Genetic and molecular basis of the immune system in the brachiopod *Lingula anatina*

Marco Gerdol[1]*, Yi-Jyun Luo[2], Noriyuki Satoh[2], Alberto Pallavicini[1,3]

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

[2] Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan

[3] Anton Dohrn Zoological Station, Villa Comunale, 80121 Napoli,Italy

*corresponding author: mgerdol@units.it

Abstract

The extension of comparative immunology to non-model systems, such as mollusks and annelids, has revealed an unexpected diversity in the complement of immune receptors and effectors among evolutionary lineages. However, several lophotrochozoan phyla remain unexplored mainly due to the lack of genomic resources. The increasing accessibility of high-throughput sequencing technologies offers unique opportunities for extending genome-wide studies to non-model systems. As a result, the genome-based study of the immune system in brachiopods allows a better understanding of the alternative survival strategies developed by these immunologically neglected phyla. Here we present a detailed overview of the molecular components of the immune system identified in the genome of the brachiopod *Lingula anatina*. Our findings reveal conserved intracellular signaling pathways as well as unique strategies for pathogen detection and killing in brachiopods.

Keywords: Brachiopoda; Lophotrochozoa; innate immunity; immune system evolution; genome.

Abbreviations

AMP: antimicrobial peptide

AP-1: activator protein 1

BPI: bactericidal/permeability increasing protein

CARD: caspase recruitment domain

CARDIF: CARD adaptor inducing IFN-beta (CARDIF)

C1qDC: C1q domain-containing

CCF: coelomic cytolytic factor

CCP: complement control protein

cIAP1/2: cellular inhibitor of apoptosis protein 1/2

CLECT-DC: C-type lectin domain-containing

CRD: carbohydrate recognition domain

DAMPs: damage associated molecular patterns

dFADD: Fas-associated death domain-containing protein

DFD: death fold domain

IKK: IκB kinase

IL-17: interleukin-17

IL-1RL: Interleukin-1 receptor-like

IMD: immune deficiency

IPS-1: IFN-beta promoter stimulator

IRF: interferon responsive factor

ISRE: interferon-stimulated response element

FREP: fibrinogen-related protein

FReD: fibrinogen-related domain

GNBPs: Gram-negative binding proteins

LGBPs: lipopolysaccharide- and β-1,3-glucan binding proteins

LRRs: leucine-rich repeats

MACPF-DC: membrane attack complex/pore forming domain-containing

MAMPs: microbe-associated molecular patterns

MAPK: mitogen-activated protein kinase

MAVS: mitochondrial antiviral-signaling protein

mcc: multiple cysteine cluster

MIF: macrophage migration inhibitory factor

NF-κB: nuclear factor kappa-light-chain-enhancer of activated B cells

NLR: NACHT-LRR proteins

PAMPs: pathogen-associated molecular patterns

PGRP: peptidoglycan recognition protein

PGRP-PSM: peptidoglycan recognition protein putative signaling mediator

PRRs: pattern recognition receptors

RHIM: RIP homotypic interaction motif

RIG-I: retinoid inducible gene I

RIP: receptor-interacting protein

RLR: RIG-like receptor

scc: single cysteine cluster

SCR: scavenger receptor, cysteine-rich

STING: stimulator of interferon genes

SUEL: sea urchin egg lectin

TAK1: transforming-growth-factor-beta-activated kinase 1

TBK1: TANK-binding kinase

TGF-β: transforming growth factor β

TIR-DC: TIR domain-containing

TNFα: tumor necrosis factor alpha

TLR: Toll-like receptor

TEP: thioester-containing protein

TNFR: tumor necrosis factor receptor

TPM: transcripts per million

1. Introduction

The Lophotrochozoa superphylum (including mollusks, annelids, brachiopods, and other invertebrates) is one of the most diversified animal groups, only second to Ecdysozoa in numbers of living species. These organisms comprise a heterogeneous mix of monophyletic invertebrate phyla (Helmkampf et al., 2008; Philippe et al., 2005) that have virtually colonized all terrestrial and aquatic environments, displaying a remarkable biological plasticity and capability of adaptation. This diversity is exemplified by mollusks, which comprise over 20% of all marine species. Despite the ecological and evolutionary importance of lophotrochozoans, their immunological study has been limited compared to the other major groups of protostomes, such as Ecdysozoa. For what concerns Mollusca, early cellular studies (Canesi et al., 2002; Matricon-Gondran and Letocart, 1999; Takahashi and Muroga, 2008) have been followed by extensive molecular surveys in gastropods (Adema et al., 2017; Coustau et al., 2015; Pila et al., 2017), cephalopods (Castillo et al., 2015; Gestal and Castellanos-Martínez, 2015) and bivalves (Gerdol and Venier, 2015; Song et al., 2015; Zhang et al., 2015). A minor interest has been directed to Annelida, with studies targeting immune cells (Boidin-Wichlacz et al., 2011; Vetvicka and Sima, 2009) and immune genes (Altincicek and Vilcinskas, 2007; Nyholm et al., 2012; Tasiemski and Salzet, 2017) of polychaetes, oligochaetes and hirudinean worms.

Detailed data concerning the immune system of other lophotrochozoan phyla is lacking, except from occasional reports focused on specific gene families (Jeong et al., 2015). Brachiopods, comprising no more than 325 extant species, can be considered as part of these immunologically unexplored phyla. The phylogenetic position of brachiopods, in particular in relation with other lophotrochozoan phyla (i.e. Nemertea) has been a contentious matter for a very long time. The recent availability of fully sequenced genomes for brachiopods, nemerteans and phoronids finally allowed to clarify the phylogenetic position of Nemertea as a sister to the closely related phyla Brachiopoda and Phoronida (**Figure 1, panel E**) (Luo et al., 2018).

Despite their current sporadic distribution, brachiopods were extremely abundant in ancient seas, with over 30,000 species known by fossil records. While brachiopods underwent a massive reduction in their distribution, being out-competed by bivalves as a consequence of the Permian-Triassic extinction, some species survived by developing a series of morphological and physiological adaptations (Posenato et al., 2014). All extant brachiopods are marine organisms, which mostly live in continental shallow waters, usually attached to the substrate with their pedicles (**Figure 1, panel A and B**). Although brachiopods morphologically resemble bivalve mollusks, their two valves are dorso–ventrally oriented, in contrast to a lateral position in bivalves. Furthermore, like phoronids and bryozoans, they possess a horseshoe-shaped lophophore, a characteristic feeding structure which has long been used for the classification of these phyla within a taxonomic group named Lophophorata (Emig, 1997).

Anatomically, brachiopods have an open circulatory system, consisting of a dorsal vessel and several interconnected blood sinuses, which deliver nutrients and oxygen to various parts of the body (**Figure 1, panel D**). The blood and the colorless fluid contained in the coelomic cavity can mix, to some extent, as the two compartments communicate. In physiological conditions, three main types of circulating cells can be recognized in *L. anatina*: (i) erythrocytes, containing the respiratory pigment hemerythrin and abundant in the blood vessels of the mantle; (ii) spindle body cells, elongated and rich in fibres, the most abundant cell type in aquaria-maintained animals; (iii) amoebocytes, with a phagocytic activity, common in the coelomic cavity of the pedicle and characterized by the presence of electron-dense granules (Rowley and Hayward, 1985) (**Figure 1, panel C**). The presence of amoeboid coelomocytes in the pedicle coelom of brachiopods has been confirmed in different species, whereas this cell type appears to be extremely rare in the mantle cavity (Heller, 1931; Morse, 1902; Prenant, 1928; Yatsu, 1902). Although some studies have further classified amoebocytes among the hyaline, eosinophilic and basophilic subtypes (Ohuye, 1938), other researchers have

later pointed out that such morphological variations might be related to the degree of cell granulation (Rowley and Hayward, 1985). While brachiopod amoebocytes cells might be involved in immune functions (James et al., 1991), this has not been demonstrated. An immune-related function for these phagocytic cells would indeed be reminiscent to that of annelid coelomocytes, which are involved in the recognition and clearance of pathogens and waste material (Vetvicka and Sima, 2009). Brachiopod amoebocytes are also involved in the resorption of necrotic tissues (Chuang, 1983) and the shell regeneration process (Pan and Watabe, 1989).

The information available concerning pathogens and diseases affecting brachiopod species are also scarce and mostly derive from paleobiological observations. Brachiopods have been subject to intensive predation by drilling gastropods and/or polychaetes during their evolution (Baliński and Yuanlin, 2010; Baumiller and Bitner, 2004; Kiel, 2008; Leighton, 2003; Teichert, 1945) and predation by marine invertebrates appears to be rather relevant even in extant brachiopods under certain conditions (Tyler et al., 2013). Signs of recovery from shell drilling have been widely documented by fossil records (Alexander, 1986; Hiller, 2014; Peel, 2015), indicating that, despite the tissue damage sustained, brachiopods can survive and recover sub-lethal damages. Shell damage could lead to the exposure of injured tissues to the external environment, thereby enabling the direct contact with pathogens that are often responsible for disease outbreaks and even mass mortality events in other sessile marine invertebrates (Garnier et al., 2007; Paillard, 2004). Fossil records also show that Devonian brachiopods could be infested by non-boring vermiform metazoan parasites of uncertain taxonomic placement, as evidenced by the presence of tubular structures on the inner surface of valves, interpreted as the result of shell deposition on parasites infesting the mantle cavity and damaging the mantle tissue (Bassett et al., 2004; Vinn et al., 2014). Despite the little amount of reports available for parasitism in contemporary brachiopods, the observation of shells from dead animals suggests that many rhynchonelliform species living in Northern America suffer from significant parasitism by spionid polychaetes (Rodrigues, 2007). Furthermore, some brachiopod species have been also reported as occasional intermediate hosts of digenean trematodes (Cremonte et al., 2008).

At the same time, very little is known about mutualistic and commensalistic symbiosis in this phylum. Apart from epizoic foraminifers and other encrusting organisms which exploit feeding currents (Zumwalt and Delaca, 1980), a curious case of commensalism has been reported between *Laqueus rubellus* and the crab *Pinnotheres laquei*, which lives between the brachiopod valves without affecting shell deposition and mantle/lophopohore shape (Feldmann et al., 1996). The microbiota associated with brachiopod tissues represent a completely uncharted territory.

Overall, while the general knowledge of brachiopod biology is still very limited, this lophotrochozan phylum represents an interesting, yet unexplored, target for immunological studies, providing an alternative model to mollusks and roundworms, which might help to improve our understanding of the evolution of immune strategies in metazoans.

2. Materials and methods

2.1. Sequence data and identification of immune-related genes

The *L. anatina* genome and transcriptome assemblies were retrieved from a previous study (Luo et al., 2015). Briefly, the genome assembly was obtained by the use of a hybrid sequencing approach combining the outputs of Illumina, 454 Life Sciences reads and PacBio sequencing technologies. The initial assembly, performed with 454 and paired-end Illumina reads only, was carried out with Newbler. The preparation of

Illumina mate-pair libraries, together with additional 8.5 Gb PacBio data (consisting of reads spanning several Kb of sequence) were used for scaffolding, greatly improving the contiguity of the assembly, which was further refined to close gaps and remove redundancy. Methodological details about the assembly procedure can be found in the genome paper (Luo et al., 2015).

The 425-Mb *L. anatina* genome (v1.0) contains 34,105 coding genes, whose models were downloaded and imported into the CLC Genomics Workbench v10 environment (Qiagen, Hilden, Germany) for subsequent analyses. As an accessory sequence dataset, the transcriptome v1.0 (http://marinegenomics.oist.jp/), assembled with Trinity (r2013_08_14) (Grabherr et al., 2011) was also taken into account. Briefly, the Trinity assembly approach generates high quality, full-length sequences of transcripts expressed in a given tissue (or a combination of tissues, depending on the biological material available), also reporting alternatively spliced isoforms. In the *Lingula* genome paper, this data, obtained from multiple adult tissues and developmental stages (see section 2.2), was exploited in the AUGUSTUS pipeline for gene annotation, which evaluated the alignment of *de novo* assembled transcripts with the reference genome to predict coding regions (Luo et al., 2015).

Transcriptome data was used: (i) to assess the completeness of the proteins virtually predicted from gene models; (ii) to evaluate the presence of alternatively spliced isoforms; and (iii) to confirm the absence of specific genes that could not be identified within the set of predicted gene models.

Both the genome-derived gene models and transcripts were used to build sequence databases for subsequent BLAST similarity searches (Altschul et al., 1990). Immune-related sequences were identified based on the presence and organization of specific functional domains in the virtually-translated protein products identified with InterProScan 5 (Jones et al., 2014). The presence of signal peptides for secretion and transmembrane regions in the encoded proteins were evaluated, whenever necessary, with Phobius (Käll et al., 2004). Similarity-based searches were mainly performed using as queries relevant immunity-related protein sequences retrieved from representative vertebrate and invertebrate organisms, namely: *Homo sapiens* (vertebrates), *Branchiostoma floridae* (Cephalochordata), *Strongylocentrotus purpuratus* (Echinodermata), *Capitella teleta* (Lophotrochozoa, Annelida), *Crassostrea gigas* (Lophotrochozoa, Mollusca, Bivalvia), *Biomphalaria glabrata* (Lophotrochozoa, Mollusca,Gastropoda) and *Drosophila melanogaster* (Ecdysozoa, Insecta). In some cases, key sequences from additional species were also used. Matching hits were initially checked for completeness by the comparison with query sequences and further evaluated by the identification of homologous sequences in the transcriptome or the alternative genome annotation provided by the NCBI Eukaryotic Genome Annotation Pipeline (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Lingula_anatina/100/). Throughout the text, levels of sequence similarity are always intended at the amino acid level, unless differently stated.

2.2. Assessment of gene expression levels

The gene expression level of target genes of interest were calculated across seven adult tissues (lophophore, whole gut, digestive cecum, dorsal mantle, ventral mantle, pedicle and regenerated pedicle) and ten developmental stages (fertilized egg, 32-cell, 128-cell, early blastula, blastula, early gastrula, mid gastrula, late gastrula, one and two pair-of-cirri larvae) based on the RNA-seq data generated within the frame of the *L. anatina* genome sequencing project (Luo et al., 2015). Sequencing reads for each sample were imported into the CLC Genomics Workbench v10 environment. Here, they were trimmed by removing low-quality regions and residual adapter sequences, and individually mapped against the reference genome based on 0.75 and 0.98 length and similarity fraction thresholds. To enable full comparability of gene expression levels both within and among samples, raw counts were converted to transcripts per million (TPM) values, a

measure which is consistent with the relative molar concentration of each mRNA species (Wagner et al., 2012). Gene expression levels were square root-transformed to build heat maps.

2.3. Specialized downstream analyses

Whenever a sequence could not be detected in *L. anatina*, RNA-seq reads from other brachiopod were downloaded from the NCBI Sequence Read Archive (data retrieved in October 2017) including *Magellania venosa* (Jackson et al., 2015), *Laqueus californianus*, *Hemithiris psittacea*, *Glottidia pyramidata* (Halanych and Kocot, 2014), *Novocrania anomala*, *Terebratalia transversa*, *Kraussina rubra* and *Liothyrella uva* (unpublished) (**Table S1**). Paired-end reads were imported into the CLC Genomics Workbench v10 environment, trimmed based on quality and separately used to generate *de novo* transcriptome assemblies for each species, setting the *word size* and *bubble size* parameters to "automatic" and the minimum contig length to 300 nucleotides. The obtained transcriptomes were analyzed to calculate N50 statistics and used to build databases for BLAST searches.

On specific occasions, remote structural similarities were identified with HHpred (Söding et al., 2005) and structural modelling of the three-dimensional structure of *L. anatina* predicted proteins was carried out with Phyre 2 (Kelley et al., 2015) based on the templates deposited in the Protein Data Bank database.

Phylogenetic trees were constructed based on the multiple sequence alignment of the target amino acid sequences built with MUSCLE (Edgar, 2004) and refined with Gblocks v0.91b to remove non-alignable residues (Talavera and Castresana, 2007). Bayesian phylogenetic inference was performed with MrBayes v3.2 (Huelsenbeck and Ronquist, 2001), implementing the best-fitting molecular models of evolution estimated by ProtTest v2.4 (Darriba et al., 2011). The number of generations for the Markov chain Monte Carlo analyses was set based on the convergence of the two independent analyses run in parallel, evaluated by the reaching of a standard deviation of split sequences <0.05 and an effective sample size >200. All estimated parameters were inspected with Tracer v.1.6 (http://tree.bio.ed.ac.uk/software/tracer/). In the graphical representation of the phylogenetic trees, nodes supported by posterior probability values <50% were collapsed.

3. Results and discussion

3.1. General remarks

The metazoan innate immune system is based on a pathway consisting of receptors, signaling and effector molecules whose concerted action activates a coordinated response towards invading microorganisms and parasites. The initial involvement of specific immune cells and humoral factors in the recognition of microbe- and pathogen-associated molecular patterns (MAMPs and PAMPs) further reinforces the response at a systemic level, leading to the recruitment of other specialized cells at the site of infection and triggering the production of pro-inflammatory and antimicrobial factors. Depending on the nature of the invading microbes, the extent and the severity of the infection, the concomitant presence of tissue damage and the overall physiological status of the animal, the interplay between these factors direct the fate of the infected cells either towards survival or death.

The *L. anatina* genome contains extracellular putative pattern recognition receptors (PRRs), secretory lectin-like molecules characterized by extraordinary sequence diversity and recognition properties. These receptors are humoral factors acting in the extracellular space, which might be involved in the activation of a primitive complement-like system. However, the precise importance this defense mechanism

in invertebrates has not been fully elucidated yet. Soluble factors are assisted in their role by expanded families of membrane-bound receptors (including toll-like receptors) and by cytosolic PRRs. Regardless of the cellular and subcellular localization of PRRs, immune signals trigger a conserved signaling pathway, which includes the activation of the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) or the activator protein 1 (AP-1), either through the IκB kinase (IKK) complex or the mitogen-activated protein kinase (MAPK) cascade. This process leads to the production of immune effectors involved in the killing of microbes, i.e., antimicrobial peptides, lysozymes and pore-forming molecules, and of elusive cytokine-like molecules employed in the perpetration of the response at a systemic level.

Below, we provide a comparative summary of the number of genes involved in immune response in *L. anatina* and in three other representative lophotrochozoan species (the polychaete worm *C. teleta*, the gastropod mollusk *Biomphalaria glabrata*, and the bivalve mollusk *C. gigas*). As outgroups, we selected the fruit fly *D. melanogaster* and the sea urchin *S. purpuratus*, which genome has been the subject of extensive studies (Hibino et al., 2006) (**Figure 2**).

The cDNA sequences of the major molecular players of the *L. anatina* immune system are reported in **Supplementary File 1**.

3.2. Extracellular PRRs

3.2.1. Fibrinogen- related domain-containing proteins

Fibrinogen-related domain containing (FReD-C) proteins are a heterogeneous group of proteins found in all metazoans that share sequence similarity with vertebrate ficolins, soluble lectins with a collagen and a fibrinogen-like C-terminal domain, involved in the activation of the lectin pathway of the complement system (Matsushita, 2010). Some FRED-C proteins have been implicated in PAMP recognition in some invertebrates, a function that has been first demonstrated to be crucial in modulating the interplay between gastropod mollusks and parasitic trematodes in a particular subfamily of molecules named fibrinogen-related proteins (FREPs) (Adema et al., 1997; Hanington and Zhang, 2011). The immune role of these PRRs has been recently investigated also in bivalves, where they are a greatly expanded gene family and lack the N-terminal immunoglobulin domain typical of *bona fide* gastropod FREPs (Huang et al., 2015; Romero et al., 2011). Notably, gastropod FREPs are highly polymorphic. In the snail *Biomphalaria glabrata*, their diversification occurs through somatic mutation, possibly serving as an anticipatory mechanism of immunity (Adema, 2015; Pila et al., 2017). Notably, *B. glabrata* FREPs, together with C-type lectin-related proteins (CREPs) and galectin-related proteins (GREPs), create an interesting group of molecules named VIgL (Variable Immunoglobulin and Lectin domain containing molecules), due to the shared presence of Ig-superfamily domains in the N-terminal region, combined with a C-terminal lectin-like domain (Dheilly et al., 2015). Consistently with the absence of similar proteins in bivalves (Gerdol, 2017), no CREP- or GREP-like sequences could be identified in *L. anatina*, reinforcing the hypothesis that these recognition proteins are gastropod-specific innovations.

The immune role of FReD-C proteins in invertebrates is not limited to mollusks, as two proteins with a similar structure to molluscan FReD-C proteins and ficolins, named tachylectin-5A and B, contribute to approximately two-thirds of hemagglutination activity in horseshoe crabs (Kawabata et al., 2009), likely reflecting the common origin of vertebrate blood coagulation and invertebrate hemocyte agglutination.

We identified 158 *L. anatina* genes that contain fibrinogen C-terminal globular domains and display a variable domain architecture. Some *L. anatina* FReD-C proteins share a remarkable sequence similarity

(~50–55% identity at the amino acid level) with *bona fide* gastropod FREPs, tachylectins, ficolins and ryncolins (a family of snake venoms interfering with the coagulation cascade), suggesting a role possibly linked to hemagglutination. However, none of these proteins displays N-terminal immunoglobulin or collagen domains, which characterize gastropod FREPs and vertebrate ficolins, respectively. It is noteworthy that a subset of approximately 35 *L. anatina* FReD-C proteins are highly similar (>60% sequence identity at the amino acid level in pairwise alignments), and therefore likely result from a recent gene family expansion event.

3.2.2. C1q domain containing proteins

C1q is a highly plastic globular domain that has been recently identified as a PRR module in many protostomes. In particular, C1q domain-containing (C1qDC) genes have undergone massive expansion in bivalves, where they represent the major class of extracellular lectin-like molecules (Gerdol et al., 2015). However, C1qDC genes are relatively uncommon in other vertebrate groups, and independent gene family expansion events have led to the acquisition of a large complement of C1qDC genes in different taxa (Gerdol et al., 2011).The *L. anatina* C1qDC family did not undergo extensive expansion with only 27 genes identified. The most common domain architecture observed in brachiopods corresponds to secreted C1q-like proteins (bearing an N-terminal signal peptide, followed by collagen repeats and a C-terminal globular C1q domain). Although this structural organization is uncommon in bivalves, where collagen repeats are replaced by functionally homologous coiled-coil regions, it is found in many vertebrate C1qDC proteins. These includes the three major components of the complement C1q complex (C1qA, C1qB, and C1qC), which are activated in response to the interaction with antigen-bound immunoglobulins. Due to their potential role as PRRs, C1qDC proteins might be part of an ancestral complement system, where they could overcome the lack of immunoglobulins by directly recognizing PAMPs.

However, in the absence of functional data and relevant sequence similarity to other functionally characterized C1qDC genes, the role of these molecules in brachiopods remains a matter of speculation. Indeed, as demonstrated by the presence of a C1qDC gene with chemotactic properties involved in nerve growth in annelid worms (Tahtouh et al., 2009), lophotrochozoan C1qDC genes are not exclusively linked to immune functions.

