Accepted Manuscript

An updated molecular basis for mussel immunity

Marco Gerdol, Paola Venier

PII: S1050-4648(15)00066-2

DOI: 10.1016/j.fsi.2015.02.013

Reference: YFSIM 3327

To appear in: *Fish and Shellfish Immunology*

Received Date: 15 November 2014

Revised Date: 2 February 2015

Accepted Date: 9 February 2015

Please cite this article as: Gerdol M, Venier P, An updated molecular basis for mussel immunity, *Fish and Shellfish Immunology* (2015), doi: 10.1016/j.fsi.2015.02.013.

Highlights Gerdol and Venier FSIM SI 2014

Mussels are not as distresses by pathogens as other bivalves are

We analyzed all the available *Mytilus galloprovincialis* sequence data

We propose an updated molecular view of mussel immune responses

We report a number of novelties concerning the various mussel PRRs

We outlined traceable elements of the mussel immune signaling

An updated molecular basis for mussel immunity

Authors

Marco Gerdol[1], Paola Venier[2*]

[1]*Department of Life Sciences, University of Trieste, via L. Giorgeri 5, 34127 Trieste, ITALY; mgerdol@units.it*

[2]*Department of Biology, University of Padua, via U. Bassi 58/b, 35121 Padua, ITALY; TEL: +390498276284; venier@bio.unipd.it*

*corresponding author

Abstract

Non-self recognition with the consequent tolerance or immune reaction is a crucial process to succeed as living organisms. At the same time the interactions between host species and their microbiome, including potential pathogens and parasites, significantly contribute to animal life diversity. Marine filter-feeding bivalves, mussels in particular, can survive also in heavily anthropized coastal waters despite being constantly surrounded by microorganisms. Based on the first outline of the *Mytilus galloprovincialis* immunome dated 2011, the continuously growing transcript data and the recent release of a draft mussel genome, we explored the available sequence data and scientific literature to reinforce our knowledge on the main gene-encoded elements of the mussel immune responses, from the pathogen recognition to its clearance. We carefully investigated molecules specialized in the sensing and targeting of potential aggressors, expected to show greater molecular diversification, and outlined, whenever relevant, the interconnected cascades of the intracellular signal transduction.

Aiming to explore the diversity of extracellular, membrane-bound and intracellular pattern recognition receptors in mussel, we updated a highly complex immune system, comprising molecules which are described here in detail for the first time (e.g. NOD-like receptors) or which had only been partially characterized in bivalves (e.g. RIG-like receptors). Overall, our comparative sequence analysis supported the identification of over 70 novel full-length immunity-related transcripts in *M. galloprovincialis*. Nevertheless, the multiplicity of gene functions relevant to immunity, the involvement of part of them in other vital processes, and also the lack of a refined mussel genome make this work still not-exhaustive and support the development of more specific studies.

Keywords

Mytilus galloprovincialis, innate immunity, transcriptome, bivalves.

List of abbreviations

AMP: antimicrobial peptide

BD: big defensin

BIR: baculovirus inhibitor of apoptosis protein repeat

BPI: bactericidal/permeability increasing protein

C1qDC: C1q domain-containing

CARD: caspase recruitment domain

CpG-DNA: CpG oligodeoxynucleotides

CLECT: C-type lectin domain

CRD: carbohydrate recognition domain

CS-αβ: cystine-stabilized alpha-beta motif

CTL: C-type lectin

Gram+: Gram positive [staining]

Gram-: Gram negative [staining]

GNBP: Gram-negative binding protein

iE-DAP: γ-D-Glu-meso-diaminopimelic acid

IFN: interferon

IPS-1: IFN-beta promoter stimulator

IRF: interferon regulatory factors

JNK: c-JUN N-terminal kinase

LGBP: lipopolysaccharide and ß-1, 3-glucan binding proteins

LRR: leucine-rich repeats

MAP3K: mitogen-activated protein kinase kinase kinase

MAMP: microbe associated molecular pattern

MAPK: mitogen-activated protein kinase

MAPKK: mitogen-activated protein kinase kinase

MKK: mitogen-activated protein kinase kinase

MDP: muramyl dipeptide

NGS: next generation sequencing

NLR: NOD-like receptor

PAMP: pathogen associated molecular pattern

PGN: peptidoglycan

PGRP: peptidoglycan recognition protein

PO: prophenoloxidase

PRR: pattern recognition receptors

RLR: RIG-like receptor

SRCR: scavenger receptor cysteine-rich

STING: stimulator of interferon genes

TIMP: tissue inhibitor of metalloproteinases

TIR: Toll-interleukin-1-receptor

TNF: tumor necrosis factor

TLR: Toll-like receptor

1. Introduction

The study of animal species often reveals taxon-specific patterns of evolutionary diversification, according to the organism life style and related environmental niches. In particular, the evolution of innate defense systems exposes the never-ending race between the animal host and more quickly evolving microorganisms, with the development of specialized host-pathogen (or host-parasite) interactions, independent events of gene loss or gene expansion and fast diversification of molecules essential for pathogen sensing and targeting [1]. As a matter of fact, the study of species unable to mount long-term adaptive responses has highlighted fascinating aspects of animal diversity and physiology in a changing environment [2,3].

Molecules and pathways of the innate immune response have been more extensively studied in invertebrates such as the fruit fly [4], the sea urchin [2] and cnidarians [5,6]. In comparison, the repertoire of gene-encoded elements composing the lophotrochozoan immunity has still to be revealed, particularly in molluscs, which represent the second most species-rich metazoan group with about 100,000 estimated extant species [7,8]. The first molluscan genome to be sequenced, pertaining to the gastropod *Lottia gigantea,* was released only in 2007, 7 and 9 years later than the genomes of *Drosophila melanogaster* and *Caenorhabditis elegans*, respectively. As regards bivalve molluscs, a class comprising species of great ecological and commercial importance, only in recent years the increasing accessibility of next generation sequencing (NGS) technologies permitted significant advances [9]. So far, just two bivalve draft genomes (*Crassostrea gigas* and *Pinctada fucata*) have been released but RNA-seq datasets for more than 40 different species have been already produced (NCBI SRA, accessed in November 2014).

The sequence data currently available for the the common mussel (*Mytilus* spp.) are summarized in **Table 1**. The first glimpse on the complex mussel immune system was provided by Sanger EST sequencing [10], an approach which was followed by 454 Life Sciences sequencing [11–15] and by high throughput Illumina sequencing, a technology allowing a better full-length reconstruction of transcripts [16]. In 2014, a non-annotated set of genomic sequences of *Mytilus galloprovincialis* was released, a real landmark for the progression of genomic studies in this bivalve [17].

Mussels are rather tolerant to environmental changes and they are therefore used as pollution sentinels in coastal waters but, more intriguingly, they appear less affected or not harmed by syndromes and infectious agents distressing other bivalves [18,19]. How mussels govern microorganisms associations with their seasonally varying amounts of microbe-associated molecular patterns (MAMPs) and virulence factors remains to be established. For these reasons, we have undertaken a revision of sequence and literature data to update our knowledge on the gene-encoded molecules shaping the strength and peculiarities of the innate responses of mussels in the context of their fluctuating holobiome. Starting from the first "immunome" description [20] and expanding the analysis to NGS datasets related to the blue mussel [13,21] and other bivalve species [22–24] we propose a step forward in the understanding of pathogen

recognition and clearance in *M. galloprovincialis*. Since the *de novo* assembly of RNA-seq data can provide only a partial view of the genes involved in mussel immune responses, we have often used the Pacific oyster *C. gigas* genome for comparison. Considering possible drawbacks inherent to the *de novo* assembly (transcript fragmentation, misassembly, etc.) we have deposited in GenBank only selected sequences of novel full length transcripts, highly supported either in terms of read coverage or confirmed by genomic sequences.

The functional validation of the novel mussel transcripts goes beyond the purpose of this work. As well, the comprehensive characterization of single genes or gene families (especially the analysis of regulatory gene elements and splicing patterns) is not affordable in a single paper nor it is feasible in the absence of a finished genome. While updating the available knowledge on the various molecules participating in the mussel immunity, we have paid more attention to receptors and effectors which likely undergo faster evolution rate and diversification, and described in detail only the key elements of the intricate, and evolutionarily more conserved, intracellular immune signaling. In fact, significant work is still needed to disentangle the interconnected pathways of intracellular signal transduction, which often depend on post-translational protein modifications, and to adequately investigate the presence of a regulatory cytokine-like network in mussels. Hence, only the signalling pathways clearly connected to the activation of specific pattern recognition receptors (PRRs) are reported in this paper.

Species	Sample	Sequencing technology	Sequencing strategy	Sequencing effort (Gbp)	Year of release	Reference**
M. galloprovincialis	mixed tissues	Sanger	EST-seq	<0.1	2009	[10]
M. galloprovincialis	mixed tissues	454	RNA-seq	<0.1	2010	[14]
M. edulis	mixed tissues	454	RNA-seq	1.1	2012	[13]
M. galloprovincialis	digestive gland	454	RNA-seq	1.5	2013	[11]
M. edulis	mantle	454	RNA-seq	0.3	2014	[12]
M. galloprovincialis	foot	454	Targeted genome sequencing	0.6	2014	[15]
M. galloprovincialis	digestive gland	Illumina	RNA-seq	8.1	2014	[16]
M. galloprovincialis	whole body	Illumina	RNA-seq	12.4	2014	PRJNA249058
M. edulis	whole body	Illumina	RNA-seq	10.9	2014	PRJNA249058
M. trossulus	whole body	Illumina	RNA-seq	5.8	2014	PRJNA249058
M. californianus	whole body	Illumina	RNA-seq	3.9	2014	PRJNA249058
M. edulis	larvae	Illumina	RNA-seq	32.8	2014	[21]
M. galloprovincialis	mantle	Illumina	Whole genome sequencing	1.6*	2014	[17]

Table 1. Overview of the sequence resources available for *Mytilus* spp. in Nov 2014. Species, samples and sequencing details, including the total sequencing effort, are reported for each study. *This number is referred to the assembled genome size. **For unpublished data, the Bioproject accession ID is reported.

2. Materials and Methods

2.1. Identification of mussel immunity-related transcripts

The Illumina RNA-seq data available for *M. galloprovincialis* (**Table 1**) were downloaded from the NCBI Sequence Read Archive and *de novo* assembled using Trinity (release 2014.04.13) with default parameters [25], setting the minimum allowed contig length to 200 nucleotides.

Based on literature data, we systematically identified protein sequences related to innate immunity in human, *D. melanogaster* and bivalve species and downloaded them from the NCBI protein database. These sequences were imported in the CLC Genomics Workbench 7.5 environment (CLC Bio, Aarhus, Denmark) and used as queries for tBLASTn searches to identify similar mussel sequences [26]. Positive matches, initially detected with a BLAST e-value threshold of $1x10^{-5}$, were checked for the presence of a complete open reading frame (from the initial ATG to the STOP codon). Whenever possible, partial sequences were elongated to their full length by comparison and reassembly with overlapping Trinity contigs or with those obtained in an alternative transcriptome assembly (*de novo assembly* with automatic detection of the *word size* and *bubble size* parameters using the CLC Genomics Workbench). If alternatively spliced variants were detected, only the contig encoding a full-length protein compared to the BLAST query was retained.

We carefully assessed the quality of the assembled mussel trascripts by mapping all the available paired-end Illumina reads on them, using the *map reads to contigs* tool and setting *length/similarity fraction* parameters to 0.75/0.95 and *insertion/deletion/mismatch* penalties to 3/3/3. Only the sequences consolidated by uniform Illumina read coverage were considered as trustwhorty and kept for further analysis. The correct assembly of mussel transcripts was further assessed by alignment with the corresponding genomic contigs, but this was only possible for a limited subset of sequences due to the high fragmentation of the released mussel genome assembly [17].

Virtual protein translations were checked for the presence of conserved domains with InterProScan v. 5.4-47.0 [27], whereas signal peptides and transmembrane domains were detected with Signalp v. 4.1 [28] and TMHMM v.2.0 [29], respectively. Specific cases where a signal peptide could not be detected in proteins expected to be targeted to the secretory pathway were further analyzed with SecretomeP 2.0 [30]. In the present paper, we only report sequences diplaying significant BLAST matches and sequence features consistent with data previously reported in other organisms and fully confirmed by a uniform read coverage. Peculiarities of the mussel gene transcripts compared to the domain organization expected in other organisms are reported, case by case, in the text.

2.2. Comparative genomics analyses

The completely annotated genome of the Pacific oyster *C. gigas* [31] was downloaded from EnsemblMetazoa (release oyster_v9, GCA_000297895.1). Genomic sequences related to immunity were identified using a strategy similar to the one described above for *M. galloprovincialis*, combining the BLAST searches with InterPro domain analyses. In the case of incongruent results between the mussel and oyster genome datasets, the possible presence of misannotated genes was further investigated at the transcript level, by searching the sequences of interest in an oyster transcript collection obtained by *de novo* assembly of RNA-seq data from multiple tissues (SRA accession IDs: SRR334212-20). This assembly was generated with the CLC Genomics Workbench *de novo* assembly tool, using the same strategy detailed above for the mussel transcriptome.

The size of specific gene families in oyster and in other organisms was estimated by the number of typical InterPro [32] or SUPERFAMILY [33] domain detected in each genome. Whenever useful, the combination of domains and their relative position were also taken into account for the correct classification of proteins within specific families.

3. Results and discussion

The complete list of the novel sequences reported in the present paper and their respective accession IDs are summarized in **Table 2**. The fragmentation of the released mussel genome (over 2 million contigs with a N50 value of ~1000) [17] prevented systematic searches and, therefore, the description of genomic landscapes and regulatory gene features (e.g. promoter elements and alternative splicing events) is not included in this work.

Sequence name	Putative function	GenBank accession ID
allograft inflammatory factor 1 (AIF1)	proinflammatory cytokine	KP125895
arthropod defensin-like 1	antimicrobial peptide	KP125907
arthropod defensin-like 1	antimicrobial peptide	KP125908
ATP-dependent RNA helicase DDX41	double-stranded DNA sensing in the cytoplasm	KP125906
bactericidal/permeability increasing protein 2	antimicrobial effector (BPI family)	KP125896
bactericidal/permeability increasing protein 3	antimicrobial effector	KP125945
complement component C3-like	complement component	KP125947
c-Jun N-terminal kinase	Intracellular signaling (MAPK pathway)	KP713438
C-type lectin 1	extracellular PRR (C-type lectin family)	KP125897
C-type lectin 2	extracellular PRR (C-type lectin family)	KP125898
C-type lectin 3	extracellular PRR (C-type lectin family)	KP125899
C-type lectin 4	extracellular PRR (C-type lectin family)	KP125900
C-type lectin 5	extracellular PRR (C-type lectin family)	KP125901
C-type lectin 6	extracellular PRR (C-type lectin family)	KP125902
C-type lectin 7	extracellular PRR (C-type lectin family)	KP125903
C-type lectin 8	extracellular PRR (C-type lectin family)	KP125904
C-type lectin 9	extracellular PRR (C-type lectin family)	KP125944
C-type lysozyme 2	antimicrobial effector (C-type lysozyme family)	KP125905
C-type lysozyme 3	antimicrobial effector (C-type lysozyme family)	KP125943
fibrinogen-related protein 10	extracellular PRR (FREP family)	KP125911
fibrinogen-related protein 11	extracellular PRR (FREP family)	KP125912
fibrinogen-related protein 12	extracellular PRR (FREP family)	KP125913
fibrinogen-related protein 8	extracellular PRR (FREP family)	KP125909
fibrinogen-related protein 9	transmembrane PRR (FREP family)	KP125910
galectin 1	extracellular PRR (galectin family)	KP125894
galectin 2	extracellular PRR (galectin family)	KP125914
galectin 3	extracellular PRR (galectin family)	KP125915
galectin 4	extracellular PRR (galectin family)	KP125916
interferon regulatory factor 1/2-like 1	transcription facor regulating IFN response	KP125917
interferon regulatory factor 1/2-like 2	transcription facor regulating IFN response	KP125918
interferon regulatory factor 5/8-like 1	transcription facor regulating IFN response	KP125919
I-type_lysozyme	antimicrobial effector (I-type lysozyme family)	KP125920
MACPF domain-containing protein 1	Perforin	KP125921
MACPF domain-containing protein 2	Perforin	KP125922
MACPF domain-containing protein 3	Perforin	KP125923
MACPF domain-containing protein 4	Perforin	KP125924
MACPF domain-containing protein 5	Perforin	KP125925
MACPF domain-containing protein 6	Perforin	KP125926
MACPF domain-containing protein 7	Perforin	KP125927
MACPF domain-containing protein 8	Perforin	KP125928
MAP kinase kinase 3/6-like	Intracellular signaling (MAPK pathway)	KP713434
MAP kinase kinase 4-like	Intracellular signaling (MAPK pathway)	KP713435
MAP kinase kinase 7-like	Intracellular signaling (MAPK pathway)	KP713437
MAP kinase kinase kinase 1-like	Intracellular signaling (MAPK pathway)	KP713433
MAP kinase p38-like	Intracellular signaling (MAPK pathway)	KP713439
membrane-bound C-type lectin	membrane-bound PRR (C-type lectin family)	KP125930
MytiLec 2	extracellular PRR (R-type lectin family)	KP125931
MytiLec 3	extracellular PRR (R-type lectin family)	KP125932
mytilin K	antimicrobial peptide (mytilin family)	KP125933
mytilin N	antimicrobial peptide (mytilin family)	KP125934
NOD-like receptor 1	intracellular bacterial and viral sensing	KP125929

peptidoglycan recognition protein 1	membrane-bound PRR (PGRP family)	KP125935
peptidoglycan recognition protein 2	membrane-bound PRR (PGRP family)	KP125936
peptidoglycan recognition protein 3	extracellular PRR (PGRP family)	KP125946
pseudomytilin 1	antimicrobial peptide (mytilin family)	KP125937
putative serine protease A	inhibitor of pathogen protease	KP141764
putative serine protease B	inhibitor of pathogen protease	KP141761
putative serine protease C	inhibitor of pathogen protease	KP141766
putative serine protease D	inhibitor of pathogen protease	KP141763
putative serine protease E	inhibitor of pathogen protease	KP141762
putative serine protease F	inhibitor of pathogen protease	KP141765
RIG-like receptor 1	intracellular viral sensing	KP125938
stimulator of interferon genes 1	intracellular viral and bacterial sensing	KP125939
stimulator of interferon genes 2	intracellular viral and bacterial sensing	KP125948
TNF ligand-like 1	cytokine	KP125940
TNF ligand-like 2	cytokine	KP125941
TNF ligand-like 3	cytokine	KP125942
Transcription factor fos-like 1	Transcription factor (MAPK pathway)	KP713441
Transcription factor fos-like 2	Transcription factor (MAPK pathway)	KP713442
Transcription factor jun-like	Transcription factor (MAPK pathway)	KP713440

Table 2: List of novel GenBank records reporting sequences expressed in *M. galloprovincialis* and discussed in this paper.

3.1. Pattern Recognition Receptors (PRRs)

The recognition of molecular motifs exposed by host-associated microbiomes (collectively called microbe associated molecular patterns, i.e. MAMPs) and by abnormal self elements is the first essential step in the activation of a coordinated and effective immune response, especially in organisms lacking adaptive immunity. As a result, specific protein-protein interactions and post-translational modifications convert the sensing phase in reaction: a finely tuned expression of genes which provides a variety of effector molecules (i.e. antimicrobial peptides, receptors and adhesion molecules, protease and protease inhibitors, cytokines and chemokines) and shapes cell behaviour in time (e.g. migration, phagocytosis, autophagy, apoptosis). The success of this ancient defense strategy depends on pathogen pressure (amount and virulence), availability and functional plasticity of PRRs, molecular pathways based on enzymatic cascades, and regulatory circuits inside and outside the immunocytes. Rapidity and intensity of the innate defense reactions are also influenced by the functional condition of the host.

The hemocytes, freely circulating in hemolymph and tissues, are the cells actively recruited in the mussel immune responses, even though other cell types may be involved [34]. In this work, we cannot attribute specific PRRs exclusively to mussel hemocytes and different experimental approaches, such as *in situ* hybridization and proteomic analyses hold the potential to clarify their cellular context in the future.

