Bivalves Present the Largest and Most Diversified Repertoire of Toll-Like Receptors in the Animal Kingdom, Suggesting Broad-Spectrum Pathogen Recognition in Marine Waters

Amaro Saco [1], Beatriz Novoa,[1] Samuele Greco,[2] Marco Gerdol,[2] and Antonio Figueras*,[1]

[1]Institute of Marine Research (IIM), National Research Council (CSIC), Vigo, Spain

[2]Department of Life Sciences, University of Trieste, Trieste, Italy

***Corresponding author:** E-mail: antoniofigueras@iim.csic.es.

Associate editor: Xuhua Xia

Abstract

Toll-like receptors (TLRs) are the most widespread class of membrane-bound innate immune receptors, responsible of specific pathogen recognition and production of immune effectors through the activation of intracellular signaling cascades. The repertoire of TLRs was analyzed in 85 metazoans, enriched on molluscan species, an underrepresented phylum in previous studies. Following an ancient evolutionary origin, suggested by the presence of TLR genes in Anthozoa (Cnidaria), these receptors underwent multiple independent gene family expansions, the most significant of which occurred in bivalve molluscs. Marine mussels (*Mytilus* spp.) had the largest TLR repertoire in the animal kingdom, with evidence of several lineage-specific expanded TLR subfamilies with different degrees of orthology conservation within bivalves. Phylogenetic analyses revealed that bivalve TLR repertoires were more diversified than their counterparts in deuterostomes or ecdysozoans. The complex evolutionary history of TLRs, characterized by lineage-specific expansions and losses, along with episodic positive selection acting on the extracellular recognition domains, suggests that functional diversification might be a leading evolutionary force. We analyzed a comprehensive transcriptomic data set from *Mytilus galloprovincialis* and built transcriptomic correlation clusters with the TLRs expressed in gills and in hemocytes. The implication of specific TLRs in different immune pathways was evidenced, as well as their specific modulation in response to different biotic and abiotic stimuli. We propose that, in a similar fashion to the remarkable functional specialization of vertebrate TLRs, the expansion of the TLR gene family in bivalves attends to a functional specification motivated by the biological particularities of these organisms and their living environment.

Key words: **Toll-like receptor, TLR, evolution, innate immunity, bivalvia, functional diversification.**

Introduction

Toll-like receptors (TLRs) are type I transmembrane (TM) glycoproteins of the plasma membrane and endosomes that act as pattern recognition receptors (PRRs) of the innate immune system. Innate immunity relies on the recognition of conserved microbe-associated molecular patterns (MAMPs) by immune receptors such as TLRs, which trigger specific cellular and humoral responses through the activation of intracellular immune signaling cascades. TLRs were first identified in *Drosophila melanogaster* in relation with embryonic development functions (Anderson and Nüsslein-Volhard 1984; Anderson et al. 1985). Subsequent studies revealed the immune implications of this receptor (Lemaitre et al. 1996), and the recognition functions of human TLR genes were identified shortly thereafter (Medzhitov et al. 1997; Poltorak et al. 1998). Since then, notwithstanding the high diversification of the immune systems of vertebrates and invertebrates, TLRs have been retrieved and studied in virtually all major animal phyla, emerging as the most widespread MAMP sensors in Metazoa (Brennan and Gilmore 2018; Orús-Alcalde et al. 2021).

TLRs are characterized by extracellular leucine-rich repeat (LRR) domains, responsible of recognition and binding, and an intracellular Toll/interleukin-1 receptor (TIR) domain that participates in the activation of signaling pathways, interacting with other TIR-containing cytoplasmic proteins such as the myeloid differentiation factor 88 (Myd88) (Satake and Sekiguchi 2012; Tassia et al. 2017). The extracellular domains are usually flanked by N-terminal and C-terminal LRR domains (LRR-NT and LRR-CT, respectively) that are distinct in sequence and structure from the central LRRs. This domain organization, known as single cysteine cluster TLRs or sccTLRs (Leulier and Lemaitre 2008), is the only one found in vertebrate TLRs, and it is also present in *Drosophila* and other invertebrates. LRR-CT and LRR-NT domains can also appear interrupting the typical LRR array. This domain organization (multiple cysteine cluster TLRs [mccTLRs]) was the predominant structure described in *Drosophila* Toll proteins

(Leulier and Lemaitre 2008) and is commonly found in invertebrates but completely missing in vertebrates. The taxonomical spread of these two different TLR architectures originally led some authors to hypothesize that mccTLRs predated sccTLRs (Luo and Zheng 2000; Brennan and Gilmore 2018). However, the identification of sccTLRs in the ancient repertoires of cnidarians disproved this hypothesis, suggesting instead that all extant metazoan TLRs stem from three ancestral clades that underwent loss or diversification in different lineages during animal evolution (Orús-Alcalde et al. 2021).

Vertebrate TLRs have been extensively studied, classified into six large TLR families that usually present at least one orthologous gene in each species (Roach et al. 2005). Functional diversification has been proven for mammal TLRs, which are divided between cell membrane TLRs that recognize different microbial cell components (TLR 1, 2, 4, 5, 6, and 10) and endosome TLRs that recognize nucleic acids (TLR 3, 7, 8, 9, 11, 12, and 13) (Kawai and Akira 2010). The canonical signaling cascade of TLRs is the Myd88-mediated pathway, leading to the nuclear factor-κB (NF-κB) inflammatory response, which is conserved in metazoans (Gauthier et al. 2010; Zhang et al. 2013). However, TLRs can recruit other intracellular partners, being therefore involved in other pathways. For example, the vertebrate TLR3, responsible of recognizing viral double-stranded RNA (dsRNA), stimulates its signaling pathway through the TIR domain-containing (TIR-DC) adaptor protein TCAM1 or TRIF (Kawai and Akira 2010). Nevertheless, while such TIR-DC intracellular adaptors have been extensively studied in vertebrates, their diversity and functional roles remains elusive in invertebrates (Gerdol et al. 2017). Regarding the extracellular domains of vertebrate TLRs, their diversity is related to high functional specialization, with each TLR exhibiting a remarkable binding specificity for different targets, ranging from lipopolysaccharides (LPS), to flagellin and to exogenous nucleic acids (Roach et al. 2005).

Despite sharing a similar domain architecture, the TLRs of invertebrate species, mainly studied in arthropods and echinoderms, are usually highly divergent from those of vertebrates, displaying poor pairwise primary sequence homology (Satake and Sekiguchi 2012; Tassia et al. 2017). The implication of invertebrate deuterostomes TLRs in immune functions derives from the studies carried out in the sea urchin *Strongylocentrotus purpuratus*, which shows an expanded and functionally diversified TLR gene family, mirroring the specialized repertoires of vertebrates (Buckley and Rast 2012). TLRs from the arthropod *D. melanogaster* are usually mostly related to development functions (Narbonne-Reveau et al. 2011), even though Toll has been also associated to the recognition of different pathogens and the subsequent production of immune effectors, such as antimicrobial peptides (Lemaitre et al. 1997; Valanne et al. 2011). The clear importance of TLRs for the innate immune system has been observed as well for other invertebrates, such as cnidarians (Brennan et al. 2017), molluscs (Zhang et al. 2013; Gerdol et al. 2017), and annelids among others (Prochazkova et al. 2019). However, most invertebrate phyla still lack evolutionary studies focused on TLRs, and the study of the immune function of these receptors is often restricted to particular TLR genes, commonly annotated by homology, which may lead to misleading inference due to the presence of the repetitive LRR domains. The lack of a well-defined picture of TLR repertoires for such phyla prevents functional studies from being associated with specific subfamilies in an appropriate evolutionary context.

Despite molluscs being the second largest invertebrate phyla after arthropods, they have been underrepresented so far in TLR evolutionary studies. The Mollusca phylum includes eight largely diversified classes, which display unique morphological, physiological, and evolutionary adaptations, being among the most successful colonizers of both marine and wetland freshwater environments. The species belonging to the class Bivalvia display a conserved intracellular signaling pathway activated downstream of TLRs, showing at the same time a highly diversified repertoire of TLRs and uncharacterized cytosolic TIR-DC proteins (Zhang et al. 2015; Gerdol et al. 2017; Regan et al. 2021). Although various studies described the immune role of a few bivalve TLRs, these considerations can be hardly extended to the full complement of the TLR genes found in a given species (Ren et al. 2017; Wei et al. 2018; Liu et al. 2019; Priyathilaka et al. 2019; Wang et al. 2019; Xu et al. 2019; Chen et al. 2022). As the vast majority of bivalves are essentially marine filter feeders, these animals are constantly exposed to the wide array of waterborne pathogens. The TLRs of these species can be an interesting object of study, since their specialization could be expected to match the great microbial diversity encountered during filtering.

In the current work, we tracked the appearance and the evolution of TLRs throughout the evolution of Metazoa, with a particular focus on bivalve molluscs. The mussel TLR gene family was studied in depth due to the unmatched degree of expansion that these organisms displayed compared with all other species. We tested the hypothesis whether the large repertoire of mussel TLRs was the product of a lineage-specific gene family expansion event, linked with the remarkable resilience of this species toward infection. To this end, we used a comparative genomics approach to evaluate the reliability of an alternative scenario characterized by the widespread convergent occurrence of similar phenomena in Metazoa. Although bivalves generally showed expanded TLR repertoires, with the presence of conserved TLR subfamilies that constituted evolutionary novelties compared with other metazoans, we revealed that mussels represented a unique case in terms of sequence diversity, driven by the complex interplay between negative and positive selective forces, frequently acting on tandemly duplicated gene copies. The analysis of comprehensive gene expression data sets further suggested that this extreme sequence diversity might be matched by an equally extreme functional specialization of TLRs to accommodate the recognition of a very broad array of MAMPs. Functional diversification

might have been the leading evolutionary force, originating a great diversity of TLRs in bivalves, necessary to deal with the wide range of pathogens to which these marine filter-feeding invertebrates are exposed due to their vital strategy.

Materials and Methods

Whole Genome Orthology Analyses Across Metazoans

The proteins encoded in the genomes of 85 species from the main Metazoa phyla were obtained from the GFF annotation of each genome. Accession IDs can be found in supplementary material S1, Supplementary Material online.

Proteins derived from the genome annotation usually contain different protein isoforms for the same genes. Since our purpose was to work at the gene level diversity, we used the AGAT toolkit scripts (Dainat 2019) to filter those proteins in order to only keep the longest isoform of each gene (metrics can be found in supplementary material S1, Supplementary Material online). These one-protein-per-gene files from each genome were submitted to BUSCO analyses with the metazoa_odb10 database (Manni et al. 2021) to check the completeness and quality of each genome (metrics can be found in supplementary material S1, Supplementary Material online).

In order to construct the phylogenetic tree for the species included in this study, additional BUSCO analyses were performed for each genome using the eukaryota_odb10 database. The Python BUSCO_phylogenomics script (McGowan et al. 2020; McGowan and Fitzpatrick 2020) and ASTRAL software (Zhang et al. 2018) were used to create one phylogeny for each of the complete BUSCO genes and to build a consensus species phylogenetic tree.

Finally, the filtered one-protein-per-gene files were submitted to an orthology analysis using Orthofinder (Emms and Kelly 2015; Emms and Kelly 2019), and the results were parsed using BioNERO (Almeida-Silva and Venancio 2022). Two orthology analyses were conducted: The first one included all the analyzed metazoan species, whereas the second one only included molluscan species.

Retrieving TLRs From the Genomes

The filtered protein files generated in the previous section were searched with HMMER (Eddy 2011) to detect LRR and TIR domains with a minimum *e*-value of 0.001. All HMM profiles related to LRR (PF15176.9, PF15779.8, PF01463.27, PF01462.21, PF00560.36, PF18805.4, PF18837.4, PF07723.16, PF07725.15, PF12799.10, PF13306.9, PF13516.9, PF13855.9, PF14580.9, PF08191.14, PF16920.8, PF08263.15, and PF18831.4) and TIR domains (PF01582.23, PF10137.12, PF13676.9, and PF18567.4) were employed. PfamScan was used to detect and remove proteins with any other additional domain (based on a 10e−3 *P*-value cutoff), as these were deemed to be unlikely to represent bona fide TLRs. Phobius (Käll et al. 2004) was used to predict TM regions. Canonical TLRs were retrieved from all the analyzed genomes (proteins with extracellular LRR domains, TM region, and intracellular TIR domain). Finally, the orthology groups containing the detected TLR proteins were identified.

