Specifically, the top diagrams show A1, Amphimedon-like type 1; V1, vertebrate-like type 1; V2, vertebrate-like type 2; C, cnidaria-like type; L1, lophotrochozoa-like type 1; L2, lophotrochozoan-like type 2; L3, lophotrochozoan-like type 3; L4, lophotrochozoan-like type 4; L5, lophotrochozoan-like type 5; and L6, lophotrochozoan-like type 6. X implies four specific structural features (X1–X4).

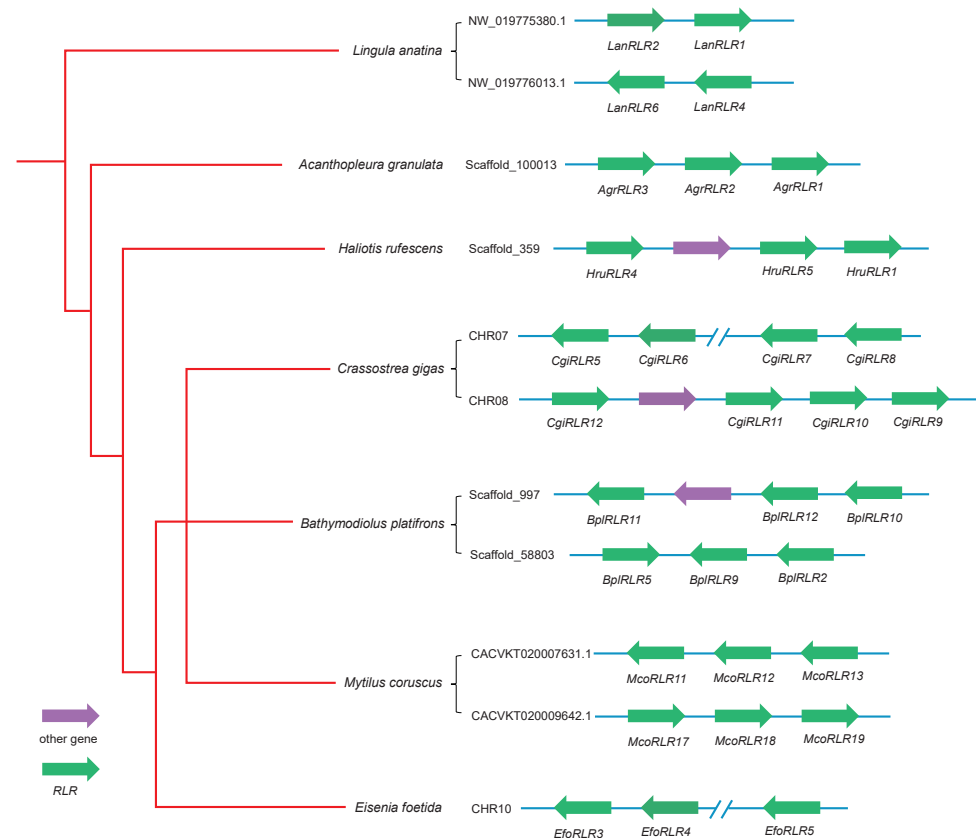


Figure 2. Local tandem duplication of *RLRs* in the seven representative lophotrochozoan species. Green arrowheads indicate *RLR* genes and their transcriptional direction; purple arrowheads indicate other genes. Phylogenetic relations among species are indicated by the red cladogram on the left.

2.2. *RLR* Domain Annotation

Canonical *RLRs* have a typical C-terminal RNA recognition domain (RD) that binds viral RNA and N-terminal CARDs to interact with the signal adaptor. We proceeded to explore the domain architecture of the *RLRs*. With the inclusion of the canonical vertebrate types (V-type), we classified metazoan *RLRs* into 11 major types based on their domain architecture (Figure 1). Most (83%) vertebrates expressed three *RLRs*, including two V1 types (*RIG-I* and *MDA5*) and one V2 type (*LGP2*). In the ancestral branch, Porifera (*A. queenslandica*), two A1-type *RLRs* with death as the N-terminal domain were annotated. Nine of the ten cnidarians were identified to possess two or three *RLRs*, and most (89%) cnidarian anthozoan *RLRs* were C (cnidarian-type), with only one CARD domain in the N-terminus. The types of *RLRs* that only occur in one species are classified as group X, which includes four species (X1–X4). Intriguingly, the domain architecture of *RLRs* showed the most striking diversity in lophotrochozoans, which contain the eleven divided *RLR* types. The N-termini of those eleven *RLR* types include: canonical V1 and V2 types, the A1 type of the death domain, the C type of the N-terminal CARD, the six L types (L1–L6) obtained from lophotrochozoans, and the X type. This observation may reflect a high level of domain grafting, resulting in significant expansion of *RLRs* in lophotrochozoans.

2.3. Phylogenetic Distribution of Three Discrete Domains

To determine the molecular evolutionary history of the *RLR* gene family in these species, we traced the phylogenetic origins of three representative domains, including RIG-I_C-RD, the DEAD/DEAH box helicase domain, and the N-terminal domain. First, a phylogenetic tree was constructed using the C-terminal conserved domain RIG-I_C-RD coding sequences of all metazoan *RLRs* (M). This tree suggested that ancestral *RLRs* duplicated in lophotrochozoans, with no *RIG-I/MDA5/LGP2* divergence detected in this phylum.

Further, most lophotrochozoan *RLRs* were found to be derived from the ancestral *RLR* gene and were divided into two groups (Cluster I and Cluster II) (Figure 3). Cluster I was clustered with the ancestral porifera and vertebrate *RLRs*, indicating that lophotrochozoan *RLRs* in Cluster I might have originated from the same ancestor as those in vertebrates. Three types of N-terminal domains were detected in the *RLRs* of Cluster I, including the canonical CARD and ancestral death type. Compared with lophotrochozoan *RLRs* in Cluster I, diversity and plasticity in the N-terminal domain architecture were observed in *RLRs* in Cluster II. Six of the seven N-terminal domains (Death, CARD, CARD-CARD, (immunoglobulin) IG, DED, (Sterile alpha motif) SAM) in this cluster were annotated. In Cluster II, the *RLRs* with N-terminal IG and DED domains were found to be independently clustered, which suggested that the L2 (DED) and L4 (IG) types of *RLRs* were independent in lophotrochozoans.

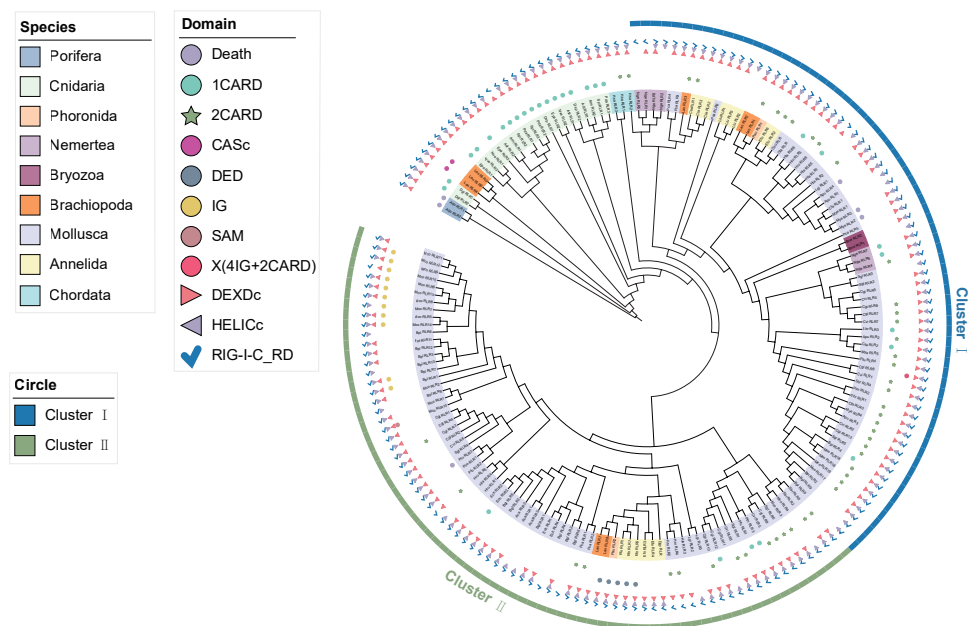


Figure 3. Phylogenetic analysis of *RLRs* based on C-terminal conserved domain RIG-I_C-RD. Maximum likelihood trees were constructed using IQ-TREE. The species used for the tree generation were 30 lophotrochozoans (as shown in Figure 1), the poriferan *A. queenslandica*, cnidaria, and the chordate *Homo sapiens*. The lophotrochozoan *RLRs* can be majorly classified into two divergent classes, Cluster I and Cluster II. The background color on the gene name represents the phylum to which it belongs, as shown in the legend ‘Species’. The different colored shapes in the outer circles represent different types of domains, as shown in the legend ‘Domain’.