PRRs are evolutionarily conserved families of extracellular, membrane-bound or cytosolic molecules whose function has been referred to a limited number of protein domains [35]. In this section, we explore and present the numerous PRRs identified in mussel, and discuss the growing body of evidence pointing out to the expansion of diverse immune receptors in marine bivalves.

3.1.1. Extracellular PRRs

Secreted PRRs constitute a large fraction of the transcriptome in the most known bivalve species. Overall, they are characterized by a dozen of different carbohydrate recognition domains (CRDs) and possibly exist in hundreds of protein variants. Domain organization and variety of the extracellular mussel PRRs are represented in **Figure 1**.

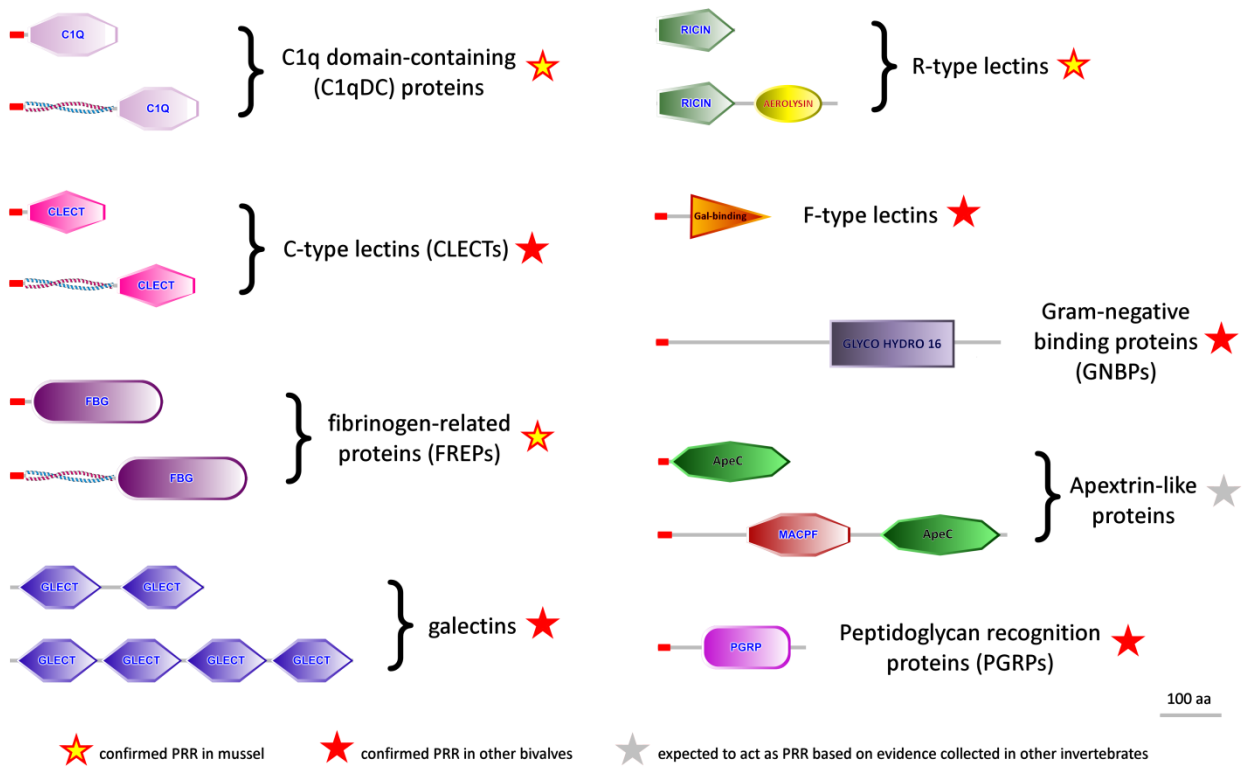


Figure 1: **Domain organization and variety of extracellular mussel PRRs.** A red segment at the N-terminus of the reported molecules indicates the signal peptide. Star symbols indicate the weight of functional evidence for each group.

3.1.1.1. C1q domain-containing (C1qDC) proteins

The C1q domain is known as a versatile PRR in many Protostomes and the widespread family of proteins displaying this domain (C1qDC proteins) probably includes one of the main, if not the largest, class of PRRs in most bivalves [36]. Even though one C1qDC protein was described in 2001 as the major component of the extrapallial fluid in *Mytilus edulis* [37], the first report of a strong over-expression of C1qDC sequences upon bacterial injection in *M. galloprovincialis* came in 2010 [38], a finding which was later confirmed also in *Mytilus coruscus* [39].

The coding sequence of mussel C1qDC proteins usually comprises an N-terminal signal peptide, a central coiled-coil region which is often missing, and a single globular C-terminal C1q domain with flexible ligand binding properties. The coiled-coil region might be functionally homologous to the collagen-like region of vertebrate C1q-like proteins (lacking in bivalves), serving as an oligomerization domain.

The expansion of this gene family was suggested by the abundance of C1qDC transcripts in the first mussel EST collection [20] and by the subsequent identification of 168 different C1qDC transcripts, with some of them being up-regulated in hemocytes after *in vivo* injection of Gram positive (Gram+) and Gram-negative (Gram-) bacteria [40]. Based on a genomic survey performed in *C. gigas*, we have reported bivalve C1qDCs with or without a coiled-coil domain (pertaining to the ghC1q and C1q-like type 2 subfamilies, respectively), with a collagen domain (C1q-like type 1, found in just a single oyster protein) and multiC1q proteins with several consecutive C1q domains [36]. In brief, the C1q gene family underwent massive expansion in Bivalvia, specifically in the Pteriomorpha and Heterodonta lineages. Consistent with the NGS-based transcriptome data of Table 1, we could identify as many as 1,274 putative C1qDC loci in the *M. galloprovincialis* draft genome. Such a remarkable diversification can explain the broad spectrum of pathogens recognized by the C1q domain in bivalves, including Gram+ and Gram- bacteria, Rickettsia-like organisms, fungi and eukaryotic parasites.

The functional characterization of bivalve C1qDC proteins is still at its early stages; however, the significant expression of many C1qDC genes in diverse tissues (e.g. digestive gland, gills and mantle) may either suggest the participation of these tissues to defense reactions or the involvement of C1qDC proteins also in processes not related to the innate immunity, like in humans.

3.1.1.2. C-type lectins

C-type lectins (CTLs) are a large class of animal lectins functioning in various biological processes, including pathogen recognition [41]. The C-type lectin domain (CLECT) characterizes both collectins and the vertebrate mannose-binding lectins able to trigger the lectin pathway of the complement system. CTLs have been associated to agglutination and opsonization of pathogens or parasites in different bivalves [42–44] and the up-regulation of their expression has been documented in the hemocytes of clams infected with *Perkinsus olseni* and in the gills of *Bathymodiolus azoricus* exposed by immersion to *Vibrio parahaemolyticus* [45,46].

Similarly to C1qDC proteins and FREPs (see Section 3.1.1.3), most bivalve CTLs are short secreted proteins with a single CLECT domain, optionally associated with a coiled-coil region, a potential oligomerization domain. However, the diversity of the CLECT domain combinations in invertebrate species is remarkable [47], often associated with other functional domains in large proteins, and its position is not always C-terminal. Based on transcriptomic data, we can confirm such a variety also in *M. galloprovincialis* and *C. gigas*.

The repertoire of bivalve CTLs is large, comprising about 350 genes in the Pacific oyster [31]. The abundance of CTLs in mussel appears to be in the same order of magnitude since we could detect 154 distinct CTL transcripts in the *M. galloprovincialis* digestive gland transcriptome [16]. Nevertheless, the specific involvement of CTLs in the mussel immune response has not been yet demonstrated. On the other hand, a number of mussel CTLs are known to take part in specific non-immune functions: two notable examples in mussel are the major acrosomal sperm proteins, which are able to dissolve the egg vitelline layer during fertilization [48,49], and the CTLs associated to particle capture during feeding [50].

3.1.1.3. Fibrinogen-related proteins (FREPs)

Hemolymph lectins bearing a C-terminal fibrinogen-like domain and similar to vertebrate ficolins (collectively named fibrinogen-related proteins or FREPs) were discovered in gastropod molluscs in 1997 [51] and possess properties other than coagulation (i.e. agglutination and antibacterial effects, developmental processes, allorecognition) [52]. As regards bivalves, an agglutinin strongly up-regulated in response to *Listonella anguillarum* challenges, AiFREP1, was recently identified as a PRR in *Argopecten irradians* [53]. A couple of years later, FREPs were identified in *M. galloprovincialis* ESTs, with sequence sets differing among and within individual mussels, clearly up-regulated in response to infection and showing opsonic properties [54,55]. These data altogether confirm the involvement of FREPs in bivalve immunity.
Mussel FREPs are simply defined by a signal peptide and a fibrinogen-like domain. Their N-terminal region sometimes contains a coiled-coil domain which could serve as an oligomerization domain, like in C1qDC proteins and similarly to collagen in ficolins. Membrane-bound FREPs are also present in mussel. Moreover, mussel FREPs differ from gastropod FREPs which show a fibrinogen-like domain associated with one or two N-terminal immunoglobulin-like domains [52,56]. Considering both full-length and partial sequences, we have detected more than 150 expressed FREPs in the transcriptome of *M. galloprovincialis,* a number consistent with previous preliminary data [16,55]. Overall, mussel FREPs represent the third most abundant class of secreted lectin-like molecules, after the C1qDC proteins and C-type lectins. Comparatively, 199 FREP genes can be identified in the *C. gigas* genome [31].

3.1.1.4. Galectins

Galectins are a widespread class of soluble animal lectins, released via the leaderless secretion pathway instead of the classical secretion pathway. They specifically bind β-galactoside sugars and, in molluscs, they are characterized by two or four tandem repeats of a galectin CRD domain. In many bivalve species, galectins have been indicated as PRRs for bacteria [57–60] and for the protozoan parasite *Perkinsus marinus* [61,62]. In *M. galloprovincialis*, galectins have already been evidenced as expressed sequences [10], even though no functional characterization has ever been carried out. In detail, we report three full-

length sequences of expressed mussel galectins with two CRDs and at least one galectin with four CRDs, an evidence which is consistent with the presence of 14 galectin genes in the draft genome of *C. gigas* [31]. Galectins primarily act as PRRs but may also represent damage-associated molecular patterns (DAMPs), able to signal the pathogen-associated tissue damage [63]. Such a hypothesis remains to be tested in molluscs.

3.1.1.5. R-type lectins

The protein family of R-type lectins is present in bacteria, plant and animals and features a CRD similar to the one found in ricin [64]. The role of these lectins in bivalve immunity has been poorly investigated so far. Overall, about 20 different genes encoding proteins with a ricin domain are present in the *C. gigas* genome: most of them pertain to the well-known class of α-N-acetylgalactosaminyltransferases, enzymes which are involved in the biosynthesis of Mucin-type O-glycans. A novel lectin named MytiLec, with globotriose-dependent cytotoxicity, has been recently identified in *M. galloprovincialis*, [65] and later a very similar lectin with antibacterial activity was identified in *Crenomytilus grayanus* [66]. These lectins share a structural motif with three very similar tandem repeats of about 50 amino acids, recognizable as a ricin-type beta trefoil domain. In *M. galloprovincialis*, we could recognize at least two other R-type lectins (we named them MytiLec 2 and 3) with an additional C-terminal pore-forming aerolysin-like domain. The combination of pathogen sensing and antimicrobial activities in the same molecule further supports the involvement of R-type lectins in pathogen clearance. The lack of a signal peptide in these mussel molecules denotes a leaderless secretory pathway, as suggested by significant SecretomeP scores.

3.1.1.6. F-type lectins

F-type lectins have been widely investigated in many invertebrates but, comparatively, their role in bivalve immunity received much less attention, as only two F-type lectins of *Pinctada* spp. involved in PAMP recognition and up-regulated in the hemocytes of challenged oysters have been identified [67,68].
The F5/8 type C domain of F-type lectins is rather common in mussel, being found in about 50 predicted proteins. Nevertheless, most of these proteins closely resemble coagulation factors or other cell adhesion-related proteins; for example the *M. edulis* bindins, important in species-specific egg/sperm recognition, pertain to this family [69]. In order to classify a bivalve protein as an immune F-type lectin, the presence of Interpro signature IPR000421 has to be coupled with functional data and therefore, for the moment, we cannot report any *bona fide* mussel F-type lectin.

3.1.1.7. Gram-negative binding proteins

Gram-negative binding proteins (GNBPs), also known as beta-glucan binding proteins, recognize β-1,3-glucans in fungi and bacteria. While GNBPs have been extensively studied in insects and crustaceans, relatively little is known about these molecules in bivalve molluscs. Arthropod GNBPs are involved in the activation of the prophenoloxidase (ProPO) system whereas some some bivalve GNBPs are reported to enhance the PO-like activity in hemocytes [70]. Bivalve GNBPs with a dual ability to bind both β-1,3-glucans and LPS have been characterized, such as in the case of *P. fucata* [71]. Proteins with such properties are usually named Lipopolysaccharide and β-1, 3-glucan binding proteins (LGBPs) and are typical of crustaceans, where, even in the absence of canonical PGRPs, they activate the ProPO system following peptidoglycan (PGN) recognition [72]. Interestingly, polymorphisms of a LGBP have been linked to increased susceptibility to *Listonella anguillarum* infections in scallops [73].

Together with the secretory peptidoglycan recognition proteins PGRP-SA, GNBP1 activates the Toll pathway and triggers melanization in response to Gram+ bacterial infections of *Drosophila*. Specifically, GNBP1 hydrolyzes the PGN of the bacterial cell wall, permitting the binding of its fragments by PGRP-SA and initiating the extracellular proteolytic cascade which results in the activation of PO, Spätzle and the Toll signaling [74,75].

Three possible GNBPs have been reported in the *M. edulis* transcriptome [13]; however, the presence of one Glycosyl hydrolases family 16 domain is not sufficient by itself to characterize a protein as a GNBP, as molluscan sequences with high similarity with GNBPs have been demonstrated to be endo-1,3-beta-D-glucanases [76,77]. Conversely, genuine GNBPs lack such an activity and act as serine proteases in the PO proteolytic cascade. Therefore, the presence of *bona fide* GNBPs and of downstream proteolytic machinery involved in a melanization cascade and in Toll signaling via a Spätzle-like molecule remains hypothetical in mussels.

3.1.1.8. Apextrin-related proteins

The apextrin C-terminal domain (ApeC) takes its name from a sea urchin protein involved in larval development. This domain has been recently recognized as a novel PRR in amphioxus, since two ApeC domain containing proteins were demonstrated to act as intra- and extra-cellular sensors of PGN and its component muramyl dipeptide (MDP) [78]. Apextrin-like proteins have been involved in pathogen recognition and inactivation also in echinoderms [79] and the over-expression of two apextrin-related transcripts in response to bacterial challenges has been reported in *M. galloprovincialis* [80]. We could predict the presence of the ApeC domain in at least 23 proteins from the Mediterranean mussel transcriptome and in 13 oyster genes. Even though ApeC is often the only domain present in proteins targeted to the secretory pathway, in 5 cases we observed an interesting association with a MACPF domain, N-terminal to ApeC (see Section 3.1.2.2). The discovery of the PRR properties of ApeC is recent and

additional functional data are needed to definitely associate apextrin-related proteins to the bivalve immune response.

3.1.2. Evidence of an ancient bivalve complement-like system

In spite of a remarkable number of reports, the literature concerning the events downstream of the activation of extracellular PRRs in bivalves is scarce. Apart from GNBPs, which are expected to trigger the Toll-like receptors signaling pathway, all the other major PRR families described in this work could potentially converge into an innate immune pathway well known in vertebrates, which fully emerged only after the divergence of Protostomes and Deuterostomes. The vertebrate complement system is based on more than 30 plasma and cell surface proteins that, through cascades of reactions, lead to pathogen neutralization and pro-inflammatory responses. Some core components of the complement system are present in Protostomes and even in cnidarians, hence, established more than 1,000 MYA [81]. Proteins showing sequence homology to some components of the complement system have been also reported in molluscs [22,23,82].

In vertebrates, the complement system can be activated through three different routes: by the activation of the C1q complex upon antigen-complexed IgMs or IgGs (classical pathway), by the spontaneous hydrolysis of the component C3 leading to PAMP recognition by the C3b fragment (alternative pathway) or by direct PAMP recognition by mannose-binding lectins or ficolins (lectin pathway). In Section 3.1.1.1 we have already presented the abundance and variability of C1qDC proteins which can directly bind pathogens and, thus, bypass the Ig-antigen recognition step of the vertebrate adaptive immune system. Furthermore, both mannose-binding lectins and ficolins potentially find their homologs in C-type lectins and FREPs, respectively (see Sections 3.1.1.2and 3.1.1.3). Thus, bivalve molluscs appear to have developed an extremely abundant and diversified repertoire of lectins, which may therefore mount the immune response through a signaling cascade similar to the vertebrate lectin pathway.

Nevertheless, bivalves lack specific serine proteases, such as C1r and C1s of the C1q complex and the MBL/ficolin-related MASP-1 and MASP-2, which are fundamental in the downstream proteolytic reactions, thus leaving a huge question mark on the molecular partners of these extracellular lectin-like PRRs.

Despite the absence of homologs for these components, convincing C3-like and C2/factor B-like sequences have been recently identified in *R. philippinarum* [83]. These findings support the existence of at least some core components of a proto-complement pathway in bivalves. Likewise, the existence of a primitive complement system resembling the alternative pathway of the mammalian complement system has been recently demonstrated in other protostomes, namely in some arthropods [84].

In the next sections we describe in detail thioester- and MACPF-domain containing proteins, possible functional homologs to the C3/C4/C5 and to the C6/C7/C8/C9 complement components, respectively. A

schematic representation of the possible bivalve homologs related to the vertebrate complement pathway is reported in **Figure 2**.

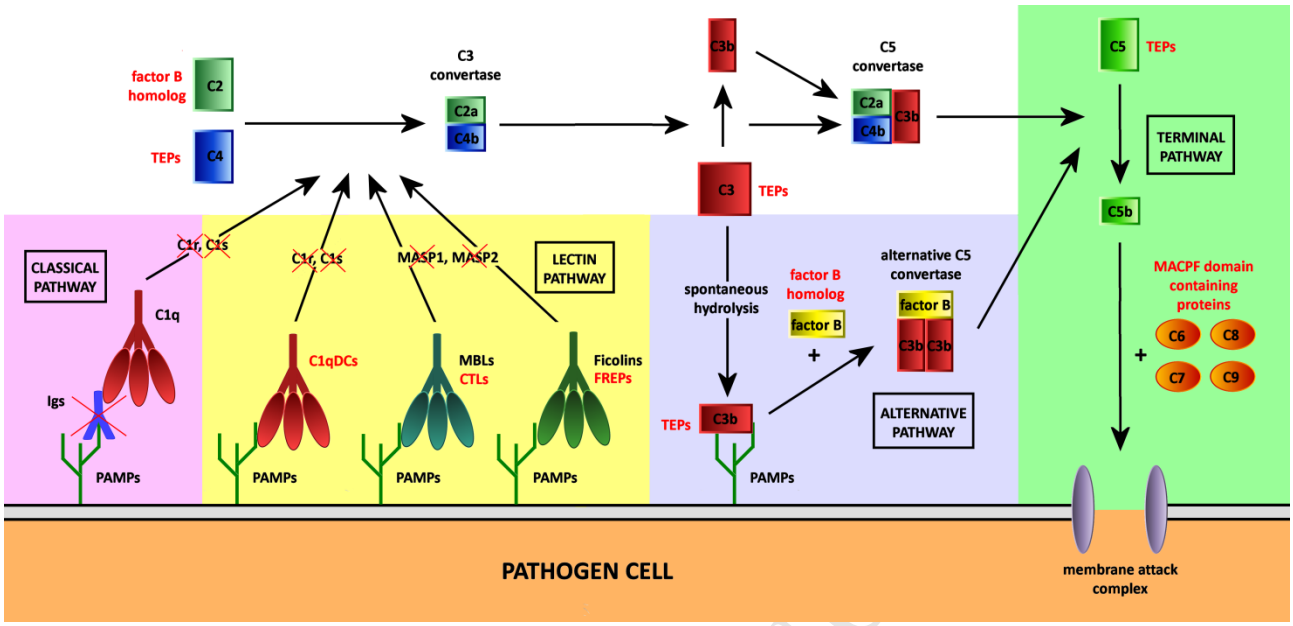


Figure 2: Comparative overview of the vertebrate complement system with mussel putative homologs. Molecules which might play a role in an ancient mussel complement system (named in red) are illustrated comparatively to those of vertebrate animals (named in black). The lack of a classical pathway mediated by immunoglobulins in vertebrates (pink background) may be compensated by the wide and effective repertoire of PRRs (e.g. C1qDC proteins). The existence of a lectin pathway is suggested by CTLs (functionally homologous to MBLs) and FREPs (functionally homologous to ficolins) (yellow background). So far, no genuine serine protease homologs of MASP proteins (evolutionarily emerging in the Cephalochordata lineage) have been traced in Bivalvia. Elements of the alternative complement pathway are present also in bivalves and other invertebrates (e.g. mussel C3 and factor B homolog) (light blue background). Following PRR activation, proteolitic cascades mediated for instance by TEPs prepare pathogen opsonization or killing (arrows conveying to the green background). Even though proteins with a MACPF domain (which characterizes the C6/7/8/9 components have been identified in mussel, their involvement the terminal pathway of the complement remain to be assessed (the reader is referred to the web version of the article for a more direct visualization of the colors used in this figure).