The limitation of this detection method could depend on the incorrect annotation of a few real TLR genes, which may therefore miss part of the predicted amino acid sequence at the N-terminus or at the C-terminus, although the chances of this occurrence were strongly reduced by the use of stringent filtering criteria (i.e., all TLRs had to display LLRs, a TM region, and the TIR domain). Some real TLRs could have been excluded from the analysis if they presented slight truncations in the genome annotation, resulting in non-detection of the required domains. To avoid systematic biases linked with low-quality genome assemblies and gene annotations, we only considered genomes with BUSCO completeness rates (metazoan database) of ∼90% for most of them and always above 70% for species such as helminths, known to have undergone massive gene loss.

De Novo Prediction of the TLR Family From Mussel Genomes

To allow a reliable comparison between complete TLR proteins encoded by the genomes of *Mytilus edulis* (GCA_019925275.1), *Mytilus coruscus* (GCA_017311375.1), and *Mytilus galloprovincialis* (GCA_900618805.1), we performed a *de novo* prediction of all TLRs genes using the same strategy. Since mussels TLRs are encoded by a single exon, emboss sixpack (Madeira et al. 2022) was used to retrieve every ORF larger than 300 codons starting with an ATG methionine-encoding codon. Then, the identification of TLRs was performed following the same methods described above for the analysis of the other metazoan genomes, but adding in this case the presence of a signal peptide detected with Phobius (Käll et al. 2004) and signalP (Teufel et al. 2022) as additional criteria since we wanted to exclude gene models unlikely to code for complete, functional TLRs. The signal peptide criterion was not used in the analysis of metazoan genomes to avoid being too strict since we were depending on the existing annotations in each genome. SMART was used to get a complete description of the TLRs domain structure for their classification (Letunic and Bork 2018; Letunic et al. 2021). The correspondence between the chromosomes of *M. edulis* and *M. coruscus* was obtained after performing whole-genome alignment with MashMap (Jain et al. 2018). *Mytilus coruscus* chromosomes were renamed and reordered according to their equivalent in *M. edulis*, and TLRs were placed in those chromosomes using CIRCOS (Krzywinski et al. 2009).

Phylogenetic and Selection Analyses

Mussel TLR phylogenetic analyses were constructed with the *de novo* predicted repertoires in the *Mytilus* genomes. Two phylogenetic analyses were performed, one with whole

TLRs and another one with only the sequence of the TIR domains, aiming for a more polished multiple sequence alignment due to the high sequence divergence of the ectodomain and the presence of a variable number of LRRs (Gerdol et al. 2017). Protein sequence alignments were performed using MAFFT (Katoh et al. 2019) and phylogenetic analysis were made with PhyML (Guindon et al. 2010), using automatic evolutionary model selection (Lefort et al. 2017) that chose JTT + G + F as the most appropriate evolutionary model for the data set (Jones et al. 1992). Another phylogenetic analysis was constructed with TLRs detected in the genome annotation of selected metazoan species from different taxa. All TLRs with LRR and TIR domains detected in these species were used, and TLR-like sequences from the poriferan *Amphimedon queenslandica*, which present extracellular immunoglobulin and intracellular TIR domains, were included as outgroups. This approach was chosen because of the need for an unambiguous outgroup to root metazoan TLRs (Orús-Alcalde et al. 2021), which could not be achieved by using TIR domain sequence only. Phylogenetic trees were annotated using iTOL (Letunic and Bork 2021).

For selection analyses, codon-aware alignments were constructed with the coding nucleotide sequences (CDS) of TLRs. Alignments were cleared using a gap tolerance of 50%, with tolerance for 1 position in the codon and deleting gaps in triplets to keep the coding information. The gap-filtered alignments were then analyzed in Datamonkey (Weaver et al. 2018) in order to detect episodic diversifying selection with MEME (Murrell et al. 2012) and negative selection with FUBAR (Murrell et al. 2013). Using the translated protein alignment, protein variability per position was calculated using the Wu–Kabat variability coefficient.

Expression Analysis

An expression analysis was constructed by mapping 275 *M. galloprovincialis* transcriptomic runs/samples (126 Illumina single-end and 149 Illumina paired-end) from 22 different Bioprojects to the "mg3" *M. galloprovincialis* reference (Gerdol et al. 2020). The IDs of all the SRA data employed can be found in supplementary material S2, Supplementary Material online (all SRA Illumina transcriptomic data that presented a minimum number of 2.5 million spots per run was retrieved). Additionally, 12 samples were added from unpublished transcriptomic data obtained from mussels under viral stimulation. These samples consisted in stimulations with spring viraemia of carp virus (SVCV), infectious pancreatic necrosis virus (IPNV), red-spotted grouper nervous necrosis virus (RGNNV), and polyinosinic:polycytidylic acid (Poly I:C) and were obtained following the same experimental design as with a viral hemorrhagic septicemia virus (VHSV) stimulation (Saco et al. 2023).

The choice of the reference sequence data set used for the RNA-seq analysis was based on preliminary investigations, which indicated that the mapping rate increased by 10% when using the "mg3" mussel genome assembly instead of the "mg10" (GCA_900618805.1) reference version (Gerdol et al. 2020). This rate was further improved by another 10% when using transcripts (i.e., including UTRs) instead of CDS. This version of the assembly is less filtered in the hemizygous genomic fraction than mg10, presenting 10% more annotated genes (Gerdol et al. 2020). According to the pangenomic nature of *M. galloprovincialis*, a gene presence/absence variation (PAV) analysis was performed for this assembly following the same pipeline described previously (Gerdol et al. 2020; Sollitto et al. 2022), allowing the detection of core transcripts that were used as the mapping reference.

All the mussel transcriptomic data was analyzed using Snakemake pipelines (Mölder et al. 2021) that included 1) trimming with fastp of both single-end and paired-end reads (Chen et al. 2018) using a threshold value of phred = 20; 2) mapping clean reads to the core reference transcripts using salmon (Patro et al. 2017) with authomatic library-type selection and default mapping parameters; 3) computing trimming and mapping statistics with MultiQC (Ewels et al. 2016); 4) generating mapping saturation curves using a python script (available in Figshare repository). Samples with a mapping rate smaller than 30% or displaying highly saturated mapping curves (i.e., low transcriptional complexity) were discarded, leading to a final data set of 252 mapped samples (supplementary material S2, Supplementary Material online).

Transcriptomic Correlation Networks and Transcriptomic Modulation

Due to the high expression differences observed among tissues (supplementary material S3, Supplementary Material online), correlation analyses were conducted independently using samples from each tissue. For gills and hemocytes, highly expressed TLRs that reached a value of 10 transcripts per million (TPM) in at least one sample were selected as nodes for constructing expression correlation networks. Pearson correlations higher than 0.7 were kept, with the exception of hemocytes, where a higher threshold (i.e., 0.8) was used due to the better correlation found in its net. Interaction nets were generated in Cytoscape (Shannon et al. 2003) using the correlation data.

Among all the transcriptomic data analyzed, samples were retrieved if they belonged to Bioprojects where mussels were treated with immune or nonself-agent stimuli (supplementary material S2, Supplementary Material online) and those transcriptomes were analyzed independently. Salmon mapping output counts were input into R using tximport (Soneson et al. 2016), using a metadata table with control and stimulated samples for each transcriptomic experiment, as shown in supplementary material S2, Supplementary Material online. DESeq2 (Love et al. 2014) was used to calculate modulated transcripts in each experiment (pipeline available in the Figshare repository).

Results

TLR Presence Across Metazoa

The set of 85 species analyzed to study the distribution of TLRs in Metazoa was enriched in molluscan species, since this phylum had been underrepresented in previous studies. The genomic annotation of the selected genomes showed satisfactory quality, as evidenced by BUSCO completeness ~90% for most species. Genomes falling under this metric presented always a completeness fraction higher than 70% and belonged to neglected phyla underrepresented in OrthoDB10 or to parasites (e.g., helminths), where massive lineage-specific gene losses were known to be present (supplementary material S1, Supplementary Material online). The number of TLR genes found in each genome (genes encoding proteins with extracellular LRR domain, TM region, and intracellular TIR domain) is shown in figure 1. The earliest branching extant metazoan taxa in which TLRs could be found were anthozoan cnidarians (TLRs were absent in other Cnidaria classes). TLRs could be identified in almost all phyla (except Bryozoa and Platyhelminthes), and the repertoires varied greatly, from a single gene to expanded families comprising hundreds of members. Significant expansions of the TLR gene family could be observed in echinoderms from the Echinoidea class (mainly in *S. purpuratus*) and in the centipede arthropod *Rhysida immarginata*, which showed a species-specific expansion unique among the generally small repertoires of Arthropoda. The most general TLR expansion could be identified in molluscs, especially in Bivalvia, while other classes as Cephalopods showed smaller repertoires. The magnitude of the expanded gene family found in the three analyzed *Mytilus* species was unmatched by any other species, and therefore these species were subjected to in-depth analyses.

A Metazoa-level orthology analysis was conducted using the 85 metazoan species. Afterwards, the orthology groups that contained TLRs shared by the three *Mytilus* mussel species were identified and displayed in figure 2. The orthogroup 1 included orthologs from 73 out of the 74 species in which TLRs were identified (the exception were the TLRs of *Brachionus calyciflorus*, placed in a species-specific orthogroup). The orthogroup 1 contained the whole TLR repertoire of several species and 100% of Chordata TLRs. While all the TLRs from Chordata were sccTLRs, orthogroup 1 included both mccTLRs and sccTLRs from other phyla. The expansions in the TLR repertoires of the sea urchin *S. purpuratus* (135 genes) and of the centipede *R. immarginata* (56 genes) were explained by 80% and 60%, respectively, by TLRs belonging to orthogroup 1 (the rest of those expansions were placed on species-specific orthogroups).

In addition, several mollusc- or bivalve-specific orthogroups were obtained, evidencing the large amount of diversity present in these species and the evolutionary novelty of most of their TLRs. Indeed, only 50% of the TLR genes found in the three mussel species were included in the orthogroup 1. The other half was distributed between orthogroups 2 and 3 (mainly Mollusca-specific mccTLRs; 12% of mussel TLR genes), orthogroups 4–6 (mainly Bivalvia-specific sccTLRs; 26% of mussel TLR genes), and orthogroups 7–12 (*Mytilus*-specific sccTLRs; 5% of mussel TLR genes).

The Expansion in Bivalves and Mussels

In order to delve into the expanded and diverse repertoire of mussels, a phylogenetic analysis was conducted with the *de novo* TLR repertoires of *M. galloprovincialis*, *M. coruscus*, and *M. edulis*. The obtained clustering based on the intracellular TIR domain (fig. 3A) was concordant with the extracellular organization of the proteins (fig. 3B) and with the analysis using whole proteins (supplementary material S4, Supplementary Material online). All mccTLRs were organized in two clusters: one conformed by single TIR mccTLRs (group SPP) and another one conformed by mccTLRs that presented two TIR domains (SPP-2TIR). The two TIR domains encoded in the SPP-2TIR TLR proteins were included separately in the analysis, and they clustered together in the same SPP-2TIR cluster, suggesting an origin by duplication of the TIR domain present in an ancestral gene with a canonical domain architecture. Additionally, several clusters of sccTLRs were obtained. These were arbitrarily named as V-type (if proteins were larger than 800 aa) and as SP-type (if proteins were shorter than 800 aa). The nomenclature system for the obtained clusters was based on the extracellular domain organization and was maintained from the last published work on bivalve TLRs (Gerdol et al. 2017). The X group (formerly known as Ls type) contained TLRs with shorter extracellular regions and lower number of predicted canonical LRR repeats interspersed by low complexity sequence. The Y cluster contained one sccTLR from each species with a TIR domain highly divergent from all the other TLRs. Figure 4A shows the phylogenetic analysis of all the Mytilidae TLRs in combination with the metazoan orthology data derived from figure 2. The TLRs identified in the genome annotation of the three *Mytilus* species were included, together with those of *Perna viridis* and *Gigantidas platifrons*. Most mussel sccTLR clusters (SP1, SP2, V1, V4, V5, and X) were included in the large metazoan orthogroup 1, which was shared by most metazoans. The SPP-2TIR subfamily (orthogroup 2) was conserved at the mollusc level, similarly to the SPP cluster (orthogroup 3) with the difference that this group was missing in cephalopods. The V3 sccTLR subfamily (orthogroups 5 and 6) shared orthology only at the Bivalvia level. Finally, there were orthogroups restricted to mussels, corresponding to divergent branches in some of the phylogenetic clusters, which most likely emerged and diversified after the split between *Mytilus* and other Mytilidae genera. Correspondence in terms of sequence classification was observed between the metazoan-computed orthogroups and the subfamilies of mussel TLRs identified by phylogenetic inference.