We proceeded to investigate the evolutionary history of the *RLR* gene family based on the intermediate helicase domain (DEXDc + HELICc domains) (Figure 4). Generally, the phylogenetic tree topology of the helicase domain is very similar to that of the RIG-I_C-RD phylogenetic tree, indicating that the integration of RIG-I_C-RD and helicase domains originated before the divergence of metazoans. In contrast to the phylogenetic tree based on RIG-I_C-RD, the lophotrochozoan *RLRs* with an N-terminal DED domain belonged to Cluster II in the helicase tree instead of Cluster I. In addition, the *RLRs* with the N-terminal death domain were within Cluster I in the helicase tree but belonged to Cluster II in the RIG-I_C-RD tree. These results support the hypothesis that linked domains of intermediate helicases and RIG-I_C-RD have ancient origins in metazoans; however, fusion events of the two domains in *RLRs* with an N-terminal DED or death domain might independently occur later under certain selection pressures.

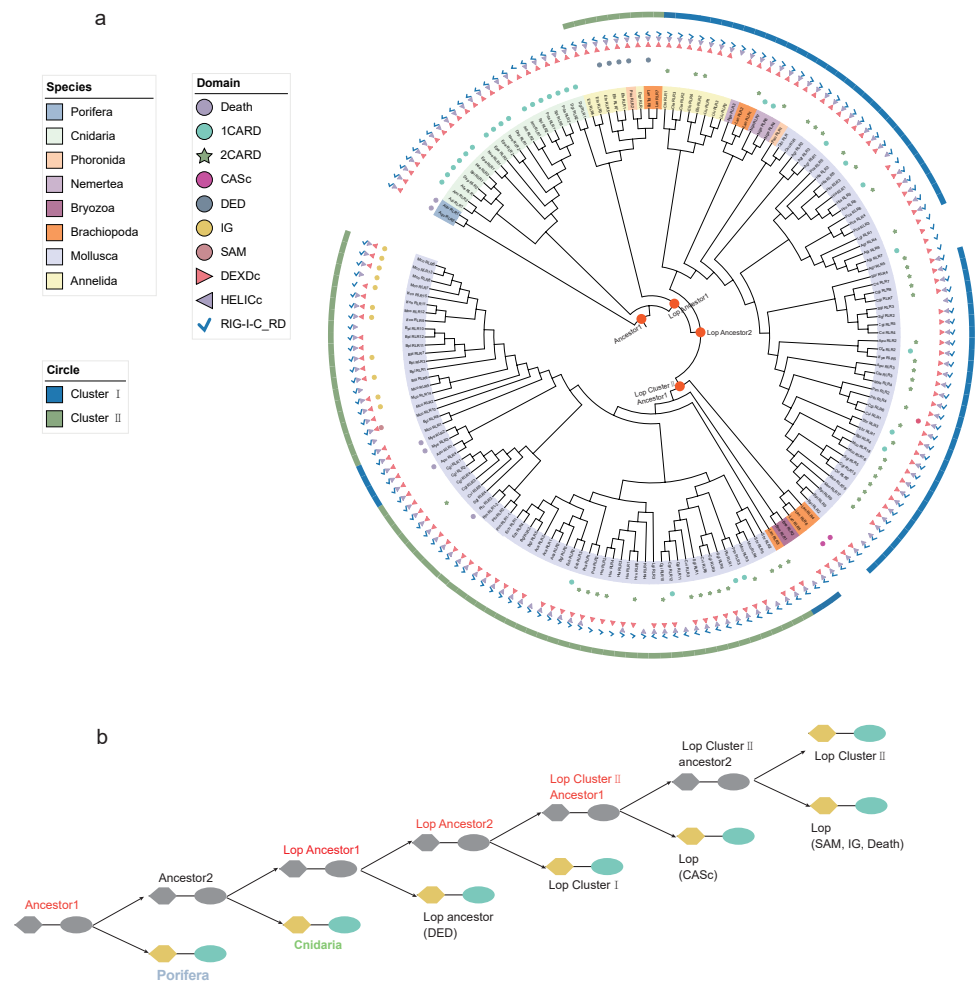


Figure 4. The evolution of intermediate helicase domains (DEXDc + HELICc domains) of the lophotrochozoan *RLRs*. **(a)** Phylogenetic tree constructed with the maximum likelihood method showing the evolution of DEXDc + HELICc domains of *RLRs* in lophotrochozoans. The different background colors of the gene names represent the phyla to which they belong, as shown in the legend ‘Species’. The different colored shapes represent different domains, as shown in the legend ‘Domain’. The blue arc in the outermost circle represents Cluster I, and the green arc represents Cluster II. The orange circles on the branches marked on the tree correspond to the evolutionary divergent nodes as shown in **(b)**. **(b)** A schematic diagram of the evolution of the intermediate helicase domains of lophotrochozoan *RLRs* based on the phylogenetic tree. Lop stands for the lophotrochozoan. Solid black lines indicate duplication events.

To clarify the evolutionary relationship of the diverse N-terminal domains of *RLRs* in lophotrochozoans, we conducted phylogenetic analysis using sequences of N-terminal domains of *RLRs* in lophotrochozoans, the ancient death domain in porifera, and the CARD domain in cnidaria (Figure 5). The two different CARD domains (CARD1 and CARD2) in the typical type of lophotrochozoan *RLRs* were separately extracted and used for phylogenetic analysis. Based on the phylogenetic tree (Figure 5a), the evolutionary histories of N-terminal domains displayed overall different patterns compared to the helicase and RIG-I-C-RD trees, especially for *RLRs* with lophotrochozoan-specific N-terminal domains. Such pheromones suggest the occurrence of multiple independent fusion events in lophotrochozoan *RLRs*. The simultaneous presence of death and helicase domains occurred in porifera, which might be the ancient state in metazoans. Further, lophotrochozoan *RLRs* with the death domain may not be orthologs of the poriferan *RLRs*; instead, these *RLRs* shared the same ancestor with cnidarian *RLRs* containing the CARD domain. Based on the evolutionary history of the CARD domain, cnidarian C-type *RLRs* encode only one CARD

instead of two CARD domains in canonical V1-type *RLRs*. The second CARD domain of lophotrochozoan *RLRs* clustered with the ancestral cnidarian CARD, indicating that the fusion of the CARD1 and CARD2 domains in the common ancestor of the lophotrochozoan ancestor might have occurred through an independent process. Of the N-terminal domains of L (lophotrochozoan)-type *RLRs*, DED and IG clustered with the CARD1 branch, while SAM and CASc clustered with the CARD2 branch. These results indicate that the highly diverse and dynamic N-terminal domains in lophotrochozoan *RLRs* could have independently emerged from domain grafting.

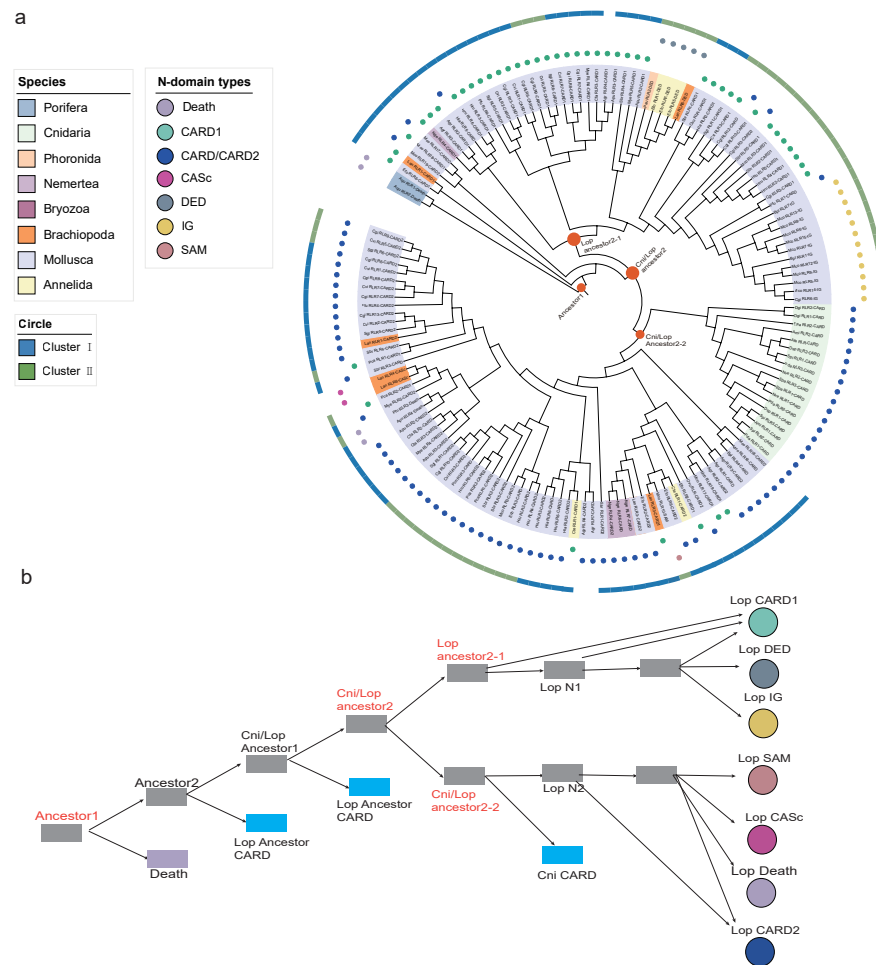


Figure 5. The evolution of N-terminal domains (CARD, DED, IG, death, CASc, SAM) of lophotrochozoan *RLRs*. **(a)** Phylogenetic tree constructed with the maximum likelihood method showing the evolution of N-terminal domains of *RLRs* in lophotrochozoans. The different background colors of the gene names represent the phyla to which they belong, as shown in the legend ‘Species’. The different colored circles represent different domains, as shown in the legend ‘Domain’. The blue arc in the outermost circle represents Cluster I, and the green arc represents Cluster II. The orange circles on the branches marked on the tree correspond to the important nodes in **(b)**. **(b)** A schematic diagram of the evolution of the N-terminal domains of *RLRs* in lophotrochozoans based on the phylogenetic tree. Cni stands for Cnidaria, and Lop stands for the lophotrochozoan. Ancestor1 indicates an ancestor where a duplication produced the first N-terminal death domain and a common ancestor of all other N-terminal domains. Cni/Lop ancestor2 indicates an ancestor where a duplication occurred that produced different N-terminal domains. Lop N1 represents the first domain of the N-terminal *RLR*, and Lop N2 represents the second. Solid black lines indicate duplication events. Colors of N-terminal domains correspond to color labels in the trees in **(a)**.