3.1.2.1. Thioester-containing proteins

Thioester-containing proteins (TEPs) comprise the vertebrate complement components C3, C4 and C5, as well as a number of invertebrate homologs, such as insect proteins functioning as opsonins and promoting phagocytosis of bacteria and melanization [85,86].

Only two studies have so far been conducted on bivalve TEPs, precisely in the scallop *Azumapecten farreri*, where the complete gene encoding the protein CfTEP was characterized, and in *R. philippinarum*, where

418 the C3-like protein mentioned in the section above has been identified [83,87]. The complex alternative
419 splicing pattern of the CfTEP mRNA as well as the differential expression of isoforms in response to diverse
420 pathogen challenges evidenced a complex regulation of its expression in the innate immune response [87].
421 TEPs have been identified in oyster and clam transcriptomes [22,23] and they are present also in mussels;
422 due to the structural similarity between TEPs and serum protease inhibitors alpha-2 macroglobulins,
423 functional analyses are necessary to evaluate the extension of this protein family in bivalves and its role in
424 bivalve immunity. Furthermore, the presence of highly similar paralogous gene products and low-
425 complexity regions hampered the reconstruction of full-length TEP transcripts from the mussel
426 transcriptome. Nevertheless, we can report the full length sequence of a complement C3 component-like
427 transcript, which is the first TEP to be ever reported in mussel (**Table 2**).
428
429 **3.1.2.2. MACPF pore-forming molecules**

430 The terminal components of the complement system can form a protein complex (the membrane attack
431 complex, or MAC) on the surface of Gram- bacteria, triggering their lysis. The MACPF superfamily (named
432 after a domain common to proteins of the mammalian membrane attack complex and to perforins) is the
433 largest family of pore-forming molecules in animals [88]. In vertebrates, perforins are produced by natural
434 killer cells and by cytotoxic T lymphocytes to trigger the killing of virus-infected cells [89]. Nevertheless, a
435 large number of other MACPF domain-containing proteins whose function is not linked to immunity have
436 been reported in mammals, including very large inducible GTPases, BRINPS and astrotactins [90].
437 Following multiple reports of perforin-like proteins in sea urchin and abalones [2,91–93], the MACPF
438 domain-containing protein Macp was identified in *M. galloprovincialis*, and it remains the only one
439 available at the present time in public databases for a bivalve mollusc [80]. Macp is a secreted protein
440 which does not present any other domain except from MACPF. The expression of its transcript was found
441 to be developmentally regulated and increased upon pathogen and PAMPs stimulation, evidence
442 suggesting its involvement in the innate immunity. Nevertheless, both domain organization and primary
443 sequence of Macp differ from those of perforins and vertebrate complement proteins; hence, Macp could
444 not be placed in any of major group of vertebrate MACPF domain-containing proteins in a phylogenetic
445 analysis. Despite the up-regulation of Macp following bacterial challenges, experimental assessment of its
446 lytic activity is still required to confirm it as a pore-forming molecule.
447 Genes for several MACPF-domain containing proteins appear to be present in bivalve genomes: indeed, a
448 total of 17 genes have been predicted in the *C. gigas* genome and we could detect 8 full length and several
449 partial MACPF transcripts in the transcriptome of *M. gallopovincialis*. Both secreted and membrane-bound
450 forms of the predicted proteins are present in mussel and oyster, but in no case they show convincing
451 sequence homology to vertebrate perforins and to the proteins C6/C7/C8/C9 of the terminal complement
452 pathway.

453 The association of ApeC with MACPF, detected in 5 predicted mussel proteins, strongly indicates the
454 combination of pathogen recognition and killing properties in the same protein sequence, as ApeC has
455 been recently functionally linked to pathogen recognition in amphioxus [78] (see Section 3.1.1.8).
456

457 **3.1.3. Membrane-bound PRRs and downstream signaling**

458 Compared to secreted PRRs, membrane-bound immune receptors appear to be less abundant in *M.*
459 *galloprovincialis,* despite their diversification and central role in the host defense against invading
460 pathogens. Unlike extracellular PRRs expected to trigger pathogen killing via the lectin-like complement
461 pathway, membrane-bound PRRs generally possess an intracellular domain which mediates the signal
462 transduction through key transcription factors and, finally, the expression of antimicrobial effectors and
463 proinflammatory cytokines (the latter perpetrate the adaptive immune response in vertebrate animals).
464 In this section, we update the knowledge on bivalve membrane-bound receptors and on the downstream
465 events triggered by PAMP recognition. An overview of these processes is provided in **Figure 3.**

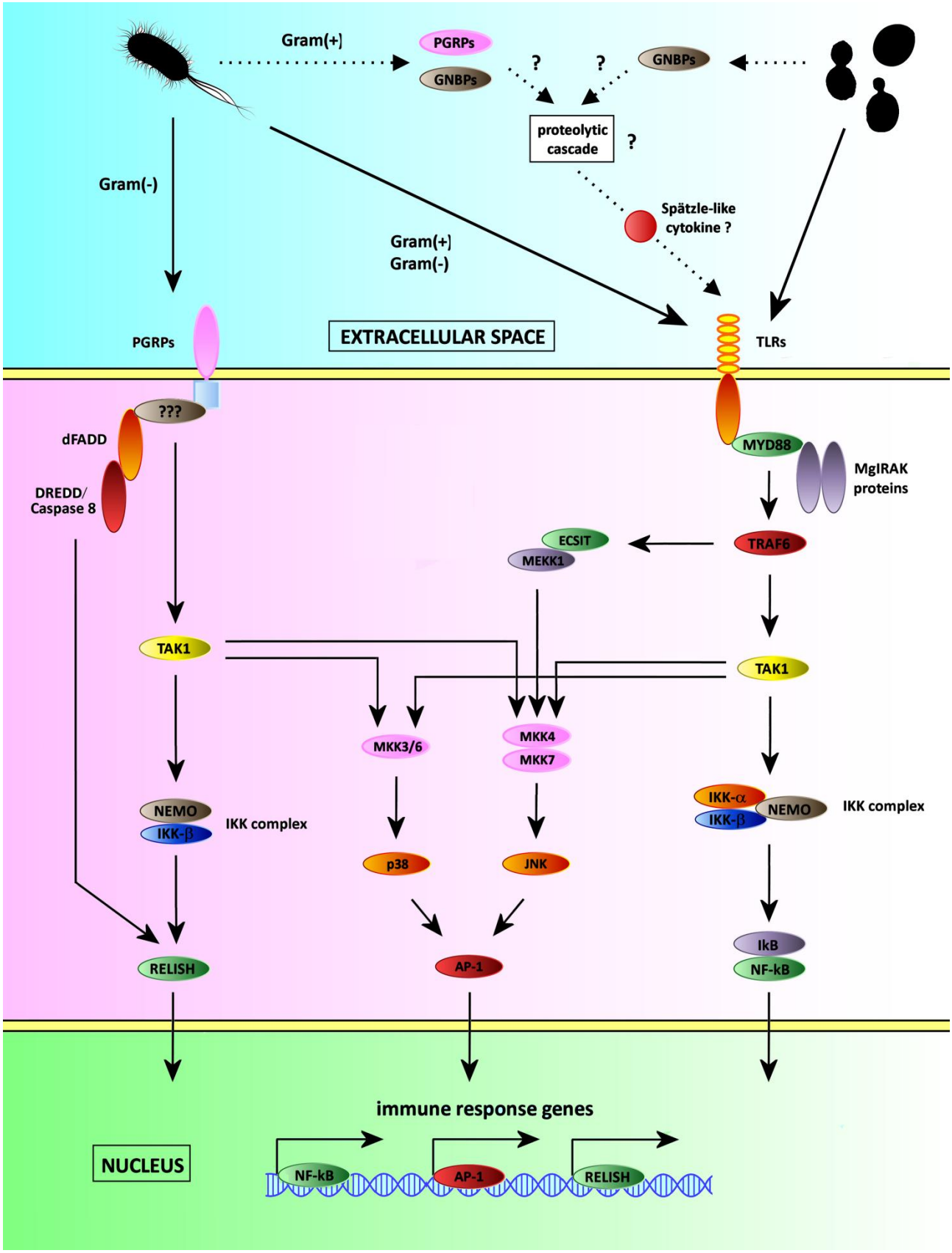


Figure 3: Membrane-bound PRRs. Once activated by PAMPs, Toll-like receptors (right) and peptidoglican recognition proteins (left) transmit the danger signal to cytosolic proteins ultimately bringing transcription factors into the cell nucleus and inducing the expression of genes essential to clear bacterial and fungal

cells. The existence of extracellular pathways leading to the activation of TLRs through extracellular PGRPs and GNBPs, well-studied in *Drosophila*, is still doubtful in bivalves, in particular due to the absence of a Spätzle-like cytokine and of protease homologs involved in the upstream cascade (see the text for details). These passages are therefore indicated by dashed arrows.

3.1.3.1. Toll-like receptors and the NF-κB signaling

Toll-like receptors (TLRs) are among the most successful PRRs of the immune response in metazoans [94]. TLRs take their name from the *Droshophila* receptor Toll, originally identified as essential in the embryo morphogenesis, but later they recognized to mediate the immune response to the fungus *Aspergillus fumigatus* [95]. Since then, several TLRs sharing a similar domain organization have been detected and extensively studied both in vertebrates and in invertebrates. Usually, they share an evolutionarily conserved intracellular Toll-interleukin-1-receptor (TIR) domain, a transmembrane region and a variable extracellular region consisting of leucine-rich repeats (LRRs) with interesting binding properties. TLRs are capable of binding a very broad range of PAMPs, including LPS, components of the bacterial cell wall, flagellin, single- and double-stranded RNA, thus potentially acting as sensors of bacteria, fungi and viruses. In response to ligand binding, TLR dimerization is expected to sequentially recruit intermediary elements such as the Myd88 adaptor, IRAK and IKK kinases among the others, thus activating transcription factors such as NF-κB and interferon-regulatory factors (IRFs) and ultimately mounting the expression of pro-inflammatory cytokines, chemokines, and anti-viral molecules [96].

Following the identification of the first bivalve TLR in *A. farreri* [97], the full repertoire of these receptors has been explored in detail in *M. galloprovincialis* and 23 TLRs, grouped in 4 different clusters according to the organization of extracellular LRRs, were identified [98]. Tissue-specific patterns of constitutive expression were reported, but only one out of the four tested mussel TLRs (MgTLR-I) was found up-regulated in response to bacterial injection, especially with Gram- bacteria. Three Myd88 adapters are expressed in mussel, with specific constitutive and inducible levels. The observed expression patterns suggested the co-regulated expression of MgTLR-I and MgMyd88-c in response to the filamentous fungus *Fusarium oxysporum* and evidenced for the first time the existence of a Toll signaling pathway in Lophotrochozoa. On the other hand, the absence of TRIF-like molecules suggests that bivalves lack a MyD88-independent pathway, homologous to the one activated downstream to human TLR3.

The subsequent identification of 15 downstream elements provided further evidence about the existence of a complete signaling pathway similar to the *Drosophila* Toll and to the mammalian TLR pathways [99]. Besides MyD88, one or more transcript variants denoting TOLLIP, IRAKs, TRAFs, TAK1, IKK, IKKγ/NEMO, IkB, Relish/p65 and NF-κB were comparatively identified in *M. galloprovincialis*, *M.edulis* and *C. gigas* (see **Figure 3**).

Many questions are still open on the bivalve TLRs: in particular, which PAMPs are specifically recognized by such receptors and how TLRs are distributed in hemocytes and other mussel cells. It is known that the human TLR7/8/9 are localized to the endosomal membrane, where they can recognize elements of bacterial and viral infections such as CpG oligodeoxynucleotides (CpG-DNA) (TLR9), single or double stranded viral RNA (TLR7/8 and TLR3, respectively), and trigger the production of interferon (IFN) and proinflammatory cytokines [100]. In other words, also mussel TLRs might recognize pathogens both in the extracellular and intracellular space.

Besides TLRs, other TIR domain-containing proteins are involved in immune signaling, since this domain is often used for homophilic interactions [101]. Interestingly, the TIR domain appears to be widespread in bivalves, with over 100 TIR domain-containg proteins predicted in the oyster genome. Despite not being as abundant as in echinoderms, such a repertoire is still wide in comparison with other invertebrates since, for instance, just about 10 TIR-domain containing proteins are present in arthropods. Further studies are necessary to reveal more details on these receptor proteins and their role in the innate immune responses of bivalves.

3.1.3.2. Peptidoglycan recognition proteins: evidence for an IMD-like signaling?

Peptidoglycan recognition receptors (PGRPs) are important PRRs present in all metazoans and able to recognize bacteria by specifically binding PGN, a major component of cell bacterial walls [102]. PGRPs have been characterized in many bivalve species, including scallops, oysters, razor clams [103–105] and the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus*, where a PGRP was found to be highly abundant in the gills, which typically host edosymbiotic bacteria [46,106]. So far, all these proteins have been regarded as short-type PGRPs for extracellular bacteria recognition. Nine PGRPs are represented in the oyster genome but, looking at the transcriptome data and considering both membrane-linked and secreted PGRPs, we could only report three full-length mussel transcripts (two membrane-bound and one secreted proteins, see **Table 2**). The expression pattern of the secreted protein PGRP3 has been comparatively investigated in *M. galloprovincialis* and *B. azoricus*, evidencing an up-regulation 12 and 24 hours after bacterial challenges in the Mediterranean mussel [107], but data concerning the regulation of membrane-bound PGRPs are still completely missing.

In insects, secreted PGRPs have a dual role in: a) recognizing Gram+ bacteria and modulating the Toll pathway (together with GNBPs) and melanization, through the proteolytic PO cascade and the cleavage of the pro-cytokine Spätzle; b) activating the IMD pathway in response to Gram- bacteria via the amidase activity of PGRP-SC, which cleaves PGN into inactive amino sugars and peptides later recognized by the transmembrane PGRP-LC [102].

As reported previously, the presence of an extracellular proteolytic cascade similar to the PO system and leading to the activation of TLRs through a Spätzle-like molecule is doubtful in bivalves (see Section 3.1.1.7).

On the contrary, the existence of an IMD pathway involved in the response to Gram- bacteria seems to be better supported by molecular data. Indeed, together with the Toll signaling pathway, Toubiana and colleagues also detected a number of sequences homologous to elements of the *Drosophila* IMD pathway in *M. galloprovincialis* (namely, TAK1, NEMO, IKK-β and Relish). Nevertheless, no plausible evidence could be found for the first key adaptor protein downstream to PGRP-LC, the IMD/RIP protein. Given that the Toll pathway was observed to be responsive to both Gram+ and Gram- bacteria, the existence of an IMD pathway was reported as uncertain in mussel [99].

However, our identification of mussel transcripts denoting transmembrane PGRPs, together with the presence in public sequence databases of dFADD and DREDD/Caspase-8, would reinforce the idea of an IMD-like pathway involved in the recognition of Gram- bacteria in bivalves (see **Figure 3**). In *Drosophila*, transmembrane PGRPs initiate the IMD signal transduction through their intracellular RIP Homotypic Interaction Motif (RHIM) domain whereas mussel PGRPs lack such domain (their intracellular region does not contain any known functional domain), so further investigations are necessary to identify the putative intracellular adaptor ptotein initiating signal transduction downstream to membrane-bound PGRPs.

3.1.3.3. MAPKs and signaling cross-talk along the TLR and IMD pathways

An alternative route which can be activated downstream of TLRs in vertebrates, and downstream of IMD in insects, is the mitogen-activated protein kinases (MAPK) pathway. The MAPK kinase cascade has been involved in the regulation of growth, differentiation and survival, and it could act both in the TLR- and in the PGRP-mediated intracellular signaling pathways in bivalves.

Various stimuli are known to activate the MAPK cascade, which is a pathway not restricted to the immune responses, and considering its complexity, in this section we just briefly outline the main signaling components possibly involved in signal transduction downstream of TAK1, a MAP kinase kinase kinase (MAP3K) acting in the intersection between these signaling cascades.

Two main routes involving the sequential phosphorylation of MAP kinase kinases (MKKs) and MAPKs are possibly activated by TAK1: a) c-JUN N-terminal kinases (JNK) via MKK4 or MKK7; b) p38 mitogen-activated protein kinases via MKK3 or MKK6. The MAPK signaling ultimately leads to the activation of the AP-1 transcription factor complex, thereby regulating the expression of various immunity- and stress-related genes.

It is important to note that the previously reported effectiveness of commercial anti-pospho-antibodies directed against MAPK, JNK and p38 in *M. galloprovincialis* and *B. azoricus*, denoted both the remarkable evolutionary sequence conservation of these molecules and the critical role of p38 and JNK kinases in the immune response of mussel hemocytes [108,109].

We could identify bivalve sequences with high similarity (in the range of 60-70% protein sequence identity) to the human MKK3/6, MKK4/7, JNKs and p38 kinases. In particular, we report the presence of MKK4,

MKK7, a single MKK3/6 and a single JNK homolog in both mussel and in oyster; moreover, compared to three p38 kinase-like genes identified in the oyster genome, only a single one was found expressed in mussel.

An alternative branch of the TLR pathway can activate MKK4/7 upstream to TAK1, through the evolutionarily conserved adaptor protein ECSIT [110], previously described also in mussel [99] and representing a bridge between TRAF6 and the MAP kinase kinase kinase MEKK-1 (**Figure 3**). MEKK-1 in turn phosphorylates and activates MKK4 and 7 [111]. We can report the presence of a highly conserved MEKK-1 in mussel which, together with ECSIT, might be involved in another point of contact between the TLR and MAPK pathways.

The signal transduction mediated by JNK and p38 determines the activation of AP-1 transcription factors, heterodimers of proteins encoded by JUN and FOS gene families, which both comprise multiple members in vertebrates. A JUN protein responsive to bacterial infections has previously been described in *Crassostrea hongkongensis* [112] and was identified in the *B. azoricus* gill transcriptome [106,107]. Based on comparative sequence analyses, we can report that a single JUN transcription factor is encoded in the Pacific oyster genome and is expressed in mussel. On the contrary, the FOS family comprises at least two different members in both species, whose relation with vertebrate FOS family proteins is unclear. Further study is required to assess the interaction between bivalve JUN and FOS members and the presence of AP-1 binding elements in the promoter of bivalve genes involved in immune and stress responses.

3.1.3.4. Scavenger receptor cysteine-rich superfamily (SRCR-SF) receptors

The scavenger receptor cysteine-rich domain defines a large and ancient superfamily of scavenger receptors collectively known as SRCR-SF receptors. They present a conserved domain, 100-110 amino acids long and stabilized by disulfide bridges, which can be found associated with a number of additional domains and co-receptors. Altogether, these features confer to the members of SRCR-SF a broad range of functions, from lipoprotein binding, to cell transport, to pathogen clearance [113]. The massive expansion the SRCR-SF in the sea urchin genome (comprising 218 genes), their marked up-regulation in response to immune challenges and their localized expression in sea urchin coelomocytes, clearly indicate the immune-related diversification of this family of receptors [2].

In the available bivalve transcriptomes the number of SRCR-SF domains appears relatively high [114], but these receptors still await a detailed characterization. The only well-characterized bivalve SRCR-SF domain-containing protein is CfSR, identified in the scallop *Azumapecten farreri* and able to bind various ligands such as LPS, PGN, mannan and zymosan [115].