A higher-resolution phylum-level orthology analysis (Roach et al. 2005; Messier-Solek et al. 2010; Buckley and

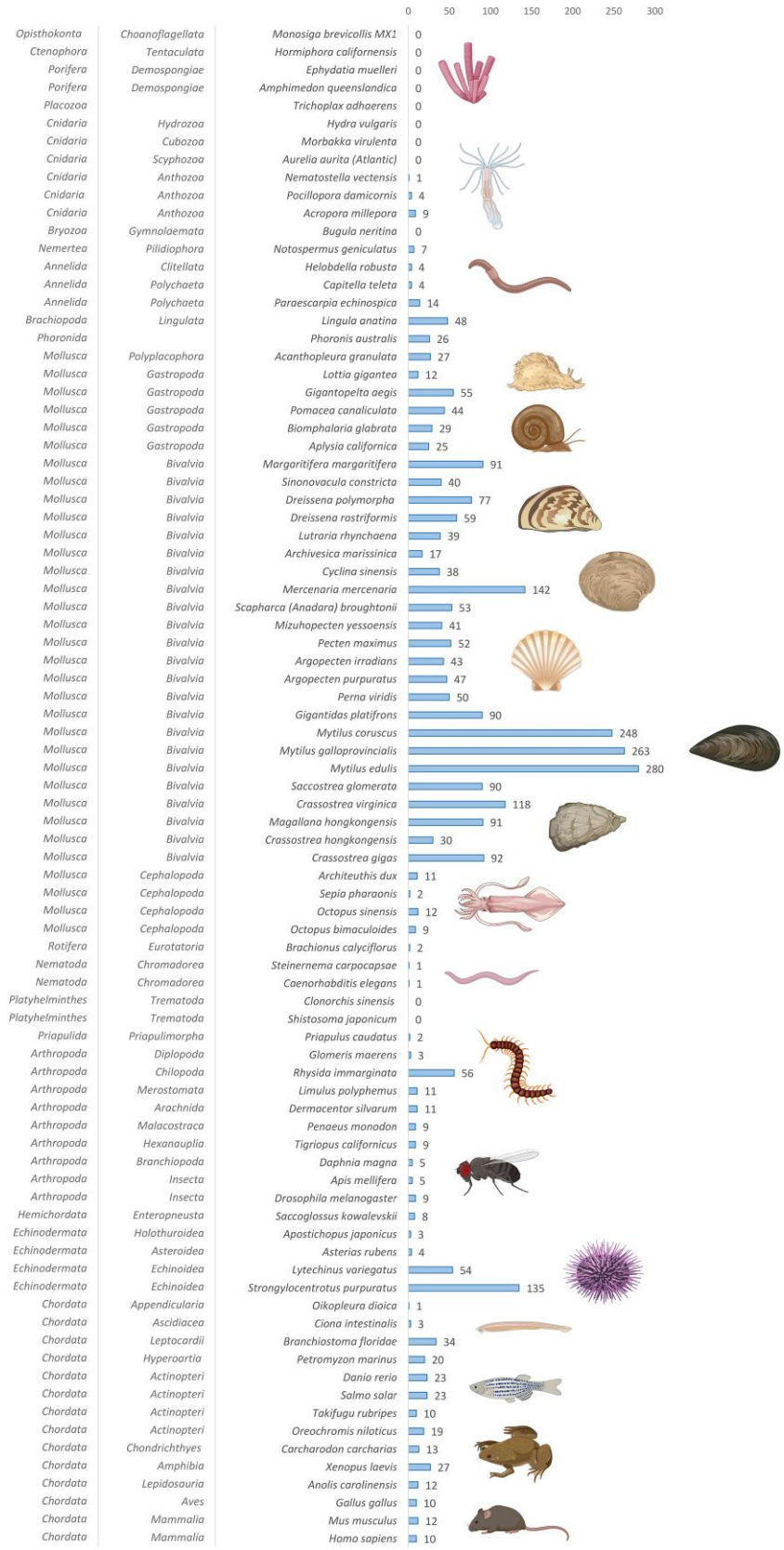

Fig. 1. TLR gene family size across Metazoa. The number of TLR genes identified in each analyzed species is shown. A general expansion can be observed in Mollusca, especially in bivalves, reaching its peak in mussel species.

Rast 2012) was performed with all molluscan species. With this analysis, almost each TLR subfamily supported by phylogenetic inference matched perfectly with a different orthology group (fig. 4*B*). Such orthogroups displayed a variable degree of taxonomical conservation, being invariably present in all the species belonging to a given molluscan class or being specific to certain classes (fig. 4*C*). The number of TLRs belonging to each orthogroup varied in

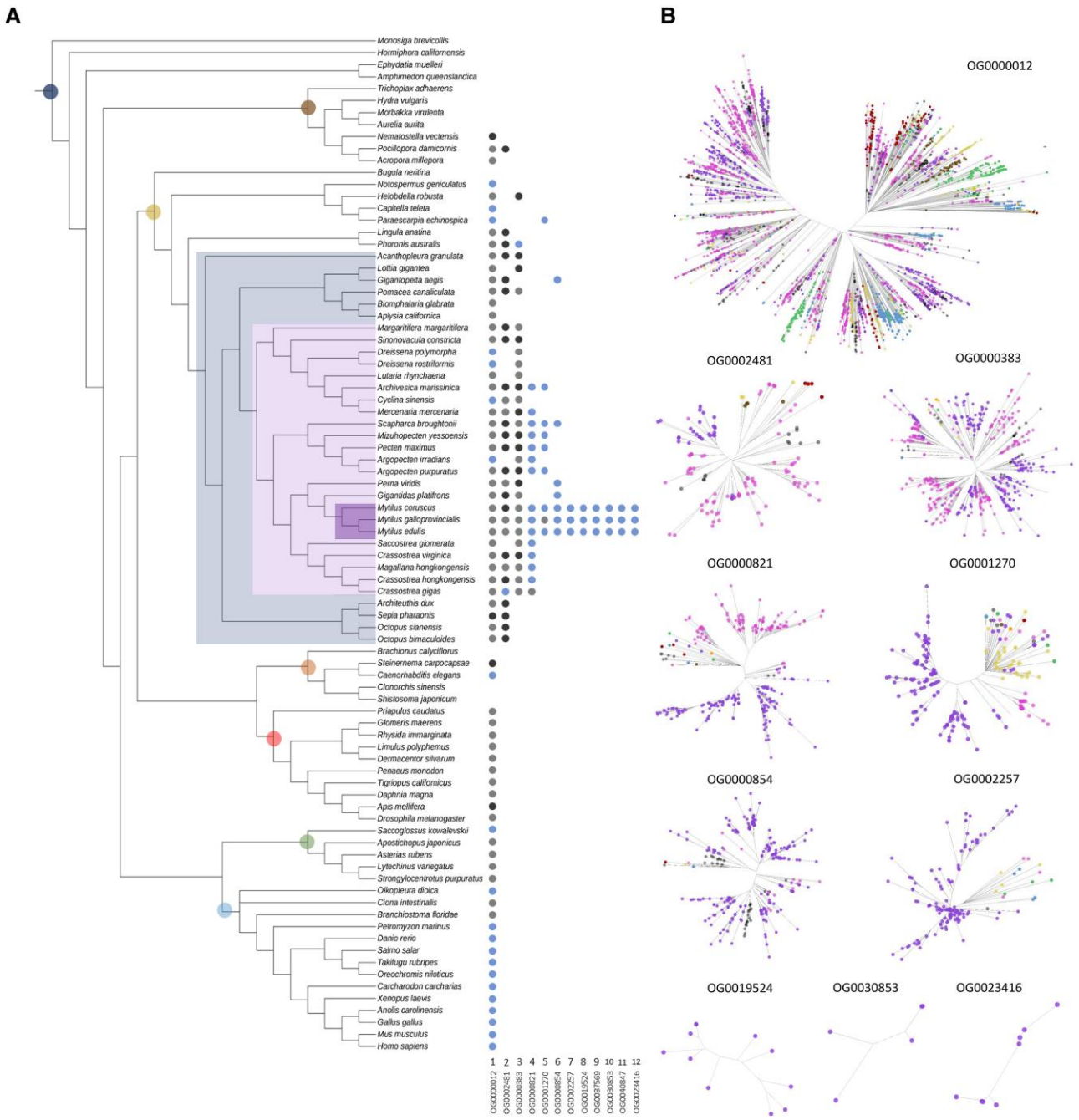

Fig. 2. Orthology analysis of mussel TLRs across the analyzed metazoan genomes. (*A*) All the orthology groups that include TLRs identified in the three *Mytilus* species are included, along with their taxonomic distribution. The structural organization of the TLRs included in each orthogroup is shown: Black dots indicate orthogroups only including mccTLRs, blue dots indicate those only including sccTLRs, and gray dots indicate those including both types. (*B*) Complete Orthofinder-resolved gene trees for each orthology group are shown, with species indicated according to the color legend reported in the branches of the panel *A* tree.

the different molluscan classes. Specific TLR subfamilies were more expanded in mussels in comparison with the sizes of the same subfamilies in other molluscs, while other subfamilies remained more similar in terms of gene number (fig. 4*D*). The conservation within molluscs of most sccTLRs (SP1, SP2, V4, V5, and X) and of SPP-2TIR subfamilies was confirmed by the phylum-level orthology analysis. SPP and V2 TLRs were found in all molluscs with the exception of cephalopods, which exhibited the smallest and least variable TLR repertoires. The Y and the V3 clusters were only found in bivalves, likely representing recent evolutionary acquisitions. Additional divergent orthogroups that were not shared above the family or genus level were found both in mussels and in other bivalves (supplementary material S5, Supplementary Material online), being often placed in divergent branches of particular subfamilies. The phylum-level orthology results were in line with what emerged from the metazoan-level analysis and