3.2.3. C-type lectins

C-type lectin domain-containing (CLECT-DC) proteins represent a large family of heterogeneous molecules characterized by the presence of a CLECT carbohydrate recognition domain (CRD), often combined with other structural motifs. While their variable domain architecture can lead to a remarkable functional diversity in Lophotrochozoa, many C-type lectins act as PRRs in marine invertebrates (Sekine et al., 2001; Yuasa et al., 1998). This function is in line with that covered by mannose-binding lectins in the vertebrate complement system (Fraser et al., 1998). Like other protostomes, the genome of *L. anatina* contains over 300 CLECT-DC genes, with a variable domain architecture and little homology (~30-35% at the amino acid level) to sequences with a well-characterized function in invertebrates. We present here a few recurrent domain architectures that might warrant further research concerning a possible involvement in the immune system (**Figure 3**).

One of the most common domain organizations comprises a short signal peptide for secretion, immediately followed by the CRD domain. In many cases, the CLECT domain is associated with an N-terminal coiled-coiled region, which might provide the functional equivalent to the collagen region of mannose-

binding lectins, enabling oligomerization. Despite low primary sequence homology, these CLECT-DC proteins show a structural organization very similar to lectins involved in PAMP recognition in other metazoans (Fraser et al., 1998; Sekine et al., 2001; Yuasa et al., 1998). The CLECT domain is often associated with other domains with diverse function, or with low complexity regions. While some of these genes (e.g. those containing von Willebrand factor type A domains) are likely to be involved in cellular adhesion as extracellular matrix proteins, the functional role of several others is still unknown. Some secreted CLECT-DC proteins are very large and comprise more than one thousand residues. In these cases, the CLECT domain is often combined with other putative carbohydrate-binding domains—e.g. the WSC domain, and with a large number of complement control protein (CCP) modules (Reid and Day, 1989), supporting their involvement in recognition and/or adhesion processes.

We found that about 20 *L. anatina* genes bear a variable number of consecutive CLECT modules (6 to 17), resulting in protein products of up to 2,500 amino acids. These large proteins, containing a single C-terminal transmembrane domain, display a remarkable structural similarity with vertebrate macrophage mannose receptors involved in the recognition of non-self and the activation of phagocytosis (Fraser et al., 1998). The main difference between the *L. anatina* mannose-receptor-like proteins and their vertebrate counterparts is the lack of the N-terminal Ricin-B-like/Fibronectin type II domain tandem. While the structural homology between the two receptors might be the result of convergent evolution (no macrophage mannose receptor has ever been identified in basal deuterostomes), such a remarkable similarity certainly warrants future functional investigations. Moreover, some of these genes combine CLECT with F-type lectin-like and LCCL domains, which have also been implicated in immune response (Trexler et al., 2000). Other unusual CLECT-DC membrane-bound receptors contain multiple CCP modules, suggesting a possible involvement in the control of the complement system.

3.2.4. Mytilectins

Mytilectins, a novel group of α-D-galactose-binding lectins recently described in mussels (*Mytilus* spp., Mollusca), can bind human lymphoma cells and trigger their apoptosis (Fujii et al., 2012). The structure of mytilectins is quite simple, as it comprises three tandemly-repeated carbohydrate-binding units structurally similar to the Ricin-B lectin domain, and it sometimes also includes an additional C-terminal domain that shows a suggestive resemblance to pore-forming bacterial toxins (Hasan et al., 2016). The presence of carbohydrate recognition and pore-forming domains within the same molecule has important implications on the possible involvement of these lectins in the activation of molecular machinery aimed at the killing of invading microbes similar to the terminal pathway of the complement system.

Interestingly, mytilectins appear to have a narrow taxonomical distribution, limited to some bivalve species. To the best of our knowledge, brachiopods become the second phylum where this group of lectins has been recognized. At least seven complete distinct loci exist in *L. anatina*, all encoding ~300 amino acids long protein precursors which are structurally similar to each other and include a C-terminal pore-forming domain. In spite of a limited amino acid primary sequence identity (30–35%), bivalve and brachiopod mytilectins share a highly conserved three-dimensional fold (**Figure S1**). However, in contrast with bivalves, *L. anatina* mytilectins bear a signal peptide, which would determine their targeting to the extracellular space. It is also worth mentioning that the C-terminal pore-forming domain of mytilectins, despite an insignificant primary sequence homology, is structurally similar to biomphalysin, a pore-forming toxin produced by the snail *B. glabrata* to kill *Schistosoma mansoni* larvae (Galinier et al., 2013).

3.2.5. Galectins

Galectins are a class of lectins binding β-galactoside residues, ancient widespread in metazoans (Müller et al., 1997), which hold a dual role as PRRs and damage associated molecular patterns (DAMPs) (Sato et al., 2009). Some invertebrate galectins, e.g. BgGal from the snail *B. glabrata*, have been specifically implicated in the recognition of pathogen-associated sugars (Yoshino et al., 2008). Most of the invertebrate galectins described so far and, in particular, those involved in PAMP recognition in mollusks, comprise either one, two or four consecutive galectin domains, with no evident signal peptide, as they are usually targeted to the extracellular space by the leaderless secretion pathway. The four *L. anatina* galectin genes contain a domain architecture similar to that of mollusks. One of these genes encodes a galectin with two CRDs, two other genes encode galectins with four CRDs, and one corresponds to an unusual lectin with three domains, with the first and the second one separated by a low complexity region. It is noteworthy that galectins with four CRDs are absent in the genomes of annelids and nemerteans, leaving brachiopods, phoronids and mollusks as the only three known lophotrochozoan phyla where such a domain combination is present. Due to their high sequence homology and highly conserved function in all metazoans, it is reasonable to assume that brachiopod galectins also share functional similarity with their molluscan homologs.

3.2.6. Hydrolase family 16 lectins

The lectin-like molecules characterized by the presence of a conserved glycoside hydrolase family 16 (IPR000757) share high homology with insect Gram-negative binding proteins (GNBPs). In *Drosophila* immunity, GNBPs play an important role, in collaboration with PGRPs, by recognizing bacterial components (Lys-type peptidoglycan and β-glucans) and triggering the extracellular proteolytic cascade which leads to the activation of Toll and melanization (Ferrandon et al., 2007). This function has only been elucidated so far in arthropods, whereas in other metazoans, TLRs are likely able to directly recognize PAMPs without the intervention of the cytokine Spätzle, the final product of the aforementioned proteolytic cascade. Proteins similar to GNBPs are however present in different metazoan phyla, even though in some cases they have no role in immunity, being rather used for digestive processes as glucanases (Hughes, 2012).

As multiple GNBP-like sequences have been reported in lophotrochozoans, either with a function as microbial sensors (and therefore named lipopolysaccharide- and β-1,3-glucan binding proteins, LGBPs) or as endo β-1,3-glucanases, we relied on phylogeny to assign the six genes identified in *L. anatina* to either of the two categories (**Figure 4**). None of the brachiopod sequences fell within the glucanase family, ruling out the involvement of *L. anatina* GNBPs in digestive processes. On the other hand, three sequences fell within the LGBP clade, together with some sequences previously described in mollusks as pattern recognition proteins and upregulated in response to bacterial challenges (Nikapitiya et al., 2010; Zhang et al., 2010).

The remaining three brachiopod sequences were placed in the same branch of annelid coelomic cytolytic factors (CCFs). Although the posterior probability was not significant for supporting this clade, the structural similarity between the three brachiopod sequences and CCFs, given by the lack of the N-terminal glucan-binding domain found in LGBPs and GNBPs alike, reinforces their homology. CCFs, first recognized in the coelomic fluid of *Eisenia fetida* (Bilej et al., 1995), play a fundamental role in the immunity of earthworms, where they act as broad-spectrum PRRs. Indeed, they can recognize cell wall components of Gram-positive and -negative bacteria (lipopolysaccharides and peptidoglycan), as well as fungi (β-1,3 glucans) (Šilerová et al., 2006), thereby activating the melanization cascade (Beschin et al., 1998).

While it remains to be confirmed whether *L. anatina* CCF-like proteins are functionally homologous to earthworm CCFs, the results of the phylogenetic analysis are consistent with a likely involvement of all the six identified brachiopod genes in immune processes. Although this observation finds a parallelism with the previous identification of lipopolysaccharide-binding proteins in blood cell lysates of *Terebratulina* sp.

(Bakholdina et al., 2014), the specific role of the two divergent groups of brachiopod GNBPs in PAMP recognition and their possible activation of the melanization cascade remains to be fully elucidated.

3.2.7. Other lectins possibly involved in pathogen recognition

The inference of possible functions of lectin-like molecules in brachiopods is complicated by the frequent combination of the main CRD with various other conserved domains and by limited sequence homology with well-described counterparts in other animals. Among these, F-type lectins, occasionally implicated in immune recognition in invertebrates (Vasta et al., 2012), represent an exemplifying case, as they comprise about 50 members in *L. anatina* with variable domain combinations. Similarly, the sea urchin egg lectin (SUEL) rhamnose-binding CRD, commonly found in invertebrate lectins as a system of protection for fertilized eggs from infection (Tateno, 2010), is present in *L. anatina*. The egg-protection role is covered in gastropods by H-type lectins. However, the study of this lectin family in Lophotrochozoa has been limited, and no functional information has been collected for non-molluscan invertebrates (Gerdol, 2017; Gerlach et al., 2005). Only two H-type lectin genes are present in *L. anatina*, and they share limited sequence homology with the agglutinins of land snails.

Moreover, proteins bearing an apextrin-like C-terminal (ApeC) domain recently emerged as possible PRRs in amphioxus (Huang et al., 2014) and a lectin-like function has also been tentatively assigned to apextrin-like molecules in bivalves (Gerdol and Venier, 2015). In contrast with molluscan genomes, which encode several apextrin-like proteins, only a single secreted ApeC gene was found in *L. anatina*. This gene bears homology with those of amphioxus, supporting the possible role of this lectin-like molecule as a PRR. Finally, a novel class of lectins recently described in the gastropod *Aplysia* (Carneiro et al., 2017; Motohashi et al., 2017), is also present in *L. anatina*. These carbohydrate-binding proteins, containing the conserved domain, DUF3011, are absent in the genomes of other gastropods and have not been described in other invertebrates. The genome of *L. anatina* contains a few members of this family, although their precise number could not be assessed due to uncertainties in the prediction of these gene models. They appear as secretory proteins of about 160 amino acids, which could be involved in the protection of eggs from bacterial infections, like that in *Aplysia*.

3.3. The complement-like system of brachiopods

In vertebrates, the complement system certainly represents the most important effector component of the innate immune response, which has often been described as a functional link between innate and acquired immunity (Dunkelberger and Song, 2009). This highly complex molecular mechanism, aimed at the opsonization and killing of pathogens and at the activation of inflammatory processes, can be fundamentally activated through three ways: (i) the classical pathway, mediated by the binding of C1q to an antigen-complexed immunoglobulin; (ii) the alternative pathway, triggered by the binding of microbes by the C3b complement component; and (iii) the lectin pathway, mediated by PAMP recognition by ficolins and mannose-binding lectins.

Due to the lack of immunoglobulins and other crucial molecules of the complement system in invertebrates, it has long been debated whether this system was an innovation of the vertebrate lineage or if an ancestral, simplified system was already present in the common ancestor of all metazoans. The first pieces of evidence supporting the latter hypothesis came from the observation that two key components, C3 and Factor B, are conserved in both vertebrates and invertebrates, although *bona fide* C3 components are crucially missing in several arthropods (Sekiguchi et al., 2012). This suggests the presence of both the

alternative and of the lectin pathways in the most recent common ancestor of all animals (Smith et al., 1999). These early observations were later confirmed in cnidarians, echinoderms, urochordates, cephalochordates (Pinto et al., 2007) and, more recently, also in some lophotrochozoan phyla (Prado-Alvarez et al., 2009).

Although the existence of a classical complement pathway is ruled out in all invertebrates by the lack of immunoglobulins, C1qDC proteins, often expanded and highly diversified, may potentially be involved in opsonization and pathogen clearance, like ficolins and mannose-binding lectins. However, due to the lack of convincing sequence homology and the absence of a series of proteases required for the activation of the complement lectin pathway (such as mannose-associated serine protease), the hypothesis that these lectin-like molecules are effectively involved in the activation of the brachiopod complement system remains a matter of speculation in absence of functional evidence. Regardless of the involvement of C1qDC proteins, FReD-C proteins, C-type lectins or any other PRRs in a lectin-like complement pathway, the two major central components of the complement system, Factor B and C3 are notably both present in *L. anatina* as single-copy genes. The brachiopod Factor B is highly similar to those of vertebrates, with the only difference given by the higher number of CCP modules present.

C3 and the accessory complement proteins C4, C5 and α_2 macroglobulin, pertain to the family of thioester-containing proteins (TEPs), characterized by the presence of a unique and conserved thioester motif (GCGEQ) which confers the ability to form covalent bonds with macromolecules (Levine and Dodds, 1990). As previously reported by other authors, C3 is highly conserved across most metazoans (with the notable exception of arthropods), and the brachiopod genome contains a phylogenetically supported, *bona fide*, C3 orthologous gene, which encodes a protein displaying both a conserved thioester motif and a C-terminal netrin domain, which differentiates C3/C4/C5 family proteins from other TEPs (**Figure S2**). Although the *L. anatina* C3 is the lone member of the C3/C4/C5 clade, at least seven other TEPs are encoded by the genome of this brachiopod (**Figure S2**), including a genuine α_2 macroglobulin, another large plasma protein potentially functioning as a powerful inhibitor of virulence factors and proteases produced by invading microbes (Armstrong and Quigley, 1999). The other *Lingula* TEPs were placed by the phylogenetic analysis with high confidence within the CD109/iTEP clade, which includes vertebrate CD109 proteins, negative regulators of the TGF-β signaling (Bizet et al., 2011), and invertebrate TEPs (iTEPs) that promote opsonization and phagocytosis (Blandin et al., 2008; Bou Aoun et al., 2010; Zhang et al., 2009). This observation warrants further investigations, especially taking into account the partial conservation of the thioester motif in some brachiopod sequences (**Figure S2**).

The mode of activation of the complement system in protostomes has been extensively studied in the horseshoe crab *Tachypleus tridentatus*, where it can be triggered either by Gram-positive bacterial infection in a Factor B-dependent way (Ariki et al., 2008) or by Gram-negative bacterial infection in a Factor B-independent way, through the lipopolysaccharide-sensitive clotting Factor C (Tagawa et al., 2012). A previous study, targeting Ecdysozoa, demonstrated that the Factor C-dependent pathway is only present in Chelicerata, while it was lost (along with C3 and Factor B) in Pancrustacea and most Myriapoda (Sekiguchi and Nonaka, 2015). Surprisingly, a single-copy gene produces a Factor C-like protein in *L. anatina*. Although this protein lacks the C-type lectin and LCCL domains typical of *T. tridentatus* Factor C (**Figure 5**), it presents the same domain architecture of the Factor C-like protein of the centipede *Scolopendra subspinipes*, which has also been suggested to be involved in the complement system (Sekiguchi and Nonaka, 2015). To the best of our knowledge, this is the first report of a Factor C-like protein outside arthropods, as well as the first piece of evidence supporting the existence of a Factor B-independent alternative pathway of the complement system in Lophotrochozoa. Upon further scrutiny, we could detect Factor C-like genes with a peppered taxonomical distribution in the genomes of other non-arthropod invertebrates, i.e. the lophotrochozoans *M. yessoensis* and *Phoronis australis*, the priapulid *Priapulus caudatus* and the amphioxus *Branchiostoma belcheri*. No similar gene was however present in several other genomes, revealing a. Phylogenetic inference

clarified that such genes were placed well within the Factor C clade, together with all the arthropod sequences previously described by Sekiguchi and Nonaka (2015) and well distinct from other animal serine proteases involved in blood clotting (**Figure S3**).

Overall, the *L. anatina* genome allows us to hypothesize the existence of an ancient lectin pathway of the complement system in brachiopods, similar to the one partially unveiled in other invertebrates (Pinto et al., 2007). This system would consist of C1qDC proteins, FReD-C proteins or C-type lectins, which might trigger the activation of C3 thanks to the action of presently unknown proteases. In addition to that, a Factor C-like protein might directly mediate the activation of the complement system in response to Gram-negative bacteria (**Figure 5**).

In vertebrates, the terminal pathway of the complement system consists of proteins related to the MACPF superfamily (C6-C9), which form the membrane attack complex (MAC), leading to the killing of Gram-negative bacteria through lytic mechanisms (Tschopp et al., 1986). Although a number of proteins pertaining to this superfamily have been already evidenced in different invertebrate phyla (Mah et al., 2004), they only show a remote sequence similarity to MAC components. In total, seven brachiopod genes encode proteins related to the MACPF superfamily. Three of these share high similarity to MPEG1, an evolutionarily conserved gene (Wiens et al., 2005) that is thought to be involved in the clearance of phagocytosed bacterial cells (He et al., 2011). However, the remaining four brachiopod MACPF proteins are largely divergent from any other known perforin studied so far: three are long secretory proteins that combine the pore-forming domain with multiple furin-like and EGF domains, whereas a single copy gene encodes a protein where the MACPF domain is associated with a C-terminal SCR domain.

Among the other possible alternative pore-forming effectors, mytilectins are likely to lead to the lysis of pathogenic microbes upon recognition, without the involvement of accessory molecules in the brachiopod complement system. Due to their previously demonstrated agglutinating properties (Hasan et al., 2016), mytilectins might be considered as likely accessory components of the brachiopod complement system (**Figure 5**). As previously mentioned, the C-terminal domain of mytilectins shares a remote structural similarity with biomphalysins, pore-forming molecules involved in immune defense in gastropods (Galinier et al., 2013), due to the presence of an aerolysin-like fold. However, no *bona fide* biomphalysin sequence is present in brachiopods. Another interesting finding in the context of pore-forming molecules was the presence of a few genes encoding members of the actinoporin family, sharing significant similarity with the clamlysin, a lytic protein produced by freshwater clams whose physiological function is presently unknown (Takara et al., 2011).

3.4. Membrane-bound receptors and related intracellular signaling pathways

3.4.1. Peptidoglycan recognition receptors and the IMD-like pathway

The *N*-acetylmuramoyl-L-alanine amidase domain characterizes a heterogeneous group of metazoan membrane-bound or secreted proteins collectively known as peptidoglycan recognition receptors (PGRPs). The interplay between PGRPs and Gram-negative binding proteins (GNBPs) has been well characterized in *Drosophila*, where this system can modulate the melanizing prophenoloxidase cascade, the cleavage of the Spätzle pro-cytokine and, subsequently, the Toll pathway in response to infections by Gram-positive bacteria. On the other hand, when Gram-negative bacteria infections occur, membrane-bound PGRPs are directly responsible for the activation of the immune deficiency (IMD)-mediated intracellular signaling cascade (Royet and Dziarski, 2007), which shares many molecular components with the vertebrate tumor necrosis factor

receptor (TNFR) pathway. Although PGRPs are also present in vertebrates, they are not bound to the cell membrane, and they are not able to convey the immune signal to the cytosol, rather functioning as direct bactericidal and bacteriostatic agents in the extracellular space thanks to their *N*-acetylmuramoyl-L-alanine amidase activity (Montaño et al., 2011).

While the presence of secreted PGRPs involved in pathogen recognition has been previously evidenced in lophotrochozoans (Ni et al., 2007), only fragmentary evidence exists for membrane-bound PGRPs (Gerdol and Venier, 2015). Furthermore, their involvement in an IMD-like pathway is still uncertain due to the absence of an IMD-like gene, which we can confirm also brachiopods. The genome of *L. anatina* has at least 12 PGRP genes that encode a peculiar repertoire of molecules compared to both arthropods and mollusks (Werner et al., 2000) (**Figure S4**).

In detail, two *L. anatina* genes produce secreted proteins with a single PGRP domain, orthologous to a PGRP previously described in the mussel *Mytilus galloprovincialis* (Gerdol and Venier, 2015) and only sharing remote similarity with arthropod PGRPs. Two additional brachiopod-specific genes encode complex secreted proteins with four highly similar and consecutive PGRP domains. The combination of the PGRP domain with three consecutive N-terminal scavenger receptor, cysteine-rich (SCR) domains is a domain combination uniquely found in brachiopods. The SCR domain is widespread in animals, and it displays a broad range of functions, including pathogen recognition and mediation of phagocytosis, thereby reinforcing the idea that this protein could be involved in the innate immune system (Canton et al., 2013). Brachiopods also possess membrane-bound PGRPs, as seven distinct genes of this type are present in *L. anatina*.

Arthropods use a signaling adaptor, IMD, to transduce immune signaling inside the cell thanks to the heterotypic interaction between the intracellular receptor-interacting protein (RIP) homotypic interaction motif (RHIM) domains of PGRPs with IMD. Consequently, the C-terminal DEATH domain of IMD triggers the downstream signaling cascade by interacting with the DEATH domain of *Drosophila* Fas-associated death domain-containing protein (dFADD). However, evidence concerning the existence of this PGRP/IMD/dFADD based system has so far remained elusive in lophotrochozoans, despite the conservation of all the downstream intracellular components of the signaling cascade. In our previous survey of the mussel immune genes repertoire, we suggested that, if present, such a system would need an alternative and still unknown intracellular adaptor due to the lack of an IMD homolog and of intracellular RHIM domains required for IMD recruitment in molluscan membrane-bound PGRPs (Gerdol and Venier, 2015). Like mussels, also *L. anatina* possess four membrane-bound PGRPs with an intracellular region without any recognizable conserved domain. However, the brachiopod genome also encodes three transmembrane proteins with a well-recognizable intracellular DEATH domain, which could potentially recruit dFADD-like molecules, thereby bypassing the need of accessory IMD-like adaptor proteins.

This finding is particularly important, as it suggests for the first time that a PGRP-based IMD-like signaling might exist in Lophotrochozoa. If confirmed by experimental evidence, this hypothetical system would not require the presence of any intracellular adaptor protein, relying on the direct heterotypic interaction between the DEATH domain of these novel receptors and dFADD, as depicted in **Figure 6**. From now on, we indicate these three brachiopod genes as PGRP-PSM (Peptidoglycan Recognition Protein Putative Signaling Mediator), waiting for a functional confirmation of its proposed mechanism of action. It remains to be elucidated whether the other brachiopod membrane-bound PGRPs devoid of a DEATH domain can recruit dFADD or other functionally homologous adaptors in a similar fashion. The recruitment of dFADD consequently determines the activation of the caspase DREDD, which can ultimately lead to the activation of the pro-inflammatory transcription factor Relish upon the cleavage of its C-terminal ankyrin repeats region by DREDD. Alternatively, the interaction between IMD and dFADD (and therefore, hypothetically, between PGRP-PSM and dFADD in brachiopods) can lead to the activation of the transforming-growth-factor-beta-

activated kinase 1 (TAK1), thereby converging to the signaling pathway activated downstream of Toll-like receptors.