Overall, at least 62 genes of the SRCR superfamily have been annotated in the oyster genome, with various domain organizations, and even a higher number of these receptors is evident in mussel transcriptomes.

Although not as abundant as in sea urchin, a relevant gene family expansion may have occurred in mussel compared to other model invertebrates (only 7 and 8 SRCRs are present in *D. melanogaster* and *C. elegans*, respectively).

3.1.4. Cytosolic PRRs

In addition to the greatly expanded families of secreted PRRs and membrane-bound TLRs, intracellular sensors of microbes and viruses have been more recently identified in bivalves. In this section, we provide an overview on the cytosolic PRRs present in mussel, with some indications on the downstream signaling network based on homologies to better studied organisms (**Figure 4)**.

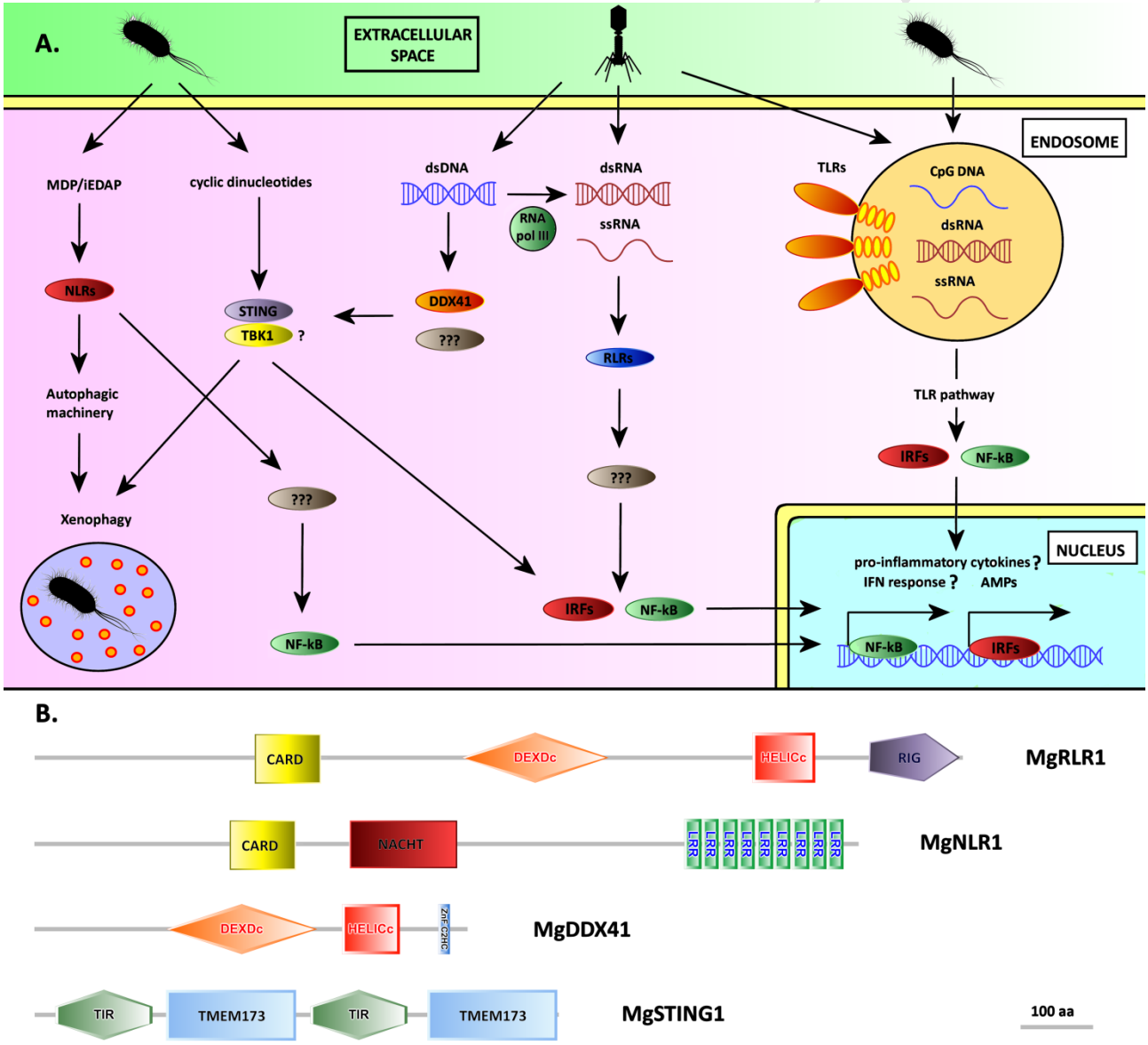


Figure 4: Cytosolic mussel PRRs. In a general view (A), structural elements of pathogens penetrated into the cell environment are exemplified by MDP/iE-DAP, cyclic dinucleotides and nucleic acids whereas NLRs, STING, DDX41, RLRs and endosomal TLRs exemplify PRRs. Only some events possibly occurring downstream are illustrated, given the lack of a robust frame of knowledge in mussel and bivalves. The domain organization of different cytosolic PRRs identified in *M. galloprovincialis* is also reported (B).

3.1.4.1. NOD-like receptors (NLRs)

The cytosolic NOD-like receptors (NLRs) reinforce the sensing of bacterial components which have been able to enter the cell, in particular the PGN-derived molecules γ-D-Glu-meso-diaminopimelic acid (iE-DAP) and MDP. About 20 NLRs have been reported in vertebrates but lineage-specific expansion to few hundred members in plants and some animal groups emphasizes their involvement in the innate immunity [116,117]. The self-assembly of NLRs in heteromeric complexes contributes to the specific recognition of PAMPs via C-terminal LRRs whereas the recruitment of downstream molecules is mediated by homophilic interactions of their variable N-terminal effector domain: usually a DEATH, a pyrin, a caspase recruitment (CARD) or a baculovirus inhibitor of apoptosis protein repeat (BIR) domain classify NLRs within the NOD, NALP or NAIP subfamilies. NLRs typically contain a central NACHT nucleoside triphosphatase domain. Despite all early diverging metazoans do already possess a rather large number of highly diversified NACHT domain-containing proteins [118], no NLR sequence has ever been reported in bivalves so far, even though the existence of NLRs was hypothesized in the blue mussel [13].

We now report the presence of at least one NLR-like sequence with a canonical tripartite domain organization in *M. galloprovincialis* (MgNLR1). This putative NLR contains a single N-terminal CARD effector domain, followed by a central NACHT domain and a C-terminal region dominated by LRRs (see **Figure 4B**). We could also identify other partial transcripts encoding NACHT-domain containing proteins, in some cases in association with C-terminal tetratricopeptide repeats, which are a common evidence in other invertebrate NLRs [119]. Since we did not find any BIR and pyrin domains associated with NACHT, the subfamilies NALPs and NAIPs appear to be absent in *M. galloprovincialis*.

Mammalian NLRs, NALPs in particular, are involved in the organization of the inflammasome complex. There are no previous reports of inflammasomes in invertebrates [120] and the absence of pyrin domains (characterizing both NALPs and the key adaptor protein PYCARD) in invertebrates, as well the absence of a caspase-1 homologous sequence in mussel implies that a similar system for the activation of the inflammatory response, if existing, is highly divergent in molluscs. Furthermore, NLRs seem to have a fundamental role in the activation of the autophagic machinery (specifically via the ATG16L1 protein) at the site of bacterial entry, thus promoting the elimination of bacterial pathogens through xenophagy [121].

In conclusion, a limited number of NLR-like sequences are present in mussel; comparatively, no NLRs genes have been annotated in the oyster genome, and one possible NLR sequence emerges from the mussel transcriptome data. Further investigations could provide more information on the involvement of mussel NLRs in the cytosolic PAMP recognition and the related downstream signaling pathway.

3.1.4.2. Retinoic acid inducible gene-I like RNA helicases (RLRs) and downstream signaling

Retinoic acid inducible gene-I like RNA helicases are more simply known as RIG-like receptors (RLRs); they are intracellular sensors of viral 5'-triphosphate (5'ppp)-single- and double-stranded RNA and they are therefore of the utmost importance in antiviral responses. In general, RLRs have been linked to the indirect detection of DNA viruses, mediated by DNA-dependent RNA polymerase III which synthetizes 5'-ppp-ssRNA from a double-stranded viral DNA template [122].

RLRs are capable of initiating an intracellular signaling cascade which ultimately leads to the production of interferon and pro-inflammatory cytokines through the activation of NF-κB and interferon regulatory factors (IRFs) [123–125]. Vertebrate RLRs are organized with one or two N-terminal CARD domains, two central DExD/H box helicase domains and a C-terminal RIG repressor domain.

The presence of RLRs in invertebrates has been a long-debated issue: indeed, *Drosophila* lacks RLRs and relies on Dicer-2 for an homologous function in antiviral response [126,127]. Nevertheless, the presence of RIG-like genes in echinoderms and cnidarians suggests an ancient origin for RLRs [128] and the antiviral response of *C. elegans* is apparently mediated by a RIG-like protein devoid of a CARD domain [129]. About bivalves, the analysis of the oyster genome reveals at least eight different RIG-domain containing proteins, four of them with a domain architecture identical to vertebrate RLRs, except for the presence of a single N-terminal CARD domain (instead of two) [130].

Philipp and colleagues reported the partial sequences of two putative RLRs in the blue mussel *M. edulis* [13] and, despite the presence of highly similar paralogous genes complicates the reconstruction of full length RLR sequences in *M. galloprovincialis*, we can report the full-length transcript of a RIG-like receptor, very similar to those of oyster (**Figure 4B**).

The finding of a mussel RLR with a canonical domain organization, consistent with vertebrate proteins, suggests competence for double-stranded viral RNA sensing in bivalve mollusks and implies the existence of a similar downstream signaling. In vertebrates, the first step in the helicase-mediated viral RNA recognition is the interaction of the CARD domain with a downstream adaptor protein, the IFN-beta promoter stimulator (IPS-1), also known as CARD adaptor inducing IFN-beta (CARDIF). IPS-1 then interacts with TRAF3 to activates TBK-1 and IKKε, thus inducing the interferon response, or with FADD, which activates the production of inflammatory cytokines via NF-κB upon its interaction with DREDD/Caspase-8 [124].

Most of the proteins involved in RLR signaling are common to the TLR and IMD pathways (see sections 3.1.3.1 and 3.1.3.2), and an almost complete RLR pathway could be identified in *M. edulis* [13].

Nevertheless, in the same study it was not possible to identify the key component IPS-1, and this is not surprising given the high sequence divergence of this protein in basal deuterostomes [2]. In essence, the evidence of vertebrate-like cytoplasmic RIG-like receptors supports the existence of the RLR signaling in bivalve molluscs, although some key signaling elements are expected to be highly divergent and have still to be identified.

3.1.4.3. Intracellular foreign DNA and bacteria sensing via the STING pathway

Besides RLRs, various cytosolic sensors of exogenous DNA exist in vertebrates, but homologous molecules seem absent in mussel or difficult to identify in invertebrates, likely due to high sequence divergence. For example, the pyrin domain characterizing the vertebrate sensors AIM2 and IFI16 could not be identified in any analyzed genome or transcriptome of bivalves, mussel included.

Nevertheless, the signals transduced by different sensors such as DAI, DDX41, and IFI16 converge in a single downstream crucial molecule named “Stimulator of interferon genes” (STING), which is known to stimulate the production of IFN, via IRF3, and proinflammatory cytokines, via NF-κB, once they are phosphorylated by the TBK1 kinase [131,132]. Though DAI and IFI16 are absent in mussel, we could identify a sequence homologous to DDX41, a member of the DEXDc helicases family which has been recently demonstrated to act as an intracellular viral DNA sensor [133]. The high sequence conservation between the mouse and mussel proteins (71% sequence identity) strongly suggests functional conservation also in bivalves, but the hypothesis requires experimental testing.

STING is a transmembrane protein which can indirectly detect the presence of bacteria by sensing conserved signaling molecules produced by bacteria (the cyclic dinucleotides c-di-GMP and c-di-AMP), thus playing an important role against bacteria and exogenous viral DNA (in collaboration with DDX41). The STING protein is evolutionarily conserved, as its origins can be traced back to Choanoflagellates, even though it was lost during metazoan evolution in nematodes and flatworms [134]. While no sequence records are available for this protein in molluscs, we ascertained the presence of at least two full-length STING-like proteins in *M. galloprovincialis* and estimated the presence of 5 genes in the *C. gigas* genome.

Quite surprisingly, the STING homologs found in mussel have a peculiar domain organization which is not found in other invertebrate genomes [134]. Like in insects, the deduced mussel STING proteins lack the N-terminal transmembrane domains anchoring vertebrate STING globular domain to the endoplasmic reticulum membrane, suggesting a different subcellular localization compared to vertebrates. The duplication of the STING globular domain and the association with two TIR domains, N-terminal to STING (**Figure 4B**), are the most striking features of mussel STING proteins. Human STING is known to be active as a dimer [135], and the presence of two STING domains within the same protein in bivalves could possibly provide a functional analogy without need of dimerization. On the other hand, the presence of TIR domains

is likely relevant to signal transduction, which is guaranteed by an alternative C-terminal extension present in vertebrates but absent in mussel.

In addition to the coordination of IFN and proinflammatory cytokines production, STING has a crucial role in the induction of an autophagic-like response following bacterial infections, leading to the ubiquitination of bacterial cells and their selective elimination by xenophagy [136] and a similar behavior has been also observed in response to α-herpesviruses [137]. A well-developed autophagic machinery is also present in mussel [13], but a detailed characterization of this pathway, which is not uniquely related to immune functions, goes beyond the scope of this paper.

As discussed in section 3.1.4.1, the STING signaling is not the only immune pathway expected to stimulate xenophagy in response to pathogen invasion, as NLRs hold a similar potential. The role of xenophagy in the innate immune response is gaining an increasing recognition [138] and the presence of multiple STING homologs, NLRs and a fully functional autophagic system in mussel suggest an interesting interplay of molecular networks in support to pathogen clearance.

3.2. Effectors molecules

Upon pathogen recognition, cross-talking signaling pathways allow the activation of specific transcription factors which, in turn, are expected to reinforce the innate immune response, via proinflammatory cytokines and interferons, and to stimulate the production of diverse humoral effectors directly involved in pathogen killing, such as, for instance, antimicrobial peptides (AMPs)

3.2.1. Antimicrobial peptides (AMPs)

Antimicrobial peptides (AMPs) are a widespread group of heterogeneous gene-encoded molecules with antibiotic functions, which are classified in different subgroups, based on their structure, amino acid composition and properties. Seven different AMP families have been identified so far in *Mytilus* spp., all pertaining to the cysteine-rich AMP subgroup (**Table 3**).

AMP class	Domain organization	Cysteine residues	Cysteine array
defensins (arthropod-like)	SP-CR	6	C-C-C---C-C-C
defensins	SP-CR-C-terminal extension*	8	C-C-C-C-C-C-C-C
mytilins	SP-CR-C-terminal extension	8	C-C-C-C-C-C-C-C
myticins	SP-CR-C-terminal extension	8	C-C-C-C-C-C-C-C
mytimacins (type-1)	SP-CR	8	C-C-C----C---C-C-C
mytimacins (type-2)	SP-CR	10	C-C-C----C-C-C-C-C-C
mytimacins (type-3)	SP-CR	12	C-C-C-C-CC-C-C-C-C-C-C
big defensins	SP-propeptide-CR	6	C-C-C-C-CC
mytimycins (type-1)	SP-CR-EF hand	12	CC-C-C--C-C--C-C-C-C-C-C

mytimycins (type-2)	SP-CR-EF hand	12	CC-C-C-CC-C-CC-C-C-C
mytimycins (type-3)	SP-CR-EF hand	14	CC-C-C-CC-C-CC-C-C-C-C-C
myticusins	SP-propeptide-CR	10	C-C-C-C-C-C-C-CCC
mytiCRP-I	SP-propeptide-CR	6	C-C-CC-C-C
mytiCRP-II	SP-CR	8	C-C-C-CC-C-C-C
mytiCRP-III	SP-CR	6	C-C-C-CC-C
mytiCRP-IV	SP-CR	10	C-C-C-CC-C-C-C-C-C

Table 3: List of cysteine-rich AMP families identified in *Mytilus* spp. Domain organization and cysteine arrays are indicated. MytiCRP-I, -II, -III and –IV represent families whose antimicrobial properties have not been demosrated yet (see section 3.2.1.7 for details). SP: signal peptide; CR: cysteine-rich domain; *: the C-terminal extension may be missing in defensins.

3.2.1.1. Defensins

Defensins are structurally characterized by a cysteine-stabilized alpha-beta motif (CS-αβ) and are almost ubiquitous in Eukaryotes. Among the invertebrate animals, defensins have been mostly studied in arthropods, where they invariably have 6 conserved cysteine residues arranged in three disulfide bridges. In 1996, the first bivalve defensin molecules were isolated from mussel hemocytes, in both *M. edulis* and in *M. galloprovincialis* [139,140]. The structure of mussel defensins is similar to that of arthropod defensins despite the presence of an additional pair of cysteines arranged in a fourth disulfide bridge [141]. Furthermore, the loop connecting the two antiparallel beta-strands of the CS-αβ motif identified as fundamental to the antibacterial and antifungal activities [142,143].

Transcriptome analyses helped to define the complete sequence of 8-cysteine defensin precursors, and three additional defensins, namely MGD3, MGD4 and MGD5, have been reported [20]. The presence of a C-terminal extension after the 8th cysteine residue in most defensins (with the exceptions of MGD3 and MGD5) is a recurring scheme in other mussel AMPs families (mytilins and myticins) which could have evolved from defensins through exon shuffling [144]. For comparison, the *C. gigas* defensins display a remarkable sequence diversity, which seems to be in turn originated from a limited number of defensin genes (three have been identified so far) [145].

We report the sequences of two novel defensin-like sequences with only six cysteines, consistent with previous reports of “arthropod-like” defensins in other molluscs [146–148]. However, functional studies are necessary to confirm the involvement of such molecules in the mussel innate immune response.

3.2.1.2. Mytilins

Mytilins are a class of AMPs strongly expressed in mussel hemocytes. Initially isolated by HPLC techniques in *M. edulis* [139], the mytilin family was later discovered to comprise different isotypes [149]. The

organization of mytilin genes and protein precursors is similar to that of most defensins and myticins, as the signal peptide and the mature Cys-rich regions are encoded by two separate exons whereas the C-terminal extension, cleaved off in the mature peptide, is encoded by a third exon [150]. The presence of the CS-αβ motif in the tridimensional structure of mytilins is also strongly reminiscent of defensins [151].

Five different mytilin sequences, named mytilin B, C, D, F and G1 have been so far identified in *M. galloprovincialis* [20]. Compared to the other most studied mussel AMP families (defensins and myticins), mytilin precursors show a minimal inter-individual sequence variability [152].

Until recent times, the purification of highly expressed peptides from active fractions of the hemolymph and EST sequencing have been the main strategies applied to the identification of AMPs in non-model species, including mussel. However, thanks to the recent high-throughput sequencing approaches, we can report some additional mussel mytilin-like sequences with peculiar variations. Most notably, the position of the 5th cysteine residue in not canonical in the novel *M. galloprovincialis* sequences of mytilin K and mytilin N, and this variation is associated with an insertion of four amino acids in the alpha helix of the CS-αβ between the first and the second cysteine residues (**Figure 5**).

On the other hand, another mytilin-like sequence named pseudomytilin 1 shows a canonical disulfide organization but displays a completely different C-terminal extension. The functional meaning of these variations and the expression pattern of these novel mytilin-like sequences remain to be fully explored.