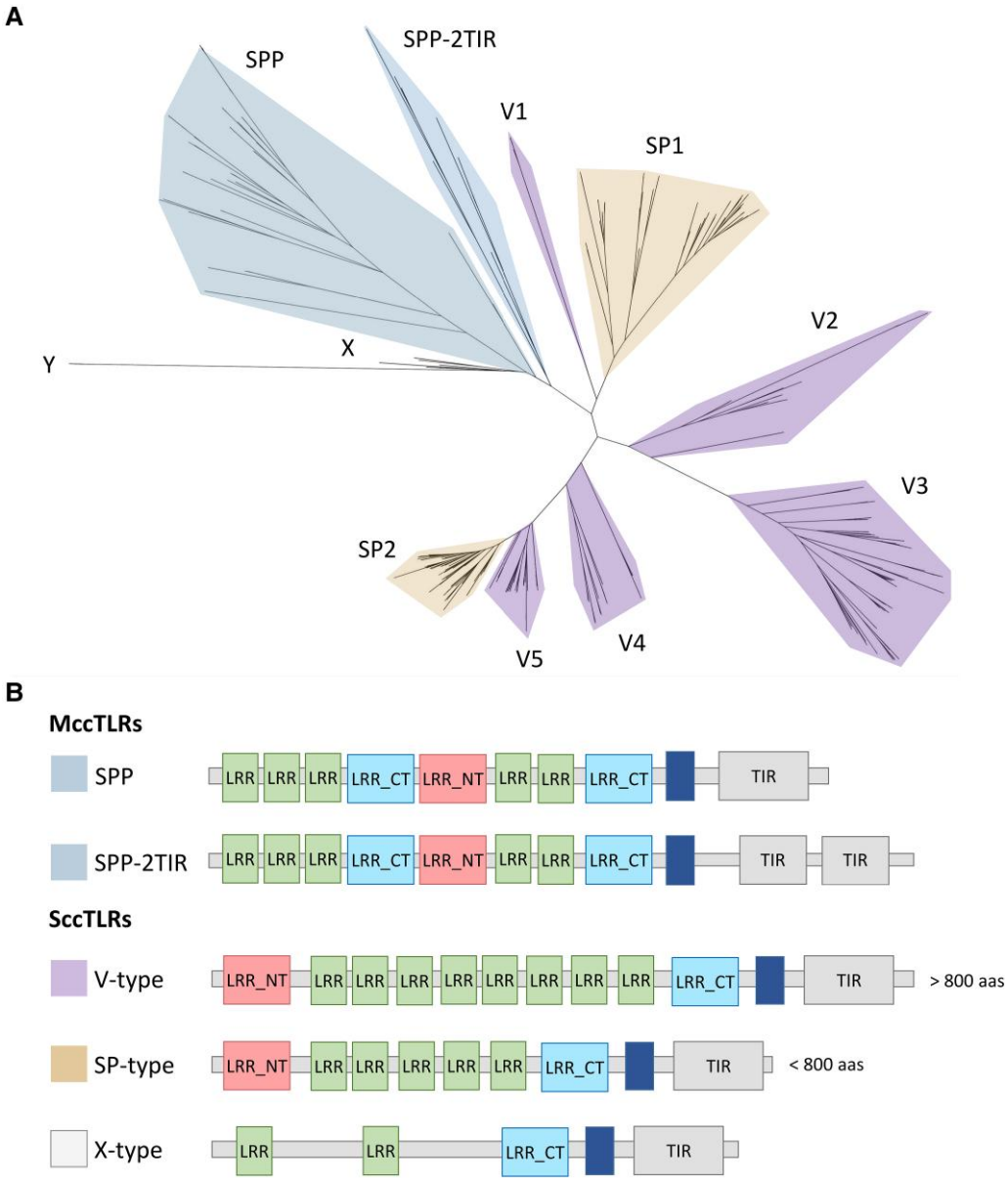

Fig. 3. Phylogenetic analysis of mussel TLRs. (*A*) Phylogenetic inference was based on the multiple sequence alignment of the TIR domains from *M. edulis*, *M. coruscus*, and *M. galloprovincialis de novo predicted* TLRs. The sequence clusters obtained match the extracellular organization of TLR ectodomains and therefore were named following previously suggested nomenclature (Gerdol et al. 2017). mccTLRs were named SPP or SPP-2TIR, depending on the presence of one or two TIR domains; sccTLRs clusters were named SP (protein size < 800 aas) or V (protein size > 800 aas); the X cluster contained TLRs with particularly short extracellular regions, and the Y cluster contained one orthologous sscTLR from each mussel species with a highly divergent TIR domain. (*B*) Graphical scheme of the domain structure of the different TLR subfamilies.

confirmed that the expansion occurred in common mollusc TLR subfamilies that present different degrees of conservation, divergence, and duplication among the different molluscan classes and species.

The Diversity of the Expansion in the Metazoan Context

The whole TLR repertoires from selected metazoan species representing the analyzed taxa were used to build the phylogenetic tree shown in figure 5. Poriferan TLR-like proteins (receptors with the intracellular TIR domain but lacking extracellular LRRs) were used as an outgroup, as they may represent ancestral TLR progenitors (Brennan and Gilmore 2018; Nie et al. 2018). As expected, phylogenetic inference clearly identified three major Metazoan TLR clades, α, β, and γ, named as previously described (Orús-Alcalde et al. 2021). Interestingly, bivalve expanded clusters were placed in all these three clades. Namely, *Mytilus* mccTLRs (SPP and SPP-2TIR) belonged to clade α along with all the mccTLRs from the other species. However, this clade is not mcc specific, since it included all cnidarian TLRs, regardless of their variable ectodomain organization. *Mytilus* sccTLRs clusters were distributed in the scc-specific clades β (subfamilies SP1 and V1) and γ

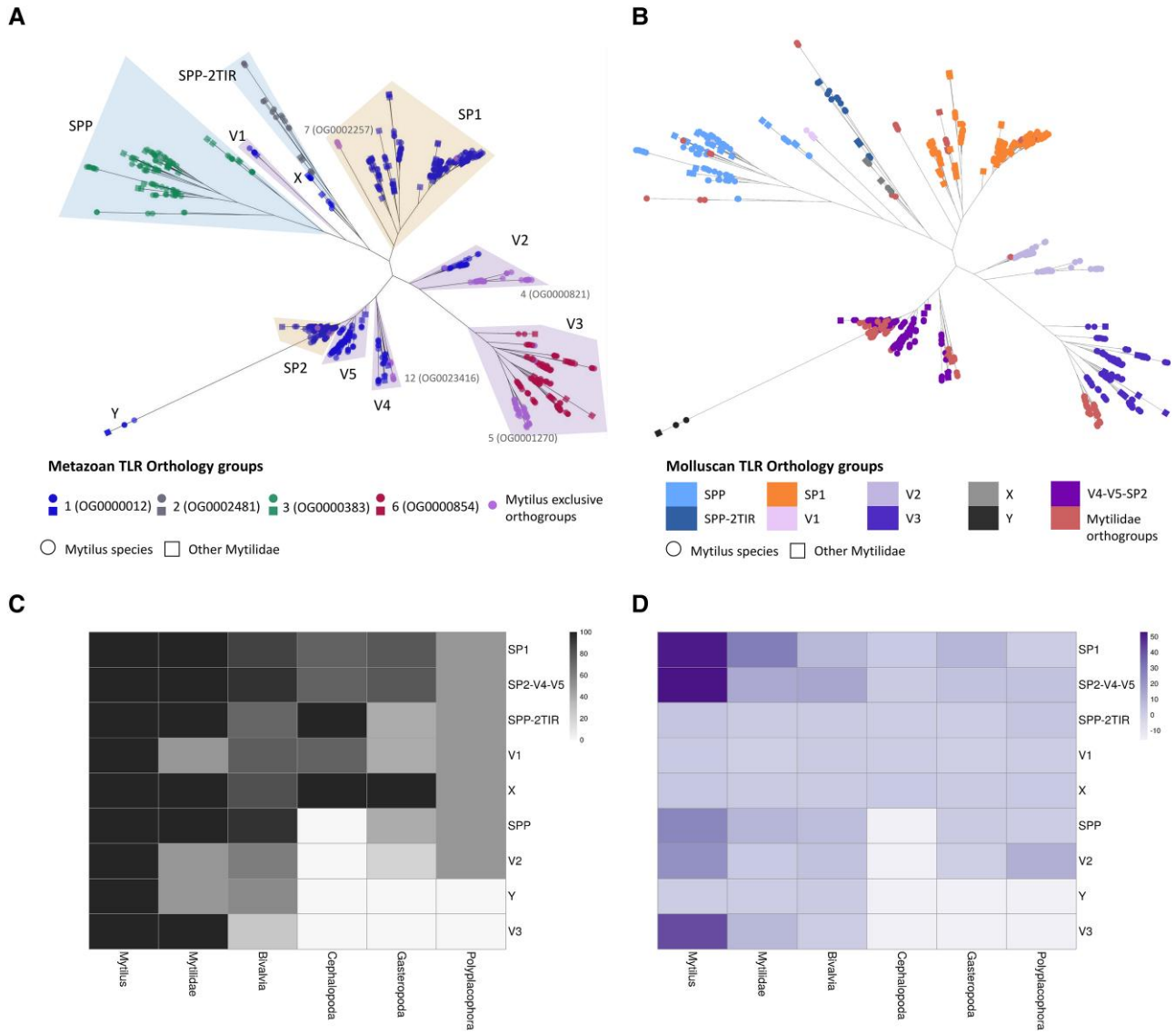

Fig. 4. Orthology analyses of mussel TLRs. (*A*) Phylogenetic tree constructed based on the multiple sequence alignment of the TIR domains of Mytilidae species (*M. galloprovincialis*, *M. edulis*, *M. coruscus*, *P. viridis*, and *G. platifrons*), combined with the Metazoa-level orthology information derived from figure 2. Tip colors indicate the orthogroup assignment of each sequence, revealing a good agreement between orthogroups and the phylogenetic relatedness of sequences. Tip shapes also differentiate the TLR sequences from *Mytilus* and other Mytilidae species. The taxonomical spread of each orthogroup within Metazoa is shown in figure 2. (*B*) The same Mytilidae TLR phylogenetic tree was annotated with the orthology groups obtained from a Mollusca-level orthology analysis. The main orthogroups matched perfectly the previously identified phylogenetic clusters. Each TLR orthogroup was indicated in a different color and named after the TLR subfamily that it includes. (*C*) Degree of conservation of TLR subfamilies–orthogroups within different molluscan classes derived from the Mollusca orthology analysis. The heat map displays the fraction of species belonging to each class (percentages shown in the legend) that presented at least one orthologous TLR for each subfamily–orthogroup. (*D*) Heat map showing the average number of proteins classified with each subfamily–orthogroup for each molluscan class.

(subfamilies V2, V3, V4, V5, and SP2). The great variability in the clade distribution of the *Mytilus* expansion clashes with the situation observed for the other TLR expansion events found in nonbivalve species, such as the sea urchin *S. purpuratus*. Although this sea urchin presents sequences belonging to each of the three clades, the expansion exclusively concerned a group of sequences belonging to clade β. Consistently with this observation, clade β was the most abundantly represented in deuterostomes, and all vertebrate TLRs were attributed to this clade.

Evolutionary Dynamics Underpinning Mussel-Specific TLR Expansions

In order to study the evolutionary dynamics underpinning the *Mytilus* TLR expansion, the occurrence of gene tandem duplications, the impact of gene presence-absence variation, and the presence of signatures of positive and negative selection were investigated.

Selection analyses were separately performed in different *Mytilus* TLR subfamilies, under the assumptions that 1) the current repertoire of mussel TLRs may result from