As briefly mentioned above, the *Drosophila* IMD signaling pathway is thought to cover a homologous function to that of the TNFα signaling pathway of vertebrates, even though the recognition receptors are different (Lemaitre and Hoffmann, 2007). Indeed, given the absence of IMD-like molecules and the different function of PGRPs in vertebrates, a different class of receptors, called TNFRs, activates this system. These cytokine receptors are primarily linked to the regulation of apoptosis and inflammation upon the binding with their ligands, some of which will be described in the following sections. A thorough discussion of the molecular mechanisms possibly leading to apoptosis through TNFRs in brachiopods goes beyond the scope of this manuscript. However, it is worth briefly exploring whether such receptors are present in the genome of *L. anatina*, or if the activation of TAK1 can be reasonably assumed to occur only in response to signals conveyed to the cytosol by PGRP-PSM. First, similarly to what has been previously reported in mollusks (Li et al., 2009), no membrane-bound protein shows remarkable sequence similarity with the vertebrate TNF receptors TNFR1 and TNFR2, even though three brachiopod proteins share the same domain organization, with extracellular TNFR repeats and an intracellular DEATH domain for signal transduction. A single-copy gene encoding a receptor with TNFR ectodomains and an intracellular TIR domain, which is typically involved in immune signaling, provides a potential candidate. Therefore, although some TNFR ligands are encoded by the *L. anatina* genome, the nature and the precise function of their receptors presently remains uncertain.

Consistently with the lack of a convincing TNFR1/2 homolog, TRADD, a direct interactor with the intracellular DEATH domain of TNFRs and a key player in the recruitment of the components of the TNFR associated complex I, is missing, not just in brachiopods but also in the genome of all other protostomes. However, despite the lack of TRADD and clear TNFR homologs, other major components of TNFR-associated complexes are present and highly conserved: for example, as mentioned above, many different TRAF genes are encoded by the *L. anatina* genome, including TRAF2 homologs. Several homologs to the cellular inhibitor of apoptosis protein 1/2 (cIAP1/2), responsible of the ubiquitination of the key molecule RIP-1 kinase (RIPK1) are also present and evolutionarily well conserved. Crucially, a RIPK1 sequence appears to be present and relatively well conserved when compared to that of amphioxus (Li et al., 2011). RIPK1 is the key regulator of the TNFR signaling, as it acts as a major switch between cell death and survival. Indeed, the complex formed by TNFR, TRADD, TRAF2 and cIAP1/2 recruits and activates TAK1, switching on canonical NF-κB signaling and leading to survival. On the other hand, the activity of the linear ubiquitin chain assembly complex (LUBAC) leads to the recruitment of other cytosolic factors to form a macromolecular complex (the TNFR-associated complex II), which triggers apoptosis (Conrad et al., 2016). While previous efforts failed to identify a RIPK1-like sequence in mollusks (Gerdol and Venier, 2015) and echinoderms (Hibino et al., 2006), the report of this molecule in brachiopods is corroborated by the presence of a similar protein in the genomes of the segmented worm *C. teleta* and the phoronid *P. australis*. This finding indicates that RIPK1 might have emerged as an important molecular switch between cell death and survival in basal protostomes, even in the absence of vertebrate-like TNFRs and TRADD, whose functional homologs might be highly divergent or even missing in protostomes.

A homolog to Relish, the transcription factor activated in the final steps of IMD signaling in Drosophila, is present in *L. anatina*. As previously suggested by other authors, the presence of Relish in Lophotrochozoa might provide support to the existence of an IMD-pathway (Zhang and Coultas, 2011). However, we show that the brachiopod Relish homolog (and, by extension, the orthologous sequences in mollusks and annelids) shares a higher degree of similarity with deuterostome p100/p105-like proteins, which is also confirmed by the presence of a C-terminal DEATH domain, absent in arthropods (**Figure S5**).

Overall, the investigation of IMD/TNFR signaling-related genes in *L. anatina* highlights that: (i) an IMD-like system, similar to that of *Drosophila*, may be activated downstream of membrane-bound PGRPs

without the involvement of an IMD-like adaptor, and (ii) a similar signaling complex may be activated downstream of TNFR-like receptors and without the involvement of TRADD, as suggested by the presence of highly conserved TRAF2 and RIPK1. However, the identity of these receptors and the hypothetical adaptor molecules bridging TNFRs with RIPK1 is still uncertain.

3.4.2. Toll-like receptors and activation of the pro-inflammatory response through the IKK complex and MAPK

Toll-like receptors (TLRs) are the most successful type of metazoan PRRs. These membrane-bound receptors can recognize a broad range of PAMPs through extracellular leucine-rich repeats (LRRs) and trigger immune signaling by the heterotypic interaction of their TIR domain with intracellular adaptor proteins. This signaling cascade finally results in the activation of the pro-inflammatory response and the transcription of effector genes involved in antimicrobial defense. The TLR gene family appears to have independently undergone great expansion in different invertebrate phyla (Dishaw et al., 2012; Gerdol et al., 2017a; Hibino et al., 2006; Zhang et al., 2015), broadening the range of PAMPs recognized and conferring an highly complex and specific immune response.

A previously published study focused on the identification of TLRs in neglected lophotrochozoan taxa reported the presence of a limited number of TLRs in the transcriptomes of some brachiopod species (Halanych and Kocot, 2014). More recently, we have highlighted the presence of over 30 membrane-bound TIR-domain-containing proteins in *L. anatina*, including several TLRs (Gerdol et al., 2017a). The in-depth analysis of the genome performed in this study permitted to refine this analysis and to identify 49 different complete TLR genes and a handful of truncated pseudogenes, which overall place brachiopods somewhere in between annelids (which have less than 10 TLRs) and mollusks (which may even possess over 100 TLRs in some cases). *L. anatina* TLRs can be conveniently divided into two categories based on the architecture of their ectodomains: single cysteine cluster (scc) and multiple cysteine cluster (mcc) TLRs. The former group comprises receptors similar to vertebrate TLRs, with typical LRRs flanked by N-terminal and C-terminal LRR domains. The latter includes evolutionarily ancient receptors similar to *Drosophila* Toll, which display two separated clusters of N-terminal and C-terminal LRRs. Forty-one *L. anatina* receptors can be classified as sccTLRs, consistently with previous reports about the expansion of this TLR subgroup in lophotrochozoans (Gerdol et al., 2017a).

However, brachiopod sccTLRs clustered in two largely divergent phylogenetic groups (**Figure 7**) that are also characterized by a different gene architecture; the members of the first and largest group are encoded by intronless genes, like in sea urchin (Hibino et al., 2006). The second group, on the other hand, consists of genes bearing a single intron located in a conserved position, dividing the exon encoding LRRs and the transmembrane domain from the one encoding the TIR domain (**Figure S6**). Both types of sccTLRs appear to have evolved through recent rounds of gene duplication, as evidenced by the presence of several highly conserved gene copies in a cluster organization within the same genomic scaffolds. The high degree of similarity among all two-exons/one-intron sccTLRs further suggests that all these genes have been originated by a relatively recent event. The seven *L. anatina* mccTLRs displayed a higher diversity and clustered in an independent clade (**Figure 7**).

It has been possible to fully delineate an almost complete machinery for the transduction of innate immune signals from the extracellular environment to the nucleus in a lophotrochozoan, the Mediterranean mussel *M. galloprovincialis* (Toubiana et al., 2014). This sophisticated system, consisting of TLRs, intracellular adaptors and kinases, and transcription factors, finally triggers the expression of pro-inflammatory cytokines and defense molecules either via the IKK complex or the MAPK pathway. Overall, the remarkable sequence homology of its molecular components revealed a significant overlap with the pathway activated

downstream of Toll in arthropods, and even more with the pathway activated downstream of TLRs in deuterostomes (Gerdol and Venier, 2015). First, it has to be taken into account that the arthropod Toll signaling is activated by the interaction between the extracellular LRR region of Toll and the pro-inflammatory cytokine Spätzle, which does not have any homolog neither in brachiopods nor other lophotrochozoans (Gerdol and Venier, 2015). The activation of Spätzle is the result of a complex extracellular proteolytic cascade governed by the interplay between secreted PGRPs and GNBPs or, in the case of fungal PAMPs, by the activation of the Persephone protease. The secreted PGRPs found in brachiopods only share a very remote sequence similarity with those of *Drosophila*. As explained in detail in the previous sections, no sequence homologous to arthropod GNBPs is present in the *L. anatina* genome, and the possible involvement of CCF-like proteins in the melanization cascade remains to be investigated. Moreover, no clear homolog to Persephone could be identified among the many brachiopod serine proteases. Overall, based on the data presently available, it appears more likely that brachiopod TLRs act similarly to vertebrate and echinoderm receptors, i.e., by directly binding PAMPs at the cell membrane site.

The link between TLR diversification and PAMP-binding specificity in invertebrates is still far from being understood, but this is not the only aspect of the TLR immune signaling cascade that remains somewhat obscure. For example, it is presently unclear whether the massive diversification of TLRs also implies the use of different intracellular adaptors. Previous molecular surveys failed to identify a TRIF homolog in mollusks, ruling out the existence of a MyD88-independent signaling pathway homolog to that activated downstream of TLR3 in human. However, it is certainly possible that other divergent intracellular TIR domain-containing (TIR-DC) adaptors hold a similar role in lophotrochozoans, due to the massive expansion of cytosolic TIR-DC proteins (Gerdol et al., 2017a).

Quite surprisingly only a single MyD88 gene is present in *L. anatina*, in stark contrast with bivalve mollusks, where multiple MyD88 genes are always present (Ning et al., 2015). The presence of a unique MyD88, the first adaptor molecule interacting with the C-terminal cytosolic TIR domain of TLRs, might indicate the possible convergence of the signals of all brachiopod TLRs to a single adaptor unless the presence of a MyD88-independent signaling is assumed. SARM, an evolutionarily conserved negative regulator of TLR signaling, is also present in the brachiopod genome as a single copy gene encoding a protein with a well-conserved structural architecture (**Figure 7**) (Peng et al., 2010).

All the other molecular components involved in the signaling downstream of MyD88 and SARM, and possibly leading to the activation of NF-κB (through the IKK complex) and/or AP-1 (through the MAPK pathway), are present in the *L. anatina* genome, with no major exception (**Figure 8**), even though some differences with vertebrates need to be pointed out. TOLLIP, an ubiquitin-binding protein associated with TLRs, which mediates inflammation (primarily by regulating IL-1R and IRAK turnover), is found in a single copy, which presents a canonical domain organization, with an N-terminal C2 and a C-terminal CUE module. *L. anatina* also displays four IRAK (Interleukin-1 receptor-associated kinase)-like genes. In vertebrates, IRAK1 and IRAK4 are recruited by MyD88, leading in turn to the recruitment of TRAF6. Two out of the four brachiopod IRAK-like molecules show high similarity to IRAK4; one, like in bivalves (Gerdol and Venier, 2015), is equally similar to both IRAK1 and IRAK4, whereas the fourth one is unlikely to be involved in signal transduction due to the lack of the N-terminal DEATH domain. TRAF6, one of the key mediators of the immune signaling due to its role in forming a complex with TAK1, is encoded by two paralogous gene copies highly homologous to that of vertebrates. TRAF6 is also an important switch for the activation of the MAPK cascade, as an alternative to the activation of the IKK complex (**Figure 8**). The activation of this alternative route is guaranteed by the interaction with ECSIT, an evolutionarily conserved adaptor (Lin et al., 2017), which is also present in two paralogous gene copies in *L. anatina*.

A clear brachiopod homolog of TAK1 is also present, as well as its accessory proteins TAB1 and TAB2 that, together with TRAF6, form a complex which translocases from the proximity of the inner face of the

plasma membrane to the cytosol. Here, TAK1 is activated, switching on the activity of the IKK signalosome. The IKK complex, composed in vertebrates by the two kinases IKKα, IKKβ and by the regulatory subunit IKKγ/NEMO, is the central regulator of NF-κB activation. Only a single IKKα/β homolog could be identified in *Lingula*, exactly like in mussels (Toubiana et al., 2014), suggesting that the function of the two highly homologous vertebrate IKK proteins (Adli et al., 2010) is covered by a single kinase in lophotrochozoans. Despite a lower structural conservation, a NEMO homolog could also be detected with high confidence. The final steps of the signaling involve the phosphorylation of the inhibitor protein IKβ/cactus, which leads to its ubiquitination and degradation, finally allowing the NF-κB transcription factor to enter the nucleus, inducing the expression of its target pro-inflammatory genes. Brachiopods possess three well-conserved IKβ homologs and, as mentioned in the previous section, two proteins belonging to the class I (p-100/p105-like) and class II (p65-like) NF-κB transcription factor family (**Figure S5**).

As briefly mentioned above and shown in **Figure 8**, TRAF6 can also activate the MAPK pathway thanks to the interaction with ECSIT. The MAPK pathway is involved in multiple biological processes not strictly correlated with immunity, including cellular differentiation, growth, and survival. The brachiopod MAPK signaling machinery appears to be complete and highly similar to that previously reported in bivalve mollusks (Gerdol and Venier, 2015). This signaling includes a phosphorylation cascade that starts from TAK1 and sequentially involves the MAPKKK MEKK1, either the MAPKK MKK4 and MKK7 (both present as single copy genes) or two sequences homologous to MKK3/6 and, finally, several JNK/p38-like kinases. This complex signaling cascade ultimately leads to the nuclear translocation of the AP-1 transcription factor, which complements NF-κB in triggering the expression of pro-inflammatory cytokines and effector molecules. The AP-1 transcription factor is a heterodimer composed of c-JUN and c-FOS family proteins, which are both found in multiple copies in the *L. anatina* genome.

Recent findings suggest that LRRs may contribute to the recognition of PAMPs in another distinct class of membrane-bound receptors, provisionally named LRRIGs. These novel receptors, identified so far only in oysters, show an immunoglobulin ectodomain associated with LRRs, can bind with high affinity LPS, PGN and poly(I:C) and appear to mediate the production of cytokines and phagocytosis in cultured hemocytes (Huang et al., 2018; Wang et al., 2017). Interestingly, two genes pertaining to this family of receptors are present in *L. anatina* (**Figure S7**), and the finding of a LRRIG-like gene also in the annelid *C. teleta* suggest that these novel receptors might be broadly distributed among Lophotrochozoa. Considering the promising results obtained in oyster, LRRIGs emerge as interesting targets for immunological research in brachiopods., especially due to their domain architecture, highly similar to that of Variable Lymphocyte Receptors found in jawless fish (Huang et al., 2018). We can also report that the intracellular domain of brachiopod LRRIGs, reported to lack any recognizable conserved domain in oysters, presents a structural arrangement similar to the Four-helix Bundle of apolipoproteins.

3.4.3. Possible involvement of other TIR-domain containing proteins in immune signal transduction

Besides TLRs, MyD88 and SARM, several other TIR-DC proteins are present in brachiopods. This is in line with previous reports in the genomes of oyster, amphioxus and other invertebrate metazoans, where over 50 "orphan" TIR-DC genes have been described (Dishaw et al., 2012; Huang et al., 2008). In a recent comparative study, we explored the variability of TIR-DC proteins in bivalve mollusks, trying to issue an improved classification system which might better capture the structural diversity of these possible mediators of the intracellular immune signaling, either downstream of TLRs or other intracellular PRRs (Gerdol et al., 2017a). Based on the previously proposed classification scheme, we could detect members of several TIR-DC gene families in *L. anatina*. Namely, we identified:

(i) Interleukin-1 receptor-like (IL-1RL) proteins: in vertebrates, these membrane-bound receptors, bearing immunoglobulin ectodomains, convey IL-1 signals to the IKK complex, activating NF-κB. While two IL-1RL genes were identified in *L. anatina*, the orthologous state of vertebrate and invertebrate IL-1RL genes is still a matter of debate due to limited sequence homology and therefore their function cannot be assigned with certainty.

(ii) receptors with EGF ectodomains (as a single-copy gene): this family of membrane-bound proteins was described for the first time in our previous study.

(iii) ecTIR-DC 3, found with five distinct gene copies. This is an ancient gene family widespread in metazoans, characterized by the contemporary presence of ARM repeats, a central TIR domain, and a degenerated C-terminal SAM domain.

(iv) ecTIR-DC 5 (single copy gene), another ancient protein family bearing 3 TIR domains, with the first and the second one separated by a SAM domain.

(v) ecTIR-DC 7, characterized by a series of TPR repeats and a C-terminal TIR domain. We could identify three distinct ecTIR-DC 7 genes in *L. anatina*.

(vi) ecTIR-DC 8, a widespread metazoan family of cytoplasmic proteins with an N-terminal Macro domain and 3 consecutive TIR domains, present in three copies in the brachiopod genome.

(vii) ecTIR-DC 9, a large intracellular protein with a neuralized domain, followed by a ROC/COR tandem, a death domain, and a C-terminal TIR domain. The central ROC/COR tandem is also found in ecTIR-DC family 14, very large proteins with complex domain architectures and two consecutive TIR domains, which were found in relatively large number (6 genes) in *L. anatina*.

(viii) ecTIR-DC 11, very ancient adaptors with a TIR domain followed by HEAT repeats. Three ecTIR-DC genes are present in the genome of *L. anatina*.

(ix) ecTIR-DC 12, a single-copy gene, which could only be found in bivalves and brachiopods, even though the two share low sequence homology. In this case, the TIR domain is followed by TRAF repeats.

(x) ecTIR-DC 15, a single copy gene of a widespread metazoan family, which encodes a large protein with two consecutive TIR-domains and a Rossman-like fold domain.

All the above-mentioned families, whose domain organization is summarized in **Figure S8**, are conserved in most lophotrochozoans, and sometimes their origin can be traced back to basal multicellular metazoans. Several other previously unreported TIR-DC proteins are however present in *L. anatina*, although in most cases they were taxonomically restricted to brachiopods. The only case fitting the criteria required to gain a classification as a novel evolutionarily conserved TIR-DC family was that of a protein containing an N-terminal TIR domain, followed by a NB-ARC domain (typical of proteins involved in cell death regulation) and by several WD40 repeats. This protein family also found in Cephalochordata and Hemichordata (but not in lophotrochozoans other than brachiopods) was therefore named ecTIR-DC family 16 (**Figure S8**). In addition to the aforementioned families, a few membrane-bound receptors lacking recognizable ectodomains were also present in *L. anatina*.

3.5. Cytosolic PRRs and related signaling pathways

3.5.1. The NLR system

NACHT-LRR proteins (NLRs) are important intracellular receptors involved in a multitude of functions, which include intracellular PAMP recognition, inflammation and apoptosis (Fritz et al., 2006). Vertebrate NLRs usually have a conserved organization, consisting of an N-terminal death fold domain (DFD, including the

DEATH, DED, CARD or PYD subfamily), a central NACHT domain for oligomerization and C-terminal LRRs for recognition. The diversity has been recently elucidated also in cnidarians, echinoderms and amphioxus, where they constitute a greatly expanded gene family (Hamada et al., 2013; Hibino et al., 2006; Huang et al., 2008). Genomic data point out that a similar gene expansion event happened in some lophotrochozoans, such as in the annelid *C. teleta* (Simakov et al., 2013), but not in others, i.e., bivalve mollusks (Zhang et al., 2015).

Our survey permitted to evidence the presence of only a single canonical NLR in *L. anatina*, displaying high similarity with the single NLR previously described in *M. galloprovincialis* (Gerdol and Venier, 2015) and with orthologous sequences in other mollusks and annelids, comprising an N-terminal caspase recruitment domain (CARD) and canonical C-terminal LRRs (**Figure 9A**). The DFD/NACHT/LRR architecture, typical of the expanded sea urchin NLR gene family, was overall not common in brachiopods. On the other hand, the *L. anatina* genome encodes three additional proteins with N-terminal CARD domains, a central NACHT module and a C-terminal region lacking any recognizable conserved domain (**Figure 9B**). All the aforementioned proteins display a short accessory domain associated to NACHT, named NAD (NACHT Associated Domain), which is fundamental for the oligomeric assembly of NLRs (Singer et al., 2014). Although these four brachiopod NLRs do not find clear homology to any of the vertebrate NLRs, they display an identical domain architecture to that of NOD1 (Chen and Pedra, 2009).

In addition to CARD-containing NLRs, we could also identify three additional putative NLRs with poorly recognizable N-terminal DFDs, but structurally modeled as non-canonical DEATH folds, associated to a central NACHT/NAD tandem and LRRs in the C-terminal region (**Figure 9C**). While the lack of sequence homology with vertebrates prevents the assignment of a putative function to such proteins, their structure once again resembles that of *bona fide* NLRs. The downstream signal transduction by NLRs requires the interaction of the N-terminal signaling DF domain with adaptor proteins. In vertebrates, inflammasome-related NLRs activate ICE caspases, which in turn activate IL-1. However, since both the key adaptor involved into this process (PYCARD) and PYRIN domains themselves are missing in invertebrates (Huang et al., 2008), the existence of inflammasome complexes in protostomes has been previously brought into question (Latz et al., 2013). Alternatively, NOD1/2 NLRs are capable of transmitting the sensing of peptidoglycan components in the cytosol through the RIP kinase 2 homolog (also known as RICK), which is apparently also absent in brachiopods, despite the presence of several RIP kinase-like sequences with variable C-terminal domains. While the possible role of such molecules in mediating NLR signaling remains to be investigated, it has been previously hypothesized that DFD-containing adaptors divergent from those of vertebrates might be held responsible for NLR signal transduction in invertebrates (Huang et al., 2008). Notwithstanding these major uncertainties, in agreement with current molecular models the activation of NLRs is thought to trigger the expression of proinflammatory cytokines, IFN and AMPs either through NF-κB, IRF3/7 or AP-1 transcription factor complexes, using pathways which partially overlap those previously described downstream of PGRP and TLR membrane-bound receptors (**Figure 8**).

3.5.2. The RLR pathway

RIG-like receptors (RLRs) are among the most important sensors of viral infection in the cytosol. These molecules with helicase activity take their name from the Retinoid Inducible Gene I (RIG-I), which in human is responsible for the detection of 5′ triphosphate uncapped double-stranded RNAs produced during viral replication (Yoneyama and Fujita, 2007). The caspase recruitment domain of RLRs activates a downstream signaling cascade which results in the activation of the pro-inflammatory transcription factors NF-κB and IRF3, thereby effectively mounting antiviral response.

L. anatina possesses at least seven RLR-like genes, which all share significant homology either to RIG-I or to MDA-5, another RLR which is involved in the sensing of longer dsDNA molecules (Kato et al., 2008). These include genes with clear homology to human RLR-I, which also displays an identical domain architecture (**Figure 9D**). Others, despite sharing significant homology, present a different number of CARD or C-terminal RIG domains or a death effector domain (DED) instead of CARD (**Figure 9E**). However, these two types of domains are correlated and share a similar function (Weber and Vincenz, 2001). Perhaps more surprising is the presence of a gene encoding an RLR/caspase hybrid protein, which lacks CARD/DED N-terminal domains, which are replaced by a caspase catalytic domain (**Figure 9F**). This unexpected domain combination, never reported before in metazoans, could imply a direct role in triggering apoptosis upon viral RNA recognition.