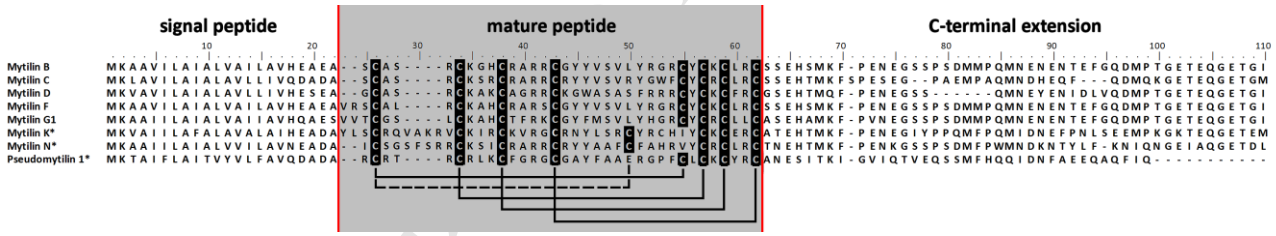


Figure 5: sequence alignment of *M. galloprovincialis* mytlin proteins. The experimentally determined disulfide bridges organization of mytilin B is indicated by solid lines, whereas the hypothetical disulfide bridge connecting cysteine 1 to cysteine 5 in mytilin K and N is shown as a dased line. * indicates novel sequences.

3.2.1.3. Myticins

The identification of myticins A and B in *M. galloprovincialis* dates 1999 [153]. These AMPs, displaying 8 cysteines and a C-terminal extension like defensins and mytilins, were found to be highly active against Gram+ bacteria. The expression of myticin B was demonstrated in a hemocyte subpopulation of small granulocytes and its expression kinetics was thoroughly investigated (for a comprehensive review see [154]). A third sequence, named myticin C, was later identified through EST sequencing [155] and, in spite

of a limited variability at a genomic level, an extreme variability was observed both among individuals and within a same individual at the transcript level [156,157]. Though at a lower level than myticin C, also mytilin A and mytilin B show an inter-individual sequence variability, mainly generated by single nucleotide changes [152]. Such a high sequence diversity could be partially justified by gene duplication, as the presence of at least two myticin C gene copies has been recently indicated [158].

Myticin C has been shown to act not only as an AMP, but also as an immuno-modulating molecule, due to its chemotactic properties [159]. Based on transcriptomic evidence, we can confirm that no additional myticin-like sequences are expressed in mussel.

3.2.1.4. Mytimycins

Owing to their strict antifungal activity, mytimycins clearly differ from mussel defensins, mytilins and myticins. The mytimycin peptide, first purified from hemocytes of *M. edulis* alongside defensins and mytilins, displayed growth-inhibiting activity towards *Neurospora crassa* and *Fusarium culmorum* [139]. Compared to defensins and mytilins, the study of mytimicin was neglected until the description of its full gene sequence in 2012 [160]. The mytimicin precursor is composed by a signal peptide, followed by a central Cys-rich domain and a C-terminal EF-hand domain (the latter is cleaved from the mature peptide). Subsequently, the expression pattern of mytimicin and the timing of its up-regulation in response to filamentous fungi challenges were investigated also in individual mussels [161,162]. The increase of transcript data now allows to explore in more detail the variability of mussel mytimicin-sequences previously only represented by a few ESTs [20]. They now appear as a group of expressed sequences with a highly variable cysteine array which can comprise either 12 or 14 cysteines (allegedly organized into 6 or 7 disulfide bridges), with the novel cysteine array of type-3 mytimycins (mytimycin F, G and I) being described for the first time in the present paper (**Table 3** and **Figure 6**).

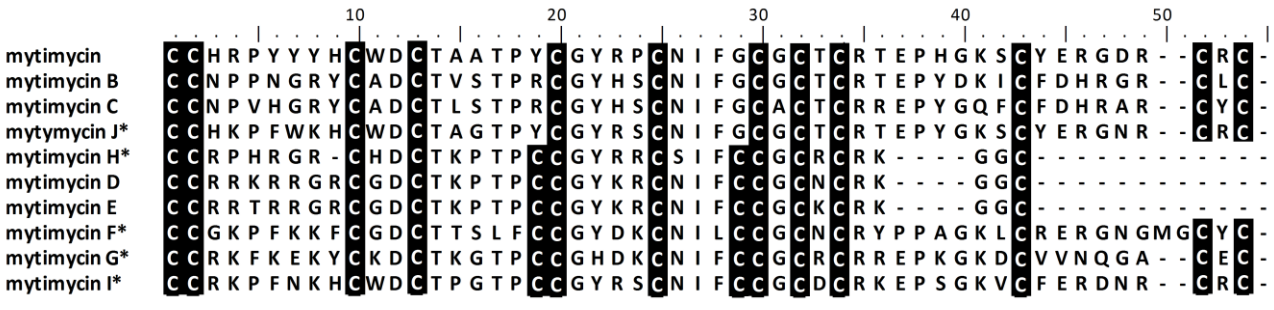


Figure 6: sequence alignment of *M. galloprovincialis* mytimycin predicted proteins. Only the cysteine-rich region corresponding to the mature peptide is shown. * indicates novel sequences.

3.2.1.5. Mytimacins

Macins represent an emerging family of secreted, positively charged AMPs, relatively poorly studied but widespread in protostomes and combining antimicrobial and nerve-repair activities [163]. Five mussel macins (mytimacins) have been described so far in *M. galloprovincialis* [164], with the number of cysteines and disulfide bridges ranging from 8 (4 bridges) to 12 (6 bridges) (**Table 3**). While 8- and 10-Cys macins have been reported in other taxa and their disulfide connectivity has been experimentally determined [165], the only 12-Cys macin described so far is the mussel mytimacin 5. Contrary to most mussel AMPs, mytimacins are not expressed in hemocytes, being instead detected in the digestive gland, gills and mantle. Macins of land invertebrates seem to exert their activity in the mucus produced by tissues in contact with the external surface [166–168]; bivalves also secrete mucus which covers their pallial tissues, with a function thought to be mainly related to filter-feeding and particle selection [169,170]. To date, no study has been carried out to investigate the antimicrobial potential of bivalve mucus and it is therefore impossible to ascertain whether macins are used in a similar fashion to other land invertebrates as mucus defense molecules. This remains, however, an interesting lead for future studies.

3.2.1.6. Big Defensins

Unlike arthropod-like defensins, big defensins (BDs) are composed by two separate domains. The N-terminal region is mainly alpha-helical and hydrophobic, while the C-terminal region is cysteine-rich and structurally similar to vertebrate beta-defensins. BDs are indeed thought to be the ancestral form of beta-defensins in invertebrates, which gave origin to vertebrate beta-defensins through exon shuffling and intronization of exonic sequence [171]. First described in *Tachipleus tridentatus* [172], BDs are apparently restricted to a few taxonomic classes, including Merostomata, Mollusca and basal cephalochordates [164]. In bivalves, a relevant number of BDs have been characterized, starting from the first report in *Argopecten irradians* [173]. A total of eight different BDs sequences have been reported in *M. galloprovincialis* and we can here report the presence of an additional sequence pertaining to this class. Differently from the BDs of other bivalves [173–175], none of mussel BDs is highly expressed in hemocytes, being instead localized in a broad range of tissues, like the mytimacins.

3.2.1.7. Other cysteine-rich antimicrobial peptides

Sequence analyses performed on the *M. galloprovincialis* transcriptome indicated the presence of a large number of short secreted cysteine-rich peptides. The lack of similarity to known sequences and the difficulty at obtaining purified peptides prevent functional tests but, most often, a positive net charge and the hypervariability of the Cys-rich domain are strongly suggestive of possible antimicrobial properties. As an example, a novel Cys-rich putative AMP (myticusin) has been recently identified in the hemocytes of

Mytilus coruscus. This AMP, bearing 10 cysteine residues arranged in an unusual disulfide pattern, was demonstrated to be active against Gram+bacteria [176].

These novel Cys-rich peptide families of mussel will likely be better characterized in the near future and, according to an upcoming report (manuscript in preparation), we propose a provisional naming scheme as follows: each novel mussel Cys-rich peptide family should be named Mytilus Cystein-Rich (MytiCRP)-*n*, where *n* is a progressive number. As an example, we have most recently characterized the large MytiCRP-I family, which comprises over 50 members of peptides whose precursors include both a conserved signal peptide and a propeptide region. Nevertheless, the mature region, bearing 8 cysteines arranged in a conotoxin-like C-C-CC-C-C array is extremely variable and subject to positive selection. Although the function of these peptides is currently unknown, we hypothesize that they may play a role in defense towards eukaryotic pathogens. The cysteine arrays of other, still uncharacterized, MytiCRP families are shown in **Table 3**.

3.2.1.8. Other classes of AMPs

Linear amphipathic and alpha-helical AMPs have been documented in vertebrates as well as in insects, where they are secreted by the fat body and released in the hemolymph [177], but so far no helical AMPs has ever been described in bivalves. Another important class of AMPs consists of peptides rich in specific amino acidic residues. The only example of this kind in bivalves is represented by an oyster Proline-rich peptide (Cg-Prp), which is expressed in hemocytes in response to bacterial challenges and displays synergic effect with defensins, even though devoid by itself of antimicrobial activity [178]. BLAST searches excluded the presence of similar peptides in *Mytilus* spp. and thus cysteine-rich peptides remain the only known class of AMPs to date.

Overall, given the high primary sequence variability, methods based on sequence similarity have a poor predictive power and other strategies could be more effective in the *de novo* prediction of additional AMPs in mussel (e.g. analyses based on the calculation of positive net charge or on the identification of amphipathic alpha helices).

3.2.2. Antimicrobial effectors with chitin-binding domains

Recently, a 6 KDa peptide with 6 cysteines, named mytichitin-A, has been reported in *Mytilus coruscus* [179]. Mainly expressed in the gonad, this AMP was strongly up-regulated in response to bacterial challenges and the C-terminal region of a rather large chitoriosidase/chitinase-like precursor was determined to be responsible of the antimicrobial activity, mainly against Gram+ bacteria.

Chitin-binding domains characterize different AMP families in invertebrates, including penaeidins and tachycitins [180,181], but this is the first report of a AMP of this class in molluscs. Given the relatively high

occurrence of chitin-binding domains in bivalves (estimated to be present in 76 oyster proteins as deduced from genome analysis), further study will be required to elucidate the antimicrobial potential of this class of mussel proteins.

3.2.3. Lysozymes

Lysozymes are antimicrobial proteins among the most well-known and studied in metazoans. Able to hydrolyze 1,4-beta-glycosidic bonds between N-acetylmuramic acid and N-acetylglucosamine in PGN, lysozymes are particularly active against Gram+ bacteria and have a role in both digestion and antibacterial defense. Animal lysozymes show limited primary sequence homology but a close resemblance of tridimensional structure; nevertheless, they can be classified in three groups, namely chicken-type (C-type), goose-type (G-type) and invertebrate-type (I-type) [182]. Molecular and phylogenetic studies indicated that C-type and I-type sequences have likely originated from ancestral gene duplications [183], with both types simultaneously present in both molluscs and arthropods.

The class of I-type lysozymes was the first one to be characterized in bivalves; their study in *M. galloprovincialis* revealed up-regulation in response to bacterial challenges [184,185]. In addition to their function in innate immunity, this class of lysozymes has also been linked to digestion processes in other bivalves [186]. In agreement with previous reports, we could identify the two known I-type lysozymes in *M. galloprovincialis* and a novel transcript sequence.

Compared to the I-type class, C-type lysozymes have been far less studied in bivalves. Following our first report of a C-type lysozyme sequence in *M. galloprovincialis* [20], a study demonstrated its up-regulation in experimental challenges with both Gram+ and Gram- bacteria [187]. We can now report at least other two C-type lysozyme sequences expressed in *M. galloprovincialis*.

G-type lysozymes were originally thought to be exclusive of vertebrates, due their absence in nematode and arthropod genomes. Since their detection in scallops, several other G-type lysozymes have been identified in bivalves [188]. Due to the achievement of an optimal activity at different pH, the two G-type lysozymes known in *M. galloprovincialis* seem to have different specialized roles in digestion and immune defense [189]. Our analyses confirmed the two G-type lysozyme transcript sequences already deposited in public databases for the Meditarranean mussel.

Recently, a phage-type lysozyme has been identified in the clam *R. philippinarum*, thus possibly expanding the lysozyme repertoire of bivalves to four different families [190], but we could not detect any orthologous sequence in the mussel transcriptome nor in the oyster genome.

3.2.4. Bactericidal/permeability increasing proteins (BPIs)

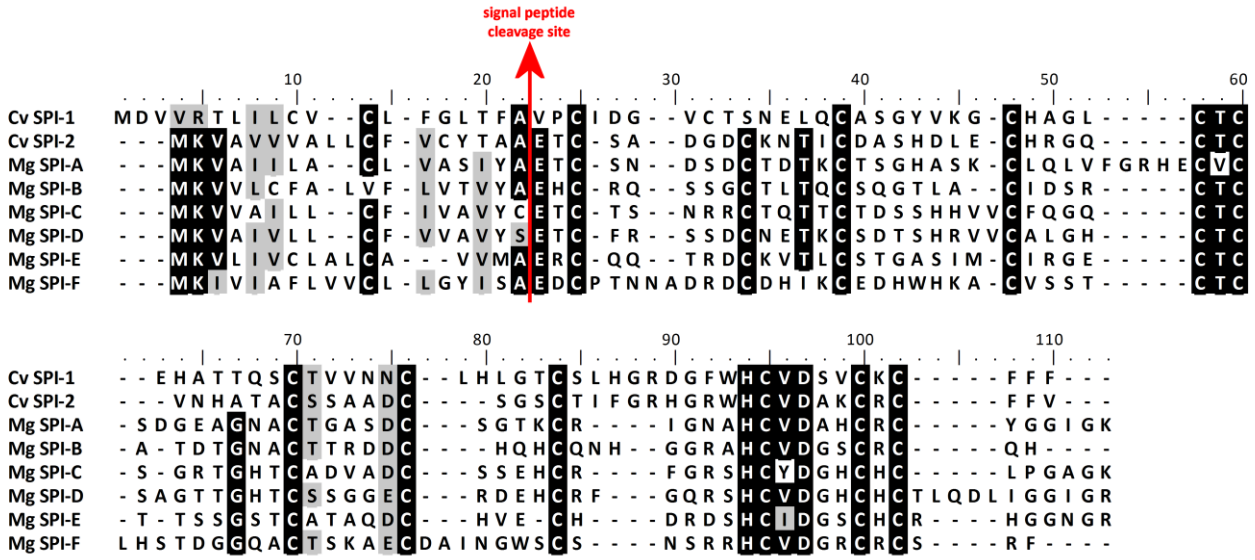
Bactericidal/permeability increasing proteins (BPIs) are evolutionarily conserved proteins present in molluscs as well in vertebrates, which can bind LPS and cause bacterial killing by increased permeability of the bacterial cytoplasmic membrane [191,192]. One BPI sequence from *M. galloprovincialis* has been deposited in 2012 in GenBank and the presence of secreted as well as membrane-bound forms of BPIs has been predicted from the transcriptome of *M. edulis* [13]. Due to the high sequence similarity with vertebrate and invertebrate BPIs, it is plausible that mussel BPIs retain an identical function. We could detect two additional full length BPI transcripts in *M. galloprovincialis*, but we could not confirm the existence of the membrane-bound isoforms reported elsewhere [13].

3.2.5. Protease inhibitors

Many pathogens produce proteases able to modulate host immunity at different levels, from recognition receptors to immune effectors [193]. The inactivation of these exogenous proteases is an important determinant of the host defense and a broad range of protease inhibitors can be expressed also in bivalves to counteract the variants of proteases produced by invading microbes.

In 2001, the first molluscan protease inhibitor, a tissue inhibitor of metalloproteinases (TIMP) was characterized in *C. gigas* and connected to an innate immune function [194,195]. Then, other accounts of different protease inhibitors of several different bivalves were published, including TIMPs, Kazal-type [196–198] and Kunitz-type [199] protease inhibitors. However, sequence homology by itself does not functionally link these protein families to the immune defense, as they are potentially involved in a wide range of processes, including embryonic development, morphogenesis and nacre formation, among others [200–202]. Consistent with data gathered from genomic analysis of all invertebrates, a rather high number of protease inhibitors has been previously reported in bivalve transcriptomes [16,22,23]. We can now report gene models for 36 TIMPs, 49 Kazal-type and 27 Kunitz-type protease inhibitors in the oyster genome and a similar abundance of these classes in the mussel transcriptome. Not being restricted to the immune responses, an adequate description of mussel protease inhibitors requires further study.

Moreover, completely novel proteins not fitting the current classification of protease inhibitors have been related to the bivalve immune responses. In particular, a novel serine protease inhibitor (cvSPI-1) with no homology to any other known sequence was shown to inhibit the major extracellular protease produced by the pathogen *Perkinsus marinus* in *Crassostrea virginica*, and to limit the proliferation of this protozoan [203,204]. This fact is of great interest, as six proteins pertaining to the same family are expressed at exceptionally high levels in the mussel digestive gland [16]. The sequence alignment between mussel and oyster serine protease inhibitors is shown in **Figure 7**.



964

965 **Figure 7**: sequence alignment of immune-related serine protease inhibitors identified in *Mytilus*
966 *galloprovincialis* (Mg) and *Crassostrea virginica* (Cv). Conserved residues are shadowed in black, a red arrow
967 indicates the signal peptide cleavage site.

968

969 **3.3. Modulators of the immune response**

970 Cytokines are a large and heterogeneous group of regulatory molecules which comprise interleukins,
971 interferons, tumor necrosis factor (TNF) and chemokines. These crucial mediators of immune response and
972 inflammation are often produced in different defense phases in response to specific transcription factors
973 (mostly those pertaining to the NF-κB and IRF families). Their pleiotropic effects attract immune cells and
974 enhance pathogen clearance, thereby perpetuating the host reaction also through adaptive mechanisms in
975 vertebrate animals.
976 Due to the general skepticism of the scientific community, investigations on invertebrate cytokines were
977 not even imagined for a very long time, until very recent years [205,206]. Conversely, the signal
978 transduction pathways leading to their activation are relatively well conserved (e.g. the TLR signaling).
979 Given the presence of a complete TLR/NF-κB pathway (see Section 3.1.3.1) and of a complex system of
980 cytosolic PRRs whose donstream signaling converges either on NF-κB or on IRFs (see Section 3.1.4), it
981 would be reasonable to expect the production of proinflammatory cytokines also in bivalves.
982 Even though there are uncertainties and gaps on the pathogen-dependend activation of interferon
983 regulatory factors in mussel (e.g. absence of the TRIF adaptor and doubtful presence of a Myd88-
984 independent TLR signal transduction), the identification of IRF-like expressed sequences in *P. fucata* and
985 *Hyriopsis cumingii*, indicated the presence of interferon-sensitive response elements (ISRE) in bivalve
986 genomes [207,208]. In addition to the three IRF-like genes present in the *C. gigas* genome, we can now
987 report at least four different IRF-like transcripts in mussel, whose similarity to vertebrate IRFs is strictly
988 limited to the DNA-binding domain. Three of them vaguely resemble human IRF1/2 whereas the fourth one

989 is more similar to IRF5 nd IRF8. Whether and how these transcription factors act in PRR recognition-
990 triggered pathways remains to to be fully elucidated.
991 For clearness, mussels do not possess interferon-like sequence and, in general, the taxonomic distribution
992 of the four-helical cytokine-like domain which characterizes the large majority of vertebrate interleukins is
993 limited to chordates. In most cases, the invertebrate molecules reminiscent of a cytokine-like function do
994 no share any sequence similarity to the vertebrate functional homologs and have probably undergone
995 independent evolution, thus making their recognition by sequence homology impossible [209]. In 1990, the
996 responsiveness of oyster cells to human IL-1 and TNF was evocative of the existence of bivalve cytokine-like
997 proteins [210]. In 2008, an interleukin-17-like transcript was unexpectedly reported and found highly and
998 rapidly induced in response to bacterial exposure in *C. gigas* hemocytes [211]. More recently, five novel IL-
999 17 homologs were identified in the *C. gigas* genome, costitutively expressed in some oyster tissues and
1000 significantly up-regulated in hemocytes after different immunostimulation trials [212]. The overall evidence
1001 pointed to the diversification of IL-17 from a common ancestor and to the idea of an IL-17-sustained AMP
1002 production in oyster. A preliminary survey performed in the mussel transcriptome has evidenced multiple
1003 IL-17-like sequences which need to be confirmed and validated with additional study also in comparison
1004 with the oyster IL-17 homologs.
1005 Two other inflammatory cytokines have been identified by similarity searches in bivalves: the macrophage
1006 migration inhibitory factor (MIF), which is present with at least three distinct gene products in *M.*
1007 *galloprovincialis* [213], and the allograft inflammatory factor (AIF), able to stimulate the phagocytic activity
1008 of granulocytes, which has been identified in different bivalve species [214–216] and is now confirmed also
1009 in *M. galloprovincialis*.
1010 In 2009, a TNF*a*-like sequence was reported for the first time in a mollusc (the abalone *Haliotis discus*) [217]
1011 and, in agreement to earlier reports on the existence of TNF receptors in scallops [218,219], more recent
1012 research led to the identification of this multifunctional cytokine also in *Ostrea edulis* [214]. We now report
1013 at least three mussel transcripts pertaining to the TNF ligand superfamily with the canonical presence of an
1014 N-terminal transmembrane region. Functional data are necessary to confirm that these molluscan TNF-like
1015 proteins function similarly to vertebrate TNFs. Some intracellular components of a hypothetical TNFR
1016 pathway in mussel are common to the IMD pathway (see FADD and Caspase-8 in **Figure 3**) but genuine
1017 homologs to RIP, TRADD and TRAF2 still have to be identified.
1018 In essence, most bivalve cytokines remains elusive due to a lack of similarity with vertebrates and a similar
1019 role could be shared by completely different molecules such as myticin C: for long time considered to
1020 strictly act as an AMP, experimental data suggested its cytokine function, since extracts of cells expressing
1021 myticin C were able to attract hemocytes [159]. In brief, the available evidence for bivalve cytokines is
1022 fragmentary and their dentification is still mostly limited to a few evolutionarily conserved molecules.
1023

4. Conclusion

The immune systems of invertebrates are still largely unknown but the peculiar expansion of genes for recognition and effector molecules in certain invertebrate lineages suggests the co-evolution of innate defense mechanisms in response to the selective pressure imposed by fast-evolving microbial communities and single infectious agents or parasites. This work confirms both the variety and multiplicity of the gene-encoded molecules participating in the *M. galloprovincialis* innate immunity. We expect that the overall sequencing resources, gene expression analyses extended to regulatory non coding RNAs and experimental studies in normal and stress conditions will provide a robust comparative basis for the identification of heritable traits involved in the resistance to bivalve diseases and multifactorial bivalve mortality, with the simultaneous identification of genome-spread or gene-related molecular markers supporting the selection of vigorous broodstocks.