2.4. Intron–Exon Structure Analysis of Lophotrochozoan RLRs with Diverse N-Terminal Domains

Previously, exon shuffling was believed to be one of the major forces driving domain grafting [47]. Accordingly, we aimed to determine whether exon–intron structure-related mechanisms contribute to the highly diverse N-terminal domains of lophotrochozoan-specific RLRs. Our hypothesis was tested by identifying sibling paralogs/orthologs with the highest sequence similarity based on the RIG-I_C-RD phylogenetic tree and comparing their exon–intron architecture (Figure 6). We first manually optimized all the shown gene models by transcriptomic read mapping to make sure the sequences of these sibling paralogs/orthologs had high confidence. Exon–intron structure divergence was found to be prevalent in the RLRs in all N-specific domains studied, including domains CASc, DED, Death, and IG. Further, the exon–intron structure was classified into three types, including ‘gain/loss of exon/intron,’ ‘exonization/pseudoexonization,’ and ‘intraexonic insertion/deletion’ [48]. Among the examples of the nine homologous RLR groups with a difference in the N-terminal domain in Figure 6, ‘gain/loss of exon/intron’ was found to be prevalent in four cases, ‘exonization/pseudoexonization’ in five cases, and ‘intraexonic insertion/deletion’ in only one case.

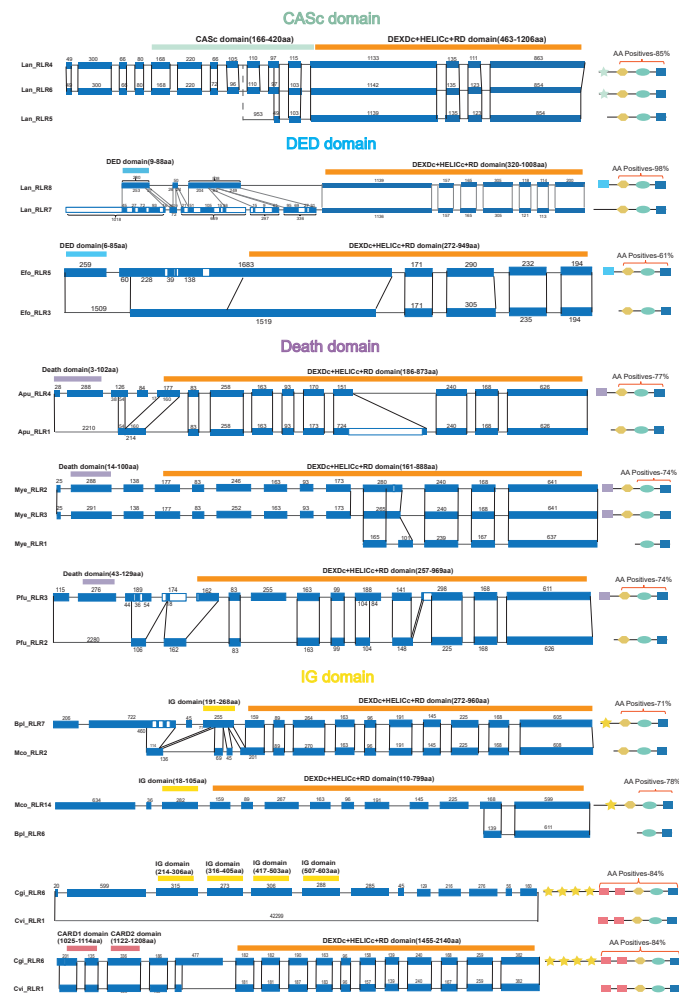


Figure 6. Exon–intron structure analysis of RLRs with diverse N-terminal domains in lophotrochozoans. The exon–intron structures (left) and the schematic diagrams of the domain composition (right) of the RLRs with non-canonical N-terminal domains of nine groups of homologous genes. The blue rectangle represents the exon, and the solid line represents the uncoded sequence in the alignment. The number marks the length of the exon or uncoded sequence. White bars in exons depict the indels that have resulted from insertion/deletion events.

The type 'gain/loss of exon/intron' is a process through which an entire partial exon/intron is inserted/deleted. For the sibling paralogs *LanRLR7* and *LanRLR8* in *L. anatina*, the domain structure of *LanRLR7* is DEXDc-HELICc-RD and does not encode the interaction N-terminal domains. Comparatively, an exon gain event with 253 amino acids was inferred in *LanRLR8*, leading to an additional N-terminal DED domain. This gain of exon did not cause a shift in the reading frame. The other corresponding exonic sequences of this sibling paralog could still be aligned with high confidence, except for the gained exon, with 98% sequence similarity at the amino acid level. The structure type of 'gain/loss of exon/intron' was found in the domain grafting of death in the *Mizuhopecten yessoensis* sibling paralogs *MyeRLR2*, *MyeRLR3*, and *MyeRLR1*. The same structure type was found in the domain grafting of the IG domain in the lophotrochozoan-specific sibling ortholog pairs *BplRLR7-McoRLR2* and *McoRLR14-BplRLR6*. The mechanism of 'exonization/pseudoexonization' is a process that leads to interchanges between exonic and non-exonic sequences. For the sibling ortholog pairs *CgiRLR6* and *CviRLR1* in oyster *RLRs*, exonization of 13 exons was characterized in *CgiRLR6*, which led to a lophotrochozoan-specific *RLR* with four IG domains in the N-terminal sequence. A similar phenomenon was found in the sibling paralogs *LanRLR4-LanRLR5-LanRLR6* and *EfoRLR5-EfoRLR3*, and the sibling ortholog *PfuRLR3-ApuRLR4*, which led to the domain grafting of CASC, DED, and death in the lophotrochozoan-specific *RLRs*, respectively (Figure 6). Finally, intraexonic insertion/deletion was found in the domain shuffling of DED in the sibling paralog *LanRLR7-LanRLR8*. Of note, the three types of structural divergence were not mutually exclusive.

By analyzing six cases of sibling paralogs and three cases of sibling orthologs, we found that the most predominant type of mechanism for structural variation in the *RLR* genes with lophotrochozoan-specific N-terminal domains was 'exonization and pseudoexonization,' which was observed in five of the nine cases studied. The second most predominant was 'gain/loss of exon/intron,' which was observed in four cases. Notably, the three types of exon-intron structure-related mechanisms were prevalent in the canonical V1- and C-type *RLRs*. In contrast to lophotrochozoan *RLRs* encoding diverse N-terminal domains, the type 'loss/gain of exon/intron' is most predominant in these canonical *RLRs* instead of the type 'exonization/intronization' (Supplementary Figure S1).

2.5. Expression Profiles of the Lophotrochozoan *RLRs*

As large expansions and highly diverse domain structures were observed in lophotrochozoan *RLRs*, we determined whether these lophotrochozoan *RLRs* are functional. We collected all published lophotrochozoan tissue transcriptome data from the NCBI GEO database (up to June 2021) and calculated the tissue expression levels of *RLR* genes in ten evolutionary representative species (Figure 7a and Supplementary Table S4), including the phoronidan *Phoronis australis*; the nemertean *Notospermus geniculatus*; the brachiopod *L. anatina*; the mollusks *Octopus bimaculoides*, *Haliotis rufescens*, *C. gigas*, *M. coruscus*, *M. yessoensis*, and *Chlamys farreri*; and the annelida *Eisenia foetida*. Among the 71 *RLRs* studied in the 10 species, 65 were found to be expressed in at least one tissue. Fourteen of the twenty *RLRs* encoding specific N-terminal domains showed expression based on the tissue expression profiles. This result suggests that most lophotrochozoan *RLRs*, both the canonical ones and those with various N-terminal domains, were potentially functional.