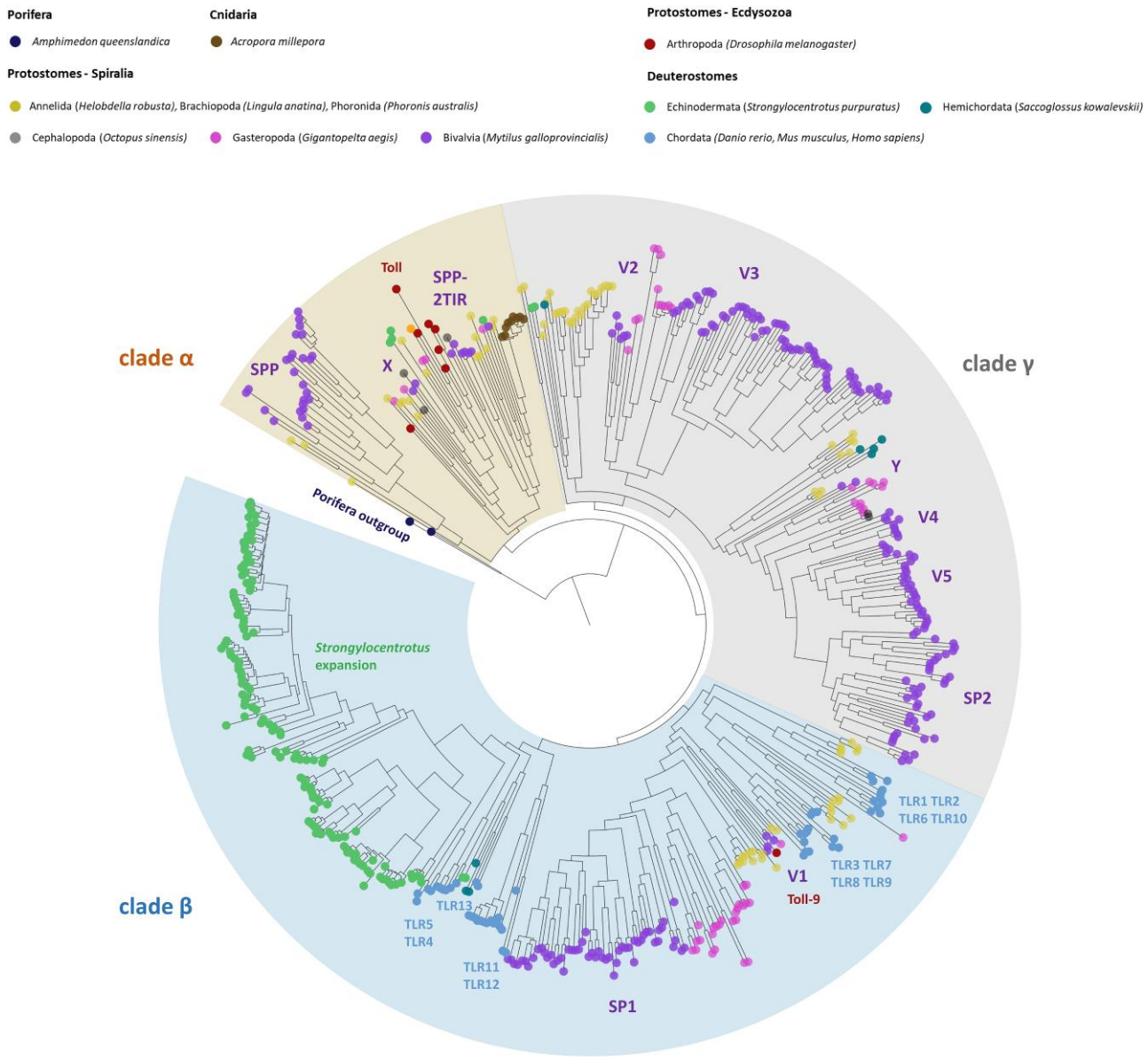

Fig. 5. Metazoan TLR clade distribution. Phylogenetic tree displaying the evolutionary relationships among all TLRs from selected representative species of the different Metazoa phyla analyzed in the current work. Poriferan TLR-like sequences were added as an outgroup for tree-rooting purposes. Three major metazoan TLR clades can be seen, with the first split in the branches corresponding to the differentiation between clade α and clade β/γ and a second split separating the β and γ sister clades. Tip colors indicate different species, evidencing the different distribution of the TLRs of each phyla across the three clades. Mussel TLR subfamilies are also indicated in the tree, as well as selected examples like the sea urchin expansion, *Drosophila* Toll proteins, and the mammalian TLR repertoire.

multiple independent expansion events and 2) the different mussel TLR subfamilies may have been shaped by different selective forces, depending on their ligands and functional specialization (fig. 6A). In general, signatures of pervasive negative selection were always stronger in the intracellular TIR domain, involved in signal transduction, compared with the extracellular LRR domains, involved in ligand binding. The opposite trend was observed for the signatures of episodic positive selection, which were stronger in the extracellular domains (fig. 6B). Consistently with this evidence of diversifying selection, a higher protein sequence variability was found in the extracellular regions. Selection analyses were also performed with the whole TLR repertoires from selected species representative of major metazoan taxa (fig. 6C). Episodic positive selection and protein variability were higher in *M. galloprovincialis* in comparison with *M. edulis* and *M. coruscus*, and the protein variability found in molluscs was usually higher than most other taxa (including *S. purpuratus*, which, on the other hand, showed the largest rate of episodic positive selection).

The chromosomic assembly of *M. edulis* and *M. coruscus* allowed detecting tandem duplications contributing to the TLR expansion in these species. The genomic location of TLRs was retrieved from the chromosomes of *M. edulis* (fig. 7A) and *M. coruscus* (fig. 7B). TLR synteny was not totally

A

	de novo seqs	lenght	pervasive negative selection			episodic positive selection			Extracellular variability	TIR domain variability
			FUBAR	% Extracellular	% TIR	MEME	% Extracellular	% TIR		
SP1	112	713	536	65.9	99.3	113	20.9	1.4	41.25	14.33
SP2	106	707	528	67.6	96.5	107	17.2	9.1	44.13	13.36
SPP	81	733	515	69	83.3	72	11.6	5.5	32.68	27.45
SPP-2TIR	7	1128	368	32.5	38.8 - 36	16	1.6	0.8 - 0.7	8.45	9.1 - 7.54
V1	8	872	264	29.3	37.6	16	1.1	3.3	1.62	1.59
V2	30	827	468	51.7	74.8	49	7	2	13.97	9.79
V3	119	841	662	74.5	93	107	13.9	7.8	36.97	25.28
V4	24	894	638	69.6	79	35	4	4.2	13.39	10.23
V5	52	848	563	60.4	90.9	58	8.1	2.8	24.94	8.85

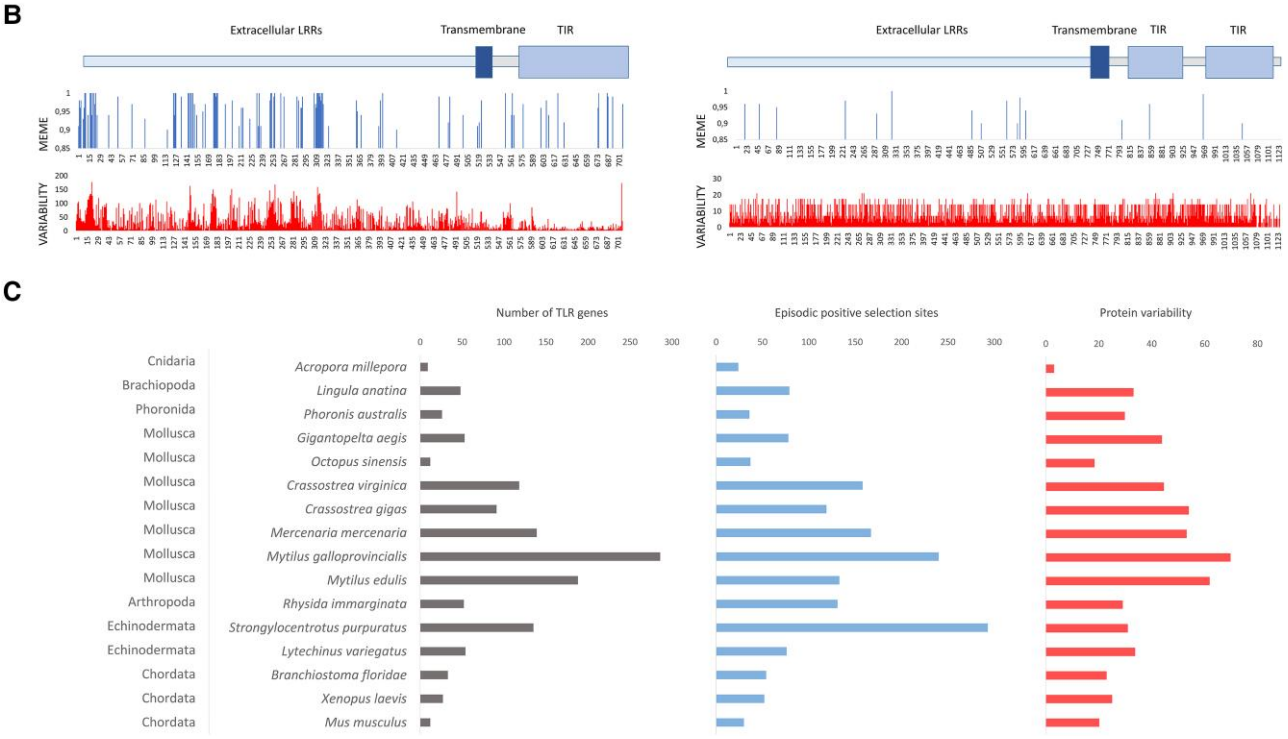

Fig. 6. TLR selection analyses. (*A*) Pervasive negative selection (FUBAR), episodic positive selection (MEME), and protein variability were calculated for each *Mytilus* TLR subfamily. Results were also calculated separately in the intracellular and extracellular regions, revealing a higher diversifying pressure and protein variability in the extracellular region. (*B*) Distribution of episodic positive selection sites, along with protein variability metrics per position for two representative *Mytilus* TLR clusters (SP2 and SPP-2TIR). (*C*) Selection and variability metrics calculated using the complete repertoire of TLR proteins from selected species, representing the diversity of analyzed phyla. Although these values correlate with the size of each repertoire, differences can be seen between the expanded repertoires, since not all of them have the same level of variability.

conserved between the two mussels. Some TLRs conserved the same position (as in the clear example of cluster Y, with just one gene per species in the same location in chromosome 5). The examples reported in figure 7*C* and *D* show that, for two pairs of homologous chromosomes, the position of some genes is clearly not conserved, while it is highly conserved in both species for others. However, even the orthologous genes found in conserved positions are often present with a different number of paralogous gene copies, derived from species-specific tandem duplications of TLRs belonging the same subfamily.

In order to study how does the mussel open pangenome affect the TLR gene family, the gene PAV of TLR genes was studied in *M. galloprovincialis*. The TLR genes detected in the reference genome were analyzed in all the resequenced genomes presently available in order to detect which TLRs were always present and which ones were absent from the genomes of some individuals (fig. 8). Thirty percent of the mussel TLRs were subjected to PAV, in line with the overall dispensable gene content of the genome. Therefore, the TLR gene family was neither enriched nor depleted in the dispensable fraction of the genome, even though SP2 TLRs seemed to be more affected by PAV than other subfamilies. Gene expression data were retrieved for different *M. galloprovincialis* tissues from the SRA data set (fig. 8). Most PAV-affected TLRs did not reach the arbitrarily set gene expression threshold of 5 TPM, especially when their dispensability (rate of genomes in which a particular gene was absent) was higher. This was expected, since the more dispensable a gene was among the analyzed genomes, the less likely it was to be present or expressed among the different transcriptomic samples. The number of TLRs that reached the gene expression threshold in each mussel tissue was 38 in hemocytes, 49 in gills, 42 in mantle, 26 in digestive gland, 10 in muscle, and 1 in mid trochophore larvae. The highest expression levels were retrieved from hemocytes and gills. Some TLRs were expressed in all tissues, while others did only