Acknowledgments

This work is supported to the research units of University of Trieste and University of Padova by PRIN 2010-11 (20109XZEPR) and PRAT 2012 (CPDA128951/12). We are also grateful to many colleagues of the whole scientific community for the common efforts towards a multidisciplinary understanding of the bivalve immunity.

References

[1] Loker ES. Macroevolutionary Immunology: A Role for Immunity in the Diversification of Animal life. Front Immunol 2012;3:25. doi:10.3389/fimmu.2012.00025.

[2] Hibino T, Loza-Coll M, Messier C, Majeske AJ, Cohen AH, Terwilliger DP, et al. The immune gene repertoire encoded in the purple sea urchin genome. Dev Biol 2006;300:349–65. doi:10.1016/j.ydbio.2006.08.065.

[3] Armitage SAO, Peuß R, Kurtz J. Dscam and pancrustacean immune memory – A review of the evidence. Dev Comp Immunol 2015;48:315-23. doi:10.1016/j.dci.2014.03.004.

[4] Lemaitre B, Hoffmann J. The host defense of *Drosophila melanogaster*. Annu Rev Immunol 2007;25:697–743. doi:10.1146/annurev.immunol.25.022106.141615.

[5] Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, Funayama N, et al. The innate immune repertoire in cnidaria--ancestral complexity and stochastic gene loss. Genome Biol 2007;8:R59. doi:10.1186/gb-2007-8-4-r59.

[6] Bosch TCG, Augustin R, Anton-Erxleben F, Fraune S, Hemmrich G, Zill H, et al. Uncovering the evolutionary history of innate immunity: the simple metazoan *Hydra* uses epithelial cells for host defence. Dev Comp Immunol 2009;33:559–69. doi:10.1016/j.dci.2008.10.004.

[7] Bieler R, Mikkelsen PM, Collins TM, Glover EA, González VL, Graf DL, et al. Investigating the Bivalve Tree of Life – an exemplar-based approach combining molecular and novel morphological characters. Invertebr Syst 2014;28:32–115.

[8] Encycl Life EOL. http://eol.org.

[9] Suárez-Ulloa V, Fernández-Tajes J, Manfrin C, Gerdol M, Venier P, Eirín-López JM. Bivalve Omics: State of the Art and Potential Applications for the Biomonitoring of Harmful Marine Compounds. Mar Drugs 2013;11:4370–89. doi:10.3390/md11114370.

[10] Venier P, Pittà CD, Bernante F, Varotto L, Nardi BD, Bovo G, et al. MytiBase: a knowledgebase of mussel (*M. galloprovincialis*) transcribed sequences. BMC Genomics 2009;10:72. doi:10.1186/1471-2164-10-72.

[11] Suarez-Ulloa V, Fernandez-Tajes J, Aguiar-Pulido V, Rivera-Casas C, Gonzalez-Romero R, Ausio J, et al. The CHROMEVALOA Database: A Resource for the Evaluation of Okadaic Acid Contamination in the Marine Environment Based on the Chromatin-Associated Transcriptome of the Mussel Mytilus galloprovincialis. Mar Drugs 2013;11:830–41. doi:10.3390/md11030830.

[12] Freer A, Bridgett S, Jiang J, Cusack M. Biomineral proteins from *Mytilus edulis* mantle tissue transcriptome. Mar Biotechnol N Y N 2014;16:34–45. doi:10.1007/s10126-013-9516-1.

[13] Philipp EER, Kraemer L, Melzner F, Poustka AJ, Thieme S, Findeisen U, et al. Massively Parallel RNA Sequencing Identifies a Complex Immune Gene Repertoire in the lophotrochozoan *Mytilus edulis*. PLoS ONE 2012;7:e33091. doi:10.1371/journal.pone.0033091.

[14] Craft JA, Gilbert JA, Temperton B, Dempsey KE, Ashelford K, Tiwari B, et al. Pyrosequencing of *Mytilus galloprovincialis* cDNAs: tissue-specific expression patterns. PloS One 2010;5:e8875. doi:10.1371/journal.pone.0008875.

[15] Rosani U, Domeneghetti S, Pallavicini A, Venier P. Target Capture and Massive Sequencing of Genes Transcribed in *Mytilus galloprovincialis*. BioMed Res Int 2014;2014:e538549. doi:10.1155/2014/538549.

[16] Gerdol M, De Moro G, Manfrin C, Milandri A, Riccardi E, Beran A, et al. RNA sequencing and de novo assembly of the digestive gland transcriptome in *Mytilus galloprovincialis* fed with toxinogenic and non-toxic strains of *Alexandrium minutum*. BMC Res Notes 2014;7:722. doi:10.1186/1756-0500-7-722.

[17] Nguyen TTT, Hayes BJ, Ingram BA. Genetic parameters and response to selection in blue mussel (*Mytilus galloprovincialis*) using a SNP-based pedigree. Aquaculture 2014;420–421:295–301. doi:10.1016/j.aquaculture.2013.11.021.

[18] Domeneghetti S, Varotto L, Civettini M, Rosani U, Stauder M, Pretto T, et al. Mortality occurrence and pathogen detection in *Crassostrea gigas* and *Mytilus galloprovincialis* close-growing in shallow waters (Goro lagoon, Italy). Fish Shellfish Immunol 2014;41:37–44. doi:10.1016/j.fsi.2014.05.023.

[19] Gestal C, Roch P, Renault T, Pallavicini A, Paillard C, Novoa B, et al. Study of diseases and the immune system of bivalves using molecular biology and genomics. Rev Fish Sci 2008;16:131–54. doi:10.1080/10641260802325518.

[20] Venier P, Varotto L, Rosani U, Millino C, Celegato B, Bernante F, et al. Insights into the innate immunity of the Mediterranean mussel *Mytilus galloprovincialis*. BMC Genomics 2011;12:69. doi:10.1186/1471-2164-12-69.

[21] Bassim S, Genard B, Gauthier-Clerc S, Moraga D, Tremblay R. Ontogeny of bivalve immunity: assessing the potential of next-generation sequencing techniques. Rev Aquac 2014:n/a – n/a. doi:10.1111/raq.12064.

[22] Zhang L, Li L, Zhu Y, Zhang G, Guo X. Transcriptome analysis reveals a rich gene set related to innate immunity in the Eastern oyster (*Crassostrea virginica*). Mar Biotechnol N Y N 2014;16:17–33. doi:10.1007/s10126-013-9526-z.

[23] Moreira R, Balseiro P, Planas JV, Fuste B, Beltran S, Novoa B, et al. Transcriptomics of in vitro immune-stimulated hemocytes from the Manila clam *Ruditapes philippinarum* using high-throughput sequencing. PloS One 2012;7:e35009. doi:10.1371/journal.pone.0035009.

[24] Pauletto M, Milan M, Moreira R, Novoa B, Figueras A, Babbucci M, et al. Deep transcriptome sequencing of *Pecten maximus* hemocytes: a genomic resource for bivalve immunology. Fish Shellfish Immunol 2014;37:154–65. doi:10.1016/j.fsi.2014.01.017.

[25] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011;29:644–52. doi:10.1038/nbt.1883.

[26] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10. doi:10.1016/S0022-2836(05)80360-2.

[27] Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinforma Oxf Engl 2014;30:1236–40. doi:10.1093/bioinformatics/btu031.

[28] Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 2011;8:785–6. doi:10.1038/nmeth.1701.

[29] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–80. doi:10.1006/jmbi.2000.4315.

[30] Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel PEDS 2004;17:349–56. doi:10.1093/protein/gzh037.

[31] Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 2012;490:49–54. doi:10.1038/nature11413.

[32] Mitchell A, Chang H-Y, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res 2015;43:D213–21. doi:10.1093/nar/gku1243.

[33] Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, et al. SUPERFAMILY--sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res 2009;37:D380–6. doi:10.1093/nar/gkn762.

[34] Mitta G, Vandenbulcke F, Noël T, Romestand B, Beauvillain JC, Salzet M, et al. Differential distribution and defence involvement of antimicrobial peptides in mussel. J Cell Sci 2000;113 (Pt 15):2759–69.

[35] Pålsson-McDermott EM, O'Neill L a. J. Building an immune system from nine domains. Biochem Soc Trans 2007;35:1437–44. doi:10.1042/BST0351437.

[36] Gerdol M, Venier P, Pallavicini A. The genome of the Pacific oyster *Crassostrea gigas* brings new insights on the massive expansion of the C1q gene family in Bivalvia. Dev Comp Immunol 2015;49:59–71. doi:10.1016/j.dci.2014.11.007.

[37] Hattan SJ, Laue TM, Chasteen ND. Purification and characterization of a novel calcium-binding protein from the extrapallial fluid of the mollusc, *Mytilus edulis*. J Biol Chem 2001;276:4461–8. doi:10.1074/jbc.M006803200.

[38] Gestal C, Pallavicini A, Venier P, Novoa B, Figueras A. MgC1q, a novel C1q-domain-containing protein involved in the immune response of *Mytilus galloprovincialis*. Dev Comp Immunol 2010;34:926–34. doi:10.1016/j.dci.2010.02.012.

[39] Liu H-H, Xiang L-X, Shao J-Z. A novel C1q-domain-containing (C1qDC) protein from *Mytilus coruscus* with the transcriptional analysis against marine pathogens and heavy metals. Dev Comp Immunol 2014;44:70–5. doi:10.1016/j.dci.2013.11.009.

[40] Gerdol M, Manfrin C, De Moro G, Figueras A, Novoa B, Venier P, et al. The C1q domain containing proteins of the Mediterranean mussel *Mytilus galloprovincialis*: a widespread and diverse family of immune-related molecules. Dev Comp Immunol 2011;35:635–43. doi:10.1016/j.dci.2011.01.018.

[41] Zelensky AN, Gready JE. The C-type lectin-like domain superfamily. FEBS J 2005;272:6179–217. doi:10.1111/j.1742-4658.2005.05031.x.

[42] Gourdine J-P, Smith-Ravin EJ. Analysis of a cDNA-derived sequence of a novel mannose-binding lectin, codakine, from the tropical clam *Codakia orbicularis*. Fish Shellfish Immunol 2007;22:498–509. doi:10.1016/j.fsi.2006.06.013.

[43] Yang J, Wang L, Zhang H, Qiu L, Wang H, Song L. C-type lectin in *Chlamys farreri* (CfLec-1) mediating immune recognition and opsonization. PloS One 2011;6:e17089. doi:10.1371/journal.pone.0017089.

[44] Wang L, Wang L, Yang J, Zhang H, Huang M, Kong P, et al. A multi-CRD C-type lectin with broad recognition spectrum and cellular adhesion from *Argopecten irradians*. Dev Comp Immunol 2012;36:591–601. doi:10.1016/j.dci.2011.10.002.

[45] Kim JY, Adhya M, Cho SK, Choi KS, Cho M. Characterization, tissue expression, and immunohistochemical localization of MCL3, a C-type lectin produced by *Perkinsus olseni*-infected

Manila clams (*Ruditapes philippinarum*). Fish Shellfish Immunol 2008;25:598–603. doi:10.1016/j.fsi.2008.07.015.

[46] Bettencourt R, Rodrigues M, Barros I, Cerqueira T, Freitas C, Costa V, et al. Site-related differences in gene expression and bacterial densities in the mussel Bathymodiolus azoricus from the Menez Gwen and Lucky Strike deep-sea hydrothermal vent sites. Fish Shellfish Immunol 2014;39:343–53. doi:10.1016/j.fsi.2014.05.024.

[47] Wood-Charlson EM, Weis VM. The diversity of C-type lectins in the genome of a basal metazoan, *Nematostella vectensis*. Dev Comp Immunol 2009;33:881–9. doi:10.1016/j.dci.2009.01.008.

[48] Lima TG, McCartney MA. Adaptive evolution of M3 lysin--a candidate gamete recognition protein in the *Mytilus edulis* species complex. Mol Biol Evol 2013;30:2688–98. doi:10.1093/molbev/mst165.

[49] Takagi T, Nakamura A, Deguchi R, Kyozuka K. Isolation, characterization, and primary structure of three major proteins obtained from *Mytilus* edulis sperm. J Biochem (Tokyo) 1994;116:598–605.

[50] Pales Espinosa E, Perrigault M, Allam B. Identification and molecular characterization of a mucosal lectin (MeML) from the blue mussel *Mytilus edulis* and its potential role in particle capture. Comp Biochem Physiol A Mol Integr Physiol 2010;156:495–501. doi:10.1016/j.cbpa.2010.04.004.

[51] Adema CM, Hertel LA, Miller RD, Loker ES. 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 1997;94:8691–6. doi:10.1073/pnas.94.16.8691.

[52] Hanington PC, Zhang S-M. The primary role of fibrinogen-related proteins in invertebrates is defense, not coagulation. J Innate Immun 2011;3:17–27. doi:10.1159/000321882.

[53] Zhang H, Wang L, Song L, Song X, Wang B, Mu C, et al. A fibrinogen-related protein from bay scallop *Argopecten irradians* involved in innate immunity as pattern recognition receptor. Fish Shellfish Immunol 2009;26:56–64. doi:10.1016/j.fsi.2008.07.019.

[54] Gorbushin AM, Iakovleva NV. A new gene family of single fibrinogen domain lectins in *Mytilus*. Fish Shellfish Immunol 2011;30:434–8. doi:10.1016/j.fsi.2010.10.002.

[55] Romero A, Dios S, Poisa-Beiro L, Costa MM, Posada D, Figueras A, et al. Individual sequence variability and functional activities of fibrinogen-related proteins (FREPs) in the Mediterranean mussel (*Mytilus galloprovincialis*) suggest ancient and complex immune recognition models in invertebrates. Dev Comp Immunol 2011;35:334–44. doi:10.1016/j.dci.2010.10.007.

[56] Dheilly NM, Duval D, Mouahid G, Eman R, Allienne J-F, Galinier R, et al. A family of variable immunoglobulin and lectin domain containing molecules in the snail *Biomphalaria glabrata*. Dev Comp Immunol 2015;48:234-43. doi:10.1016/j.dci.2014.10.009.

[57] Bao Y, Shen H, Zhou H, Dong Y, Lin Z. A tandem-repeat galectin from blood clam *Tegillarca granosa* and its induced mRNA expression response against bacterial challenge. Genes Genomics 2013;35:733–40. doi:10.1007/s13258-013-0123-3.

[58] Song X, Zhang H, Zhao J, Wang L, Qiu L, Mu C, et al. An immune responsive multidomain galectin from bay scallop *Argopectens irradians*. Fish Shellfish Immunol 2010;28:326–32. doi:10.1016/j.fsi.2009.11.016.

[59] Wang Z, Jian J, Lu Y, Wang B, Wu Z. A tandem-repeat galectin involved in innate immune response of the pearl oyster *Pinctada fucata*. Mar Genomics 2011;4:229–36. doi:10.1016/j.margen.2011.06.004.

[60] Zhang D, Jiang S, Hu Y, Cui S, Guo H, Wu K, et al. A multidomain galectin involved in innate immune response of pearl oyster Pinctada fucata. Dev Comp Immunol 2011;35:1–6. doi:10.1016/j.dci.2010.08.007.

[61] Kim JY, Kim YM, Cho SK, Choi KS, Cho M. Noble tandem-repeat galectin of Manila clam *Ruditapes philippinarum* is induced upon infection with the protozoan parasite *Perkinsus olseni*. Dev Comp Immunol 2008;32:1131–41. doi:10.1016/j.dci.2008.03.002.

[62] Tasumi S, Vasta GR. A galectin of unique domain organization from hemocytes of the Eastern oyster (*Crassostrea virginica*) is a receptor for the protistan parasite *Perkinsus marinus*. J Immunol Baltim Md 1950 2007;179:3086–98.

[63] Sato S, St-Pierre C, Bhaumik P, Nieminen J. Galectins in innate immunity: dual functions of host soluble beta-galactoside-binding lectins as damage-associated molecular patterns (DAMPs) and as receptors for pathogen-associated molecular patterns (PAMPs). Immunol Rev 2009;230:172–87. doi:10.1111/j.1600-065X.2009.00790.x.

[64] Cummings RD, Etzler ME. R-type Lectins. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essent. Glycobiol. 2nd ed., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.

[65] Fujii Y, Dohmae N, Takio K, Kawsar SMA, Matsumoto R, Hasan I, et al. A lectin from the mussel *Mytilus galloprovincialis* has a highly novel primary structure and induces glycan-mediated cytotoxicity of globotriaosylceramide-expressing lymphoma cells. J Biol Chem 2012;287:44772–83. doi:10.1074/jbc.M112.418012.

[66] Kovalchuk SN, Chikalovets IV, Chernikov OV, Molchanova VI, Li W, Rasskazov VA, et al. cDNA cloning and structural characterization of a lectin from the mussel *Crenomytilus grayanus* with a unique amino acid sequence and antibacterial activity. Fish Shellfish Immunol 2013;35:1320–4. doi:10.1016/j.fsi.2013.07.011.

[67] Anju A, Jeswin J, Thomas PC, Vijayan KK. Molecular cloning, characterization and expression analysis of F-type lectin from pearl oyster *Pinctada fucata*. Fish Shellfish Immunol 2013;35:170–4. doi:10.1016/j.fsi.2013.03.359.

[68] Chen J, Xiao S, Yu Z. F-type lectin involved in defense against bacterial infection in the pearl oyster (*Pinctada martensii*). Fish Shellfish Immunol 2011;30:750–4. doi:10.1016/j.fsi.2010.12.025.

[69] Wu Q, Li L, Zhang G. Crassostrea angulata bindin gene and the divergence of fucose-binding lectin repeats among three species of *Crassostrea*. Mar Biotechnol N Y N 2011;13:327–35. doi:10.1007/s10126-010-9304-0.

[70] Jayaraj SS, Thiagarajan R, Arumugam M, Mullainadhan P. Isolation, purification and characterization of beta-1,3-glucan binding protein from the plasma of marine mussel *Perna viridis*. Fish Shellfish Immunol 2008;24:715–25. doi:10.1016/j.fsi.2007.11.012.

[71] Zhang D, Ma J, Jiang J, Qiu L, Zhu C, Su T, et al. Molecular characterization and expression analysis of lipopolysaccharide and β-1,3-glucan-binding protein (LGBP) from pearl oyster *Pinctada fucata*. Mol Biol Rep 2010;37:3335–43. doi:10.1007/s11033-009-9920-9.