Owing to the prevalence of expression in duplicated lophotrochozoan *RLRs* with diverse N-terminal domains, whether these genes retained the function of immune reorganization was unknown. To clarify whether such retention existed, we compared the expression levels of *RLRs* in immune-related and unrelated tissues (Figure 7a and Supplementary Table S4) and found both tissue-prevalent and specific patterns in lophotrochozoan *RLRs* with diverse N-terminal domains. In the bivalve *C. gigas*, most *CgiRLRs* with diverse N-terminal domains were found to be highly expressed in the labial palp, mantle, male gonad, and digestive gland. In the brachiopod *L. anatina*, *LanRLR8* with an N-terminal DED domain and *LanRLR6* with an N-terminal CASC domain were found to be significantly

expressed in the digestive tissue. Molluscan *McoRLR10* with an N-terminal SAM domain and *McoRLR12* with an N-terminal IG domain were also highly expressed in the gut, and *McoRLR10* was highly expressed in the digestive gland (Figure 7a and Supplementary Table S4). These results indicate that *RLRs* might be important in mucosal immunity. However, molluscan *MyeRLR2* and *MyeRLR3* with an N-terminal death domain showed the highest expression levels in hemocytes, suggesting their potential function in hemocyte-mediated immunity. Many duplicated lophotrochozoan *RLRs* with special N-terminal domains were upregulated in both the digestive gland and hemocytes, such as *McoRLR5* and *McoRLR13* with an N-terminal IG domain, which were highly expressed in hemocytes (Figure 7a and Supplementary Table S4). Taken together, these results suggest that many *RLRs* composed of diverse N-terminal domains are highly expressed in immune-related tissues and might play a significant role in innate immune recognition.

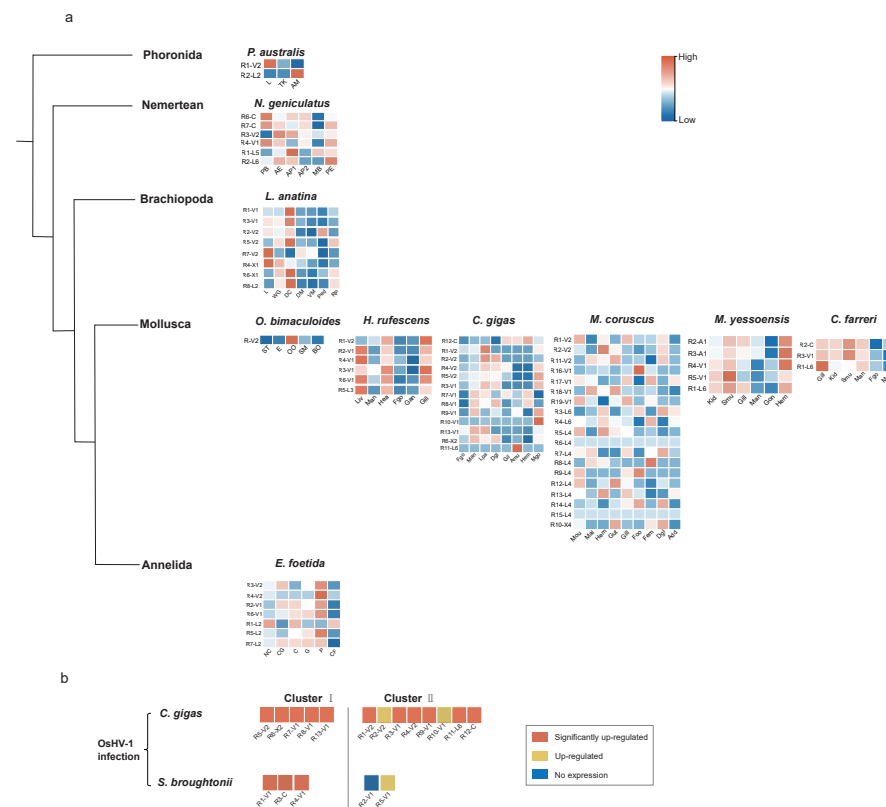


Figure 7. Expression patterns of *RLR* genes among different tissues and under virus infection. (a) Transcriptome expression heatmaps of *RLRs* of different tissues in ten Lophotrochozoa species. Ampulla (AM), trunk (TK), proboscis (PB), anterior end (AE), anterior part1 (AP1), anterior part2 (AP2), mid-body (MB), posterior end (PE), lophophore (L), whole gut tissue (WG), digestive cecum (DC), dorsal mantle (DM), ventral mantle (VM), pedicle (Ped), regenerated pedicle (RP), statocyst tissue (ST), eye (E), olfactory organ (OO), skin from mantle (SM), brain tissue from optic lobe (BO), liver (Liv), mantle (Man), heart (Hea), female gonad (Fgo), adult female ganglion (Gan), gill (Gill), labial palp (Lpa), digestive gland (Dgl), adductor muscle (Amu), hemocyte (Hem), male gonad (Mgo), mouth (Mou), male gonad (Mal), foot (Foo), female gonad (Fem), adductor muscle (Add), kidney (Kid), striated muscle (Smu), gonad (Gon), nerve cord (NC), chlorogog gut (CG), crop (C), gizzard (G), pharynx (P), coelomic fluid (CF). Heatmap displays the expression level of *RLR* genes. (b) Transcriptional change of the *RLRs* under Ostreid herpesvirus-1 (OsHV-1) infection. The orange rectangle represents significantly upregulated *RLR* genes (Log2(FC) > 1.5); the yellow rectangle represents the upregulated *RLR* genes but not significantly; the blue rectangle represents the *RLR* genes that are not expressed.

We determined whether the duplicated lophotrochozoan *RLRs* with diverse N-terminal domains would be upregulated under viral challenges. We searched for virus or virus-related PAMP infection transcriptomes in the NCBI GEO database (up to June 2021), found two published databases from bivalves *C. gigas* and *Scapharca broughtonii*, and calculated their expression profiles (Figure 7b). A total of 11 of the 13 *RLRs* in *C. gigas* were significantly upregulated ($\text{Log}_2(\text{FC}) > 1.5$) during oyster herpes virus infection. Of these *RLRs*, *CgiRLR6* encodes a lophotrochozoan-specific N-terminal IG domain. Similarly, *S. broughtonii SbrRLR3* encoding only one CARD at the N-terminus was significantly upregulated ($\text{Log}_2(\text{FC}) > 1.5$) under virus challenge. These results suggest that lophotrochozoan *RLRs* with special domains may play an important role in antiviral immune recognition.

2.6. Evidence of Positive Selection in Lophotrochozoan *RLRs* with Diverse N-Terminal Domains

As innate immune receptors are responsible for severe diseases, *RLRs* must rapidly evolve and are thus subjected to positive selection pressures [44]. We determined whether the molecular evolution of the duplicated lophotrochozoan *RLRs* with diverse N-terminal domains was driven by natural selection. Accordingly, a positive analysis was performed on lophotrochozoan *RLRs* with diverse N-terminal domains and gene expression levels enriched in the immune-related tissues mentioned above. We used the RIG-I_C-RD domain of *RLRs* for selection analysis, as this part is the virus recognition domain.

We performed positive selection analysis on five *RLRs* of *M. yessoensis*, all of which were in Cluster I of the phylogenetic tree (Figure 3). Positive selection signals ($\omega = 521.37$) could be identified in branches, including *MyeRLR2* and *MyeRLR3* (Figure 8a). Both genes encoded the N-terminal death domain. We further reconstructed the 3D structures of the five *MyeRLRs* (Figure 8c). Based on the results, the surface of the RNA-binding region of *MyeRLR2* and *MyeRLR3* is positively charged, which is consistent with human *RIG-I* [41]. Although *MyeRLR1* is positively charged, its N-terminus has no domain, which may prevent its function. However, the surface of the RNA-binding region of *MyeRLR4* and *MyeRLR5* has a less positive charge.

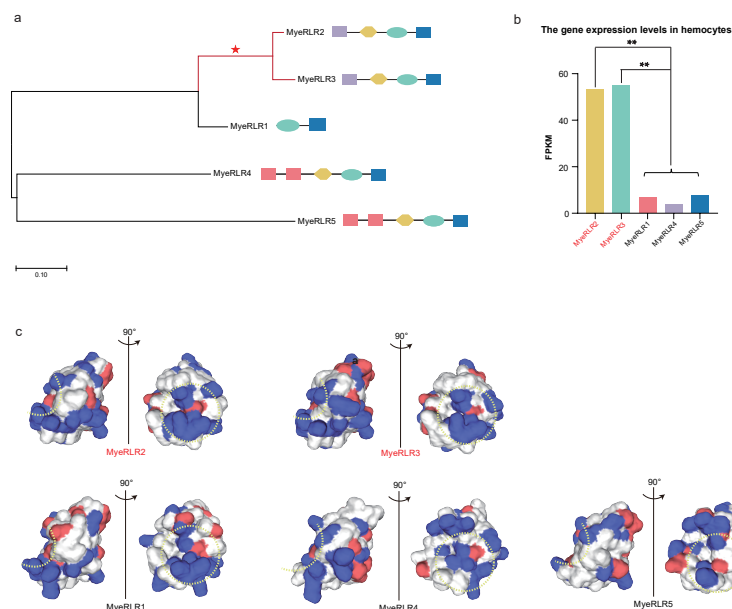


Figure 8. Positive selection analysis of representative *RLR* genes. (a) Evolutionary relationships of *MyeRLRs*. The branches with dN/dS (ω) values > 1.0 are marked with red lines. (b) The expression level of five *MyeRLRs* in hemocytes. One-way ANOVA showed that the expression levels of *MyeRLR2* and *MyeRLR3* were significantly (p -value < 0.01 , marked with **) more upregulated than those of other *MyeRLRs*. (c) Electrostatics of the RNA-binding surfaces of the *MyeRLR* RIG-I_C-RDs. The determination of RNA-binding sites refers to previous studies [49]. Positively charged surfaces are colored blue, and negatively charged surfaces are red. Dotted yellow line indicates RNA-binding loop.