The downstream molecular partner connecting RLRs to the activation of the pro-inflammatory signal, the IFN-beta promoter stimulator (IPS-1)/CARD adaptor inducing IFN-beta (CARDIF)/mitochondrial antiviral-signaling protein (MAVS) (**Figure 10**), has been recently identified in the Pacific oyster *C. gigas* (Huang et al., 2017). Unlike its vertebrate functional homolog, oyster IPS-1/CARDIF/MAVS lacks the central proline-rich region, which is replaced by a DEATH domain (**Figure S9**). *L. anatina* possesses two paralogous genes that share high sequence homology with the *C. gigas* sequence and it homologs found in the Eastern oyster and yesso scallop genomes. The only major difference between the bivalve and brachiopod sequences consists in the presence of a longer central region, which contains a mucin-like domain before the C-terminal transmembrane domain (**Figure S9**). Remarkably, no MAVS-like gene was detected in other lophotrochozoan genomes, including those of gastropod and cephalopod mollusks, annelids, phoronids and nemerteans.
It has been previously demonstrated that RLRs are completely missing in some major protostome phyla (e.g. insects), which therefore entirely base their antiviral sensing system on Dicer-2, the second copy of the evolutionarily conserved endoribonuclease which is usually involved in the synthesis of miRNA and siRNA. This system, which is also potentially involved in the antiviral response of vertebrates, is employed to load fragments of digested viral dsRNA on the RISC complex, thereby promoting the blockade of viral replication by RNA interference (Berkhout and Haasnoot, 2006). While the possible involvement of the RISC machinery in the lophotrochozoan antiviral response is far from being established, it appears evident that *L. anatina*, like mollusks (Rosani et al., 2016), displays a single Dicer gene.

3.5.3. The STING pathway

The stimulator of interferon genes (STING) pathway is a focal intracellular hub for the sensing of exogenous DNA and bacteria. Indeed, the downstream pathways of multiple DNA sensors converge here, and STING itself can recognize conserved bacterial messengers, thereby promoting the production of proinflammatory cytokines and interferon via NF-κB (Burdette and Vance, 2013). We could identify two *L. anatina* genes encoding proteins which shared the same peculiarities of bivalve STING (Gerdol and Venier, 2015), with the duplication of the STING C-terminal domain, accompanied by the presence of two TIR domains which might serve for downstream signal transduction. The same domain architecture is also present in Annelida, suggesting that the STING domain duplication may be typical of lophotrochozoans and that it may warrant full functionality without the need of dimerization (Gerdol, 2017). Unlike vertebrates, brachiopod STING proteins are devoid of transmembrane regions, suggesting that their subcellular localization might not be the mitochondrial outer membrane.

While STING can directly recognize and bind foreign nucleic acids in the cytosol, its activity is usually enhanced by accessory molecular partners that, in most cases, have been so far properly characterized only in vertebrates. Among these, cGAS is particularly relevant, as it can bind foreign nucleic acids in the cytosol, thereby producing the second messenger cGAMP and activating STING (Ma and Damania, 2016). Despite the

report of multiple cGAS-like gene copies in other lophotrochozoans (Gerdol, 2017), we could just detect a single cGAS gene in *L. anatina*. In vertebrates, STING activates the transcription of proinflammatory genes through the phosphorylation of IRFs, mediated by the TANK-binding kinase 1 (TBK1) (Tanaka and Chen, 2012), as summarized in **Figure 10**. While two TBK1 gene copies are present in the genome of *L. anatina*, as previously mentioned the presence of two TIR domains suggests that an uncharacterized TIR-DC protein adaptor is involved in the first step of the downstream signaling pathway.

3.6. Effector molecules and modulators of the immune response

3.6.1. Antimicrobial peptides

Antimicrobial peptides (AMPs) are heterogeneous small secreted peptides which are employed as the first line of defense towards pathogens by a broad range of organisms, including protostome animals. AMPs are usually classified into different families based on structural features and chemo/physical properties, which determine their spectrum and mode of activity. While many AMPs are encoded by taxonomically restricted gene families, others display a broader distribution, sometimes as a result of convergent structural evolution. Although defensin-like peptides are present in some major invertebrate taxa, including arthropods and bivalve mollusks (Froy and Gurevitz, 2003), they are absent in others, including annelids and gastropod mollusks (Gerdol, 2017), which makes their orthologous status disputed (Rodríguez de la Vega and Possani, 2005). In agreement with the peppered distribution of defensins in invertebrates, we could not identify any such sequence in *L. anatina*, nor in the transcriptome of any other brachiopod species. On the other hand, while no big defensin could be identified in *L. anatina* sequence data, it is worth of a note that transcripts encoding this β-defensin-related AMP were detected in the transcriptomes of other brachiopod species, all pertaining to Rhynchonelliformea, i.e., articulate brachiopods (**Figure S10**). Since these AMPs had so far been described only in a few distantly related invertebrate groups, including bivalve mollusks, horseshoe crabs and amphioxus (Gerdol et al., 2012), this finding is important to solve the puzzle of the evolution of beta-defensins, as these peptides might be discovered in other neglected Lophotrochozoan taxa in the future.

Another class of AMPs which appears to be widespread in invertebrates, from Cnidaria to Echinodermata, macins (Gerdol et al., 2012), was not detected in brachiopods and none of the numerous AMP families described so far in annelids (Tasiemski, 2008) could be identified either. However, one should keep in mind that AMPs are rapidly evolving molecules, which are shaped by positive selection and which are often encoded by taxonomically restricted gene families (Tennessen, 2005). Given the poor knowledge of brachiopod molecular immunity, it is not possible to describe any other AMP by sequence similarity-based screening. However, methods aimed at the *de novo* identification of AMPs from transcriptomes have already provided encouraging results in other invertebrates (Leoni et al., 2017) and they might be used in the future, in combination with functional assays, to identify which effector peptides, if any, display a significant antimicrobial activity.

3.6.2. Lysozymes and bactericidal/permeability-increasing proteins

Lysozymes are among the most widespread and well-known proteins with antimicrobial properties in metazoans, particularly effective against Gram-positive bacteria. Although lysozymes differ for their primary sequence, they all show a marked structural similarity and they can be classified into three major groups, i.e., chicken-type (C-type), goose-type (G-type) and invertebrate-type (I-type), that can be found isolated or in

combination in different animal phyla (Callewaert and Michiels, 2010). Among Lophotrochozoa, mollusks produce all the three lysozyme types (Gerdol and Venier, 2015), arthropods only produce C- and I- type, while both round- and segmented worm genomes only encode I-type genes. Surprisingly, the *L. anatina* genome bears a single G-type lysozyme gene and three additional genes encoding proteins with a lysozyme domain (Interpro: IPR023347) which could not be categorized within any of the three above-mentioned categories. Upon further scrutiny, these three protein products resulted to be highly similar to phage T4-type lysozymes, whose function is to hydrolyze the 1,4-beta-glycosidic bond between *N*-acetylmuramic acid and *N*-acetylglucosamine in the peptidoglycan contained in bacterial cell walls, in order to release mature phage particles. Curiously, this type of lysozyme appears to be extremely rare in metazoans: while, until very recently, the only official report concerned the clam *R. philippinarum* (Ding et al., 2014), it has been revealed that multiple bivalve species have acquired and integrated into their genomes bacteriophages lysozyme genes (Ren et al., 2017). Based on our observations, this intriguing evolutionary strategy can now be extended to brachiopods.

In vertebrates, the bactericidal/permeability-increasing proteins (BPI) hold a similar function against Gram-negative bacteria, by binding lipopolysaccharide upon release by lysosomal granules of neutrophil cells (Wilde et al., 1994). BPIs are highly conserved in most metazoans, but they are crucially missing in some protostomes, including arthropods and flatworms (Baron et al., 2016). Due to their high structural similarity, BPIs are thought to have a similar Gram-negative specificity of action both in all animals. Within lophotrochozoans, BPIs have been described in the coelomocytes of segmented worms (Škanta et al., 2016) and in mollusks, where they are sometimes involved in egg protection by parental transfer, as well documented in the snail *B. glabrata* (Baron et al., 2013, 2016). Overall, genome and transcriptome data point towards their widespread presence in all mollusks and annelids. Two nearly identical BPI sequence could be identified in *L. anatina*, pointing out a lower sequence diversity compared to snails (Baron et al., 2016). These two gene products showed a canonical domain organization, with a signal peptide for secretion targeting the molecule to the extracellular space, and two consecutive BPI domains.

3.6.3. Modulators of the immune response

Despite the impressive conservation of the intracellular signaling machinery activated in response to pathogen recognition and the demonstration of the responsiveness of invertebrate cells to human cytokines by pioneering studies (Quistad et al., 2014), so far only a very few cytokines have been identified in invertebrates. These short secretory molecules are major players in the modulation of the immune response, as their production is under the strict control of transcription factors belong to the NF-κB and IRF families. Cytokines mediate the recruitment of specialized immune cells at the site of infection and help perpetrating the immune signal until complete clearance of pathogens. Cytokines have been mostly studied in vertebrates, leaving the nature of their largely divergent homologous sequence in invertebrates elusive. The idea itself that such molecules were present in invertebrates had long been challenged, until recent times (Beschin et al., 2001). Our improved molecular knowledge of metazoans, which has been allowed by the popularization of genome sequencing approaches, led to major discoveries such as the ancestral origin of the transforming growth factor β (TGF-β) superfamily (Herpin et al., 2004), the presence of prokineticin-like cytokines involved in arthropod hematopoiesis (Söderhäll et al., 2005), the presence of TNF-alpha-like molecules and their LITAF mediators in crustaceans and mollusks (Sun et al., 2014; Wang et al., 2012), macrophage migration inhibitory factor (MIF) in mollusks (Garcia et al., 2010; Rosani et al., 2015) and myticins in mussels (Balseiro et al., 2011).

Despite these discoveries, sequences homologous to interferon and the other four-helix cytokines remain unknown in all invertebrates, brachiopods included. Like in mollusks, however, the presence of interferon-responsive factors (IRFs) suggests that still unidentified interferon-like cytokines and, to some

extent, conserved upstream regulatory routes also exist in invertebrates (Huang et al., 2013). In particular, four IRFs are encoded by the genome of *L. anatina*. One of these bears high similarity to IRF 4/8, while the other three show a more remote similarity to vertebrate IRF 1/2 molecules, limited to the DNA-binding domain. While the function of these four transcription factors is unclear, they could potentially regulate the expression of genes containing interferon-stimulated response element (ISRE) in their promoters.

Interleukin-17 (IL-17), which pertains to a family with a markedly different structural organization compared to four-helix cytokines, has been recently shown to be conserved across nearly all animals (Garcia et al., 2010; Rosani et al., 2015). A single IL-17 gene is present in the *L. anatina* genome, accompanied by a single well recognizable transmembrane receptor and a single CIKS intracellular adapter. This suggests the existence of a relatively simple pro-inflammatory IL-17 signaling route in brachiopods compared to that hypothesized for bivalve mollusks (Rosani et al., 2015). Consistently with multiple reports about the presence of TNFα-like molecules in other lophotrochozoans (De Zoysa et al., 2009; Zhang et al., 2015), four genes encoding membrane-bound proteins with the expected domain architecture could be identified in *L. anatina*, two of which are particularly similar to vertebrate TRAIL proteins. Like mentioned in **section 3.4.1**, some TNF receptor-like proteins, expected to cover an important role in regulators of cell survival and inflammation, are also encoded by the *L. anatina* genome. However, although they display an intracellular DEATH domain and the same domain organization of human TNFR1/2, they only share limited sequence homology with their vertebrate counterparts, making it unclear whether they can cover a homologous function in brachiopods.

Two MIF genes are also present in brachiopods. In vertebrates, this cytokine has a pivotal role in regulating the function of macrophages, thereby being a major player in the inflammation process. Although MIF genes have been previously reported in Lophotrochozoa (Cui et al., 2011; Wang et al., 2008), their mode of action and functional overlap with their vertebrate homologs is still a matter of debate due to the absence of CD74 (required for MIF binding) and the obvious cellular divergence between vertebrate macrophages and invertebrate immunocytes (Ottaviani, 2011). Despite these uncertainties, MIF certainly plays an important role in the regulation of immune response in lophotrochozoans, as evidenced by the functional studies carried out on *S. mansoni*-infected snails (Garcia et al., 2010). Due to the key role of snail MIF in the encapsulation of trematode sporocysts, it is possible that the *L. anatina* orthologs covers a similar role against yet to be identified parasites.

A gene orthologous to an important factor involved in the regulation of hemocyte development, as well as in anti-digenean defense response in gastropods, progranulin, is present in *L. anatina* (Pila et al., 2016). Since granulin homologs are multifunctional proteins with important roles in normal cell development, tumorigenesis and wound healing, the involvement of brachiopod progranulins in the maturation of immune cells needs to be confirmed by functional studies.

3.7. Tissue specificity of immune genes

The analysis of RNA-seq data can provide a detailed snapshot of the ontogeny of the immune system along embryonic development, as well as of the immune competence of the different adult tissues. As shown in **Figure 11**, the vast majority of the genes encoding molecules potentially involved in immune processes are turned off during the early phases of larval development, as they only start to be expressed once the main organs of the adult body begin to be established, i.e., at the one and two pairs-of-cirri stages.

There are, however, some notable exceptions. The most important one is certainly represented by intracellular immune signaling, which is active in the very early stages and which can be stimulated by multiple membrane-bound and cytosolic receptors. The dynamic spatial and temporal regulation of TLRs is a key factor in the establishment of the dorso–ventral polarity and organogenesis in *D. melanogaster* (Kambris et al., 2002). Although such a role has been demonstrated to be lineage-specific in dipterans (Sachs et al.,

2015), it is noteworthy that a single mccTLR was expressed at moderate levels from the fertilized egg to the blastula stage, along with MyD88 and all the downstream signaling components. This observation suggests that TLRs may play a role during early development in brachiopods. From the early gastrula stage onwards, other three mccTLR genes started to be transcribed, while the expression of the previous one progressively decreased, well matching the dynamic patterns described for arthropods. Like in oysters, no sccTLR was expressed at significant levels during larval development (Gerdol et al., 2017a).

A few other immune genes were also expressed during embryonic development. These include cytosolic PRRs (NLRs in particular), galectins (which also function as DAMPs and regulators of apoptosis) and macrophage mannose receptor-like C-type lectins (at moderate levels). Major differences could be observed among the seven adult tissues analyzed. In detail, the digestive cecum and the pedicle displayed a relatively poor expression of immune genes. The former is a highly specialized tissue, which is devoted to nutrient absorption during the digestive process. The latter is a structure mainly composed of connective tissue and muscular fibers, which is used to burrow the shell into soft substrates. Like bivalve foot, the pedicle is therefore not expected to express a broad range of immune genes (Gerdol et al., 2017b). The transcription of mRNAs encoding PRRs was much higher in the tissues that represent main surfaces of contact with the external environment. Among these, the lophophore was the organ which displayed the expression of the largest number of immune genes, followed by mantle (ventral and dorsal parts had very similar profiles) and whole gut. These tissues are, for different reasons, constantly exposed to microbes, including potentially pathogenic agents, present in the water column, either through feeding or due to injury and shell damage.

An in-depth view of expression dynamics revealed that most TLRs were not expressed at all in any of the adult tissues, nor in any of the developmental stages, mirroring the results obtained in oyster and suggesting a finely regulated mechanism of induction (Gerdol et al., 2017a). There was, however, a clear difference between mcc and sccTLRs. The former class, which as mentioned above may play a role in embryonic morphogenesis, was also broadly expressed in adult brachiopods. On the contrary, most sccTLRs were poorly expressed in adult tissues, and those achieving a significant level were, for the most part, pertaining to the expanded two-exon subclass. All the cytosolic components of the downstream signaling were consistently transcribed at steady levels in all adult tissues, reflecting the convergence of multiple signals from different receptors on the same intracellular cascade, including NLRs and RLRs, which were also broadly expressed at moderate levels in all adult tissues.

Besides these PRRs, the other immune gene families expressed ubiquitously and therefore denoting constitutively expressed defense systems, were galectins, PGRPs, complement components (factor B, factor C, and C3) and phage-type lysozymes. Other systems displayed a marked preference for the mantle and lophophore tissues, including STING/cGAS, GNBPs, BPIs and the G-type lysozyme. Overall, C1qDC proteins, FReD-C proteins and C-type lectins, large families whose involvement in immune recognition in brachiopods is uncertain, were generally poorly expressed but, at the same time, they often displayed a strong tissue specificity, similarly to what has been previously described in bivalves (Huang et al., 2015).

4. Conclusions

The release of the genome of the brachiopod *L. anatina* permitted to identify a highly complex immune system, which consists of several hundred molecular players among PRRs, signaling adaptors, effector molecules and modulators of the immune response. We showed that the brachiopod immunome shares a remarkable similarity with that of invertebrate deuterostomes lacking adaptive immunity, and a less significant overlap with to that of arthropods. Therefore, the *L. anatina* immunome confirmed the divergence between the evolutionary strategies followed by Lophotrochozoa and Ecdysozoa for the development of an efficient defense system towards invading microorganisms. However, while a significant overlap could be

detected between the immune gene repertoires of brachiopods and mollusks, the exclusive presence of some families of receptors and effector molecules in only one of these two phyla highlights the development of peculiar, lineage-specific adaptations which might be related to specific aspects of brachiopod biology.

Due to the lack of immunological studies in brachiopods, most of the observations presented in this work concerning the functional aspects of brachiopod immunity derive from homology-based inference and a comparative interpretation of data reported in other invertebrates, and they have therefore to be considered as hypothetical. Nevertheless, this general overview provides a substantial contribution to the comparative study of the evolution of the immune system in Metazoa, helping to focus future studies targeting *L. anatina* or other brachiopod species.

Figure captions

Figure 1: A) photograph of an adult *Lingula anatina* specimen. B) Schematic illustration showing the internal anatomy of the animal, with the dorsal shell removed. The main tissues are indicated. C) View of circulating cells extracted from mantle blood vessels; round-shaped erythrocytes and elongated spindle body cells are visible. D) Detail of blood vessels located in the mantle. E) Simplified phylogeny of Lophotrochozoa *sensu strictu*, highlighting the most relevant phyla, based on molecular data by Luo et al. (2017).

Figure 2: General overview of the main protein-coding genes involved in the immune system of *L. anatina* (*La*) and other representative invertebrate species: the mollusk *B. glabrata* (*Bg*) and *C. gigas* (*Cg*), the polychaete worm *C. teleta* (*Ct*), the fruit fly *D. melanogaster* (*Dm*) and the sea urchin *S. purpuratus* (*Sp*). □: presence; □: absence. The number of genes identified for a category is specified, whenever relevant. scc, single cysteine cluster; mcc, multiple cysteine cluster; TLR, Toll-like receptor; PGRP, peptidoglycan recognition protein; NLR, NACHT-LRR proteins; RLR, RIG-like receptor; STING, stimulator of interferon genes; C1qDC, C1q domain-containing; FReD-C: fibrinogen-related domain-containing proteins; FREP, *bona fide* fibrinogen-related protein; TEP, thioester-containing protein; IMD, immune deficiency; NF-κB, nuclear factor kappa-light-chain-enhancer of activated B cells; AP-1, activator protein 1; IRF, interferon responsive factor; AMP, antimicrobial peptide; MACPF-DC, membrane attack complex/pore forming domain-containing; BPI, bactericidal/permeability increasing protein.

Figure 3: Domain architecture of selected *L. anatina* C-type lectin domain containing proteins, which might possibly be involved in PAMP recognition. Repeated domains are indicated between square brackets, with the number of repeats indicated. CLECT, C-type lectin; FA58C, coagulation factor 5/8 C-terminal domain; WSC, carbohydrate-binding WSC domain; LCCL, Limulus factor C, Coch-5b2 and Lgl1 domain; TSP1, thrombospondin-1; CCP, complement control protein module; Zf-C2H2, zinc finger, C2H2-type; SR, scavenger receptor cysteine-rich domain; KR, kringle domain; EGF, epidermal growth factor.

Figure 4: Bayesian phylogeny of *L. anatina* GNBP-like proteins. GNBPs, Gram-negative binding proteins; CCFs, coelomic cytolytic factors. The sequences of *L. anatina* are marked with diamonds. The tree was built with the functionally characterized sequences from *D. melanogaster* and lophotrochozoans deposited in GenBank. Sequences are named based on the abbreviation of the genus and species names (first three letters

each). Posterior probability values are shown for the main nodes of the tree. Apocal, *Aporrectodea caliginosa*; Apoict, *Aporrectodea icterica*; Apolon, *Aporrectodea longa*; Aporos, *Aporrectodea rosea*; Azufar, *Azumapecten farreri*; Chlalb, *Chamys albidus*; Chlros, *Chlamys rosealbus*; Cragig, *C. gigas*; Denben, *Dendrobaena veneta*; Dromel, *D. melanogaster*; Eisfet, *Eisenia fetida*; Haldis, *Haliotis discus*; Haltub, *Haliotis tuberculata*; Linana, *L. anatina*; Litsic, *Littorina sitkana*; Lumrub, *Lumbricus rubellus*; Mermer, *Mercenaria mercenaria*; Mizyes, *Mizuhopecten yessoensis*; Pervir, *Perna viridis*; Pinfuc, *P. fucata*; Spisac, *Spisula sachalinensis*; Taplit, *Tapes literata*. The accession IDs of the sequences used in this phylogenetic tree are reported in **Table S2**.

Figure 5: A: Domain architecture comparison between *Tachypleus tridentatus* Factor C and *L. anatina* Factor C-like proteins. B: hypothetical functioning of the complement system in brachiopods. The recognition of Gram-positive bacteria PAMPs by secreted lectin-like molecules probably triggers the activation of the complement proteolytic cascade through the action of proteases that remain to be uncovered. The cleavage of C3 to C3b can trigger opsonization and the consequent clearance of invading bacteria by phagocytosis. As an alternative, the activation of C3 could lead to the killing of the pathogen by components homologous to those of the vertebrate terminal pathway of the complement system. The lysis of bacterial cells could be also attained by mytilectins, which combine a lectin-like domain with a pore-forming domain. The complement system is likely to be similarly activated by factor C, which could recognize lipopolysaccharide, leading to the activation of C3 in response to Gram-negative bacteria.

Figure 6: IMD pathway in arthropods and hypothetical organization of the homologous pathway in mollusks and brachiopods. Once activated by peptidoglycan (PGN) upon infection by Gram-negative bacteria, membrane bound PGRPs transmit the immune signal inside the cell, leading to the activation of an intracellular signaling cascade that ultimately activates the Relish transcription factor, which in turn regulates the production of antimicrobial peptides and pro-inflammatory genes. In arthropods, PGRPs activate the signaling cascade through the interaction of their intracellular RHIM domain with the adaptor IMD, which in turn interacts with the DEATH domain of dFADD. However, IMD is absent in both mollusks and brachiopods. While molluscan membrane-bound PGRPs also lack any recognizable intracellular domain, leaving their involvement in the activation of the intracellular immune signaling cascade a matter of speculation, some transmembrane brachiopod PGRPs, named PGRP-PSM, possess and intracellular DEATH domain which could hypothetically recruit dFADD, thereby bypassing the need for a IMD-like adaptor.

Figure 7 Panel A: circular phylogram depicting the relationship between *L. anatina* TLRs. Nodes supported by posterior probability values lower than 0.5 have been collapsed. Panel B: domain architecture of mccTLRs, sccTLRs and their cytosolic partners MyD88 and SARM. Panel C: a 35 Kb long region located on scaffold 150 of the *L. anatina* genome assembly, containing 5 sccTLR genes in a cluster organization. Arrows represent genes.