[72] Liu H, Wu C, Matsuda Y, Kawabata S-I, Lee BL, Söderhäll K, et al. Peptidoglycan activation of the proPO-system without a peptidoglycan receptor protein (PGRP)? Dev Comp Immunol 2011;35:51–61. doi:10.1016/j.dci.2010.08.005.

[73] Siva VS, Yang C, Yang J, Wang L, Wang L, Zhou Z, et al. Association of CfLGBP gene polymorphism with disease susceptibility/resistance of Zhikong scallop (*Chlamys farreri*) to *Listonella anguillarum*. Fish Shellfish Immunol 2012;32:1117–23. doi:10.1016/j.fsi.2012.03.017.

[74] Gobert V, Gottar M, Matskevich AA, Rutschmann S, Royet J, Belvin M, et al. Dual activation of the *Drosophila* toll pathway by two pattern recognition receptors. Science 2003;302:2126–30. doi:10.1126/science.1085432.

[75] Wang L, Weber ANR, Atilano ML, Filipe SR, Gay NJ, Ligoxygakis P. Sensing of Gram-positive bacteria in *Drosophila*: GNBP1 is needed to process and present peptidoglycan to PGRP-SA. EMBO J 2006;25:5005–14. doi:10.1038/sj.emboj.7601363.

[76] Kozhemyako VB, Rebrikov DV, Lukyanov SA, Bogdanova EA, Marin A, Mazur AK, et al. Molecular cloning and characterization of an endo-1,3-beta-D-glucanase from the mollusk *Spisula sachalinensis*. Comp Biochem Physiol B Biochem Mol Biol 2004;137:169–78. doi:10.1016/j.cbpc.2003.10.018.

[77] Zakharenko AM, Kusaykin MI, Kovalchuk SN, Anastyuk SD, Ly BM, Sova VV, et al. Enzymatic and molecular characterization of an endo-1,3-β-d-glucanase from the crystalline styles of the mussel *Perna viridis*. Carbohydr Res 2011;346:243–52. doi:10.1016/j.carres.2010.11.008.

[78] Huang G, Huang S, Yan X, Yang P, Li J, Xu W, et al. Two apextrin-like proteins mediate extracellular and intracellular bacterial recognition in amphioxus. Proc Natl Acad Sci 2014;111:13469–74. doi:10.1073/pnas.1405414111.

[79] Dheilly NM, Haynes PA, Bove U, Nair SV, Raftos DA. Comparative proteomic analysis of a sea urchin (*Heliocidaris erythrogramma*) antibacterial response revealed the involvement of apextrin and calreticulin. J Invertebr Pathol 2011;106:223–9. doi:10.1016/j.jip.2010.09.008.

[80] Estévez-Calvar N, Romero A, Figueras A, Novoa B. Involvement of pore-forming molecules in immune defense and development of the Mediterranean mussel (*Mytilus galloprovincialis*). Dev Comp Immunol 2011;35:1017–31. doi:10.1016/j.dci.2011.03.023.

[81] Nonaka M, Kimura A. Genomic view of the evolution of the complement system. Immunogenetics 2006;58:701–13. doi:10.1007/s00251-006-0142-1.
[82] Castellanos-Martínez S, Arteta D, Catarino S, Gestal C. De Novo Transcriptome Sequencing of the *Octopus vulgaris* Hemocytes Using Illumina RNA-Seq Technology: Response to the Infection by the Gastrointestinal Parasite Aggregata octopiana. PloS One 2014;9:e107873. doi:10.1371/journal.pone.0107873.
[83] Prado-Alvarez M, Rotllant J, Gestal C, Novoa B, Figueras A. Characterization of a C3 and a factor B-like in the carpet-shell clam, *Ruditapes decussatus*. Fish Shellfish Immunol 2009;26:305–15. doi:10.1016/j.fsi.2008.11.015.
[84] Sekiguchi R, Nonaka M. Evolution of the complement system in protostomes revealed by de novo transcriptome analysis of six species of Arthropoda. Dev Comp Immunol 2015 (in press). doi:10.1016/j.dci.2014.12.008.
[85] Blandin SA, Marois E, Levashina EA. Antimalarial Responses in *Anopheles gambiae*: From a Complement-like Protein to a Complement-like Pathway. Cell Host Microbe 2008;3:364–74. doi:10.1016/j.chom.2008.05.007.
[86] Bou Aoun R, Hetru C, Troxler L, Doucet D, Ferrandon D, Matt N. Analysis of Thioester-Containing Proteins during the Innate Immune Response of *Drosophila melanogaster*. J Innate Immun 2010;3:52–64. doi:10.1159/000321554.
[87] Zhang H, Wang L, Song L, Zhao J, Qiu L, Gao Y, et al. The genomic structure, alternative splicing and immune response of *Chlamys farreri* thioester-containing protein. Dev Comp Immunol 2009;33:1070–6. doi:10.1016/j.dci.2009.05.007.
[88] Tschopp J, Masson D, Stanley KK. Structural/functional similarity between proteins involved in complement- and cytotoxic T-lymphocyte-mediated cytolysis. Nature 1986;322:831–4. doi:10.1038/322831a0.
[89] Rosado CJ, Kondos S, Bull TE, Kuiper MJ, Law RHP, Buckle AM, et al. The MACPF/CDC family of pore-forming toxins. Cell Microbiol 2008;10:1765–74. doi:10.1111/j.1462-5822.2008.01191.x.
[90] Kondos SC, Hatfaludi T, Voskoboinik I, Trapani JA, Law RHP, Whisstock JC, et al. The structure and function of mammalian membrane-attack complex/perforin-like proteins. Tissue Antigens 2010;76:341–51. doi:10.1111/j.1399-0039.2010.01566.x.
[91] Kemp IK, Coyne VE. Identification and characterisation of the Mpeg1 homologue in the South African abalone, *Haliotis midae*. Fish Shellfish Immunol 2011;31:754–64. doi:10.1016/j.fsi.2011.07.010.
[92] Mah SA, Moy GW, Swanson WJ, Vacquier VD. A perforin-like protein from a marine mollusk. Biochem Biophys Res Commun 2004;316:468–75. doi:10.1016/j.bbrc.2004.02.073.
[93] Wang G-D, Zhang K-F, Zhang Z-P, Zou Z-H, Jia X-W, Wang S-H, et al. Molecular cloning and responsive expression of macrophage expressed gene from small abalone *Haliotis diversicolor* supertexta. Fish Shellfish Immunol 2008;24:346–59. doi:10.1016/j.fsi.2007.12.008.
[94] Leulier F, Lemaitre B. Toll-like receptors--taking an evolutionary approach. Nat Rev Genet 2008;9:165–78. doi:10.1038/nrg2303.
[95] Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. Cell 1996;86:973–83.
[96] Wang JQ, Jeelall YS, Ferguson LL, Horikawa K. Toll-Like Receptors and Cancer: MYD88 Mutation and Inflammation. Front Immunol 2014;5. doi:10.3389/fimmu.2014.00367.
[97] Qiu L, Song L, Xu W, Ni D, Yu Y. Molecular cloning and expression of a Toll receptor gene homologue from Zhikong Scallop, *Chlamys farreri*. Fish Shellfish Immunol 2007;22:451–66. doi:10.1016/j.fsi.2006.05.003.
[98] Toubiana M, Gerdol M, Rosani U, Pallavicini A, Venier P, Roch P. Toll-like receptors and MyD88 adaptors in *Mytilus*: Complete cds and gene expression levels. Dev Comp Immunol 2013;40:158–66. doi:10.1016/j.dci.2013.02.006.
[99] Toubiana M, Rosani U, Giambelluca S, Cammarata M, Gerdol M, Pallavicini A, et al. 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 2014;45:300–12. doi:10.1016/j.dci.2014.03.021.

[100] Blasius AL, Beutler B. Intracellular toll-like receptors. Immunity 2010;32:305–15. doi:10.1016/j.immuni.2010.03.012.

[101] O'Neill LAJ, Bowie AG. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. Nat Rev Immunol 2007;7:353–64. doi:10.1038/nri2079.

[102] Royet J, Dziarski R. Peptidoglycan recognition proteins: pleiotropic sensors and effectors of antimicrobial defences. Nat Rev Microbiol 2007;5:264–77. doi:10.1038/nrmicro1620.

[103] Itoh N, Takahashi KG. Distribution of multiple peptidoglycan recognition proteins in the tissues of Pacific oyster, *Crassostrea gigas.* Comp Biochem Physiol B Biochem Mol Biol 2008;150:409–17. doi:10.1016/j.cbpb.2008.04.011.

[104] Ni D, Song L, Wu L, Chang Y, Yu Y, Qiu L, et al. Molecular cloning and mRNA expression of peptidoglycan recognition protein (PGRP) gene in bay scallop (*Argopecten irradians*, Lamarck 1819). Dev Comp Immunol 2007;31:548–58. doi:10.1016/j.dci.2006.09.001.

[105] Wei X, Yang J, Yang D, Xu J, Liu X, Yang J, et al. Molecular cloning and mRNA expression of two peptidoglycan recognition protein (PGRP) genes from mollusk Solen grandis. Fish Shellfish Immunol 2012;32:178–85. doi:10.1016/j.fsi.2011.11.009.

[106] Bettencourt R, Pinheiro M, Egas C, Gomes P, Afonso M, Shank T, et al. High-throughput sequencing and analysis of the gill tissue transcriptome from the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus*. BMC Genomics 2010;11:559. doi:10.1186/1471-2164-11-559.

[107] Martins E, Figueras A, Novoa B, Santos RS, Moreira R, Bettencourt R. Comparative study of immune responses in the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus* and the shallow-water mussel *Mytilus galloprovincialis* challenged with Vibrio bacteria. Fish Shellfish Immunol 2014;40:485–99. doi:10.1016/j.fsi.2014.07.018.

[108] Canesi L, Betti M, Ciacci C, Scarpato A, Citterio B, Pruzzo C, et al. Signaling pathways involved in the physiological response of mussel hemocytes to bacterial challenge: the role of stress-activated p38 MAP kinases. Dev Comp Immunol 2002;26:325–34.

[109] Bettencourt R, Dando P, Collins P, Costa V, Allam B, Serrão Santos R. Innate immunity in the deep sea hydrothermal vent mussel *Bathymodiolus azoricus*. Comp Biochem Physiol A Mol Integr Physiol 2009;152:278–89. doi:10.1016/j.cbpa.2008.10.022.

[110] Kopp E, Medzhitov R, Carothers J, Xiao C, Douglas I, Janeway CA, et al. ECSIT is an evolutionarily conserved intermediate in the Toll/IL-1 signal transduction pathway. Genes Dev 1999;13:2059–71.

[111] Tu Z, Mooney SM, Lee FS. A subdomain of MEKK1 that is critical for binding to MKK4. Cell Signal 2003;15:65–77.

[112] Xiang Z, Qu F, Li J, Qi L, Yang Z, Kong X, et al. Activator protein-1 (AP-1) and response to pathogen infection in the Hong Kong oyster (*Crassostrea hongkongensis*). Fish Shellfish Immunol 2014;36:83–9. doi:10.1016/j.fsi.2013.10.005.

[113] Canton J, Neculai D, Grinstein S. Scavenger receptors in homeostasis and immunity. Nat Rev Immunol 2013;13:621–34. doi:10.1038/nri3515.

[114] McDowell IC, Nikapitiya C, Aguiar D, Lane CE, Istrail S, Gomez-Chiarri M. Transcriptome of American Oysters, *Crassostrea virginica*, in Response to Bacterial Challenge: Insights into Potential Mechanisms of Disease Resistance. PLoS ONE 2014;9:e105097. doi:10.1371/journal.pone.0105097.

[115] Liu L, Yang J, Qiu L, Wang L, Zhang H, Wang M, et al. A novel scavenger receptor-cysteine-rich (SRCR) domain containing scavenger receptor identified from mollusk mediated PAMP recognition and binding. Dev Comp Immunol 2011;35:227–39. doi:10.1016/j.dci.2010.09.010.

[116] Griebel T, Maekawa T, Parker JE. Nucleotide-binding oligomerization domain-like receptor cooperativity in effector-triggered immunity. Trends Immunol 2014. doi:10.1016/j.it.2014.09.005.

[117] Barbé F, Douglas T, Saleh M. Advances in Nod-like receptors (NLR) biology. Cytokine Growth Factor Rev 2014;25:681-97. doi:10.1016/j.cytogfr.2014.07.001.

[118] Lange C, Hemmrich G, Klostermeier UC, López-Quintero JA, Miller DJ, Rahn T, et al. Defining the Origins of the NOD-Like Receptor System at the Base of Animal Evolution. Mol Biol Evol 2011;28:1687–702. doi:10.1093/molbev/msq349.

[119] Hamada M, Shoguchi E, Shinzato C, Kawashima T, Miller DJ, Satoh N. The complex NOD-like receptor repertoire of the coral *Acropora digitifera* includes novel domain combinations. Mol Biol Evol 2013;30:167–76. doi:10.1093/molbev/mss213.

[120] Jearaphunt M, Noonin C, Jiravanichpaisal P, Nakamura S, Tassanakajon A, Söderhäll I, et al. Caspase-1-like regulation of the proPO-system and role of ppA and caspase-1-like cleaved peptides from proPO in innate immunity. PLoS Pathog 2014;10:e1004059. doi:10.1371/journal.ppat.1004059.

[121] Travassos LH, Carneiro LAM, Ramjeet M, Hussey S, Kim Y-G, Magalhães JG, et al. Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. Nat Immunol 2010;11:55–62. doi:10.1038/ni.1823.

[122] Chiu Y-H, Macmillan JB, Chen ZJ. RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. Cell 2009;138:576–91. doi:10.1016/j.cell.2009.06.015.

[123] Loo Y-M, Gale M. Immune signaling by RIG-I-like receptors. Immunity 2011;34:680–92. doi:10.1016/j.immuni.2011.05.003.

[124] Thompson AJV, Locarnini SA. Toll-like receptors, RIG-I-like RNA helicases and the antiviral innate immune response. Immunol Cell Biol 2007;85:435–45. doi:10.1038/sj.icb.7100100.

[125] Yoneyama M, Fujita T. Function of RIG-I-like Receptors in Antiviral Innate Immunity. J Biol Chem 2007;282:15315–8. doi:10.1074/jbc.R700007200.

[126] Deddouche S, Matt N, Budd A, Mueller S, Kemp C, Galiana-Arnoux D, et al. The DExD/H-box helicase Dicer-2 mediates the induction of antiviral activity in *Drosophila*. Nat Immunol 2008;9:1425–32. doi:10.1038/ni.1664.

[127] Galiana-Arnoux D, Dostert C, Schneemann A, Hoffmann JA, Imler J-L. Essential function in vivo for Dicer-2 in host defense against RNA viruses in *Drosophila*. Nat Immunol 2006;7:590–7. doi:10.1038/ni1335.

[128] Zou J, Chang M, Nie P, Secombes CJ. Origin and evolution of the RIG-I like RNA helicase gene family. BMC Evol Biol 2009;9:85. doi:10.1186/1471-2148-9-85.

[129] Lu R, Yigit E, Li W-X, Ding S-W. An RIG-I-Like RNA Helicase Mediates Antiviral RNAi Downstream of Viral siRNA Biogenesis in *Caenorhabditis elegans*. PLoS Pathog 2009;5:e1000286. doi:10.1371/journal.ppat.1000286.

[130] Zhang Y, Yu F, Li J, Tong Y, Zhang Y, Yu Z. The first invertebrate RIG-I-like receptor (RLR) homolog gene in the pacific oyster *Crassostrea gigas*. Fish Shellfish Immunol 2014;40:466–71. doi:10.1016/j.fsi.2014.07.029.

[131] Burdette DL, Vance RE. STING and the innate immune response to nucleic acids in the cytosol. Nat Immunol 2013;14:19–26. doi:10.1038/ni.2491.

[132] Abe T, Barber GN. Cytosolic-DNA-mediated, STING-dependent proinflammatory gene induction necessitates canonical NF-κB activation through TBK1. J Virol 2014;88:5328–41. doi:10.1128/JVI.00037-14.

[133] Zhang Z, Yuan B, Bao M, Lu N, Kim T, Liu Y-J. The helicase DDX41 senses intracellular DNA mediated by the adaptor STING in dendritic cells. Nat Immunol 2011;12:959–65. doi:10.1038/ni.2091.

[134] Wu X, Wu F-H, Wang X, Wang L, Siedow JN, Zhang W, et al. Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. Nucleic Acids Res 2014;42:8243–57. doi:10.1093/nar/gku569.

[135] Ouyang S, Song X, Wang Y, Ru H, Shaw N, Jiang Y, et al. Structural Analysis of the STING Adaptor Protein Reveals a Hydrophobic Dimer Interface and Mode of Cyclic di-GMP Binding. Immunity 2012;36:1073–86. doi:10.1016/j.immuni.2012.03.019.

[136] Watson RO, Manzanillo PS, Cox JS. Extracellular M. tuberculosis DNA targets bacteria for autophagy by activating the host DNA-sensing pathway. Cell 2012;150:803–15. doi:10.1016/j.cell.2012.06.040.

[137] Rasmussen SB, Horan KA, Holm CK, Stranks AJ, Mettenleiter TC, Simon AK, et al. Activation of autophagy by α-herpesviruses in myeloid cells is mediated by cytoplasmic viral DNA through a mechanism dependent on stimulator of IFN genes. J Immunol Baltim Md 1950 2011;187:5268–76. doi:10.4049/jimmunol.1100949.

[138] Deretic V, Saitoh T, Akira S. Autophagy in infection, inflammation and immunity. Nat Rev Immunol 2013;13:722–37. doi:10.1038/nri3532.

[139] Charlet M, Chernysh S, Philippe H, Hetru C, Hoffmann JA, Bulet P. Isolation of several cysteine-rich antimicrobial peptides from the blood of a mollusc, *Mytilus edulis*. J Biol Chem 1996;271:21808–13.

[140] Hubert F, Noel T, Roch P. A member of the arthropod defensin family from edible Mediterranean mussels (*Mytilus galloprovincialis)*. Eur J Biochem FEBS 1996;240:302–6.