Of note, the selection analysis was consistent with the gene expression patterns in *RLRs* of *M. yessoensis*. Both *MyeRLR2* and *MyeRLR3* showed a significantly higher expression in the immune-related tissue hemolymph (Figure 8b). The lophotrochozoan *RLRs* with diverse N-terminal domains might interact with other adaptors without a CARD domain, triggering the unique lophotrochozoan cascade in antiviral immunity. The selection results suggest that lophotrochozoan-specific *RLRs* were under rapid positive selection, indicating that these unique *RLRs* might play important and novel roles in innate immunity.

3. Discussion

Interaction between pathogens and hosts leads to a dynamic evolutionary arms race. Invertebrates, which lack adaptive immunity, evolved a variety of broad, unique, and complex repertoires of innate immune sensors. In this study, we explored the diversity and specificity of invertebrate innate immune recognition in the lophotrochozoan *RLR* gene family and identified the diversity of *RLRs* in lophotrochozoans, which is mainly reflected in the divergence of the N-terminal domains. By exploring the molecular evolutionary mechanism driving the diversity of domain arrangement in lophotrochozoan *RLRs*, we found that it might be due to rapid domain grafting, exon–intron structural divergence, expression diversification, and positive selection. To the best of our knowledge, this is the first systematic study of the molecular evolution of *RLRs* in lophotrochozoans.

One of the most interesting findings of this study is that lophotrochozoan *RLRs* represent the successful use of genetic linkages of N-terminal domains to expand and diversify the immune repertoire. The invertebrate immune system is innate and encoded in the germline. Extensive expansion of immune receptors has been proposed to reveal an alternative mechanism for the diversity and specificity of innate immune recognition in the absence of an adaptive immune system [8–10]. In our study, the *RLR* gene families experienced expansion, aligning with previous immune receptor studies in amphioxys [9]. Importantly, the domain arrangement was found to be highly diverse in the duplicated lophotrochozoan *RLRs*. Manual correction of all gene models with non-canonical N-terminal domains was performed, and their domain arrangements were examined, which indicated that our data can be used for further phylogenetic and molecular evolution studies. Therefore, gene models were predicted with high confidence, arguing against gene modeling errors as an explanation for our results.

In vertebrates, the N-terminal CARD domain of *RIG-I* and *MDA5* interacts with the CARD domain of the mitochondrial protein *MAVS* for signal transmission [31–33,50]. In invertebrates, the protein interaction between the N-terminus of canonical *RLR* receptors and the adaptor *MAVS* is conserved in the mollusk *C. gigas* [35]. Of note, non-canonical N-terminal domains, including the death, DED, CASc, IG, and SAM domains, were also observed in lophotrochozoan *RLRs*. Among them, the death and DED subfamily, together with CARD, constitutes the death domain (DD) superfamily. By mediating homotypic interactions within each domain subfamily, these proteins play important roles in the assembly and activation of apoptotic and inflammatory complexes [51,52]. The CASc domain represents the C-terminal conserved domain found in caspases, mainly from animals. Caspases are mainly involved in mediating cell apoptosis and are recruited as apoptosis initiators that trigger the apoptosis process, and as effectors of apoptosis [53,54]. The CASc domain is also reported to be involved in inflammatory processes [55]. Previous studies suggested that another important intracellular immune receptor, *NLR* (nucleotide oligomerization domain (NOD)-like receptor), recruited the apoptosis-related domains pyrin and baculovirus inhibitor repeats during evolution, leading to the control of the activation of inflammatory caspases in animals [56,57]. We hypothesized that lophotrochozoan *RLRs* with a CASc domain might be involved in apoptosis, inflammation, or pyroptosis. Molecules with IG-like domains are involved in a variety of immunological functions, including adaptive immune receptors, innate immune molecules, and accessory molecules [58]. Previous studies suggested that diverse forms of IG-containing molecules and their specificity of immune function in non-self recognition or interaction with endogenous molecules are

remarkable [59]. One of the well-known examples is the IG domain's variation in the snail fibrinogen-related proteins that exhibits different forms of somatic variation [60]. Although the C-terminal domain of *RLR* is responsible for initial pathogen recognition, the N-terminal domains are required for downstream signaling [24,30]. Novel immune response diversity and specificities might thus be acquired by integrating non-canonical domains into *RLRs* in lophotrochozoans. These new integrated non-canonical protein domains might play an unidentified role in innate immunity or host defense.

The phylogenetic results suggest that the lophotrochozoan *RLRs* integrated with non-canonical N-terminal domains were distributed unevenly across the *RLR* phylogeny with the dominant clade in Cluster II, and most were species-specific. Most lophotrochozoan *RLRs* clustered with vertebrate *RIG-I* instead of *MDA5/LDP2* and were divided into two clusters. The domain composition showed that the functions of the *RLRs* in Cluster II were more diverse than those in Cluster I. Previous studies have shown that in addition to *RLRs* [21], the other two gene families encoding immune receptors, *NLR* and *TLR*, underwent massive species-specific expansions and domain shuffling in various lineages [61]. In the NACHT protein family, a diversity of N-terminal domains, including death, CARD, DED, BIR, and PYRIN domains, was found. Among them, three types of domain combinations have emerged multiple times in different lineages, including death-NACHT-LRR, CARD-NACHT-LRR, and PYRIN-NACHT-LRR [61]. Studies on the TIR protein family have shown that across the metazoans, the N-terminal domain connected to TIR has the IG, death, and SAM domains, and IG-TIR has emerged multiple times in different lineages [61]. In contrast, in the specific domains of the N-terminus of *RLRs* (except CARD), the death, DED, and IG domains have emerged multiple times in different lineages, while the CASc and SAM domains have only been found in one species. Two separate domains of *TLR*, the TIR domain and the LRR domain, have been found to have a domain combination phenomenon. The combination of the P-TIR and P-LRR domains occurred in non-bilateria, while the early combination of the V-TIR and V-LRR domains for V-TLR occurred after the divergence of bilateria and non-bilateria [62]. Generally, it is speculated that the combination of the above domains is not random but follows certain rules, which might be selective pressure exerted by pathogens in a specific environment [61,62]. In addition, the phenomenon of domain shuffling in killer cell Ig-like receptors (*KIRs*) is expressed on the surface of NK cells, which leads to the production of new *KIRs* [63].

Interestingly, the exon–intron structure seems to significantly contribute to the integration of non-canonical N-terminal domains in the lophotrochozoan *RLRs*. The structure of the *RLR* protein is relatively simple in ancestral metazoan porous animals and cnidarians. However, in lophotrochozoans, the domain arrangement of the *RLR* protein showed a complex and diverse pattern with the addition of the second CARD (CARD1) and non-canonical N-terminal domains. Deletion and incorporation of specific domains have become prevalent. Based on our results, the diversity of the N-terminal domains in lophotrochozoan *RLRs* should be attributed to domain shuffling after gene duplication. As the domains are often correlated with exon boundaries, exon shuffling is believed to be one of the major forces driving domain shuffling [47]. Our study indicates that intronization/exonization is the main driving force of exon shuffling. A similar phenomenon of exon shuffling was observed in the *KIR* gene family. For *KIR* genes, the exon–intron structure correlates with the four main parts: the first three parts comprise an Ig domain, D0, D1, and D2, and the last part comprises the stem (S), transmembrane (TM), and cytoplasmic (CYT) domains [63,64]. A previous study speculated that, in addition to domain shuffling's *KIR* being more favored by natural selection, another important reason for domain shuffling is that certain introns are hot spots for recombination [63]. For the *RLR* gene, almost all specific domains were found to be fully encoded by only one exon (Figure 6), which also creates an environment in which exon shuffling occurs.

Finally, as positive selection signals and immune-related expression patterns were detected independently in different clades of lophotrochozoan non-canonical *RLRs*, these domain fusions may have a selective advantage for the organisms [62,65]. The possibility

of domain recombination to achieve functional diversity has been mentioned in many evaluations [63,66,67], and the diversity of the N-terminal domain may affect the specificity of ligand binding [66]. The study of the function of abaecin through exon shuffling suggests the role of exon shuffling in buffering the loss-of-function mutations in a gene [67]. However, at present, sufficient immune-related transcriptome data are unavailable to verify whether the new lophotrochozoan *RLRs* have new functions. Further, whether proteins with the same domain architecture share similar functions in different species is still unknown. Some examples including those from the families discussed here suggest that this notion does not always hold true. For instance, although *Drosophila* Toll-like receptors mainly perform roles in embryonic development, their mammalian homologs are key regulators of immune responses [68,69]. Therefore, extrapolation of protein function based on the domain architecture must be performed very carefully.