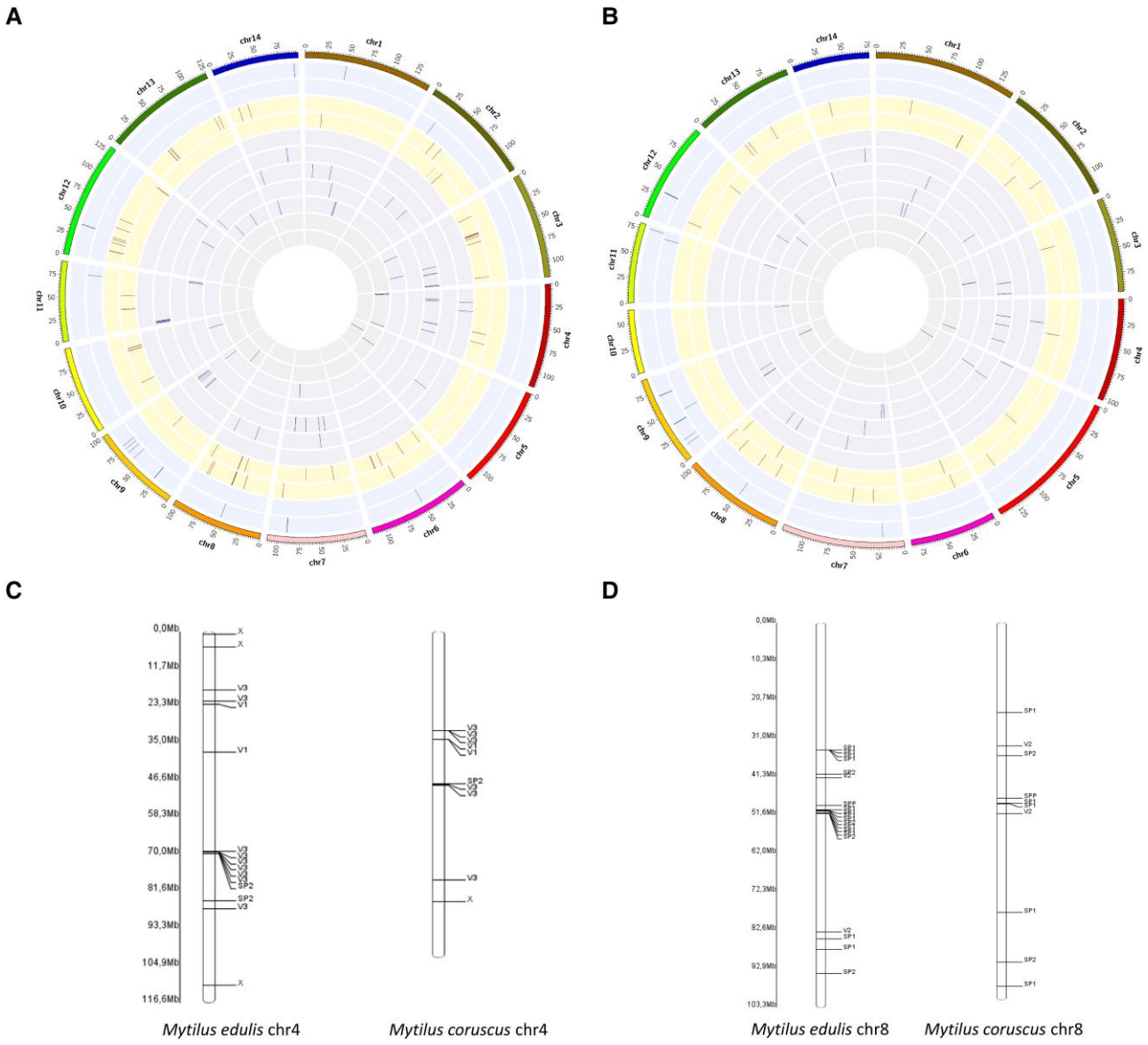

Fig. 7. Chromosomic distribution of *M. coruscus* and *M. edulis* TLRs. *Mytilus coruscus* chromosomes were renamed and reordered after their equivalent *M. edulis* chromosomes. (*A*) Distribution of TLRs divided by subfamily in *M. edulis* chromosomes. (*B*) Distribution of TLRs divided by subfamily in *M. coruscus* chromosomes. For panels *A* and *B*, TLR subfamilies in external–internal order are as follows: SPP, SPP-2TIR, SP1, SP2, V1, V2, V3, V4, V5, X, and Y. Two selected comparisons between the homologous chromosomes of the two mussel species are shown for chromosomes 4 (*C*) and 8 (*D*), where different tandem duplications of TLRs, mainly belonging to the same subfamily, occurred in both genomes independently.

reach the expression threshold in samples from a single tissue, indicating possible tissue specificity.

Expression Analyses and Functional Diversification

TLRs expressed in gills or hemocytes were selected as nodes to construct transcriptomic correlation networks using the transcriptomic expression database. Transcripts with annotations consistent with the immunological context of TLR-mediated signaling were correlated with the node TLRs in gills (fig. 9*A*), including canonical Myd88 proteins as well as other functionally uncharacterized TIR-containing proteins that may act as cytoplasmic interactors of TLRs. Several TLR nodes (1–8) were correlated with NF-κB-related transcripts, especially with the inhibitor of NF-κB (IκB) kinase (IKK), which releases the active form of NF-κB from its cytoplasmic repressor. These TLRs, NF-κB-related transcripts, and other likely immune interactors were interconnected in a clear correlation network displayed in figure 9*B*. Most of the TLRs which expression was correlated with NF-κB were SPP type and included 7 out of the 11 upregulated TLRs in mussel gills stimulated in vivo with a waterborne bacterial infection with *Vibrio splendidus* (fig. 9*C*). Other TLRs conformed smaller expression clusters and were expressed both in the *Vibrio splendidus* infection transcriptome

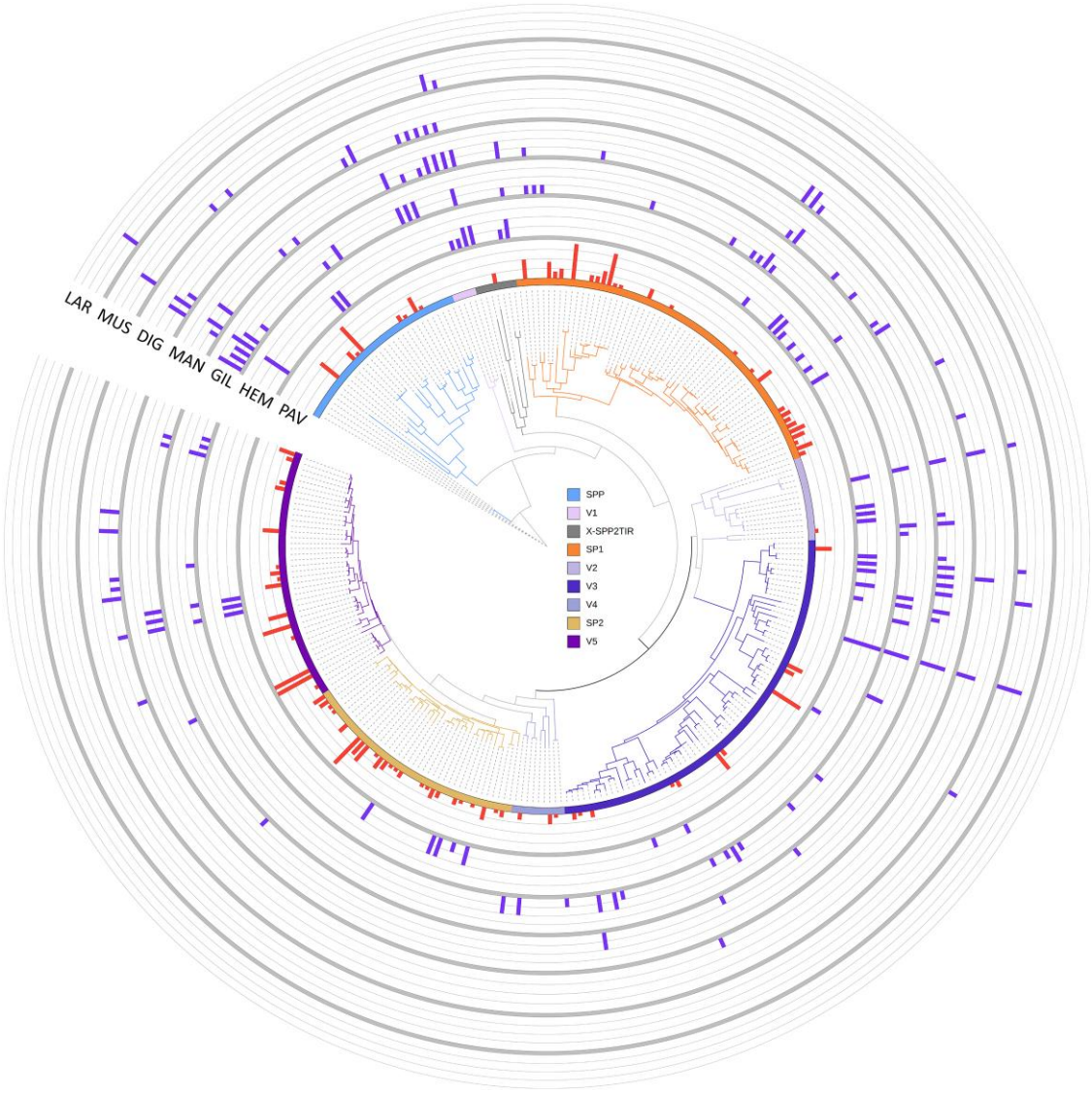

Fig. 8. Gene PAV and tissue specific expression of mussel TLRs. Circular phylogram summarizing the evolution of *M. galloprovincialis* TLRs, indicating the different sequence clusters. Different panels are stacked over the tree. The red histogram (PAV) shows the genomic dispensability of each TLR gene. TLRs with no visible red bar in the histogram were core genes present in all resequenced genomes. Red bars indicate dispensable TLRs and the height of the bar is equivalent to the dispensability of each TLR (i.e., the number of resequenced genomes in which that gene was absent). The other panels show the maximum expression of TLRs observed across all SRA mussel transcriptomic data set for hemocytes (HEM), gill (GIL), mantle (MAN), digestive gland (DIG), muscle (MUS), and larvae at mid trochophore stage (LAR). For each tissue, scale of the histogram has four levels according to expression levels of 5–10; 10–100; 100–1,000; and >1,000 TPMs.

and in another gill transcriptome obtained from mussels stimulated with diarrhetic shellfish poisoning (DSP) toxin, produced by *Prorocentrum lima*. Nevertheless, no clear modulation was detected for those cases.

The selected node TLRs in hemocytes allowed to retrieve a higher number of correlated transcripts than gills, including many functionally uncharacterized TIR proteins that may act as TLR intracellular mediators and other immune genes related to several immune pathways (fig. 10*A*). As in gills, a clear group of TLRs correlated with genes involved in the NF-κB pathway (TLRs 1–7). Another TLR group was correlated with antiviral response related proteins (TLRs 8–11). Myticins, key defense peptides that have the highest expression levels in mussel hemocytes, were correlated with a single node TLR (TLR 13). The large majority of node TLRs and their correlated transcripts were interconnected in two large networks (fig. 10*B*), one enriched in TLRs correlated with NF-κB genes and the other one including most of the TLRs correlated with other immune genes (fig. 1*C*). TLR modulation in hemocytes was analyzed in samples belonging to a bacterial stimulation experiment with two consecutive *Vibrio splendidus* infections, carried out in the same mussels within a week. In this case, modulated TLRs were mainly sccTLRs from clusters SP1, V2, and V3. For some cases, the same genes were among the six upregulated TLRs in response to the first infection and among the 12 downregulated TLRs in response to the second infection (fig. 10*D*).