[141] Yang YS, Mitta G, Chavanieu A, Calas B, Sanchez JF, Roch P, et al. Solution structure and activity of the synthetic four-disulfide bond Mediterranean mussel defensin (MGD-1). Biochemistry (Mosc) 2000;39:14436–47.
[142] Romestand B, Molina F, Richard V, Roch P, Granier C. Key role of the loop connecting the two beta strands of mussel defensin in its antimicrobial activity. Eur J Biochem FEBS 2003;270:2805–13.
[143] Gueguen Y, Herpin A, Aumelas A, Garnier J, Fievet J, Escoubas J-M, et al. Characterization of a defensin from the oyster *Crassostrea gigas*. Recombinant production, folding, solution structure, antimicrobial activities, and gene expression. J Biol Chem 2006;281:313–23. doi:10.1074/jbc.M510850200.
[144] Froy O, Gurevitz M. Arthropod and mollusk defensins – evolution by exon-shuffling. Trends Genet 2003;19:684–7. doi:10.1016/j.tig.2003.10.010.
[145] Schmitt P, Gueguen Y, Desmarais E, Bachère E, Lorgeril J de. Molecular diversity of antimicrobial effectors in the oyster *Crassostrea gigas*. BMC Evol Biol 2010;10:23. doi:10.1186/1471-2148-10-23.
[146] Peng K, Wang J, Sheng J, Zeng L, Hong Y. Molecular characterization and immune analysis of a defensin from freshwater pearl mussel, *Hyriopsis schlegelii*. Aquaculture 2012;334–337:45–50. doi:10.1016/j.aquaculture.2011.12.039.
[147] Seo J-K, Crawford JM, Stone KL, Noga EJ. Purification of a novel arthropod defensin from the American oyster, *Crassostrea virginica*. Biochem Biophys Res Commun 2005;338:1998–2004. doi:10.1016/j.bbrc.2005.11.013.
[148] Xu W, Faisal M. Defensin of the zebra mussel (*Dreissena polymorpha*): molecular structure, in vitro expression, antimicrobial activity, and potential functions. Mol Immunol 2010;47:2138–47. doi:10.1016/j.molimm.2010.01.025.
[149] Mitta G, Vandenbulcke F, Hubert F, Salzet M, Roch P. Involvement of Mytilins in Mussel Antimicrobial Defense. J Biol Chem 2000;275:12954–62. doi:10.1074/jbc.275.17.12954.
[150] Mitta G, Hubert F, Dyrynda EA, Boudry P, Roch P. Mytilin B and MGD2, two antimicrobial peptides of marine mussels: gene structure and expression analysis. Dev Comp Immunol 2000;24:381–93. doi:10.1016/S0145-305X(99)00084-1.
[151] Roch P, Yang Y, Toubiana M, Aumelas A. NMR structure of mussel mytilin, and antiviral-antibacterial activities of derived synthetic peptides. Dev Comp Immunol 2008;32:227–38. doi:10.1016/j.dci.2007.05.006.
[152] Rosani U, Varotto L, Rossi A, Roch P, Novoa B, Figueras A, et al. Massively parallel amplicon sequencing reveals isotype-specific variability of antimicrobial peptide transcripts in *Mytilus galloprovincialis*. PloS One 2011;6:e26680. doi:10.1371/journal.pone.0026680.
[153] Mitta G, Hubert F, Noël T, Roch P. Myticin, a novel cysteine-rich antimicrobial peptide isolated from haemocytes and plasma of the mussel *Mytilus galloprovincialis*. Eur J Biochem FEBS 1999;265:71–8.
[154] Li H, Parisi M-G, Parrinello N, Cammarata M, Roch P. Molluscan antimicrobial peptides, a review from activity-based evidences to computer-assisted sequences. Invertebr Surviv J 2011;8:85–97.
[155] Venier P, Pallavicini A, De Nardi B, Lanfranchi G. Towards a catalogue of genes transcribed in multiple tissues of *Mytilus galloprovincialis*. Gene 2003;314:29–40. doi:10.1016/S0378-1119(03)00708-X.
[156] Costa MM, Dios S, Alonso-Gutierrez J, Romero A, Novoa B, Figueras A. Evidence of high individual diversity on myticin C in mussel (*Mytilus galloprovincialis*). Dev Comp Immunol 2009;33:162–70. doi:10.1016/j.dci.2008.08.005.
[157] Pallavicini A, Costa M del M, Gestal C, Dreos R, Figueras A, Venier P, et al. High sequence variability of myticin transcripts in hemocytes of immune-stimulated mussels suggests ancient host-pathogen interactions. Dev Comp Immunol 2008;32:213–26. doi:10.1016/j.dci.2007.05.008.
[158] Vera M, Martínez P, Poisa-Beiro L, Figueras A, Novoa B. Genomic Organization, molecular diversification, and evolution of antimicrobial peptide myticin-c genes in the mussel (*Mytilus galloprovincialis*). PLoS ONE 2011;6. doi:10.1371/journal.pone.0024041.
[159] Balseiro P, Falcó A, Romero A, Dios S, Martínez-López A, Figueras A, et al. *Mytilus galloprovincialis* Myticin C: A Chemotactic Molecule with Antiviral Activity and Immunoregulatory Properties. PLoS ONE 2011;6:e23140. doi:10.1371/journal.pone.0023140.

[160] Sonthi M, Toubiana M, Pallavicini A, Venier P, Roch P. Diversity of Coding Sequences and Gene Structures of the Antifungal Peptide Mytimycin (MytM) from the Mediterranean Mussel, *Mytilus galloprovincialis*. Mar Biotechnol 2011;13:857–67. doi:10.1007/s10126-010-9345-4.

[161] Cantet F, Toubiana M, Parisi M-G, Sonthi M, Cammarata M, Roch P. Individual variability of mytimycin gene expression in mussel. Fish Shellfish Immunol 2012;33:641–4. doi:10.1016/j.fsi.2012.06.020.

[162] Sonthi M, Cantet F, Toubiana M, Trapani M-R, Parisi M-G, Cammarata M, et al. Gene expression specificity of the mussel antifungal mytimycin (MytM). Fish Shellfish Immunol 2012;32:45–50. doi:10.1016/j.fsi.2011.10.017.

[163] Jung S, Sönnichsen FD, Hung C-W, Tholey A, Boidin-Wichlacz C, Haeusgen W, et al. Macin family of antimicrobial proteins combines antimicrobial and nerve repair activities. J Biol Chem 2012;287:14246–58. doi:10.1074/jbc.M111.336495.

[164] Gerdol M, De Moro G, Manfrin C, Venier P, Pallavicini A. Big defensins and mytimacins, new AMP families of the Mediterranean mussel *Mytilus galloprovincialis*. Dev Comp Immunol 2012;36:390–9. doi:10.1016/j.dci.2011.08.003.

[165] Hung C-W, Jung S, Grötzinger J, Gelhaus C, Leippe M, Tholey A. Determination of disulfide linkages in antimicrobial peptides of the macin family by combination of top-down and bottom-up proteomics. J Proteomics 2014;103:216–26. doi:10.1016/j.jprot.2014.04.001.

[166] Jung S, Dingley AJ, Augustin R, Anton-Erxleben F, Stanisak M, Gelhaus C, et al. Hydramacin-1, structure and antibacterial activity of a protein from the basal metazoan *Hydra*. J Biol Chem 2009;284:1896–905. doi:10.1074/jbc.M804713200.

[167] Tasiemski A, Vandenbulcke F, Mitta G, Lemoine J, Lefebvre C, Sautière P-E, et al. Molecular Characterization of Two Novel Antibacterial Peptides Inducible upon Bacterial Challenge in an Annelid, the Leech *Theromyzon tessulatum*. J Biol Chem 2004;279:30973–82. doi:10.1074/jbc.M312156200.

[168] Zhong J, Wang W, Yang X, Yan X, Liu R. A novel cysteine-rich antimicrobial peptide from the mucus of the snail of *Achatina fulica*. Peptides 2013;39:1–5. doi:10.1016/j.peptides.2012.09.001.

[169] Beninger PG, St-Jean SD. The role of mucus in particle processing by suspension-feeding marine bivalves: Unifying principles. Mar Biol 1997;129:389–97. doi:10.1007/s002270050179.

[170] Dufour SC, Beninger PG. A functional interpretation of cilia and mucocyto distributions on the abfrontal surface of bivalve gills. Mar Biol 2001;138:295–309. doi:10.1007/s002270000466.

[171] Zhu S, Gao B. Evolutionary origin of β-defensins. Dev Comp Immunol 2013;39:79–84. doi:10.1016/j.dci.2012.02.011.

[172] Saito T, Kawabata S, Shigenaga T, Takayenoki Y, Cho J, Nakajima H, et al. A novel big defensin identified in horseshoe crab hemocytes: isolation, amino acid sequence, and antibacterial activity. J Biochem (Tokyo) 1995;117:1131–7.

[173] Zhao J, Song L, Li C, Ni D, Wu L, Zhu L, et al. Molecular cloning, expression of a big defensin gene from bay scallop *Argopecten irradians* and the antimicrobial activity of its recombinant protein. Mol Immunol 2007;44:360–8. doi:10.1016/j.molimm.2006.02.025.

[174] Rosa RD, Santini A, Fievet J, Bulet P, Destoumieux-Garzón D, Bachère E. Big defensins, a diverse family of antimicrobial peptides that follows different patterns of expression in hemocytes of the oyster *Crassostrea gigas*. PloS One 2011;6:e25594. doi:10.1371/journal.pone.0025594.

[175] Zhao J, Li C, Chen A, Li L, Su X, Li T. Molecular characterization of a novel big defensin from clam *Venerupis philippinarum*. PloS One 2010;5:e13480. doi:10.1371/journal.pone.0013480.

[176] Liao Z, Wang X, Liu H, Fan M, Sun J, Shen W. Molecular characterization of a novel antimicrobial peptide from *Mytilus coruscus*. Fish Shellfish Immunol 2013;34:610–6. doi:10.1016/j.fsi.2012.11.030.

[177] Bulet P, Stöcklin R, Menin L. Anti-microbial peptides: from invertebrates to vertebrates. Immunol Rev 2004;198:169–84.

[178] Gueguen Y, Bernard R, Julie F, Paulina S, Delphine D-G, Franck V, et al. Oyster hemocytes express a proline-rich peptide displaying synergistic antimicrobial activity with a defensin. Mol Immunol 2009;46:516–22. doi:10.1016/j.molimm.2008.07.021.

[179] Qin C-L, Huang W, Zhou S-Q, Wang X-C, Liu H-H, Fan M-H, et al. Characterization of a novel antimicrobial peptide with chiting-biding domain from *Mytilus coruscus*. Fish Shellfish Immunol 2014;41:362–70. doi:10.1016/j.fsi.2014.09.019.

[180] Destoumieux D, Bulet P, Loew D, Van Dorsselaer A, Rodriguez J, Bachère E. Penaeidins, a new family of antimicrobial peptides isolated from the shrimp *Penaeus vannamei* (Decapoda). J Biol Chem 1997;272:28398–406.

[181] Kawabata S, Nagayama R, Hirata M, Shigenaga T, Agarwala KL, Saito T, et al. Tachycitin, a small granular component in horseshoe crab hemocytes, is an antimicrobial protein with chitin-binding activity. J Biochem (Tokyo) 1996;120:1253–60.

[182] Callewaert L, Michiels CW. Lysozymes in the animal kingdom. J Biosci 2010;35:127–60.

[183] Bachali S, Jager M, Hassanin A, Schoentgen F, Jollès P, Fiala-Medioni A, et al. Phylogenetic Analysis of Invertebrate Lysozymes and the Evolution of Lysozyme Function. J Mol Evol 2002;54:652–64. doi:10.1007/s00239-001-0061-6.

[184] Costa MM, Prado-Alvarez M, Gestal C, Li H, Roch P, Novoa B, et al. Functional and molecular immune response of Mediterranean mussel (*Mytilus galloprovincialis*) haemocytes against pathogen-associated molecular patterns and bacteria. Fish Shellfish Immunol 2009;26:515–23.

[185] Li H, Parisi M-G, Toubiana M, Cammarata M, Roch P. Lysozyme gene expression and hemocyte behaviour in the Mediterranean mussel, *Mytilus galloprovincialis*, after injection of various bacteria or temperature stresses. Fish Shellfish Immunol 2008;25:143–52. doi:10.1016/j.fsi.2008.04.001.

[186] Xue Q, Hellberg ME, Schey KL, Itoh N, Eytan RI, Cooper RK, et al. A new lysozyme from the eastern oyster, Crassostrea virginica, and a possible evolutionary pathway for i-type lysozymes in bivalves from host defense to digestion. BMC Evol Biol 2010;10:213. doi:10.1186/1471-2148-10-213.

[187] Wang Q, Wang C, Mu C, Wu H, Zhang L, Zhao J. A novel C-type lysozyme from *Mytilus galloprovincialis*: insight into innate immunity and molecular evolution of invertebrate C-type lysozymes. PloS One 2013;8:e67469. doi:10.1371/journal.pone.0067469.

[188] Zhao J, Song L, Li C, Zou H, Ni D, Wang W, et al. Molecular cloning of an invertebrate goose-type lysozyme gene from *Chlamys farreri*, and lytic activity of the recombinant protein. Mol Immunol 2007;44:1198–208. doi:10.1016/j.molimm.2006.06.008.

[189] Wang Q, Zhang L, Zhao J, You L, Wu H. Two Goose-Type Lysozymes in *Mytilus galloprovincialis*: Possible Function Diversification and Adaptive Evolution. PLoS ONE 2012;7. doi:10.1371/journal.pone.0045148.

[190] Ding J, Wang R, Yang F, Zhao L, Qin Y, Zhang G, et al. Identification and characterization of a novel phage-type like lysozyme from Manila clam, *Ruditapes philippinarum*. Dev Comp Immunol 2014;47:81–9. doi:10.1016/j.dci.2014.06.013.

[191] Gonzalez M, Gueguen Y, Destoumieux-Garzón D, Romestand B, Fievet J, Pugnière M, et al. Evidence of a bactericidal permeability increasing protein in an invertebrate, the *Crassostrea gigas* Cg-BPI. Proc Natl Acad Sci U S A 2007;104:17759–64. doi:10.1073/pnas.0702281104.

[192] Mao Y, Zhou C, Zhu L, Huang Y, Yan T, Fang J, et al. Identification and expression analysis on bactericidal permeability-increasing protein (BPI)/lipopolysaccharide-binding protein (LBP) of ark shell, *Scapharca broughtonii*. Fish Shellfish Immunol 2013;35:642–52. doi:10.1016/j.fsi.2013.05.025.

[193] Donnelly S, Dalton JP, Robinson MW. How pathogen-derived cysteine proteases modulate host immune responses. Adv Exp Med Biol 2011;712:192–207. doi:10.1007/978-1-4419-8414-2_12.

[194] Montagnani C, Avarre JC, de Lorgeril J, Quiquand M, Boulo V, Escoubas JM. First evidence of the activation of Cg-timp, an immune response component of pacific oysters, through a damage-associated molecular pattern pathway. Dev Comp Immunol 2007;31:1–11. doi:10.1016/j.dci.2006.04.002.

[195] Montagnani C, Le Roux F, Berthe F, Escoubas JM. Cg-TIMP, an inducible tissue inhibitor of metalloproteinase from the Pacific oyster *Crassostrea gigas* with a potential role in wound healing and defense mechanisms(1). FEBS Lett 2001;500:64–70.

[196] Zhu L, Song L, Chang Y, Xu W, Wu L. Molecular cloning, characterization and expression of a novel serine proteinase inhibitor gene in bay scallops (*Argopecten irradians*, Lamarck 1819). Fish Shellfish Immunol 2006;20:320–31. doi:10.1016/j.fsi.2005.05.009.

[197] Zhang D, Ma J, Jiang S. Molecular characterization, expression and function analysis of a five-domain Kazal-type serine proteinase inhibitor from pearl oyster *Pinctada fucata*. Fish Shellfish Immunol 2014;37:115–21. doi:10.1016/j.fsi.2013.12.011.

[198] Maldonado-Aguayo W, Núñez-Acuña G, Valenzuela-Muñoz V, Chávez-Mardones J, Gallardo-Escárate C. Molecular characterization of two kazal-type serine proteinase inhibitor genes in the surf clam *Mesodesma donacium* exposed to *Vibrio anguillarum*. Fish Shellfish Immunol 2013;34:1448–54. doi:10.1016/j.fsi.2013.03.356.

[199] Wei X, Yang J, Yang J, Liu X, Liu M, Yang D, et al. A four-domain Kunitz-type proteinase inhibitor from *Solen grandis* is implicated in immune response. Fish Shellfish Immunol 2012;33:1276–84. doi:10.1016/j.fsi.2012.09.015.

[200] Xiong X, Chen L, Li Y, Xie L, Zhang R. Pf-ALMP, a novel astacin-like metalloproteinase with cysteine arrays, is abundant in hemocytes of pearl oyster *Pinctada fucata*. Biochim Biophys Acta BBA - Gene Struct Expr 2006;1759:526–34. doi:10.1016/j.bbaexp.2006.09.006.

[201] Yan F, Jiao Y, Deng Y, Du X, Huang R, Wang Q, et al. Tissue inhibitor of metalloproteinase gene from pearl oyster *Pinctada martensii* participates in nacre formation. Biochem Biophys Res Commun 2014;450:300–5. doi:10.1016/j.bbrc.2014.05.118.

[202] Rimphanitchayakit V, Tassanakajon A. Structure and function of invertebrate Kazal-type serine proteinase inhibitors. Dev Comp Immunol 2010;34:377–86. doi:10.1016/j.dci.2009.12.004.

[203] Xue Q-G, Waldrop GL, Schey KL, Itoh N, Ogawa M, Cooper RK, et al. A novel slow-tight binding serine protease inhibitor from eastern oyster (*Crassostrea virginica*) plasma inhibits perkinsin, the major extracellular protease of the oyster protozoan parasite *Perkinsus marinus*. Comp Biochem Physiol B Biochem Mol Biol 2006;145:16–26. doi:10.1016/j.cbpb.2006.05.010.

[204] La Peyre JF, Xue Q-G, Itoh N, Li Y, Cooper RK. Serine protease inhibitor cvSI-1 potential role in the eastern oyster host defense against the protozoan parasite *Perkinsus marinus*. Dev Comp Immunol 2010;34:84–92. doi:10.1016/j.dci.2009.08.007.

[205] Kaiser P, Rothwell L, Avery S, Balu S. Evolution of the interleukins. Dev Comp Immunol 2004;28:375–94. doi:10.1016/j.dci.2003.09.004.

[206] Malagoli D, Sacchi S, Ottaviani E. Lectins and cytokines in celomatic invertebrates: two tales with the same end 2010.

[207] Wang G, Li X, Li J. Association between SNPs in interferon regulatory factor 2 (IRF-2) gene and resistance to Aeromonas hydrophila in freshwater mussel *Hyriopsis cumingii.* Fish Shellfish Immunol 2013;34:1366–71. doi:10.1016/j.fsi.2013.02.006.

[208] Huang X-D, Liu W-G, Wang Q, Zhao M, Wu S-Z, Guan Y-Y, et al. Molecular characterization of interferon regulatory factor 2 (IRF-2) homolog in pearl oyster *Pinctada fucata*. Fish Shellfish Immunol 2013;34:1279–86. doi:10.1016/j.fsi.2013.02.003.

[209] Beschin A, Bilej M, Torreele E, De Baetselier P. On the existence of cytokines in invertebrates. Cell Mol Life Sci CMLS 2001;58:801–14.

[210] Hughes TK, Smith EM, Chin R, Cadet P, Sinisterra J, Leung MK, et al. Interaction of immunoactive monokines (interleukin 1 and tumor necrosis factor) in the bivalve mollusc *Mytilus edulis*. Proc Natl Acad Sci U S A 1990;87:4426–9.

[211] Roberts S, Gueguen Y, de Lorgeril J, Goetz F. Rapid accumulation of an interleukin 17 homolog transcript in *Crassostrea gigas* hemocytes following bacterial exposure. Dev Comp Immunol 2008;32:1099–104. doi:10.1016/j.dci.2008.02.006.

[212] Li J, Zhang Y, Zhang Y, Xiang Z, Tong Y, Qu F, et al. Genomic characterization and expression analysis of five novel IL-17 genes in the Pacific oyster, *Crassostrea gigas*. Fish Shellfish Immunol 2014;40:455–65. doi:10.1016/j.fsi.2014.07.026.

[213] Parisi M-G, Toubiana M, Mangano V, Parrinello N, Cammarata M, Roch P. MIF from mussel: Coding sequence, phylogeny, polymorphism, 3D model and regulation of expression. Dev Comp Immunol 2012;36:688–96. doi:10.1016/j.dci.2011.10.014.

[214] Martín-Gómez L, Villalba A, Carballal MJ, Abollo E. Molecular characterisation of TNF, AIF, dermatopontin and VAMP genes of the flat oyster *Ostrea edulis* and analysis of their modulation by diseases. Gene 2014;533:208–17. doi:10.1016/j.gene.2013.09.085.

[215] Li J, Chen J, Zhang Y, Yu Z. Expression of allograft inflammatory factor-1 (AIF-1) in response to bacterial challenge and tissue injury in the pearl oyster, *Pinctada martensii*. Fish Shellfish Immunol 2013;34:365–71. doi:10.1016/j.fsi.2012.11.012.

[216] Zhang Y, Li J, Yu F, He X, Yu Z. Allograft inflammatory factor-1 stimulates hemocyte immune activation by enhancing phagocytosis and expression of inflammatory cytokines in *Crassostrea gigas*. Fish Shellfish Immunol 2013;34:1071–7. doi:10.1016/j.fsi.2013.01.014.

[217] De Zoysa M, Jung S, Lee J. First molluscan TNF-alpha homologue of the TNF superfamily in disk abalone: molecular characterization and expression analysis. Fish Shellfish Immunol 2009;26:625–31. doi:10.1016/j.fsi.2008.10.004.

[218] Li L, Qiu L, Song L, Song X, Zhao J, Wang L, et al. First molluscan TNFR homologue in Zhikong scallop: molecular characterization and expression analysis. Fish Shellfish Immunol 2009;27:625–32. doi:10.1016/j.fsi.2009.07.009.

[219] Su J, Qiu L, Li L, Liu L, Wang L, Siva VS, et al. cDNA cloning and characterization of a new member of the tumor necrosis factor receptor family gene from scallop, *Chlamys farreri*. Mol Biol Rep 2011;38:4483–90. doi:10.1007/s11033-010-0578-0.