4. Materials and Methods

4.1. Data Collection

To identify the *RLR* genes, the genome and protein sequence datasets of 58 metazoan species at different evolutionary nodes were collected from the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>, accessed on 1 June 2021) or from other databases (Supplementary Figure S2; Table S1). Transcriptome data for *RLR* gene expression analysis were obtained from the NCBI Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra/>, accessed on 1 June 2021). Multiple biotic stress data included PRJNA450478 (*S. broughtonii* after Ostreid herpesvirus-1 infection), and PRJNA146329 (*C. gigas* larvae infected with OsHV-1). Adult tissue data included PRJNA393252 (*P. australis*); PRJNA393252 (*N. geniculatus*); PRJNA286275 (*L. anatina*); PRJNA658966 (*O. bimaculoides*); PRJNA488641 (*H. rufescens*); PRJNA146329 (*C. gigas*); PRJNA578350 (*M. coruscus*); PRJNA185465 (*C. farreri*); PRJNA259405 (*M. yessoensis*); PRJNA608692 (*E. foetida*).

4.2. *RLR* Gene Identification and Phylogenetic Analysis

We used domain prediction and a sequence homology search of the *RLR* genes in 58 metazoan species (Figure S2). First, a local version of HMMER version3.1b2 (Howard Hughes Medical Institute; Cambridge, UK, 2015) [70], available from <http://hmmer.org/download.html> (accessed on 1 June 2021), was used to identify the *RLR* genes by screening the RIG-I_C-RD (PF11648) domain on the genomes of all species. Second, regions of each genome potentially harboring *RLR* genes were identified using TBLASTN (National Center for Biotechnology Information, Bethesda, MD, USA) [71] with *RLR* proteins from model organisms as query sequences. When we used the above two methods to find the loss of *RLRs* in some species, we used PSI-BLAST (<http://blast.ncbi.nlm.nih.gov/>, accessed on 4 March 2022) to further confirm this conclusion. Third, according to the structural characteristics of *RLRs* in a broad sense, we used the visualization software SMART (<http://smart.embl-heidelberg.de/>, accessed on 1 September 2021) [72] to screen the *RLR* genes of each species containing an intermediate DEAD/DEAH box helicase domain and a RIG-I_C-RD domain. Finally, we corrected some *RLR* genes with suspected domain deletions or special domains using local GeneWise software (<https://www.ebi.ac.uk/Tools/psa/genewise/>, accessed on 1 September 2021) [73] and transcriptome read mapping, which greatly ensures the accuracy of each N-terminal-specific *RLR* gene model studied.

The highly conserved sequences of the RIG-I_C-RD domain of *RLRs* identified by PFAM HMM were used to conduct the phylogenetic analyses in the present study. The protein sequences encoding the RIG-I_C-RD domain were aligned using the L-INS-I strategy in the local MAFFT v7.310 (Kazutaka Katoh, Osaka, Japan, 2013) [74]. Maximum likelihood trees were generated using IQ-TREE v1.6.12 (Bui Quang Minh, ANU (Australian National University), Australia, 2011) [75], which could automatically test and select the best alternative model. In order to test branch reliability, we adopted the fast bootstrap method with 1000 replicates, which can also be integrated into IQ-TREE. Trees were handled using iTOL v6 (<https://itol.embl.de/>, accessed on 1 November 2021). In addition, in order to explain

the overall evolution of *RLRs*, the N-terminal domain coding sequences and intermediate helicase domain (DEXDc + HELICc domains) coding sequences were extracted to construct phylogenetic trees according to the same method.

4.3. Identification of *RLR* Homologs and Analysis of Gene Structures

First, the local BLAST version 2.9.0 (National Center for Biotechnology Information, Bethesda, MD, USA) was used for homologous gene pair identification by reciprocal BLAST searching on the retrieved datasets (Figure S2). Then, the above results were confirmed by phylogenetic analysis of the *RLR* gene family. However, the homologous gene pairs found by this method were often too limited, so we only used the method of phylogenetic analysis to identify *RLR* gene homologous pairs containing special lophotrochozoan *RLRs*.

To determine whether paralogous or orthologous genes have diverged in exon–intron structure, we compared their genomic sequences. Two paralogs or orthologs were regarded as structurally divergent if they had different numbers of exons, or if they had the same number of exons but the lengths of at least one pair of homologous exons were different. To understand the underlying mechanisms of structural divergence, we generated pairwise alignments for each gene pair, using the corresponding mRNAs as guidance. Intraexonic insertion/deletion was deduced when an indel was found within the aligned homologous exons. Exon/intron gain/loss was inferred if an orphan exon/intron was the result of exon duplication, exon shuffling, exon scrambling, intron insertion, or intron deletion. Exonization/pseudoexonization was identified when the corresponding exonic and non-exonic sequences could be aligned with confidence [48].

4.4. Transcriptomic Analysis of Gene Expression

The FastxToolkit pipeline (http://hannonlab.cshl.edu/fastx_toolkit/index.html, accessed on 1 November 2021) was used to process the raw reads to evaluate sequencing quality and remove low-quality reads (length threshold < 50 bp and quality threshold < 20), adaptor sequences, poly-N, and known non-coding RNAs (Figure S2). Genome-based indexing and sam file generation were achieved using local Hisat2 version 2.1.0 (University of Texas Southwestern Medical Center, Dallas, TX, USA, 2017) [76] and Bowtie2 version 2.3.5.1 (University of Maryland, College Park, MD, USA, 2019) [77], and sam files were sorted using Samtools version 1.11 (Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK) [78]. The obtained clean reads were then individually mapped to the genome of the respective species. Gene expression levels were measured by fragments per kilobase million (FPKM). The expression levels were quantified using Cufflinks version 2.2.1 (Harvard University, Cambridge, MA, USA) [79]. The differentially expressed *RLR* genes (DEGs) were identified with the edgeR tool of the R programming language with the threshold value $|\log_2FC| \geq 1.5$ (multiple of fold change, FC: difference) and $FDR \leq 0.05$.

4.5. Positive Selection Analysis

Multiple sequence alignments of *RLRs* were performed using ClustalW (<https://www.genome.jp/tools-bin/clustalw>, accessed on 1 December 2021) with default parameters, and the resulting alignments were refined with trimAl version 1.2 (Centre for Genomic Regulation, Barcelona, Spain) [80] (Figure S2). Phylogenetic trees were constructed with ML analytical approaches based on MEGA7. The robustness of the inferred trees was assessed using bootstrapping with 1000 replicates in the phylogenetic tree. Phylogenetic trees were visualized using iTOL and used for subsequent positive selection analysis.

PAML version 4.9j (University of California, Berkeley, USA) [81] was used for comparing the rate per site of dN (nonsynonymous) to the rate per site of dS (synonymous) mutations. The recommended subset of four M-series models of M1a (nearly neutral), M2a (positive selection), M7 (beta), and M8 (beta and ω) coupled with Bayesian empirical Bayes (BEB) methods was implemented. The log-likelihood values (lnL) of M2aM1a and M8-M7 were from explicit tests for the presence of positively selected sites. The *p* values were corrected by a multiple testing correction method. Furthermore, the probabilities of sites

under positive selection were assessed by their posterior probabilities calculated with the BEB method. The amino acid site was considered as a positively selected site if the value of $dN/dS > 1$ appeared in the LRT and the posterior probability exceeded 90%. Finally, SWISS-MODEL (<http://swissmodel.expasy.org/>, accessed on 10 December 2021) was used to locate and visualize the 3D structure of *RLRs*.

5. Conclusions

In this study, we systematically described the evolutionary history of the *RLR* gene family in lophotrochozoans by investigating their domain architecture, phylogeny, exon–intron structure, expression profiles, and selection patterns. Our study revealed many previously unknown N-terminal domain fusions in lophotrochozoan *RLRs*, which might result in the diversity and specificity of the innate immune response. We traced the exon–intron structure of these non-canonical lophotrochozoan *RLRs* and found that the mechanism of exonization/pseudoexonization might drive the formation of these *RLRs*. Many non-canonical lophotrochozoan *RLRs* exhibit positive selection signals and immune-related expression patterns, indicating that non-canonical lophotrochozoan *RLRs* might have a selective advantage for organisms in the innate immune response. Overall, our findings suggest that the complex and unique domain arrangement of lophotrochozoan *RLRs* might result from rapid domain grafting, exon–intron structural divergence, expression diversification, and positive selection, which may have led to functionally distinct paralogs or orthologs in the innate immune response. Our research provides new insights into the molecular evolution of innate receptors in invertebrates in the absence of antibody-mediated adaptive immunity.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms23073415/s1>.

Author Contributions: Conceived and designed experiments: L.Z.; data analyses: S.Y., J.C. and L.Z.; contributed reagents/materials/computer resources: L.Z. and Y.X.; wrote the paper: S.Y., L.Z., J.C. and S.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China 41976088 to L.Z., the Strategic Priority Research Program of the Chinese Academy of Sciences XDB42000000 to L.Z., and the Key Development Project of Centre for Ocean Mega-Research of Science, Chinese Academy of Science COMS2019R01 to L.Z.

Institutional Review Board Statement: Ethical review and approval were not applicable for studies not involving humans or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Yongnan Li for discussions and the High-Performance Computing Center of the Institute of Oceanology, CAS.