Figure 8. Schematic representation of the TLR signaling pathway in *L. anatina* (left) and domain architecture of the main components involved (right). The cytosolic adaptor MyD88, which recruits IRAK kinases (whose turnover is regulated by TOLLIP) and TRAF6, mediates the infection signal transmitted by TLRs. This molecular

complex can either activate TAK1 and the associated proteins TAB1 and TAB2, or can associate with ECSIT, thereby leading to the activation of the MAP kinase signaling cascade. In the former case, TAK1 turns on the activity of the IKK signalosome, which phosphorylates IκB, leading to its ubiquitation and degradation, allowing the translocation of NF-κB to the nucleus. On the other hand, the MAPK pathway leads to the activation of the AP-1 transcription factor through p38 kinases. The coordinated action of NF-κB and AP-1 leads to the transcription of immune effectors.

Figure 9: Schematic domain organization of NLR-like and RLR proteins in *L. anatina*. A: NLR displaying the canonical CARD/NACHT/NAD/LRR domain organization. B: Atypical NLR-like proteins lacking detectable C-terminal LRRs. C: Atypical NLR-like proteins with a highly degenerated N-terminal Death Fold Domain. D: Canonical RLR proteins displaying either one or two N-terminal CARD domains. E. RLR-like proteins presenting a poorly detectable N-terminal Death Effector Domain (DED), associated with one or two C-terminal RIG modules. F: hybrid RLR/Caspase protein displaying an N-terminal caspase effector domain.

Figure 10: Summary of mode of action of the main cytosolic PRRs in *L. anatina*. Missing or yet uncharacterized components are indicated by question marks and the name of the functionally homologous protein in vertebrates in shown in red. Briefly, NLRs can recognize the bacterial peptidoglycan components MDP and iE-DAP, triggering the expression of immune effectors through the kinase RIP2, which does not show any clear homolog in brachiopods. RLRs are specialized in the recognition of viral RNAs, and the consequent step in this antiviral signaling involves MAVS/IPS-1/CARDIF. Finally, STING is a multifunctional receptor that can either directly recognize the bacterial second messenger c-di-GMP or indirectly collect signals from other cytosolic partners, including cGAS, which synthetizes cGAMP from double-stranded exogenous DNA templates. STING-mediated immune signals are mediated by TBK1, even though the presence of TIR domains in STING suggests the possible involvement of other accessory proteins. The nature of the adaptor proteins involved in the several steps of the transduction of the immune signal to the nucleus in brachiopods are presently unknown.

Figure 11: Heat map summarizing expression profiles of immune genes along the embryonic development of *L. anatina* and in seven adult tissues. Genes which did not reach significant expression level (i.e. TPM < 10) in at least one sample were removed. Due to the high number of C-type lectin and FReD-C genes, only those pertaining to the macrophage mannose receptor-like and molluscan tachylectin/ficolin-like subfamilies are shown, respectively. FE, fertilized egg; 32: 32-cell stage; 128, 128-cell stage; EB, early blastula; B, blastula; EG, early gastrula; MG, mid gastrula; LG, late gastrula; 1PCL, one pair-of-cirri larvae; 2PCL, two pairs-of-cirri larvae; DC, digestive cecum; WG, whole gut; PD, pedicle; RP, regenerated pedicle; VM, ventral mantle; DM, dorsal mantle; LP, lophophore; C1qDC, C1q domain-containing protein; FReD, fibrinogen-related domain; PRRs, pattern recognition receptors; PGRPs, peptidoglycan recognition proteins; sccTLRs and mccTLRs, single- and multiple-cysteine cluster Toll-like receptors. Gene expression levels are available in **Table S4**. The complete cDNA sequences used for this analysis are provided in **Supplementary File 1**.

Acknowledgements

We thank Dr. Samuele Greco for graphical design assistance in Figure 1. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Adema, C.M. (2015). Fibrinogen-Related Proteins (FREPs) in Mollusks. Results Probl. Cell Differ. *57*, 111–129.

Adema, C.M., Hertel, L.A., Miller, R.D., and Loker, E.S. (1997). A family of fibrinogen-related proteins that precipitates parasite- derived molecules is produced by an invertebrate after infection. Proc. Natl. Acad. Sci. U. S. A. *94*, 8691–8696.

Adema, C.M., Hillier, L.W., Jones, C.S., Loker, E.S., Knight, M., Minx, P., Oliveira, G., Raghavan, N., Shedlock, A., Amaral, L.R. do, et al. (2017). Whole genome analysis of a schistosomiasis-transmitting freshwater snail. Nat. Commun. *8*, 15451.

Adli, M., Merkhofer, E., Cogswell, P., and Baldwin, A.S. (2010). IKKα and IKKβ Each Function to Regulate NF-κB Activation in the TNF-Induced/Canonical Pathway. PLOS ONE *5*, e9428.

Alexander, R.R. (1986). Resistance to and Repair of Shell Breakage Induced by Durophages in Late Ordovician Brachiopods. J. Paleontol. *60*, 273–285.

Altincicek, B., and Vilcinskas, A. (2007). Analysis of the immune-related transcriptome of a lophotrochozoan model, the marine annelid Platynereis dumerilii. Front. Zool. *4*, 18.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Ariki, S., Takahara, S., Shibata, T., Fukuoka, T., Ozaki, A., Endo, Y., Fujita, T., Koshiba, T., and Kawabata, S. (2008). Factor C Acts as a Lipopolysaccharide-Responsive C3 Convertase in Horseshoe Crab Complement Activation. J. Immunol. *181*, 7994–8001.

Armstrong, P.B., and Quigley, J.P. (1999). Alpha2-macroglobulin: an evolutionarily conserved arm of the innate immune system. Dev. Comp. Immunol. *23*, 375–390.

Bakholdina, S.I., Naberezhnykh, G.A., Gorbach, V.I., Isaeva, M.P., and Solov'eva, T.F. (2014). Marine invertebrates of the Sea of Okhotsk as a new source of lypopolysaccharide-binding proteins. Russ. J. Mar. Biol. *40*, 59–65.

Baliński, A., and Yuanlin, S. (2010). Tubular Shell Infestations in Some Mississippian Spirilophous Brachiopods. Acta Palaeontol. Pol. *55*, 689–694.

Balseiro, P., Falcó, A., Romero, A., Dios, S., Martínez-López, A., Figueras, A., Estepa, A., and Novoa, B. (2011). Mytilus galloprovincialis Myticin C: A Chemotactic Molecule with Antiviral Activity and Immunoregulatory Properties. PLoS ONE *6*, e23140.

Baron, O.L., van West, P., Industri, B., Ponchet, M., Dubreuil, G., Gourbal, B., Reichhart, J.-M., and Coustau, C. (2013). Parental transfer of the antimicrobial protein LBP/BPI protects Biomphalaria glabrata eggs against oomycete infections. PLoS Pathog. *9*, e1003792.

Baron, O.L., Deleury, E., Reichhart, J.-M., and Coustau, C. (2016). The LBP/BPI multigenic family in invertebrates: Evolutionary history and evidences of specialization in mollusks. Dev. Comp. Immunol. *57*, 20–30.

Bassett, M.G., Popov, L.E., and Holmer, L.E. (2004). The oldest-known metazoan parasite? J. Paleontol. *78*, 1214–1216.

Baumiller, T.K., and Bitner, M.A. (2004). A case of intense predatory drilling of brachiopods from the Middle Miocene of southeastern Poland. Palaeogeogr. Palaeoclimatol. Palaeoecol. *214*, 85–95.

Berkhout, B., and Haasnoot, J. (2006). The interplay between virus infection and the cellular RNA interference machinery. FEBS Lett. *580*, 2896–2902.

Beschin, A., Bilej, M., Hanssens, F., Raymakers, J., Dyck, E.V., Revets, H., Brys, L., Gomez, J., Baetselier, P.D., and Timmermans, M. (1998). Identification and Cloning of a Glucan- and Lipopolysaccharide-binding Protein from Eisenia foetidaEarthworm Involved in the Activation of Prophenoloxidase Cascade. J. Biol. Chem. *273*, 24948–24954.

Beschin, A., Bilej, M., Torreele, E., and De Baetselier, P. (2001). On the existence of cytokines in invertebrates. Cell. Mol. Life Sci. CMLS *58*, 801–814.

Bilej, M., Brys, L., Beschin, A., Lucas, R., Vercauteren, E., Hanušová, R., and De, B. (1995). Identification of a cytolytic protein in the coelomic fluid of Eisenia foetida earthworms. Immunol. Lett. *45*, 123–128.

Bizet, A.A., Liu, K., Tran-Khanh, N., Saksena, A., Vorstenbosch, J., Finnson, K.W., Buschmann, M.D., and Philip, A. (2011). The TGF-β co-receptor, CD109, promotes internalization and degradation of TGF-β receptors. Biochim. Biophys. Acta *1813*, 742–753.

Blandin, S.A., Marois, E., and Levashina, E.A. (2008). Antimalarial Responses in Anopheles gambiae: From a Complement-like Protein to a Complement-like Pathway. Cell Host Microbe *3*, 364–374.

Boidin-Wichlacz, C., Vergote, D., Slomianny, C., Jouy, N., Salzet, M., and Tasiemski, A. (2011). Morphological and functional characterization of leech circulating blood cells: role in immunity and neural repair. Cell. Mol. Life Sci. *69*, 1717–1731.

Bou Aoun, R., Hetru, C., Troxler, L., Doucet, D., Ferrandon, D., and Matt, N. (2010). Analysis of Thioester-Containing Proteins during the Innate Immune Response of Drosophila melanogaster. J. Innate Immun. *3*, 52–64.

Burdette, D.L., and Vance, R.E. (2013). STING and the innate immune response to nucleic acids in the cytosol. Nat. Immunol. *14*, 19–26.

Callewaert, L., and Michiels, C.W. (2010). Lysozymes in the animal kingdom. J. Biosci. *35*, 127–160.

Canesi, L., Gallo, G., Gavioli, M., and Pruzzo, C. (2002). Bacteria-hemocyte interactions and phagocytosis in marine bivalves. Microsc. Res. Tech. *57*, 469–476.

Canton, J., Neculai, D., and Grinstein, S. (2013). Scavenger receptors in homeostasis and immunity. Nat. Rev. Immunol. *13*, 621–634.

Carneiro, R.F., Torres, R.C.F., Chaves, R.P., de Vasconcelos, M.A., de Sousa, B.L., Goveia, A.C.R., Arruda, F.V., Matos, M.N.C., Matthews-Cascon, H., Freire, V.N., et al. (2017). Purification, Biochemical Characterization, and Amino Acid Sequence of a Novel Type of Lectin from Aplysia dactylomela Eggs with Antibacterial/Antibiofilm Potential. Mar. Biotechnol. N. Y. N *19*, 49–64.

Castillo, M.G., Salazar, K.A., and Joffe, N.R. (2015). The immune response of cephalopods from head to foot. Fish Shellfish Immunol. *46*, 145–160.

Chen, G., and Pedra, J.H.F. (2009). The Inflammasome in Host Defense. Sensors *10*, 97–111.

Chuang, S.H. (1983). Brachiopoda. In Vol I. Oogenesis, Oviposition and Oosorption, K.G. Adiyoki, and R.D. Adiyodi, eds. (Chichester, New York: Wiley & Sons Ltd.), pp. 571–584.

Conrad, M., Angeli, J.P.F., Vandenabeele, P., and Stockwell, B.R. (2016). Regulated necrosis: disease relevance and therapeutic opportunities. Nat. Rev. Drug Discov. *15*, 348–366.

Coustau, C., Gourbal, B., Duval, D., Yoshino, T.P., Adema, C.M., and Mitta, G. (2015). Advances in gastropod immunity from the study of the interaction between the snail Biomphalaria glabrata and its parasites: A review of research progress over the last decade. Fish Shellfish Immunol. *46*, 5–16.

Cremonte, F., Vázquez, N., and Ituarte, C. (2008). The development of Gymnophallus australis Szidat, 1962 (Digenea: Gymnophallidae) from the Patagonian coast (Argentina) from metacercaria to adult, with an amended diagnosis of Gymnophallus Odhner, 1905. Syst. Parasitol. *69*, 23–31.

Cui, S., Zhang, D., Jiang, S., Pu, H., Hu, Y., Guo, H., Chen, M., Su, T., and Zhu, C. (2011). A macrophage migration inhibitory factor like oxidoreductase from pearl oyster Pinctada fucata involved in innate immune responses. Fish Shellfish Immunol. *31*, 173–181.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics btr088.

De Zoysa, M., Jung, S., and Lee, J. (2009). First molluscan TNF-alpha homologue of the TNF superfamily in disk abalone: molecular characterization and expression analysis. Fish Shellfish Immunol. *26*, 625–631.

Dheilly, N.M., Duval, D., Mouahid, G., Emans, R., Allienne, J.-F., Galinier, R., Genthon, C., Dubois, E., Du Pasquier, L., Adema, C.M., et al. (2015). A family of variable immunoglobulin and lectin domain containing molecules in the snail Biomphalaria glabrata. Dev. Comp. Immunol. *48*, 234–243.

Ding, J., Wang, R., Yang, F., Zhao, L., Qin, Y., Zhang, G., and Yan, X. (2014). Identification and characterization of a novel phage-type like lysozyme from Manila clam, Ruditapes philippinarum. Dev. Comp. Immunol. *47*, 81–89.

Dishaw, L.J., Haire, R.N., and Litman, G.W. (2012). The amphioxus genome provides unique insight into the evolution of immunity. Brief. Funct. Genomics *11*, 167–176.

Dunkelberger, J.R., and Song, W.-C. (2009). Complement and its role in innate and adaptive immune responses. Cell Res. *20*, 34–50.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Emig, C.C. (1997). Les Lophophorates constituent-ils un embranchement? Bull. Société Zool. Fr. *122*, 279–288.

Feldmann, R.M., Mackinnon, D.I., Endo, K., and Chirino-Galvez, L. (1996). Pinnotheres laquei Sakai (Decapoda: Pinnotheridae), a tiny crab commensal within the brachiopod Laqueus rubellus (Sowerby) (Terebratulida: Laqueidae). J. Paleontol. *70*, 303–311.

Ferrandon, D., Imler, J.-L., Hetru, C., and Hoffmann, J.A. (2007). The Drosophila systemic immune response: sensing and signalling during bacterial and fungal infections. Nat. Rev. Immunol. *7*, 862–874.

Fraser, I.P., Koziel, H., and Ezekowitz, R.A.B. (1998). The serum mannose-binding protein and the macrophage mannose receptor are pattern recognition molecules that link innate and adaptive immunity. Semin. Immunol. *10*, 363–372.

1335 Fritz, J.H., Ferrero, R.L., Philpott, D.J., and Girardin, S.E. (2006). Nod-like proteins in immunity, inflammation
1336 and disease. Nat. Immunol. *7*, 1250–1257.

1337 Froy, O., and Gurevitz, M. (2003). Arthropod and mollusk defensins – evolution by exon-shuffling. Trends
1338 Genet. *19*, 684–687.

1339 Fujii, Y., Dohmae, N., Takio, K., Kawsar, S.M.A., Matsumoto, R., Hasan, I., Koide, Y., Kanaly, R.A., Yasumitsu,
1340 H., Ogawa, Y., et al. (2012). A lectin from the mussel Mytilus galloprovincialis has a highly novel primary
1341 structure and induces glycan-mediated cytotoxicity of globotriaosylceramide-expressing lymphoma cells. J.
1342 Biol. Chem. *287*, 44772–44783.

1343 Galinier, R., Portela, J., Moné, Y., Allienne, J.F., Henri, H., Delbecq, S., Mitta, G., Gourbal, B., and Duval, D.
1344 (2013). Biomphalysin, a New β Pore-forming Toxin Involved in Biomphalaria glabrata Immune Defense
1345 against Schistosoma mansoni. PLOS Pathog. *9*, e1003216.

1346 Garcia, A.B., Pierce, R.J., Gourbal, B., Werkmeister, E., Colinet, D., Reichhart, J.-M., Dissous, C., and Coustau,
1347 C. (2010). Involvement of the Cytokine MIF in the Snail Host Immune Response to the Parasite Schistosoma
1348 mansoni. PLOS Pathog. *6*, e1001115.

1349 Garnier, M., Labreuche, Y., Garcia, C., Robert, M., and Nicolas, J.L. (2007). Evidence for the involvement of
1350 pathogenic bacteria in summer mortalities of the Pacific oyster Crassostrea gigas. Microb. Ecol. *53*, 187–
1351 196.

1352 Gerdol, M. (2017). Immune-related genes in gastropods and bivalves: a comparative overview. Invertebr.
1353 Surviv. J. *14*, 95–111.

1354 Gerdol, M., and Venier, P. (2015). An updated molecular basis for mussel immunity. Fish Shellfish Immunol.
1355 *46*, 17–38.

1356 Gerdol, M., Manfrin, C., De Moro, G., Figueras, A., Novoa, B., Venier, P., and Pallavicini, A. (2011). The C1q
1357 domain containing proteins of the Mediterranean mussel Mytilus galloprovincialis: a widespread and
1358 diverse family of immune-related molecules. Dev. Comp. Immunol. *35*, 635–643.

1359 Gerdol, M., De Moro, G., Manfrin, C., Venier, P., and Pallavicini, A. (2012). Big defensins and mytimacins,
1360 new AMP families of the Mediterranean mussel Mytilus galloprovincialis. Dev. Comp. Immunol. *36*, 390–
1361 399.

1362 Gerdol, M., Venier, P., and Pallavicini, A. (2015). The genome of the Pacific oyster Crassostrea gigas brings
1363 new insights on the massive expansion of the C1q gene family in Bivalvia. Dev. Comp. Immunol. *49*, 59–71.

1364 Gerdol, M., Venier, P., Edomi, P., and Pallavicini, A. (2017a). Diversity and evolution of TIR-domain-
1365 containing proteins in bivalves and Metazoa: New insights from comparative genomics. Dev. Comp.
1366 Immunol. *70*, 145–164.

1367 Gerdol, M., Fujii, Y., Hasan, I., Koike, T., Shimojo, S., Spazzali, F., Yamamoto, K., Ozeki, Y., Pallavicini, A., and
1368 Fujita, H. (2017b). The purplish bifurcate mussel Mytilisepta virgata gene expression atlas reveals a
1369 remarkable tissue functional specialization. BMC Genomics *18*, 590.

1370 Gerlach, D., Schlott, B., Zähringer, U., and Schmidt, K.-H. (2005). N-acetyl-D-galactosamine/N-acetyl-D-
1371 glucosamine--recognizing lectin from the snail Cepaea hortensis: purification, chemical characterization,
1372 cloning and expression in E. coli. FEMS Immunol. Med. Microbiol. *43*, 223–232.

1373 Gestal, C., and Castellanos-Martínez, S. (2015). Understanding the cephalopod immune system based on
1374 functional and molecular evidence. Fish Shellfish Immunol. *46*, 120–130.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

Halanych, K.M., and Kocot, K.M. (2014). Repurposed Transcriptomic Data Facilitate Discovery of Innate Immunity Toll-Like Receptor (TLR) Genes Across Lophotrochozoa. Biol. Bull. *227*, 201–209.

Hamada, M., Shoguchi, E., Shinzato, C., Kawashima, T., Miller, D.J., and Satoh, N. (2013). The complex NOD-like receptor repertoire of the coral Acropora digitifera includes novel domain combinations. Mol. Biol. Evol. *30*, 167–176.

Hanington, P.C., and Zhang, S.-M. (2011). The primary role of fibrinogen-related proteins in invertebrates is defense, not coagulation. J. Innate Immun. *3*, 17–27.

Hasan, I., Gerdol, M., Fujii, Y., Rajia, S., Koide, Y., Yamamoto, D., Kawsar, S.M.A., and Ozeki, Y. (2016). cDNA and Gene Structure of MytiLec-1, A Bacteriostatic R-Type Lectin from the Mediterranean Mussel (Mytilus galloprovincialis). Mar. Drugs *14*, 92.

He, X., Zhang, Y., and Yu, Z. (2011). An Mpeg (macrophage expressed gene) from the Pacific oyster Crassostrea gigas: molecular characterization and gene expression. Fish Shellfish Immunol. *30*, 870–876.

Heller, M. (1931). Exkretorische Tatitigkeit der Brachiopoden. Z. Für Morphol. Ökol. Tiere *24*, 238–258.

Helmkampf, M., Bruchhaus, I., and Hausdorf, B. (2008). Phylogenomic analyses of lophophorates (brachiopods, phoronids and bryozoans) confirm the Lophotrochozoa concept. Proc. R. Soc. Lond. B Biol. Sci. *275*, 1927–1933.

Herpin, A., Lelong, C., and Favrel, P. (2004). Transforming growth factor-β-related proteins: an ancestral and widespread superfamily of cytokines in metazoans. Dev. Comp. Immunol. *28*, 461–485.

Hibino, T., Loza-Coll, M., Messier, C., Majeske, A.J., Cohen, A.H., Terwilliger, D.P., Buckley, K.M., Brockton, V., Nair, S.V., Berney, K., et al. (2006). The immune gene repertoire encoded in the purple sea urchin genome. Dev. Biol. *300*, 349–365.

Hiller, N. (2014). Drill holes and shell repair in brachiopods from a Late Cretaceous (Maastrichtian) oyster reef, North Canterbury, New Zealand. Cretac. Res. *49*, 83–90.

Huang, B., Zhang, L., Li, L., Tang, X., and Zhang, G. (2015). Highly diverse fibrinogen-related proteins in the Pacific oyster Crassostrea gigas. Fish Shellfish Immunol. *43*, 485–490.

Huang, B., Zhang, L., Du, Y., Xu, F., Li, L., and Zhang, G. (2017). Characterization of the Mollusc RIG-I/MAVS Pathway Reveals an Archaic Antiviral Signalling Framework in Invertebrates. Sci. Rep. *7*, 8217.

Huang, G., Huang, S., Yan, X., Yang, P., Li, J., Xu, W., Zhang, L., Wang, R., Yu, Y., Yuan, S., et al. (2014). Two apextrin-like proteins mediate extracellular and intracellular bacterial recognition in amphioxus. Proc. Natl. Acad. Sci. *111*, 13469–13474.

Huang, Q., Yu, M., Chen, H., Zeng, M., Sun, Y., Saha, T.T., and Chen, D. (2018). LRFN (leucine-rich repeat and fibronectin type-III domain-containing protein) recognizes bacteria and promotes hemocytic phagocytosis in the Pacific oyster Crassostrea gigas. Fish Shellfish Immunol. *72*, 622–628.

Huang, S., Yuan, S., Guo, L., Yu, Y., Li, J., Wu, T., Liu, T., Yang, M., Wu, K., Liu, H., et al. (2008). Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. Genome Res. *18*, 1112–1126.

Huang, X.-D., Liu, W.-G., Wang, Q., Zhao, M., Wu, S.-Z., Guan, Y.-Y., Shi, Y., and He, M.-X. (2013). Molecular characterization of interferon regulatory factor 2 (IRF-2) homolog in pearl oyster Pinctada fucata. Fish Shellfish Immunol. *34*, 1279–1286.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinforma. Oxf. Engl. *17*, 754–755.