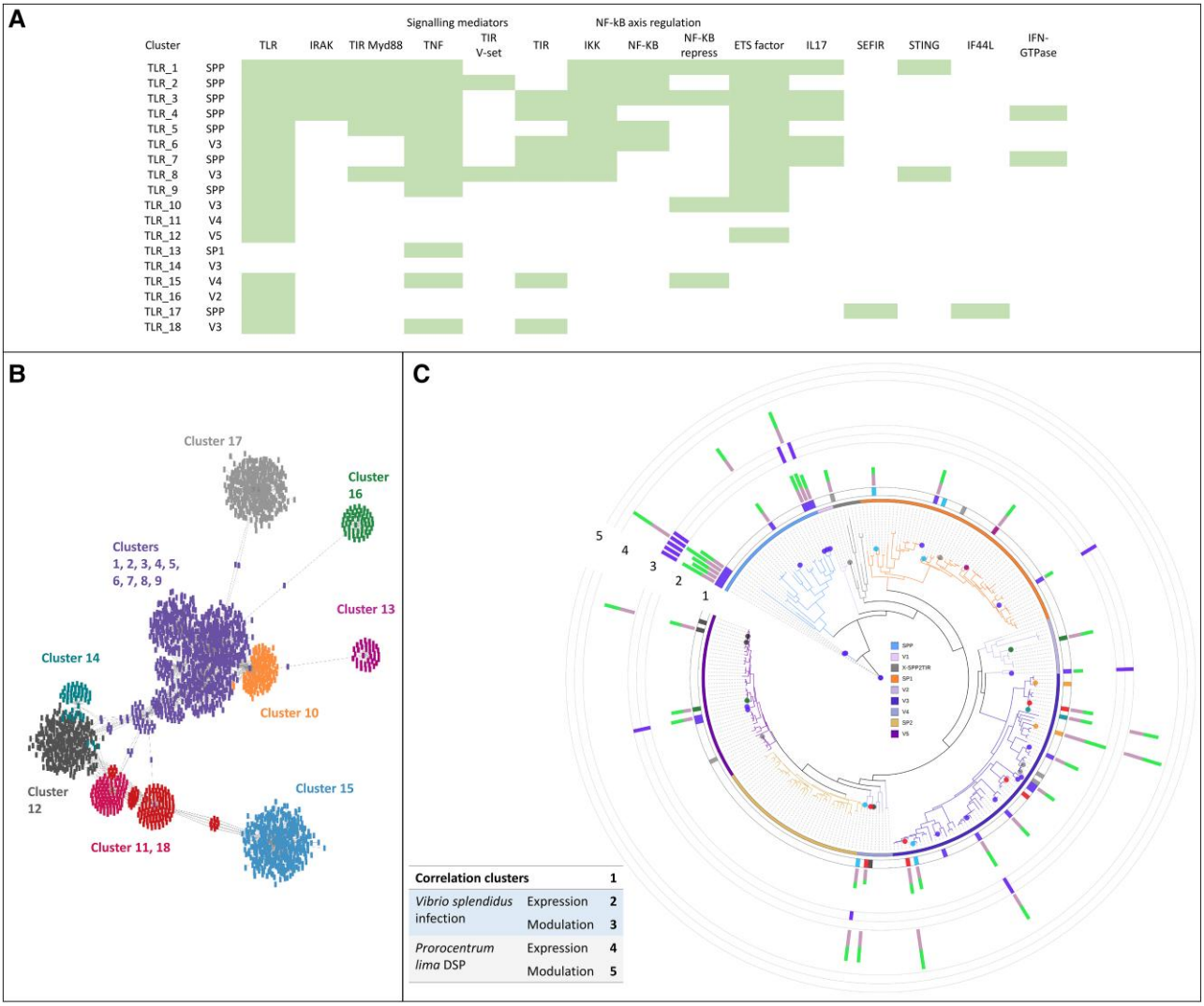

FIG. 9. TLR expression analyses in mussel gills. (*A*) Correlation clusters (0.7) calculated for each node TLR (TLRs that reached 10 TPMs in gill samples). The presence of immune-related transcripts relevant in the context or TLR-mediated signaling is indicated for each correlation cluster. (*B*) Correlation network including all the correlation clusters from panel *A*. The interconnection of all NF-κB-related clusters (1–8) is shown in the center of the net. (*C*) Phylogenetic tree of *M. galloprovincialis* TLRs indicating expression (minimum 5 TPMs) in different gill transcriptomic experiments that could involve external agent recognition. Graph 1 indicates which TLRs were classified into the expression correlation clusters, using the color legend of panel *B*. For a bacterial bath infection (PRJNA638821), control/stimulated expression levels are indicated in red and green, respectively (graph 2), and modulation of gene expression is indicated in graph 3 (bar = 2 for upregulation and bar = 1 for downregulation). The absence of bars in graph 3 indicates lack of statistically significant modulation. Expression (graph 4) and modulation (graph 5) are represented in the same way for gills stimulated by DSP produced by *P. lima* (PRJNA486919).

Mussel transcriptomes designed with different viral stimulations were also analyzed, evidencing three modulated TLRs, of which one was upmodulated and belonged to the antiviral expression correlation cluster (fig. 10*E*).

Discussion

In the present work, we focused on elucidating the origins and diversification of the bivalve TLRs, which display the largest repertoire of these receptors in the animal kingdom. To this end, we investigated TLR gene distribution in the broader context of metazoan evolution. Although both poriferans and cnidarians (except Anthozoa) lack canonical TLRs, they present TLR-like receptors with an intracellular TIR domain but no extracellular LRRs. This would support the hypothesis that the ancestral metazoan TLRs may have emerged by gene fusion of preexisting TIR-only and LRR-only genes in the eumetazoan ancestor, after the separation of poriferans, but before the split between cnidarians and bilaterians (Leulier and Lemaitre 2008; Brennan and Gilmore 2018; Nie et al. 2018). Nevertheless, while TLRs were absent from the choanoflagellate studied in this work (*Monosiga brevicollis*), they have been previously described in other choanoflagellates (Richter et al. 2018), free-living unicellular eukaryotes that are considered as the closest relatives of Metazoa. This would suggest that the origin of TLRs predates metazoans and that the metazoan ancestor to all animals would

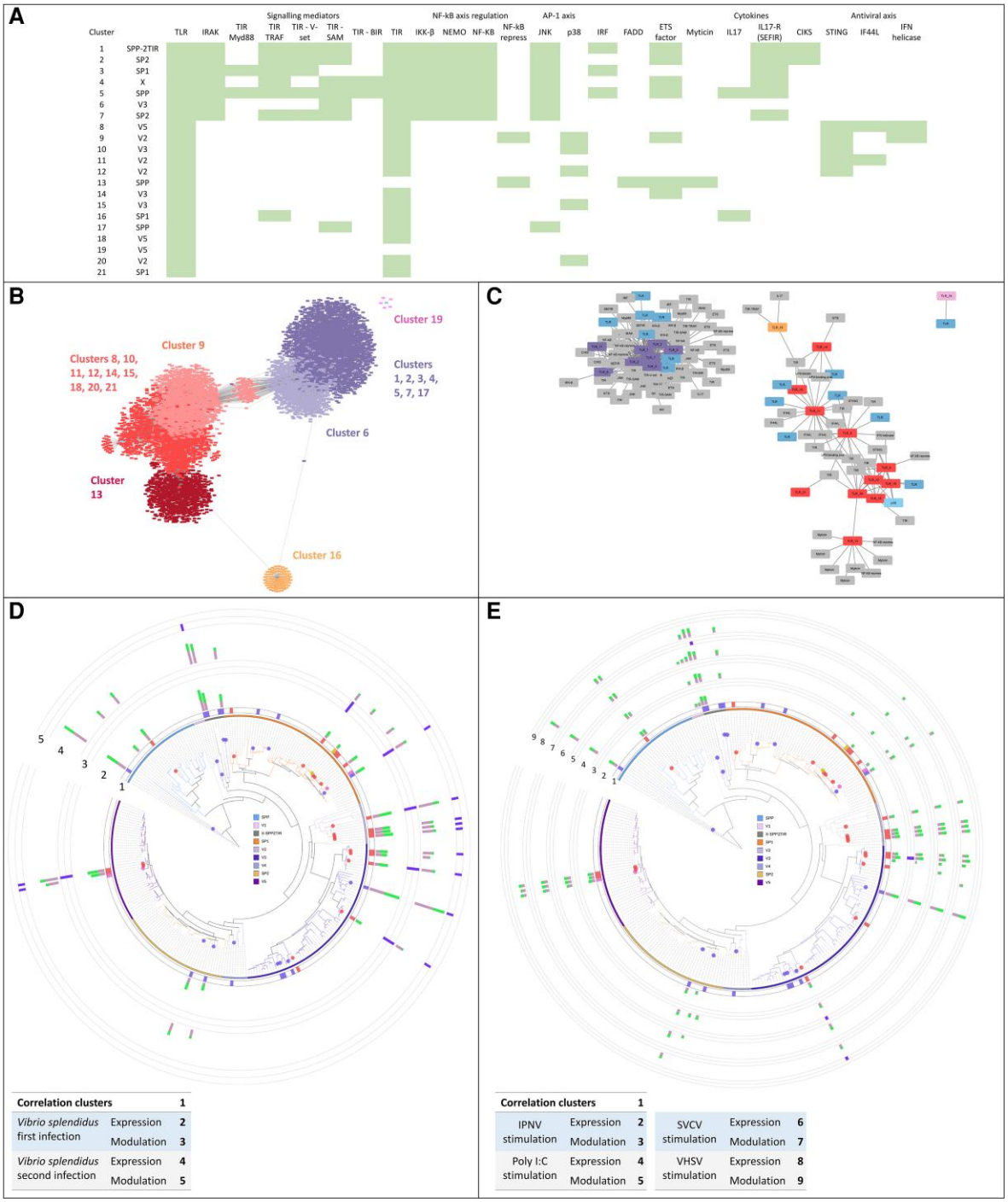

Fig. 10. TLR expression analyses in mussel hemocytes. (*A*) Correlation clusters (0.8) calculated for each node TLR (TLRs that reached 10 TPMs in hemocytes samples). The presence of immune-related transcripts relevant in the context or TLR-mediated signaling is indicated for each correlation cluster. (*B*) Correlation network including all the correlation clusters from panel *A*. Two main groups are formed in which most of the node TLRs, and their correlated genes are blended into two bigger correlation clusters (indicated in purple and red). Only two node TLRs with their clusters of correlated transcripts are more far in the net. (*C*) Correlation net filtered by annotation to represent only TLRs and the TLR-related immune genes shown in panel *A*. Two large clusters can be seen. One of these includes the TLRs correlated with NF-κB axis genes (purple TLR nodes), whereas the other one grouped almost all the others (red TLR nodes). Correlated target TLRs are displayed in blue and correlated target immune genes are displayed in grey. (*D*) Expression analysis of TLRs in mussel hemocytes stimulated with two consecutive bacterial infections (PRJNA466718). The phylogenetic relationships among mussel TLRs are summarized by a circular phylogram, and their classification into the correlation clusters is indicated with the same colors used in panels *B* and *C* (graph 1). For the first infection, control/stimulated expression levels are indicated in red and green respectively (graph 2), and modulation is indicated in graph 3 (bar = 2 for upregulation and bar = 1 for downregulation). The absence of bars in graph 3 indicates lack of statistically significant modulation. For the second infection, expression (graph 4) and modulation (graph 5) are represented in the same way. (*E*) Expression analysis of TLRs in mussel hemocytes after different viral stimulations (unpublished data). Graph 1 indicates the same expression cluster classification. TLRs expressed and modulated against the different viral stimuli are shown for infectious pancreatic necrosis virus (IPNV) (graphs 2 and 3), polyinosinic:polycytidylic acid (Poly I:C) (graphs 4 and 5), spring viraemia of carp virus (SVCV) (graphs 6 and 7), and viral hemorrhagic septicemia virus (VHSV) (graphs 8 and 9).

already present a TLR gene encoding a receptor with a canonical organization. Hence, the metazoan ancestor would be predicted to carry at least an ancestral clade α TLR gene and an ancestral clade β/γ TLR gene, which diversified afterwards giving place to the three TLRs clades (α, β, and γ) in which all metazoan TLRs are classified (Orús-Alcalde et al. 2021). Our results are in line with this hypothesis. The metazoan orthology analysis indeed revealed a common TLR orthology group shared across the 85 metazoan species analyzed, pointing to a shared ancestral origin. Additionally, the three major metazoan TLR clades were retrieved in the phylogenetic tree constructed with the TLR repertoires of selected species from the taxa analyzed in this study.

Based on our data, most early-branching metazoans phyla, that is, poriferans, ctenophores, and cnidarians (except Anthozoa), would have lost ancestral TLR genes, only retaining noncanonical TLR-like receptors lacking LRR ectodomains. Cnidarian anthozoans would have kept only the clade α TLR, which was the lone contributor to the generation of their TLR repertoires through gene duplication. In line with previous data (Miller et al. 2007), we identified a single α TLR (mccTLR) in the cnidarian *Nematostella vectensis*, but other anthozoans, such as *Acropora millepora*, were clearly endowed with more complex clade α TLR repertoires that included both mccTLRs and sccTLRs. In bilaterians, the ancestral TLR β/γ would have been duplicated before the split between the deuterostome and the protostome lineages, resulting in clade β and the clade γ TLRs. Previous data had established an exclusive classification of deuterostomes TLRs with clade β (Orús-Alcalde et al. 2021). We observed that vertebrate chordates would have indeed lost clades α and γ, presenting only sccTLRs belonging to clade β, but this was not the general case for all deuterostomes. Indeed, the echinoderm sea urchin *S. purpuratus* and the hemichordate *Saccoglossus kowalevskii* had both sccTLRs from clade γ and mccTLRs from clade α, even though most of their TLRs were clade β sccTLRs. *Strongylocentrotus purpuratus* presents a highly expanded TLR repertoire dominated by a huge expansion of paralogous clade β genes, conserving few mccTLRs from clade α and clade γ sccTLRs. A less expanded repertoire was retrieved from the other Echinoidea species *Lytechinus variegatus*, while a great reduction of TLRs was observed in the other echinoderms included in our study. However, previous data had evidenced that the expansion in *S. purpuratus* would be also occurring in the same magnitude in more closely related sea urchins as *Allocentrotus fragilis* and *Strongylocentrotus franciscanus* (Buckley and Rast 2012). In light of the data we collected concerning the presence and evolution of TLRs in deuterostomes, the event that led to the loss of clades α and γ TLRs is predicted to have occurred more recently, perhaps in the latest common ancestor of Olfactores, as suggested by the massive reduction in the TLR repertoires of urochordates.