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

References

1. Koonin, E.V.; Dolja, V.V. A virocentric perspective on the evolution of life. *Curr. Opin. Virol.* **2013**, *3*, 546–557. [[CrossRef](#)] [[PubMed](#)]
2. TenOever, B.R. The Evolution of Antiviral Defense Systems. *Cell Host Microbe* **2016**, *19*, 142–149. [[CrossRef](#)] [[PubMed](#)]
3. Halanych, K.M. Lophotrochozoa, Diversification of. In *Encyclopedia of Evolutionary Biology*; Kliman, R.M., Ed.; Academic Press: London, UK, 2016; pp. 405–408.
4. Salzet, M.; Tasiemski, A.; Cooper, E. Innate immunity in lophotrochozoans: The annelids. *Curr. Pharm. Des.* **2006**, *12*, 3043–3050. [[CrossRef](#)]
5. Zhang, L.; Li, L.; Zhu, Y.; Zhang, G.; Guo, X. Transcriptome analysis reveals a rich gene set related to innate immunity in the Eastern oyster (*Crassostrea virginica*). *Mar. Biotechnol.* **2014**, *16*, 17–33. [[CrossRef](#)] [[PubMed](#)]
6. Guo, X.; He, Y.; Zhang, L.; Lelong, C.; Jouaux, A. Immune and stress responses in oysters with insights on adaptation. *Fish Shellfish Immunol.* **2015**, *46*, 107–119. [[CrossRef](#)]
7. Wang, L.; Song, X.; Song, L. The oyster immunity. *Dev. Comp. Immunol.* **2018**, *80*, 99–118. [[CrossRef](#)]

8. Rast, J.P.; Smith, L.C.; Loza-Coll, M.; Hibino, T.; Litman, G.W. Genomic insights into the immune system of the sea urchin. *Science* **2006**, *314*, 952–956. [[CrossRef](#)]
9. Huang, S.; Yuan, S.; Guo, L.; Yu, Y.; Li, J.; Wu, T.; Liu, T.; Yang, M.; Wu, K.; Liu, H.; et al. Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res.* **2008**, *18*, 1112–1126. [[CrossRef](#)]
10. Zhang, L.; Li, L.; Guo, X.; Litman, G.W.; Dishaw, L.J.; Zhang, G. Massive expansion and functional divergence of innate immune genes in a protostome. *Sci. Rep.* **2015**, *5*, 8693. [[CrossRef](#)]
11. Hoffmann, J.A., Jr.; Kafatos, F.C.; Janeway, C.A.; Ezekowitz, R.A. Phylogenetic perspectives in innate immunity. *Science* **1999**, *284*, 1313–1318. [[CrossRef](#)]
12. Rosenstiel, P.; Philipp, E.E.; Schreiber, S.; Bosch, T.C. Evolution and function of innate immune receptors—insights from marine invertebrates. *J. Innate Immun.* **2009**, *1*, 291–300. [[CrossRef](#)] [[PubMed](#)]
13. Boller, T.; Felix, G. A renaissance of elicitors: Perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* **2009**, *60*, 379–406. [[CrossRef](#)] [[PubMed](#)]
14. Brubaker, S.W.; Bonham, K.S.; Zanoni, I.; Kagan, J.C. Innate immune pattern recognition: A cell biological perspective. *Annu. Rev. Immunol.* **2015**, *33*, 257–290. [[CrossRef](#)]
15. Chan, J.; Wang, L.; Li, L.; Mu, K.; Bushek, D.; Xu, Y.; Guo, X.; Zhang, G.; Zhang, L. Transcriptomic Response to Perkinsus marinus in Two Crassostrea Oysters Reveals Evolutionary Dynamics of Host-Parasite Interactions. *Front. Genet.* **2021**, *12*, 795706. [[CrossRef](#)] [[PubMed](#)]
16. Janeway, C.A., Jr.; Medzhitov, R. Innate immune recognition. *Annu. Rev. Immunol.* **2002**, *20*, 197–216. [[CrossRef](#)] [[PubMed](#)]
17. Kumar, H.; Kawai, T.; Akira, S. Pathogen recognition by the innate immune system. *Int. Rev. Immunol.* **2011**, *30*, 16–34. [[CrossRef](#)] [[PubMed](#)]
18. Loo, Y.M.; Gale, M., Jr. Immune signaling by RIG-I-like receptors. *Immunity* **2011**, *34*, 680–692. [[CrossRef](#)] [[PubMed](#)]
19. Yoneyama, M.; Fujita, T. RNA recognition and signal transduction by RIG-I-like receptors. *Immunol. Rev.* **2009**, *227*, 54–65. [[CrossRef](#)]
20. Yoneyama, M.; Kikuchi, M.; Natsukawa, T.; Shinobu, N.; Imaizumi, T.; Miyagishi, M.; Taira, K.; Akira, S.; Fujita, T. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat. Immunol.* **2004**, *5*, 730–737. [[CrossRef](#)]
21. Sarkar, D.; Desalle, R.; Fisher, P.B. Evolution of MDA-5/RIG-I-dependent innate immunity: Independent evolution by domain grafting. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17040–17045. [[CrossRef](#)]
22. 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](#)]
23. Lewandowska, M.; Sharoni, T.; Admoni, Y.; Aharoni, R.; Moran, Y. Functional Characterization of the Cnidarian Antiviral Immune Response Reveals Ancestral Complexity. *Mol. Biol. Evol.* **2021**, *38*, 4546–4561. [[CrossRef](#)] [[PubMed](#)]
24. Yoneyama, M.; Kikuchi, M.; Matsumoto, K.; Imaizumi, T.; Miyagishi, M.; Taira, K.; Foy, E.; Loo, Y.M.; Gale, M., Jr.; Akira, S.; et al. Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J. Immunol.* **2005**, *175*, 2851–2858. [[CrossRef](#)] [[PubMed](#)]
25. Kato, H.; Sato, S.; Yoneyama, M.; Yamamoto, M.; Uematsu, S.; Matsui, K.; Tsujimura, T.; Takeda, K.; Fujita, T.; Takeuchi, O.; et al. Cell type-specific involvement of RIG-I in antiviral response. *Immunity* **2005**, *23*, 19–28. [[CrossRef](#)] [[PubMed](#)]
26. Kato, H.; Takeuchi, O.; Sato, S.; Yoneyama, M.; Yamamoto, M.; Matsui, K.; Uematsu, S.; Jung, A.; Kawai, T.; Ishii, K.J.; et al. Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **2006**, *441*, 101–105. [[CrossRef](#)]
27. Gitlin, L.; Barchet, W.; Gilfillan, S.; Cella, M.; Beutler, B.; Flavell, R.A.; Diamond, M.S.; Colonna, M. Essential role of mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8459–8464. [[CrossRef](#)]
28. Kang, D.C.; Gopalkrishnan, R.V.; Wu, Q.; Jankowsky, E.; Pyle, A.M.; Fisher, P.B. mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 637–642. [[CrossRef](#)]
29. Su, Z.Z.; Sarkar, D.; Emdad, L.; Barral, P.M.; Fisher, P.B. Central role of interferon regulatory factor-1 (IRF-1) in controlling retinoic acid inducible gene-I (RIG-I) expression. *J. Cell. Physiol.* **2007**, *213*, 502–510. [[CrossRef](#)]
30. Saito, T.; Hirai, R.; Loo, Y.M.; Owen, D.; Johnson, C.L.; Sinha, S.C.; Akira, S.; Fujita, T.; Gale, M., Jr. Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 582–587. [[CrossRef](#)]
31. Kawai, T.; Takahashi, K.; Sato, S.; Coban, C.; Kumar, H.; Kato, H.; Ishii, K.J.; Takeuchi, O.; Akira, S. IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nat. Immunol.* **2005**, *6*, 981–988. [[CrossRef](#)]
32. Meylan, E.; Curran, J.; Hofmann, K.; Moradpour, D.; Binder, M.; Bartenschlager, R.; Tschopp, J. Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **2005**, *437*, 1167–1172. [[CrossRef](#)] [[PubMed](#)]
33. Seth, R.B.; Sun, L.; Ea, C.K.; Chen, Z.J. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF- κ B and IRF 3. *Cell* **2005**, *122*, 669–682. [[CrossRef](#)] [[PubMed](#)]
34. Honda, K.; Yanai, H.; Negishi, H.; Asagiri, M.; Sato, M.; Mizutani, T.; Shimada, N.; Ohba, Y.; Takaoka, A.; Yoshida, N.; et al. IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* **2005**, *434*, 772–777. [[CrossRef](#)]
35. Huang, B.; Zhang, L.; Du, Y.; Xu, F.; Li, L.; Zhang, G. Characterization of the Mollusc RIG-I/MAVS Pathway Reveals an Archaic Antiviral Signalling Framework in Invertebrates. *Sci. Rep.* **2017**, *7*, 8217. [[CrossRef](#)] [[PubMed](#)]