Hughes, A.L. (2012). Evolution of the βGRP/GNBP/β-1,3-glucanase family of insects. Immunogenetics *64*, 549–558.

Jackson, D.J., Mann, K., Häussermann, V., Schilhabel, M.B., Lüter, C., Griesshaber, E., Schmahl, W., and Wörheide, G. (2015). The Magellania venosa Biomineralizing Proteome: A Window into Brachiopod Shell Evolution. Genome Biol. Evol. *7*, 1349–1362.

James, M.A., Ansell, A.D., and Curry, G.B. (1991). Functional morphology of the gonads of the articulate brachiopodTerebratulina retusa. Mar. Biol. *111*, 401–410.

Jeong, C.-B., Lee, J.-H., Lee, J.-S., and Rhee, J.-S. (2015). Early expansion and expression of the lipopolysaccharide (LPS)-induced TNF-α factor (LITAF) gene family in the LPS-exposed monogonont rotifer Brachionus koreanus. Comp. Biochem. Physiol. B Biochem. Mol. Biol. *188*, 15–23.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinforma. Oxf. Engl. *30*, 1236–1240.

Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2004). A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. *338*, 1027–1036.

Kambris, Z., Hoffmann, J.A., Imler, J.-L., and Capovilla, M. (2002). Tissue and stage-specific expression of the Tolls in Drosophila embryos. Gene Expr. Patterns GEP *2*, 311–317.

Kato, H., Takeuchi, O., Mikamo-Satoh, E., Hirai, R., Kawai, T., Matsushita, K., Hiiragi, A., Dermody, T.S., Fujita, T., and Akira, S. (2008). Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid–inducible gene-I and melanoma differentiation–associated gene 5. J. Exp. Med. *205*, 1601–1610.

Kawabata, S., Koshiba, T., and Shibata, T. (2009). The lipopolysaccharide-activated innate immune response network of the horseshoe crab. Invertebr. Surviv. J. *6*, 59–77.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858.

Kiel, S. (2008). Parasitic polychaetes in the Early Cretaceous Hydrocarbon seep-restricted brachiopod Peregrinella Multicarinata. J. Paleontol. *82*, 1215–1217.

Latz, E., Xiao, T.S., and Stutz, A. (2013). Activation and regulation of the inflammasomes. Nat. Rev. Immunol. *13*, 397–411.

Leighton, L.R. (2003). Predation on Brachiopods. In Predator—Prey Interactions in the Fossil Record, P.H. Kelley, M. Kowalewski, and T.A. Hansen, eds. (Springer US), pp. 215–237.

Lemaitre, B., and Hoffmann, J. (2007). The host defense of Drosophila melanogaster. Annu. Rev. Immunol. *25*, 697–743.

Leoni, G., De Poli, A., Mardirossian, M., Gambato, S., Florian, F., Venier, P., Wilson, D.N., Tossi, A., Pallavicini, A., and Gerdol, M. (2017). Myticalins: A Novel Multigenic Family of Linear, Cationic Antimicrobial Peptides from Marine Mussels (Mytilus spp.). Mar. Drugs *15*, 261.

Levine, R.P., and Dodds, A.W. (1990). The Thioester Bond of C3. In The Third Component of Complement, (Springer, Berlin, Heidelberg), pp. 73–82.

Li, J., Yuan, S., Qi, L., Huang, S., Huang, G., Yang, M., Xu, L., Li, Y., Zhang, R., Yu, Y., et al. (2011). Functional conservation and innovation of amphioxus RIP1-mediated signaling in cell fate determination. J. Immunol. Baltim. Md 1950 *187*, 3962–3971.

Li, L., Qiu, L., Song, L., Song, X., Zhao, J., Wang, L., Mu, C., and Zhang, H. (2009). First molluscan TNFR homologue in Zhikong scallop: molecular characterization and expression analysis. Fish Shellfish Immunol. *27*, 625–632.

Lin, Y.H., Zhang, W., Li, J.W., Zhang, H.W., and Chen, D.Y. (2017). Amphioxus ortholog of ECSIT, an evolutionarily conserved adaptor in the Toll and BMP signaling pathways. Mol. Biol. (Mosk.) *51*, 42–49.

Luo, Y.-J., Takeuchi, T., Koyanagi, R., Yamada, L., Kanda, M., Khalturina, M., Fujie, M., Yamasaki, S., Endo, K., and Satoh, N. (2015). The Lingula genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. Nat. Commun. *6*, 8301.

Luo, Y.-J., Kanda, M., Koyanagi, R., Hisata, K., Akiyama, T., Sakamoto, H., Sakamoto, T., and Satoh, N. (2018). Nemertean and phoronid genomes reveal lophotrochozoan evolution and the origin of bilaterian heads. Nat. Ecol. Evol. *2*, 141.

Ma, Z., and Damania, B. (2016). The cGAS-STING Defense Pathway and Its Counteraction by Viruses. Cell Host Microbe *19*, 150–158.

Mah, S.A., Moy, G.W., Swanson, W.J., and Vacquier, V.D. (2004). A perforin-like protein from a marine mollusk. Biochem. Biophys. Res. Commun. *316*, 468–475.

Matricon-Gondran, M., and Letocart, M. (1999). Internal Defenses of the Snail Biomphalaria glabrata: I. Characterization of Hemocytes and Fixed Phagocytes. J. Invertebr. Pathol. *74*, 224–234.

Matsushita, M. (2010). Ficolins: complement-activating lectins involved in innate immunity. J. Innate Immun. *2*, 24–32.

Montaño, A.M., Tsujino, F., Takahata, N., and Satta, Y. (2011). Evolutionary origin of peptidoglycan recognition proteins in vertebrate innate immune system. BMC Evol. Biol. *11*, 79.

Morse, E.S. (1902). Observations on living Brachiopoda (Boston : Boston Society of Natural History).

Motohashi, S., Jimbo, M., Naito, T., Suzuki, T., Sakai, R., and Kamiya, H. (2017). Isolation, Amino Acid Sequences, and Plausible Functions of the Galacturonic Acid-Binding Egg Lectin of the Sea Hare Aplysia kurodai. Mar. Drugs *15*, 161.

Müller, W.E.G., Blumbach, B., Wagner-Hülsmann, C., Lessel, U., and author_in_Japanese (1997). Galectins in the Phylogenetically Oldest Metazoa, the Sponges [Porifera]. Trends Glycosci. Glycotechnol. *9*, 123–130.

Ni, D., Song, L., Wu, L., Chang, Y., Yu, Y., Qiu, L., and Wang, L. (2007). Molecular cloning and mRNA expression of peptidoglycan recognition protein (PGRP) gene in bay scallop (Argopecten irradians, Lamarck 1819). Dev. Comp. Immunol. *31*, 548–558.

Nikapitiya, C., De Zoysa, M., Oh, C., Lee, Y., Ekanayake, P.M., Whang, I., Choi, C.Y., Lee, J.-S., and Lee, J. (2010). Disk abalone (Haliotis discus discus) expresses a novel antistasin-like serine protease inhibitor: Molecular cloning and immune response against bacterial infection. Fish Shellfish Immunol. *28*, 661–671.

Ning, X., Wang, R., Li, X., Wang, S., Zhang, M., Xing, Q., Sun, Y., Wang, S., Zhang, L., Hu, X., et al. (2015). Genome-wide identification and characterization of five MyD88 duplication genes in Yesso scallop (Patinopecten yessoensis) and expression changes in response to bacterial challenge. Fish Shellfish Immunol. *46*, 181–191.

Nyholm, S.V., Song, P., Dang, J., Bunce, C., and Girguis, P.R. (2012). Expression and Putative Function of Innate Immunity Genes under in situ Conditions in the Symbiotic Hydrothermal Vent Tubeworm Ridgeia piscesae. PLoS ONE *7*, e38267.

Ohuye, T. (1938). On thc coelomic corpuscles in the body fluid of some invertebrates VIII. Supplementary note on the formed elements in the coelomic fluid of some Brachiopoda. Sci. Rep. Tohoku Imp. Univ. *12*, 241–253.

Ottaviani, E. (2011). Immunocyte: The invertebrate counterpart of the vertebrate macrophage. Invertebr. Surviv. J. *8*, 1–4.

Paillard, C. (2004). A short-review of brown ring disease, a vibriosis affecting clams, Ruditapes philippinarum and Ruditapes decussatus. Aquat. Living Resour. *17*, 467–475.

Pan, C.-M., and Watabe, N. (1989). Periostracum Formation and Shell Regeneration in the Lingulid Glottidia pyramidata (Brachiopoda: Inarticulata). Trans. Am. Microsc. Soc. *108*, 283–298.

Peel, J.S. (2015). Failed predation, commensalism and parasitism on lower Cambrian linguliformean brachiopods. Alcheringa Australas. J. Palaeontol. *39*, 149–163.

Peng, J., Yuan, Q., Lin, B., Panneerselvam, P., Wang, X., Luan, X.L., Lim, S.K., Leung, B.P., Ho, B., and Ding, J.L. (2010). SARM inhibits both TRIF- and MyD88-mediated AP-1 activation. Eur. J. Immunol. *40*, 1738–1747.

Philippe, H., Lartillot, N., and Brinkmann, H. (2005). Multigene Analyses of Bilaterian Animals Corroborate the Monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Mol. Biol. Evol. *22*, 1246–1253.

Pila, E.A., Gordy, M.A., Phillips, V.K., Kabore, A.L., Rudko, S.P., and Hanington, P.C. (2016). Endogenous growth factor stimulation of hemocyte proliferation induces resistance to Schistosoma mansoni challenge in the snail host. Proc. Natl. Acad. Sci. U. S. A. *113*, 5305–5310.

Pila, E.A., Li, H., Hambrook, J.R., Wu, X., and Hanington, P.C. (2017). Schistosomiasis from a Snail's Perspective: Advances in Snail Immunity. Trends Parasitol. *33*, 845–857.

Pinto, M.R., Melillo, D., Giacomelli, S., Sfyroera, G., and Lambris, J.D. (2007). Ancient origin of the complement system: emerging invertebrate models. Adv. Exp. Med. Biol. *598*, 372–388.

Posenato, R., Holmer, L.E., and Prinoth, H. (2014). Adaptive strategies and environmental significance of lingulid brachiopods across the late Permian extinction. Palaeogeogr. Palaeoclimatol. Palaeoecol. *399*, 373–384.

Prado-Alvarez, M., Rotllant, J., Gestal, C., Novoa, B., and Figueras, A. (2009). Characterization of a C3 and a factor B-like in the carpet-shell clam, Ruditapes decussatus. Fish Shellfish Immunol. *26*, 305–315.

Prenant, L. (1928). Notes histologiques sur Terebratulina caput-serpentis L. Bull. Société Zool. Fr. *53*, 113–125.

1529 Quistad, S.D., Stotland, A., Barott, K.L., Smurthwaite, C.A., Hilton, B.J., Grasis, J.A., Wolkowicz, R., and
1530 Rohwer, F.L. (2014). Evolution of TNF-induced apoptosis reveals 550 My of functional conservation. Proc.
1531 Natl. Acad. Sci. *111*, 9567–9572.

1532 Reid, K.B., and Day, A.J. (1989). Structure-function relationships of the complement components. Immunol.
1533 Today *10*, 177–180.

1534 Ren, Q., Wang, C., Jin, M., Lan, J., Ye, T., Hui, K., Tan, J., Wang, Z., Wyckoff, G.J., Wang, W., et al. (2017). Co-
1535 option of bacteriophage lysozyme genes by bivalve genomes. Open Biol. *7*.

1536 Rodrigues, S.C. (2007). Biotic interactions recorded in shells of recent rhynchonelliform brachiopods from
1537 san juan island, usa. J. Shellfish Res. *26*, 241–252.

1538 Rodríguez de la Vega, R.C., and Possani, L.D. (2005). On the evolution of invertebrate defensins. Trends
1539 Genet. TIG *21*, 330–332.

1540 Romero, A., Dios, S., Poisa-Beiro, L., Costa, M.M., Posada, D., Figueras, A., and Novoa, B. (2011). Individual
1541 sequence variability and functional activities of fibrinogen-related proteins (FREPs) in the Mediterranean
1542 mussel (Mytilus galloprovincialis) suggest ancient and complex immune recognition models in
1543 invertebrates. Dev. Comp. Immunol. *35*, 334–344.

1544 Rosani, U., Varotto, L., Gerdol, M., Pallavicini, A., and Venier, P. (2015). IL-17 signaling components in
1545 bivalves: Comparative sequence analysis and involvement in the immune responses. Dev. Comp. Immunol.
1546 *52*, 255–268.

1547 Rosani, U., Pallavicini, A., and Venier, P. (2016). The miRNA biogenesis in marine bivalves. PeerJ *4*, e1763.

1548 Rowley, A.F., and Hayward, P.J. (1985). Blood cells and coelomocytes of the inarticulate brachiopod Lingula
1549 anatina. J. Zool. *205*, 9–18.

1550 Sachs, L., Chen, Y.-T., Drechsler, A., Lynch, J.A., Panfilio, K.A., Lässig, M., Berg, J., and Roth, S. (2015).
1551 Dynamic BMP signaling polarized by Toll patterns the dorsoventral axis in a hemimetabolous insect. eLife *4*,
1552 e05502.

1553 Sato, S., St-Pierre, C., Bhaumik, P., and Nieminen, J. (2009). Galectins in innate immunity: dual functions of
1554 host soluble beta-galactoside-binding lectins as damage-associated molecular patterns (DAMPs) and as
1555 receptors for pathogen-associated molecular patterns (PAMPs). Immunol. Rev. *230*, 172–187.

1556 Sekiguchi, R., and Nonaka, M. (2015). Evolution of the complement system in protostomes revealed by de
1557 novo transcriptome analysis of six species of Arthropoda. Dev. Comp. Immunol. *50*, 58–67.

1558 Sekiguchi, R., Fujito, N.T., and Nonaka, M. (2012). Evolution of the thioester-containing proteins (TEPs) of
1559 the arthropoda, revealed by molecular cloning of TEP genes from a spider, Hasarius adansoni. Dev. Comp.
1560 Immunol. *36*, 483–489.

1561 Sekine, H., Kenjo, A., Azumi, K., Ohi, G., Takahashi, M., Kasukawa, R., Ichikawa, N., Nakata, M., Mizuochi, T.,
1562 Matsushita, M., et al. (2001). An ancient lectin-dependent complement system in an ascidian: novel lectin
1563 isolated from the plasma of the solitary ascidian, Halocynthia roretzi. J. Immunol. Baltim. Md 1950 *167*,
1564 4504–4510.

1565 Šilerová, M., Procházková, P., Josková, R., Josens, G., Beschin, A., De Baetselier, P., and Bilej, M. (2006).
1566 Comparative study of the CCF-like pattern recognition protein in different Lumbricid species. Dev. Comp.
1567 Immunol. *30*, 765–771.

Simakov, O., Marletaz, F., Cho, S.-J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D.-H., Larsson, T., Lv, J., Arendt, D., et al. (2013). Insights into bilaterian evolution from three spiralian genomes. Nature *493*, 526–531.

Singer, H., Biswas, A., Zimmer, N., Messaed, C., Oldenburg, J., Slim, R., and El-Maarri, O. (2014). NLRP7 inter-domain interactions: the NACHT-associated domain is the physical mediator for oligomeric assembly. Mol. Hum. Reprod. *20*, 990–1001.

Škanta, F., Procházková, P., Roubalová, R., Dvořák, J., and Bilej, M. (2016). LBP/BPI homologue in Eisenia andrei earthworms. Dev. Comp. Immunol. *54*, 1–6.

Smith, L.C., Azumi, K., and Nonaka, M. (1999). Complement systems in invertebrates. The ancient alternative and lectin pathways. Immunopharmacology *42*, 107–120.

Söderhäll, I., Kim, Y.-A., Jiravanichpaisal, P., Lee, S.-Y., and Söderhäll, K. (2005). An Ancient Role for a Prokineticin Domain in Invertebrate Hematopoiesis. J. Immunol. *174*, 6153–6160.

Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. *33*, W244-248.

Song, L., Wang, L., Zhang, H., and Wang, M. (2015). The immune system and its modulation mechanism in scallop. Fish Shellfish Immunol. *46*, 65–78.

Sun, Y., Zhou, Z., Wang, L., Yang, C., Jianga, S., and Song, L. (2014). The immunomodulation of a novel tumor necrosis factor (CgTNF-1) in oyster Crassostrea gigas. Dev. Comp. Immunol. *45*, 291–299.

Tagawa, K., Yoshihara, T., Shibata, T., Kitazaki, K., Endo, Y., Fujita, T., Koshiba, T., and Kawabata, S. (2012). Microbe-Specific C3b Deposition in the Horseshoe Crab Complement System in a C2/Factor B-Dependent or -Independent Manner. PLOS ONE *7*, e36783.

Tahtouh, M., Croq, F., Vizioli, J., Sautiere, P.-E., Van Camp, C., Salzet, M., Daha, M.R., Pestel, J., and Lefebvre, C. (2009). Evidence for a novel chemotactic C1q domain-containing factor in the leech nerve cord. Mol. Immunol. *46*, 523–531.

Takahashi, K.G., and Muroga, K. (2008). Cellular Defense Mechanisms in Bivalve Molluscs. Fish Pathol. *43*, 1–17.

Takara, T., Nakagawa, T., Isobe, M., Okino, N., Ichinose, S., Omori, A., and Ito, M. (2011). Purification, molecular cloning, and application of a novel sphingomyelin-binding protein (clamlysin) from the brackishwater clam, Corbicula japonica. Biochim. Biophys. Acta *1811*, 323–332.

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. *56*, 564–577.

Tanaka, Y., and Chen, Z.J. (2012). STING specifies IRF3 phosphorylation by TBK1 in the cytosolic DNA signaling pathway. Sci. Signal. *5*, ra20.

Tasiemski, A. (2008). Antimicrobial peptides in annelids. Invertebr. Surviv. J. *5*, 75–82.

Tasiemski, A., and Salzet, M. (2017). Neuro-immune lessons from an annelid: The medicinal leech. Dev. Comp. Immunol. *66*, 33–42.

Tateno, H. (2010). SUEL-related lectins, a lectin family widely distributed throughout organisms. Biosci. Biotechnol. Biochem. *74*, 1141–1144.

Teichert, C. (1945). Parasitic worms in Permian brachiopod and pelecypod shells in Western Australia. Am. J. Sci. *243*, 197–209.

Tennessen, J.A. (2005). Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. J. Evol. Biol. *18*, 1387–1394.

Toubiana, M., Rosani, U., Giambelluca, S., Cammarata, M., Gerdol, M., Pallavicini, A., Venier, P., and Roch, P. (2014). Toll signal transduction pathway in bivalves: complete cds of intermediate elements and related gene transcription levels in hemocytes of immune stimulated Mytilus galloprovincialis. Dev. Comp. Immunol. *45*, 300–312.

Trexler, M., Bányai, L., and Patthy, L. (2000). The LCCL module. Eur. J. Biochem. *267*, 5751–5757.

Tschopp, J., Masson, D., and Stanley, K.K. (1986). Structural/functional similarity between proteins involved in complement- and cytotoxic T-lymphocyte-mediated cytolysis. Nature *322*, 831–834.

Tyler, C.L., Leighton, L.R., Carlson, S.J., Huntley, J.W., and Kowalewski, M. (2013). Predation on Modern and Fossil Brachiopods: Assessing Chemical Defenses and Palatability. PALAIOS *28*, 724–735.

Vasta, G.R., Ahmed, H., Bianchet, M.A., Fernández-Robledo, J.A., and Amzel, L.M. (2012). Diversity in recognition of glycans by F-type lectins and galectins: molecular, structural, and biophysical aspects. Ann. N. Y. Acad. Sci. *1253*, E14–E26.

Vetvicka, V., and Sima, P. (2009). Origins and functions of annelide immune cells: the concise survey. Invertebr. Surviv. J. *6*, 138–143.

Vinn, O., Wilson, M.A., and Toom, U. (2014). Earliest rhynchonelliform brachiopod parasite from the Late Ordovician of northern Estonia (Baltica). Palaeogeogr. Palaeoclimatol. Palaeoecol. *411*, 42–45.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. Theor. Den Biowissenschaften *131*, 281–285.

Wang, G.-D., Zhang, K.-F., Zhang, Z.-P., Zou, Z.-H., Jia, X.-W., Wang, S.-H., Lin, P., and Wang, Y.-L. (2008). Molecular cloning and responsive expression of macrophage expressed gene from small abalone Haliotis diversicolor supertexta. Fish Shellfish Immunol. *24*, 346–359.

Wang, P.-H., Wan, D.-H., Pang, L.-R., Gu, Z.-H., Qiu, W., Weng, S.-P., Yu, X.-Q., and He, J.-G. (2012). Molecular cloning, characterization and expression analysis of the tumor necrosis factor (TNF) superfamily gene, TNF receptor superfamily gene and lipopolysaccharide-induced TNF-α factor (LITAF) gene from Litopenaeus vannamei. Dev. Comp. Immunol. *36*, 39–50.

Wang, X., Wang, M., Xu, Q., Xu, J., Lv, Z., Wang, L., and Song, L. (2017). Two novel LRR and Ig domain-containing proteins from oyster Crassostrea gigas function as pattern recognition receptors and induce expression of cytokines. Fish Shellfish Immunol. *70*, 308–318.

Weber, C.H., and Vincenz, C. (2001). The death domain superfamily: a tale of two interfaces? Trends Biochem. Sci. *26*, 475–481.

Werner, T., Liu, G., Kang, D., Ekengren, S., Steiner, H., and Hultmark, D. (2000). A family of peptidoglycan recognition proteins in the fruit fly Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. *97*, 13772–13777.

Wiens, M., Korzhev, M., Krasko, A., Thakur, N.L., Perović-Ottstadt, S., Breter, H.J., Ushijima, H., Diehl-Seifert, B., Müller, I.M., and Müller, W.E.G. (2005). Innate immune defense of the sponge Suberites domuncula

against bacteria involves a MyD88-dependent signaling pathway. Induction of a perforin-like molecule. J. Biol. Chem. *280*, 27949–27959.

Wilde, C.G., Seilhamer, J.J., McGrogan, M., Ashton, N., Snable, J.L., Lane, J.C., Leong, S.R., Thornton, M.B., Miller, K.L., and Scott, R.W. (1994). Bactericidal/permeability-increasing protein and lipopolysaccharide (LPS)-binding protein. LPS binding properties and effects on LPS-mediated cell activation. J. Biol. Chem. *269*, 17411–17416.

Yatsu, N. (1902). Notes On Histology Of Lingula Anatina Brugière... J Coll Sci Imp Univ Tokyo *17*, 1–29.

Yoneyama, M., and Fujita, T. (2007). Function of RIG-I-like Receptors in Antiviral Innate Immunity. J. Biol. Chem. *282*, 15315–15318.

Yoshino, T.P., Dinguirard, N., Kunert, J., and Hokke, C.H. (2008). Molecular and functional characterization of a tandem-repeat galectin from the freshwater snail Biomphalaria glabrata, intermediate host of the human blood fluke Schistosoma mansoni. Gene *411*, 46–58.

Yuasa, H.J., Furuta, E., Nakamura, A., and Takagi, T. (1998). Cloning and sequencing of three C-type lectins from body surface mucus of the land slug, Incilaria fruhstorferi. Comp. Biochem. Physiol. B Biochem. Mol. Biol. *119*, 479–484.

Zhang, S.-M., and Coultas, K.A. (2011). Identification and characterization of five transcription factors that are associated with evolutionarily conserved immune signaling pathways in the schistosome-transmitting snail Biomphalaria glabrata. Mol. Immunol. *48*, 1868–1881.

Zhang, D., Ma, J., Jiang, J., Qiu, L., Zhu, C., Su, T., Li, Y., Wu, K., and Jiang, S. (2010). Molecular characterization and expression analysis of lipopolysaccharide and β-1,3-glucan-binding protein (LGBP) from pearl oyster Pinctada fucata. Mol. Biol. Rep. *37*, 3335–3343.

Zhang, H., Wang, L., Song, L., Zhao, J., Qiu, L., Gao, Y., Song, X., Li, L., Zhang, Y., and Zhang, L. (2009). The genomic structure, alternative splicing and immune response of Chlamys farreri thioester-containing protein. Dev. Comp. Immunol. *33*, 1070–1076.

Zhang, L., Li, L., Guo, X., Litman, G.W., Dishaw, L.J., and Zhang, G. (2015). Massive expansion and functional divergence of innate immune genes in a protostome. Sci. Rep. *5*.

Zumwalt, G.S., and Delaca, T.E. (1980). Utilization of Brachiopod Feeding Currents by Epizoic Foraminifera. J. Paleontol. *54*, 477–484.

Genetic and molecular basis of the immune system in the brachiopod *Lingula anatina*

Marco Gerdol, Yi-Jyun Luo, Noriyuki Satoh, Alberto Pallavicini

Supplementary figures and tables

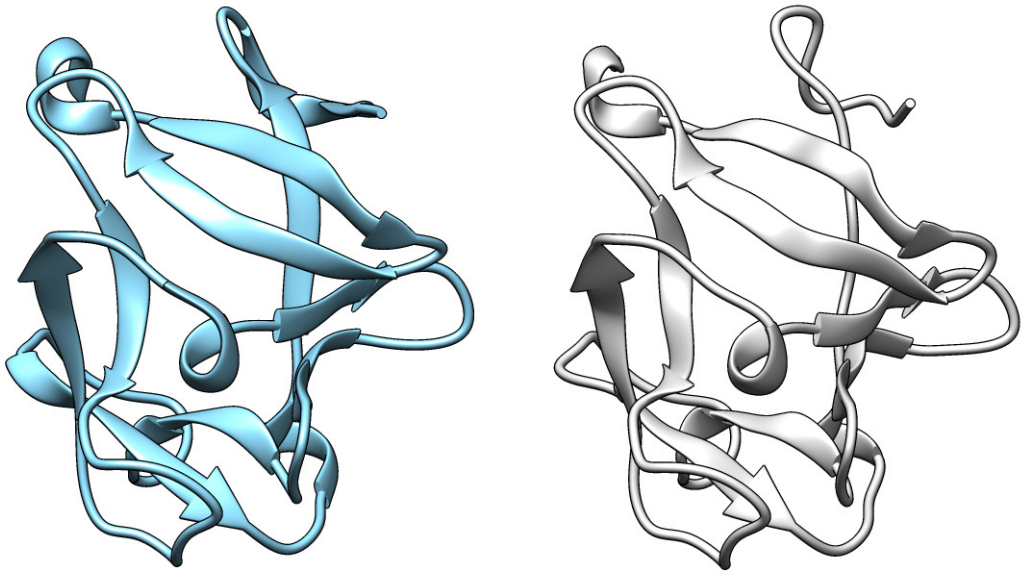
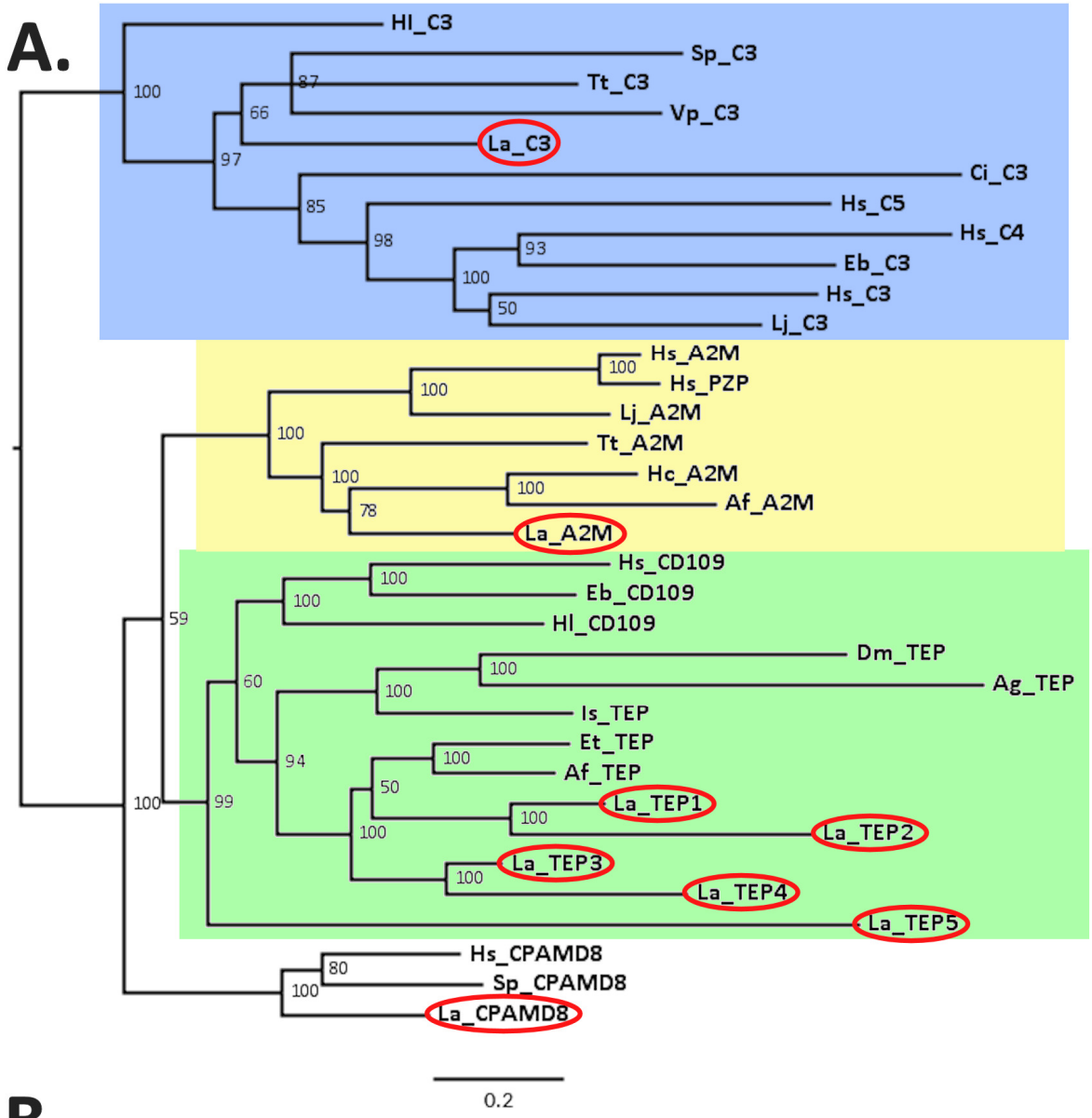
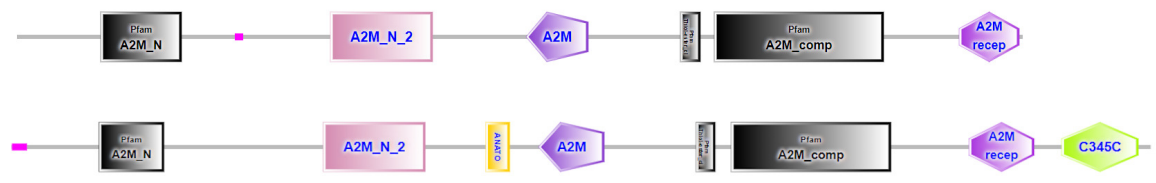


Figure S1: Three-dimensional structure of *Mytilus galloprovincialis* MytiLec-1 (PDB accession ID: 3WMV, left) and folding of the *Lingula anatina* mytilectin XP_013383925.1, as predicted by Phyre2 (confidence: 97.8%; coverage: 43%), displayed as a ribbon structure (right). Only the region corresponding to amino acids 36-145 is displayed for MytiLec-1. Note that brachiopod mytilectins also possess an N-terminal signal peptide regions and a C-terminal pore-forming domain that could not be modeled based on MytiLec-1.


A.
Hl_C3
Sp_C3
Tt_C3
Vp_C3
La_C3
Ci_C3
Hs_C5
Hs_C4
Eb_C3
Hs_C3
Lj_C3
Hs_A2M
Hs_PZP
Lj_A2M
Tt_A2M
Hc_A2M
Af_A2M
La_A2M
Hs_CD109
Eb_CD109
Hl_CD109
Dm_TEP
Ag_TEP
Is_TEP
Et_TEP
Af_TEP
La_TEP1
La_TEP2
La_TEP3
La_TEP4
La_TEP5
Hs_CPAMD8
Sp_CPAMD8
La_CPAMD8
100
87
66
97
85
98
93
100
50
100
100
100
100
78
59
100
60
100
94
99
50
100
100
100
100
80
100
0.2


B.
Pfam
A2M_N
A2M_N_2
A2M
Pfam
A2M_comp
A2M
recep
Pfam
A2M_N
A2M_N_2
A2M
Pfam
A2M_comp
A2M
recep
C345C

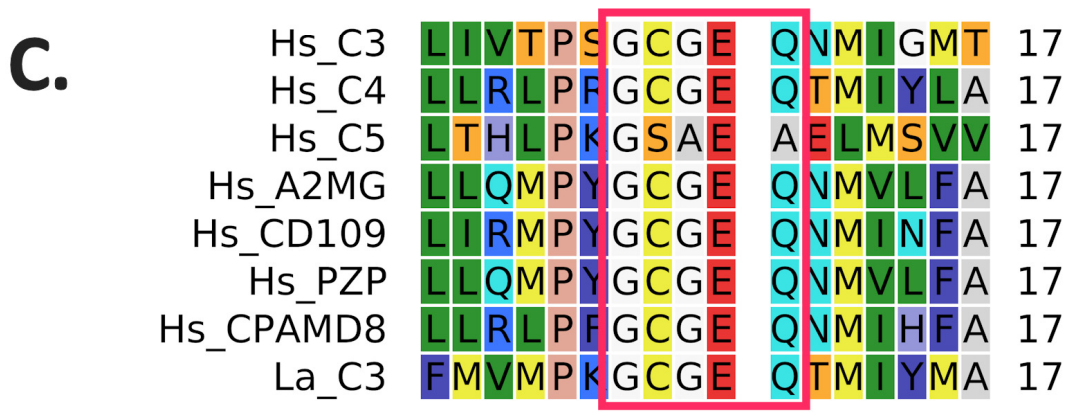

C.
Hs_C3 LIVTPSGCGEQNMIGMT 17
Hs_C4 LLRLPRGCGEQTMIYLA 17
Hs_C5 LTHLPKGSAEAELMSVV 17
Hs_A2MG LLQMPYGCGEQNMVLFA 17
Hs_CD109 LIRMPYGCGEQNMINFA 17
Hs_PZP LLQMPYGCGEQNMVLFA 17
Hs_CPAMD8 LLRLPRGCGEQNMIHFA 17
La_C3 FMVMPKGCGEQTMIYMA 17

Figure S2: Panel A: Bayesian phylogeny of metazoan thioester containing proteins. The analysis was carried out with MrBayes, based on a WAG+G+I model of molecular evolution. The analysis was stopped after the reaching of convergence of all estimated parameters (100,000 generations). The sequences from *Lingula anatina* are highlighted with a red circle. Posterior probability support values for all nodes are shown. The sequences clustered within the C3/C4/C5 family clade (light blue background), the alpha-2 macroglobulin clade (yellow background) and the CD109/iTEP clade (light green background). Sequences are named based on the initial letter of the genus and species scientific names. The accession IDs of the sequences used for this analysis are: *Homo sapiens* C3, C4A, C5, A2M, CD109, CPAMD8, and PZP:NP_000055, P0C0L4, AAA51925, P01023, NP_598000, NP_056507 and CAA38255. *Strongylocentrotus purpuratus* C3 and CPAMD8: NP_999686 and XP_785018. *Ciona intestinalis* C3-1, and CPAMD8: NP_001027684, and XP_002124325. *Lethenteron japonicum* C3 and A2M: Q00685 and BAA02762. *Eptatretus burgeri* C3 and CD109: P98094 and BAD12264. *Tachypleus tridentatus* C3 and A2M: BAH02276 and BAA19844. *Venerupis decussatus* C3: ACN37845; *Hyriopsis cumingii*: A2M: ABJ89824. *Haliplanella lineata* C3-1 and CD109: BAJ05269 and BAJ05272. Apis mellifera A2M: XP_392454. *Drosophila melanogaster* iTEP: NP_523578. *Anopheles gambiae* iTEP: AAG00600. *Ixodes scapularis* iTEP: XP_002409560. *Euphaedusa tau* TEP: BAE44110. *Azumapecten farreri* TEP: ADA77515; *Lingula anatina* C3, A2M, CAPMD8, TEP1, TEP2, TEP3, TEP4 and TEP: XP_013420336, XP_013397771, XP_013402650, XP_013389531, XP_013387199, XP_013383486, XP_013420587.

Panel B: domain architecture of iTEP, CD109 and A2M proteins (top), compared to complement C3/C4/C5 family members (bottom). The *Lingula anatina* C3 protein shows the C-terminal netrin domain, as expected.

Panel C: sequence alignment of the conserved GCGEQ thioester motif in human thioester proteins. This domain is also present, with no variations, in *Lingula anatina* C3.

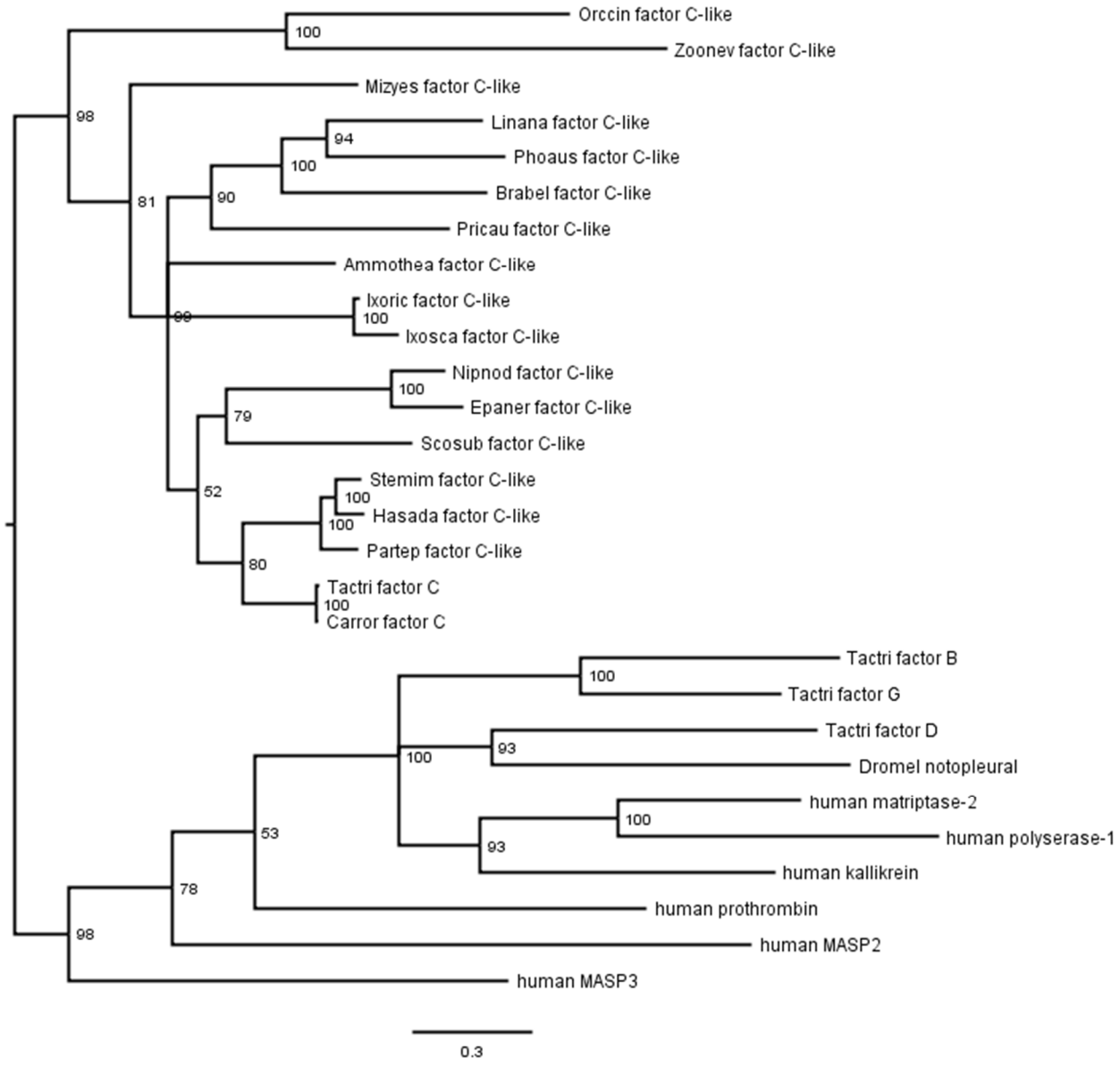


Figure S3: Bayesian phylogeny of Factor C and Factor C-like proteins from arthropods and lophotrochozoans, including the Factor C-like protein identified in *Lingula anatina*. The tree includes the sequences previously reported By Sekiguchi and Nonaka (2015) and other relevant animal sequences from human (MASP2/3, kallikrein, prothrombin, polyserase-1 and matriptase-2), *Drosophila melanogaster* (notopleural) and horseshoe crabs (Factor B, D and D). The phylogenetic tree is based on the multiple sequence alignment of the serine-protease domain shared by all sequences. Posterior probability values are shown for the main nodes of the tree. List of accession IDs of the sequences used: P28175.1, Q27081.1, BAA13312.1, BAA04045.1, XP_013394674.1, XP_021349166.1, XP_019627213.1, XP_011511292.1, CAC15568.1, CAC85953.1, NP_000497.1, XP_011526284.1, NP_001305325.1, XP_014664365.1, BAR45632.1, AAB34362.1, KFM72697.1, BAR45633.1, XP_015930211.1, AII02148.1, XP_002410620.1, XP_020999971.1, BAR45635.1, BAR45636.1, BAR45634.1, ODM96450.1, XP_021919270.1, NP_610437.2. The sequence from *Phoronis australis* was retrieved from the OIST *Phoronis australis* genome browser (http://marinegenomics.oist.jp/pau_v2/viewer/info?project_id=51), accession ID: TRINITY_DN303042_c8_g1_i1.

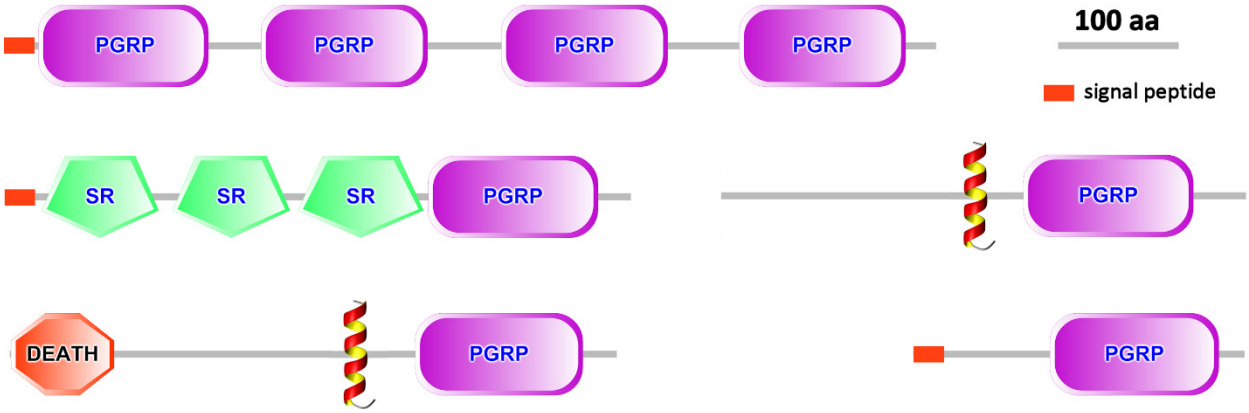


Figure S4: Domain architecture of peptidoglycan recognition proteins (PGRPs) encoded by the genome of *L. anatina*. SR: Scavenger receptor Cys-rich.