36. Bamming, D.; Horvath, C.M. Regulation of signal transduction by enzymatically inactive antiviral RNA helicase proteins MDA5, RIG-I, and LGP2. *J. Biol. Chem.* **2009**, *284*, 9700–9712. [[CrossRef](#)]
37. Satoh, T.; Kato, H.; Kumagai, Y.; Yoneyama, M.; Sato, S.; Matsushita, K.; Tsujimura, T.; Fujita, T.; Akira, S.; Takeuchi, O. LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1512–1517. [[CrossRef](#)]
38. Venkataraman, T.; Valdes, M.; Elsby, R.; Kakuta, S.; Caceres, G.; Saijo, S.; Iwakura, Y.; Barber, G.N. Loss of DExD/H box RNA helicase LGP2 manifests disparate antiviral responses. *J. Immunol.* **2007**, *178*, 6444–6455. [[CrossRef](#)]
39. Li, X.; Ranjith-Kumar, C.T.; Brooks, M.T.; Dharmiaiah, S.; Herr, A.B.; Kao, C.; Li, P. The RIG-I-like receptor LGP2 recognizes the termini of double-stranded RNA. *J. Biol. Chem.* **2009**, *284*, 13881–13891. [[CrossRef](#)]
40. Zou, J.; Chang, M.; Nie, P.; Secombes, C.J. Origin and evolution of the RIG-I like RNA helicase gene family. *BMC Evol. Biol.* **2009**, *9*, 85. [[CrossRef](#)]
41. Mukherjee, K.; Korithoski, B.; Kolaczowski, B. Ancient origins of vertebrate-specific innate antiviral immunity. *Mol. Biol. Evol.* **2014**, *31*, 140–153. [[CrossRef](#)]
42. Majzoub, K.; Wrensch, F.; Baumert, T.F. The Innate Antiviral Response in Animals: An Evolutionary Perspective from Flagellates to Humans. *Viruses* **2019**, *11*, 758. [[CrossRef](#)] [[PubMed](#)]
43. Fan, Y.; Huang, Z.Y.; Cao, C.C.; Chen, C.S.; Chen, Y.X.; Fan, D.D.; He, J.; Hou, H.L.; Hu, L.; Hu, X.T.; et al. Genome of the Chinese tree shrew. *Nat. Commun.* **2013**, *4*, 1426. [[CrossRef](#)] [[PubMed](#)]
44. Xu, L.; Yu, D.; Fan, Y.; Peng, L.; Wu, Y.; Yao, Y.G. Loss of RIG-I leads to a functional replacement with MDA5 in the Chinese tree shrew. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 10950–10955. [[CrossRef](#)] [[PubMed](#)]
45. Zheng, W.; Satta, Y. Functional Evolution of Avian RIG-I-Like Receptors. *Genes* **2018**, *9*, 456. [[CrossRef](#)]
46. Chen, J.; Shang, S.; Wu, X.; Zhong, H.; Zhao, C.; Wei, Q.; Zhang, H.; Xia, T.; Chen, Y.; Zhang, H.; et al. Genomic analysis and adaptive evolution of the RIG-I-like and NOD-like receptors in reptiles. *Int. J. Biol. Macromol.* **2019**, *134*, 1045–1051. [[CrossRef](#)]
47. Liu, M.; Grigoriev, A. Protein domains correlate strongly with exons in multiple eukaryotic genomes—vidence of exon shuffling? *Trends Genet. TIG* **2004**, *20*, 399–403. [[CrossRef](#)]
48. Xu, G.; Guo, C.; Shan, H.; Kong, H. Divergence of duplicate genes in exon-intron structure. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1187–1192. [[CrossRef](#)]
49. Lu, C.; Ranjith-Kumar, C.T.; Hao, L.; Kao, C.C.; Li, P. Crystal structure of RIG-I C-terminal domain bound to blunt-ended double-strand RNA without 5' triphosphate. *Nucleic Acids Res.* **2010**, *39*, 1565–1575. [[CrossRef](#)]
50. Xu, L.G.; Wang, Y.Y.; Han, K.J.; Li, L.Y.; Zhai, Z.; Shu, H.B. VISA is an adapter protein required for virus-triggered IFN-beta signaling. *Mol. Cell* **2005**, *19*, 727–740. [[CrossRef](#)]
51. Wajant, H. Death receptors. *Essays Biochem.* **2003**, *39*, 53–71.
52. Bhardwaj, A.; Aggarwal, B.B. Receptor-mediated choreography of life and death. *J. Clin. Immunol.* **2003**, *23*, 317–332. [[CrossRef](#)] [[PubMed](#)]
53. Yang, J.Y.; Widmann, C. Antiapoptotic signaling generated by caspase-induced cleavage of RasGAP. *Mol. Cell. Biol.* **2001**, *21*, 5346–5358. [[CrossRef](#)] [[PubMed](#)]
54. Salvesen, G.S.; Abrams, J.M. Caspase activation—stepping on the gas or releasing the brakes? Lessons from humans and flies. *Oncogene* **2004**, *23*, 2774–2784. [[CrossRef](#)] [[PubMed](#)]
55. Jiménez Fernández, D.; Lamkanfi, M. Inflammatory caspases: Key regulators of inflammation and cell death. *Biol. Chem.* **2015**, *396*, 193–203. [[CrossRef](#)]
56. Kelley, N.; Jeltama, D.; Duan, Y.; He, Y. The NLRP3 Inflammasome: An Overview of Mechanisms of Activation and Regulation. *Int. J. Mol. Sci.* **2019**, *20*, 3328. [[CrossRef](#)]
57. Kofoed, E.M.; Vance, R.E. Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity. *Nature* **2011**, *477*, 592–595. [[CrossRef](#)]
58. Van Lookeren Campagne, M.; Verschoor, A. Pathogen clearance and immune adherence “revisited”: Immuno-regulatory roles for CRIG. *Semin. Immunol.* **2018**, *37*, 4–11. [[CrossRef](#)]
59. Barclay, A.N. Membrane proteins with immunoglobulin-like domains—a master superfamily of interaction molecules. *Semin. Immunol.* **2003**, *15*, 215–223. [[CrossRef](#)]
60. Zhang, S.M.; Adema, C.M.; Kepler, T.B.; Loker, E.S. Diversification of Ig superfamily genes in an invertebrate. *Science* **2004**, *305*, 251–254. [[CrossRef](#)]
61. Zhang, Q.; Zmasek, C.M.; Godzik, A. Domain architecture evolution of pattern-recognition receptors. *Immunogenetics* **2010**, *62*, 263–272. [[CrossRef](#)]
62. Wu, B.; Huan, T.; Gong, J.; Zhou, P.; Bai, Z. Domain combination of the vertebrate-like TLR gene family: Implications for their origin and evolution. *J. Genet.* **2011**, *90*, 401–408. [[CrossRef](#)] [[PubMed](#)]
63. Rajalingam, R.; Parham, P.; Abi-Rached, L. Domain shuffling has been the main mechanism forming new hominoid killer cell Ig-like receptors. *J. Immunol.* **2004**, *172*, 356–369. [[CrossRef](#)] [[PubMed](#)]
64. Martin, A.M.; Freitas, E.M.; Witt, C.S.; Christiansen, F.T. The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster. *Immunogenetics* **2000**, *51*, 268–280. [[CrossRef](#)]
65. Hamada, M.; Shoguchi, E.; Shinzato, C.; Kawashima, T.; Miller, D.J.; Satoh, N. The complex NOD-like receptor repertoire of the coral *Acropora digitifera* includes novel domain combinations. *Mol. Biol. Evol.* **2013**, *30*, 167–176. [[CrossRef](#)] [[PubMed](#)]

66. Zhu, K.; Gao, B.; Zhu, S. Characterization of a chimeric antimicrobial peptide uncovers evolutionary significance of exon-shuffling. *Biochem. Biophys. Res. Commun.* **2012**, *428*, 360–364. [[CrossRef](#)]
67. Ranji, A.; Boris-Lawrie, K. RNA helicases: Emerging roles in viral replication and the host innate response. *RNA Biol.* **2010**, *7*, 775–787. [[CrossRef](#)] [[PubMed](#)]
68. Zambounis, A.; Elias, M.; Sterck, L.; Maumus, F.; Gachon, C.M. Highly dynamic exon shuffling in candidate pathogen receptors . . . what if brown algae were capable of adaptive immunity? *Mol. Biol. Evol.* **2012**, *29*, 1263–1276. [[CrossRef](#)]
69. Kambris, Z.; Hoffmann, J.A.; Imler, J.L.; Capovilla, M. Tissue and stage-specific expression of the Tolls in *Drosophila* embryos. *Gene Expr. Patterns GEP* **2002**, *2*, 311–317. [[CrossRef](#)]
70. Leulier, F.; Lemaitre, B. Toll-like receptors—taking an evolutionary approach. *Nat. Rev. Genet.* **2008**, *9*, 165–178. [[CrossRef](#)]
71. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)]
72. Letunic, I.; Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **2018**, *46*, D493–D496. [[CrossRef](#)] [[PubMed](#)]
73. Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.N.; Potter, S.C.; Finn, R.D.; et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641. [[CrossRef](#)] [[PubMed](#)]
74. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
75. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)] [[PubMed](#)]
76. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [[CrossRef](#)] [[PubMed](#)]
77. Langmead B|JG|b: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25. [[CrossRef](#)]
78. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *Gigascience* **2021**, *10*, giab008. [[CrossRef](#)]
79. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R.; Pimentel, H.; Salzberg, S.L.; Rinn, J.L.; Pachter, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **2012**, *7*, 562–578. [[CrossRef](#)]
80. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [[CrossRef](#)]
81. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)]