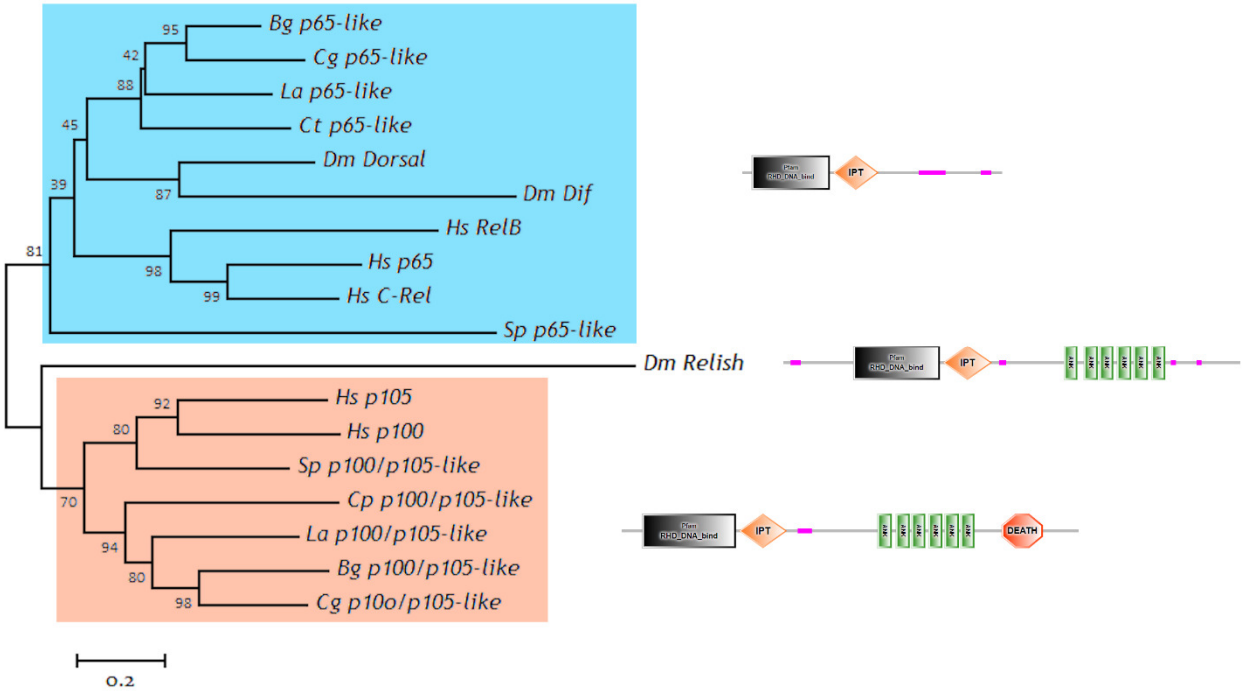


Figure S5: Phylogenetic relationship between class I (light pink) and class II (light blue) NF-κB family members from human (Hs), *Lingula anatina* (la), *Biomphalaria glabrata* (Bg), *Crassostrea gigas* (Cg), *Capitella teleta* (Ct), *Drosophila melanogaster* (Dm), *Carcinoscorpius rotundicauda* (Cr) and *Strongylocentrotus purpuratus* (Sp). Class I members present C-terminal ankyrin repeats, which are followed by a DEATH domain in deuterostome and lophotrochozoan p100/p105-lie protein. This domain is not present in *Drosophila* Relish (shown with a white background). Accession IDs: P15330.2 (Dm_Dorsal), Q94527.1 (Dm_relish), P98149.2 (Dm_Dif), P19838.2 (Hs_p105), Q00653.4 (Hs_p100), Q04206.2 (Hs_p65), Q04864.1 (Hs_C-Rel), Q01201.2 (Hs_RelB), ELU08708.1 (Ct_p100/p105-like), ELU04547.1 (Ct_p65-like), XP_011684092.1 (Sp_p100/p105-like), XP_780741.2 (Sp_p65-like), XP_011423502.1 (Cg_p100/p105-like), NP_001292261.1 (Cg_p65-like), XP_013405334.1 (La_p100/p105-like), XP_013391679.1 (La_p65-like), ABC75034.1 (Cr_p100/p105-like), AAZ40333.1 (Cr_p65-like), ACZ25560.1 (Bm_p100/p105-like), ACZ25559.1 (Bm_p65-like).

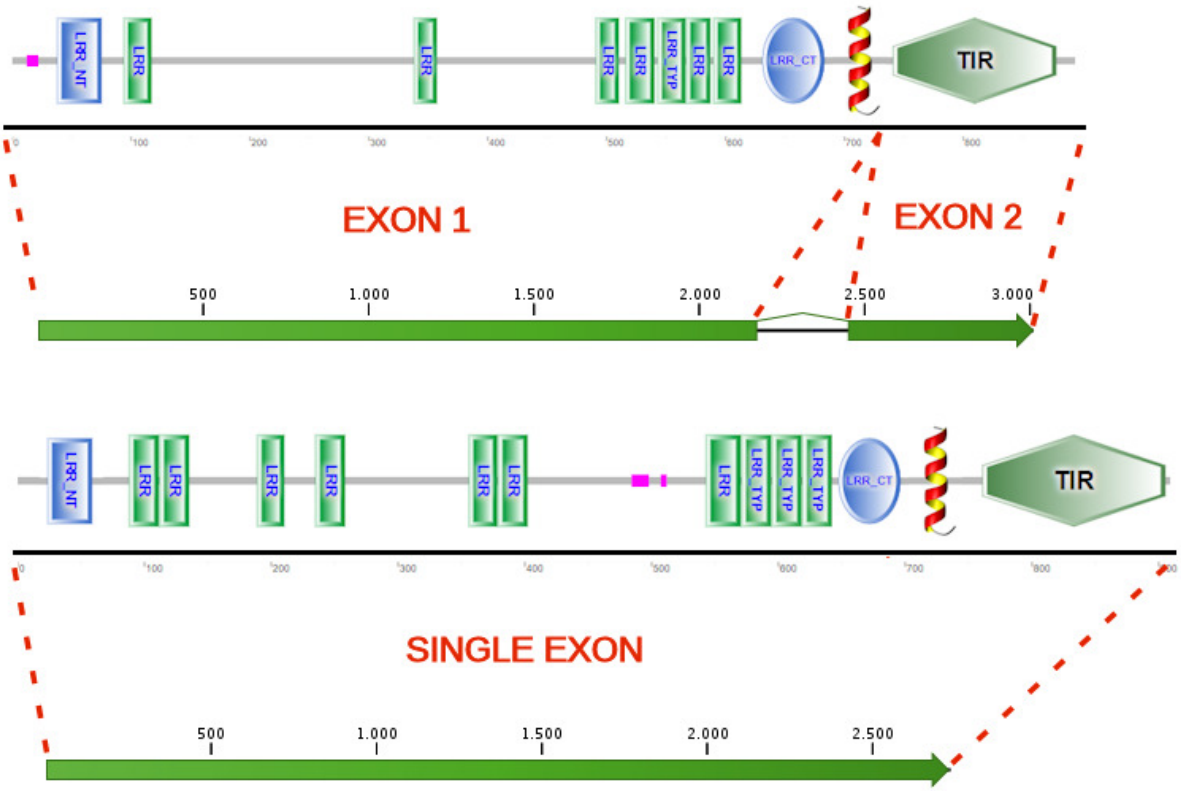


Figure S6: Exemplified genomic architecture of *L. anatina* sccTLR genes. Above, two-exons single-cysteine cluster TLRs comprise two exons; the first one encodes the entire N-terminal region comprising leucine-rich repeats and the transmembrane domain, whereas the TIR domain is entirely encoded by the second exon. Below, intronless single-cysteine cluster TLRs contain the entire open reading frame in a single exon.

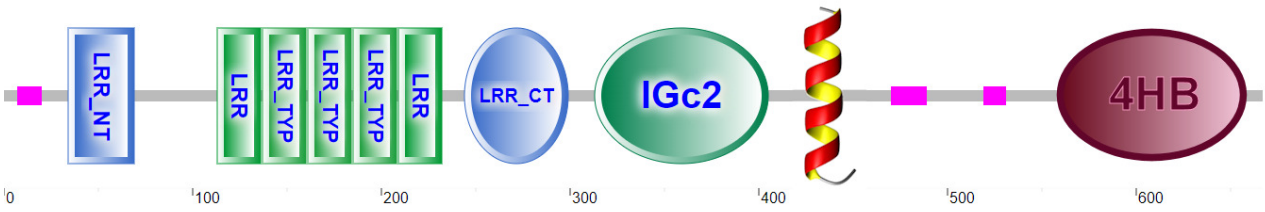


Figure S7: Domain organization of *Lingula anatina* LRRIGs, as exemplified by the sequence XP_013384809.1. Briefly, extracellular LRR domains are followed by an immunoglobulin-like (IGc2) domain. The intracellular part of the receptors presents a four helix bundle (4HB) domain structurally similar to apolipoproteins.

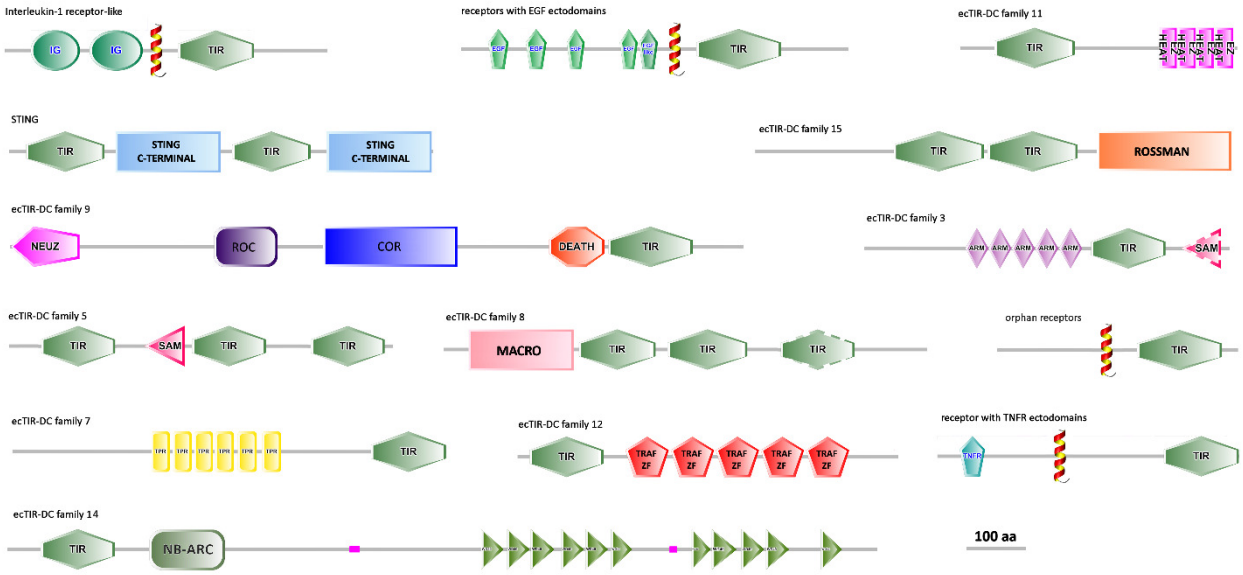


Figure S8: Schematic representation of brachiopod evolutionarily conserved TIR-domain-containing proteins, named based on the scheme proposed by Gerdol et al. 2017a. Non-canonical domains are dashed. IG: immunoglobulin; EGF: epidermal growth factor; EZ_HEAT: E-Z type HEAT repeats; ROSSMAN: Rossman-like fold domain; NEUZ: naturalized domain; ROC: ROC domain; COR: C-terminal of Roc domain; DEATH: Death domain; ARM: armadillo-type repeats; TRAF ZF: TRAF-like zinc finger domain; TNFR: TNF-receptor domain; TPR: tetratricopeptide repeats; WD40: WD40 repeats; NB-ARC: NB-ARC domain.

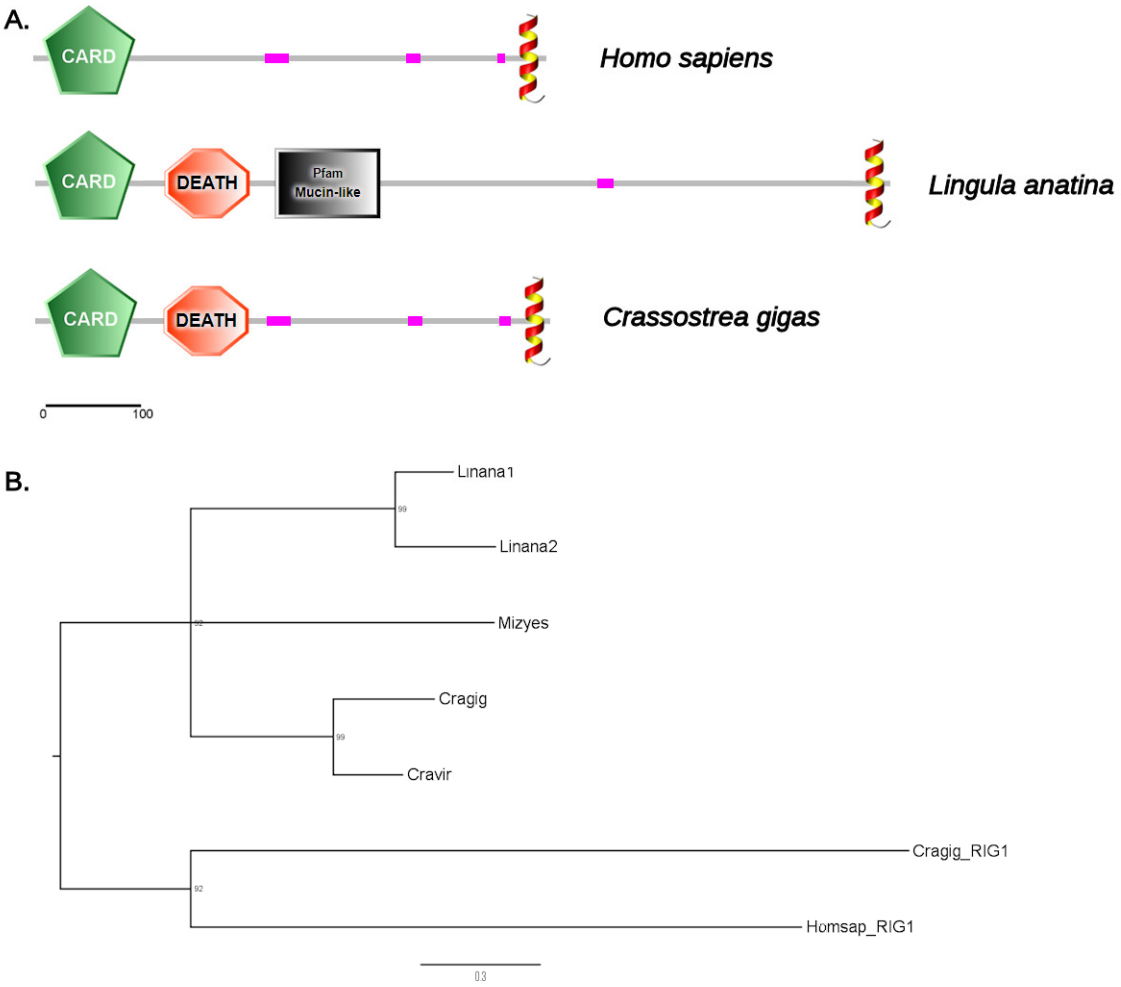
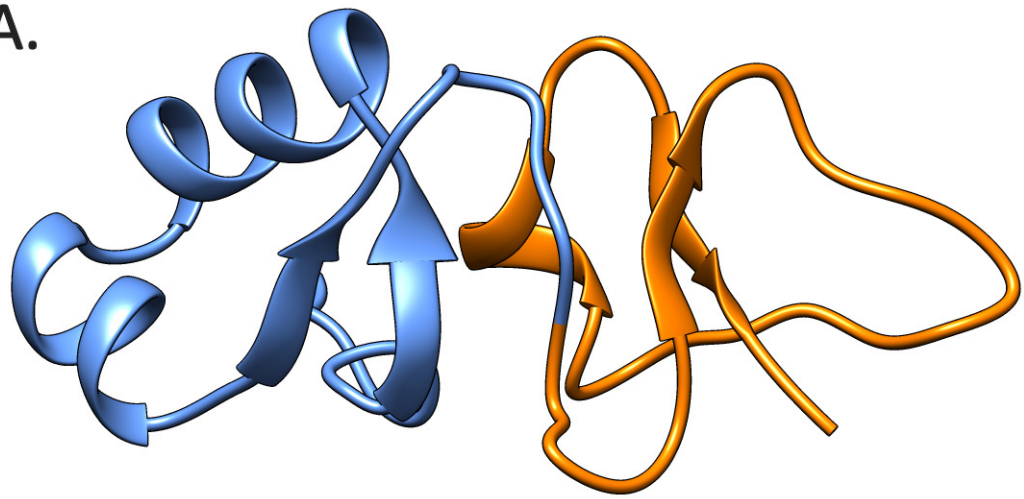


Figure S9: A: domain architecture of human MAVS/CARDIF/IPS-1, its functional homolog CgMAVS from *C. gigas* and the related sequence from *L. anatina*. B: Bayesian phylogeny of lophotrochozoan MAVS-like sequences from *C. gigas* (KY630189.1), *Crassostrea virginica* (XP_022343364.1), *Mizuhopecten yessoensis* (XP_021350730.1) and *L. anatina* (XP_013403654.1 and XP_013403698.1). Human (NP_055129.2) and oyster (KY630188) RIG-1 sequences were used as outgroups to root the tree. Nodes supported by posterior probabilities lower than 50% were collapsed.

A.



B.

```
Laqcal        CANNRGWCRPRCLSYQRA-DWYHSA-VCGSYTCC
Liouva        CANNRGWCRHRCFSHEGIYTRNDADIICGHYHCC
Krarub        CAGNRGWCRVSCFFREHS-DSYYGD-VCGKYNCC
Magven        CANNRGWCRRICFSHERVYTWNDANIICGKYDCC
Tactri        CAGNRGWCRSKCFRHEYV-DTYYSA-VCGRYFCC
Brabel        CANNRGWCRSRCFRHEYI-DSWHSD-VCGSYDCC
Asyluc        CANNRGWCRQTCFDHEYI-DWYYTD-VCGNYKCC
Hyrcum        CANNRGWCRLSCFSHEYV-DWYNTA-VCGYYKCC
Argirr        CYGNRGWCRSSCRSYEREYRGGNLG-VCGSYKCC
Cragig_BD1    CANNRGWCRPTCFSHEYT-DWFNND-VCGSYRCC
Rudphi        CANNRGWCRPTCFSHEYI-DGYHSA-ICGGNSCC
Mytgal_BD1    CANNRGWCRAICFDHEVV-DHYHSD-ICGAYKCC
Mytgal_BD3a   CANNRGWCRPNCGRGEYH-NWYHSS-TCGFYKCC
Mytgal_BD4    CANGRGFCRARCFANESM-SLYFSS-LCGAYKCC
Mimnob        CRGNSGWCRPKCARYERE-YTANLG-VCGRNKCC
Mizyes        CRGNRGWCRSTCFSHEREYRGGNLG-VCGRNKCC
Drepol        CANNRGWCRSACFSHEYV-DHYHSD-VCGYFKCC
Mermer        CANNRGWCRSTCFSHEYI-DWAHTA-VCGNYHCC
Villie        CANNRGWCRKFCSSHEYV-DWYSTD-VCGSYRCC
consensus     C  nrGwCR  C   e          vCG   CC
```

Figure S10: A: Structural model of *Magellania venosa* big defensin, predicted by Phyre 2; the N-terminal hydrophobic and the C-terminal cysteine-rich domains are shown in blue and orange, respectively. B: Sequence alignment of the cysteine array of representative big defensins sequences from Brachiopoda (marked by a box), Bivalvia, Merostomata and Cephalochordata. Laqcal: *Laqueus californianus*; Liouva: *Liothyrella uva*; Krarub: *Kraussina rubra*; Magven: *Magellania venosa*; Tactri: *Tachypleus tridentatus*; Brabel: *Branchiostoma belcheri*; Asyluc: *Asymmetron lucayanum*; Hyrcum: *Hyropsis cumingii*; Argirr: *Argopecten irradians*; Cragig: *Crassostrea gigas*; Rudphi: *Ruditapes philippinarum*; Mytgal: *Mytilus galloprovincialis*; Mimnob; *Mimachlamys nobilis*; Mizyes: *Mizuhopecten yessoensis*; Drepol: *Dreissena polymorpha*; Mermer: *Meretrix meretrix*; Villie: *Villosa lienosa*. Accession IDs are provided in **Table S3**.

Species name	Order	Reads (M)	Assembled contigs	Assembly N50	Tissue
Glottidia pyramidata	Lingulida	68	42,738	872	pedicle, mantle, and lophophore
Hemithiris psittacea	Rhynchonellida	61	23,646	1439	lophophore and mantle
Kraussina rubra	Terebratulida	45	28,468	496	soft tissues
Laqueus californianus	Terebratulida	67	37,509	846	lophophore
Liothyrella uva	Terebratulida	283	52,268	1234	whole animal
Magellania venosa	Terebratulida	420	39,872	1476	mantle
Novocrania anomala	Craniida	52	14,093	485	soft tissues
Terebratalia transversa	Terebratulida	105	38,948	1216	whole animal

Table S1: Summary of the transcriptome data used for comparative immunogenomics screening in this study.

Sequence name	Accession ID	Sequence name	Accession ID
Linana_1	XP_013384618.1	Spisac_glucanase	AAP74223.1
Linana_2	XP_013384636.1	Chlalb_glucanase	AAZ04385.1
Linana_3	XP_013400557.1	Chlros_glucanase	AAZ04386.1
Linana_4	XP_013408261.1	Pervir_glucanase	ACM68926.1
Linana_5	XP_013413498.1	Myzyes_glucanase	XP_021378165.1
Linana_6	XP_013408262.1	Taplit_glucanase	AEE89454.1
Dromel_GNBP1	NP_524142.2	Haldis_glucanase	BAH84971.1
Dromel_GNBP2	NP_524141.1	Haltub_glucanase	AFQ98375.1
Dromel_GNBP3	NP_523986.2	Litsik_glucanase	ACN22491.1
Eisfet_CCF1	AAC35887.1	Azufar_LGBP	AAP82240.1
Kumrub_CCF	AAY85746.1	Pinfuc_LGBP	ACN76701.1
Denven_CCF	AAY85745.1	Haldiv_LGBP	ACO07358.1
Aporos_CCF	AAY85744.1	Mermer_LGBP	AGT42010.1
Apolon_CCF	AAY85743.1	Cragig_LGBP	NP_001292282.1
Apoict_CCF	AAY85742.1	Taplit_LGBP	AEE89455.1
Apocal_CCF	AAY79172.1	Cragig_LGBP	NP_001292293.1
Dromel_GNBP3	NP_523986.2		

Table S2: accession number of the protein sequences used for the generation of the phylogenetic tree in Figure 4.

Sequence name	Accession ID
Argirr	Q0H293.1
Asyluc	GESY01089460.1
Brabel	Q86QN6.1
Carrot	CK086629.1
Cragig_BD1	AEE92767.1
Drepol	GFBQ01007908.1
Hyrcum	AEP26934.1
Mermer	JI261145.1
Mimnob	ANY95044.1
Mizyes	OWF47949.1
Mytgal_BD1	CCC15007.1
Mytgal_BD3a	CCC15009.1
Mytgal_BD4	CCC15012.1
Rudphi	AEK78068.1
Scabro	AFQ02696.1
Tactri	P80957.2
Villie	JR534503.1

Table S3: accession number of the sequences displayed in the multiple sequence alignment presented in Figure 10. The following sequences were obtained from *de novo* transcriptome assembly:

>Laqcal
MRQVFLSTAIWTLLLVNQMLCGAVKVKQTDHPEEMEKRVFPAIYIGAAVSPWVYAWLVAAFGAALVLANHVTIRRRNDN
HSCANNRGWCRPRCLSYQRADWYHSAVCGSYTCCRPK
>Magven
MGRIIFLAAFWTSLFVGQLTAGTLENERVHDEKEKRAIALPFWYIGVRVSPWVWRGLVALYTAAVLSRNRVVKVDSDSHSC
ANNRGWCRRICFSHERVYTWNDANIICGKYDCCVPAGGQVGR
>Krarub
MGRVFLSTVIWTLLLANQVLGAIIVKQGDQENTREKRATFAIPLIYVGAKVSLRVWAALVALYGVTALYASHVIKVSSDNHS
CAGNRGWCRVSCFFREHSDSYYGDVCGKYNCCRY
>Liouva
MCREYWSAIIWASLLVSHVTVGHSLNVRSHQTEQKNEKKQIALPYWYFGIAVSPWVWAALAVTYGGVLLLRNHVTKVDT
DSHNCANNRGWCRHRCFSHEGIYTRNDADIICGHYHCCVPKDISIG