Regarding protostomes, the exclusive presence of clade α mccTLRs and clade β sccTLRs in ecdysozoans would imply that clade γ was secondarily lost in this lineage. The ecdysozoan TLR repertoires studied in several arthropod and nematode species were generally smaller and less variable in magnitude than other metazoan groups. However, we can report for the first time a remarkable TLR expansion in *R. immarginata*, belonging to the superclass Myriapoda. This species presented both α mccTLRs and β sccTLRs, as in all other arthopods, but displayed a great expansion of paralogous genes in a branch of the α clade. The limited genomic resources available for this group of arthropods currently prevent the opportunity to study whether this was a species-specific feature of *R. immarginata*, or these considerations could be generally extended to centipedes and millipedes. The TLR gene families of the other main protostome group, Spiralia, presented a much higher size variation than ecdysozoans. Nontrochozoan spiralians either entirely lost TLRs (as in Platyhelminthes, possibly as a result of their parasitic lifestyle) or had very small repertoires (two TLR genes were found in the rotifer *B. calyciflorus*). On the other hand, trochozoan spiralians generally conserved TLRs belonging to the three clades and displayed large differences in the size of each TLR gene family. Expanded repertoires were found in the analyzed brachiopod and phoronid species, in line with previous findings (Luo et al. 2018). The marine deep ocean annelid *Paraescarpia echinospica* evidenced as well an expansion in comparison with the generally smaller families found in other annelids. TLRs were more sparsely found in Bryozoa, being present in some species, such as *Membranipora membranacea* (Orús-Alcalde et al. 2021), but entirely missing in others, like *Bugula neritina*. The most evident and generalized expansion of TLRs was found in molluscs, with large repertoires in all classes except cephalopods. The magnitude of this expansion was particularly high in bivalves, reaching an unmatched size in the three *Mytilus* mussels studied in the present work.

Bivalve repertoires were not only the largest ones but also the most diverse. In the other nonmolluscan invertebrate with the greatest expansion, the sea urchin, almost all TLRs were placed in the same metazoan-common orthology group, showing limited diversity and suggesting a relatively recent origin from a single expansion event. In stark contrast, bivalve TLRs were classified into different orthology groups, belonging to the three major metazoan TLR clades evidenced by phylogenetic inference. TLR subfamilies identified in mussels were conserved at different degrees between mollusc species. Indeed, while most of them had orthologs in all molluscan classes, others were found in bivalves, and a few others were restricted to *Mytilus* spp. Our data indicate that the main mussel TLRs subfamilies were conserved at least at Bivalvia level. These results are supported by previous phylogenetic analyses of oyster TLRs, which reported an organization in similar subfamilies (Zhang et al. 2015). We propose that the current large repertoire of mussel TLRs is the product of multiple, independent expansion events, which occurred at different times during evolution. A first expansion likely occurred in the latest common ancestor

of all molluscs, explaining the larger repertoire these metazoans show compared with other Lophotrochozoa. Subsequent independent expansions likely occurred before the radiation of bivalves, and others led to recent *Mytilus*-specific evolutionary innovations, originating the observed phylogenetic branches without orthology outside of mussels.

The relative contribution of different evolutionary forces to bivalve TLR evolution was studied in order to reveal the mechanisms underpinning their expansion and diversification. We collected genomic evidence supporting the frequent occurrence of gene tandem duplication in mussel TLRs, an evolutionary mechanism that had been previously identified as the main driver of the expanded TLR families of other invertebrates, such as the sea urchin (Hibino et al. 2006). The impact of diversifying selection, acting on the paralogous genes generated by this process, was stronger in the extracellular domains of bivalve TLRs, leading to a higher protein sequence variability the extracellular region. This observation is consistent with the well-established function of LRRs in the context of immune recognition, as they are responsible for determining ligand binding specificity. Therefore, we interpret these signatures of diversifying selection associated with a largely expanded TLR repertoire as an evolutionary successful broadening of the recognized ligand array. The expanded TLR repertoires found in other invertebrates seem to be driven by similar mechanisms, since the effect of positive selection in extracellular regions had also been observed in the sea urchin (Buckley and Rast 2012). Previous data collected about vertebrate TLRs confirmed that positive selection in the extracellular domains was focused in positions responsible of interaction with ligands (Wlasiuk and Nachman 2010; Alcaide and Edwards 2011; Areal et al. 2011), linking diversifying selection with functional diversification (Hughes and Piontkivska 2008).

Close relatives of chordates, such as the invertebrate *Ciona intestinalis*, possess few TLRs that exert "hybrid" functions, responding to multiple PAMPs that in humans would be recognized by individual different TLRs (Sasaki et al. 2009). Hence, humans and other vertebrates have developed an expanded repertoire of clade β TLR, starting from a low number of ancestral multifunctional receptors, according to a functional specification strategy (Satake and Sekiguchi 2012). Additionally, pathogen-associated molecular patterns from bacteria of marine environments cannot be recognized by the immune receptors of animals that live in different environments, such as terrestrial mammals (Gauthier et al. 2021, 2022), supporting the idea that the receptors found in each species may be adapted to specific ligands. Functional diversification was also evidenced in previous data for the sea urchin expansion (Buckley and Rast 2012).

A comprehensive expression analysis was performed in order to test the functional diversification hypothesis as the main reason for the expanded and diverse TLR repertoires in mussels. The expression of TLR genes followed different trends, forming several transcriptomic correlation clusters with cytosolic TIR-DC proteins and other immune signaling molecules and effectors. Albeit most bivalve TIR-DC proteins are functionally uncharacterized, bivalves have a large array of these cytosolic proteins that could interact with the intracellular TIR domain of TLRs, as the canonical Myd88 does, activating different intracellular pathways (Gerdol et al. 2017). Despite being more conserved than the ectodomain, the intracellular TIR domains of TLRs must evolve accordingly to allow the triggering of a specific immune response, as indicated by the fact that TLRs sharing similar LRR organizations clustered together in phylogenetic trees constructed using only TIR domains, as it had been observed previously in oyster (Zhang et al. 2015). Therefore, the specialized ligand recognition mediated by diversified extracellular LRRs is likely matched by a similar functional specialization of the intracellular TIR domains, which would allow the recruitment of different intracellular TIR-DC partners, triggering specific signaling pathways. As of note, while this specialization may allow a fine modulation of immune response toward stimuli, some TLRs might perform different functions, for example, by regulating the development process (Narbonne-Reveau et al. 2011; Buckley and Rast 2012). This is supported by the observation that, while most mussel TLRs were placed in clusters that mostly included immune genes responsive to bacterial or viral challenges in gills and hemocytes, others were not. For example, one mussel TLR was specifically expressed in mid-trochophore larvae stage in our data set, and other TLRs could be activated in the different development stages, following an expression pattern marked by the life strategy of each species to develop immunocompetence and reach adult stages (Tirapé et al. 2007; Balseiro et al. 2013; Orús-Alcalde et al. 2021).

The results of the present study suggest an extensive functional diversification as the main driver for the bivalve-specific TLR expansion. The higher expansion in bivalve TLRs with respect with the other molluscan classes may be explained by the lifestyle of these animals. Being mainly sessile, and filtering high amounts of sea water for feeding purposes, bivalves are constantly exposed to the great concentration and diversity of bacterial and viral pathogens present in the sea (Sogin et al. 2006; Gerdol et al. 2020; Regan et al. 2021). The high water concentration of marine bacteria (1×10^6 ml^{-1}, compared with an air concentration of 1×10^3 ml^{-1}; Prussin et al. 2015) has been associated, in other phyla, with peculiar evolutionary adaptations. For example, the very high prevalence of LPS in marine environments might prevent its effective detection as a meaningful signal indicative of Gram-negative bacterial infection in teleost fishes, which have consequently lost TLR4 genes, responsible for recognizing LPS in other chordates (Gauthier et al. 2022). Nevertheless, specific immune responses toward pathogenic Gram-negative bacteria have been evidenced in several bivalve species (Li et al. 2015; Jiang et al. 2017; Zhao et al. 2017), including mussels (Rey-Campos, Moreira, Valenzuela-Muñoz, et al. 2019; Saco et al. 2020). In the

present work, we collected evidence that supports the modulation of specific TLR subfamilies against bacterial pathogens, indicating a potential implication in their recognition. Repeated infections by the same bacterial pathogen in mussel induce an immune tolerance response, characterized by the downregulation, following the second encounter with the pathogen, of some genes involved in the inflammatory response to the first infection (Rey-Campos, Moreira, Gerdol, et al. 2019). Several TLRs displayed this behavior, being upregulated after the first infection and inhibited in response to the second one (fig. 10*D*). This highlights the existence of a regulatory mechanism that could be the key for explaining bivalve resilience. Indeed, these organisms might be able to recognize the high diversity of pathogens in the sea due to functional diversification of their immune receptors, avoiding at the same time an excessive and continuous activation of immune responses thanks to the enactment of a tolerization process, necessary in an environment with a high pathogen concentration.

Long lifetimes are another factor that could explain the need for the acquisition of a more complex set of immune receptors in certain invertebrates. Immunological specificity is one of the main reasons why adaptive immunity is proposed as key for the long lifespan of vertebrates (O'Connor and Cornwallis 2022). However, there are invertebrate species with only innate immunity and with long lifetimes, even surpassing vertebrates. Due to the presence of several long-lived species (Ridgway et al. 2011), including the longest living metazoan known to date, *Arctica islandica* (Butler et al. 2013), bivalves have been used as model organisms to study the basis of longevity (Abele et al. 2009; Buttemer et al. 2010; Moss et al. 2016; Blier et al. 2017). The presence of an expanded TLR repertoire in sea urchin could be similarly explained by its long lifespan by the need, shared with bivalves, to allow the specific recognition of different marine pathogens that it could face throughout its life cycle (Satake and Sekiguchi 2012). The larger expansion of the TLR family in mussels than in other bivalves could be due to the high resilience and invasiveness of this species. The success of mussels in colonizing different environments is supported by a complex genome with large lineage-level expansions and PAV that would explain its great adaptability (Gerdol et al. 2020). Larger expansions have been reported as well for other immune receptor and effector proteins in this species, in comparison with the generally expanded repertoires in other bivalves (Gerdol et al. 2011; Gerdol and Venier 2015; Saco et al. 2021). The conserved pattern of variable and expanded immune gene families in mussels implies a significant energy cost that must be justified by its functionality in supporting the lifestyle of these species (Regan et al. 2021). Further expression data and functional characterization studies may help to reveal the different recognition targets of the expanded TLR gene family in bivalves and to confirm which pathways and responses can be specifically activated by each TLR subfamily.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This research was funded by Ministerio de Ciencia e Innovación (PID2021-124955OB-I00) and by Agencia Galega de Innovación (IN607B 2022/13). A.S. was supported by a Spanish AEI/EU-FSE predoctoral contract PRE2019-090760. We thank Dr Filipe Castro and Dr André Machado for the training and advice. We thank Dr David Palatinsky for bioinformatics support. Finally, we thank the researchers whose published RNA-seq and genomic data were used in this study.

Data Availability

The genomes analyzed from metazoan species and the mussel transcriptomic reads used for the expression analysis are indicated in the supplementary materials S1 and S2, Supplementary Material online, respectively. The new TLR sequences identified for each species in the present work, and which were subsequently used for the different analyses, are available along with the pfam_scan identified domains in the supplementary material S6, Supplementary Material online. A Figshare repository (doi: 10.6084/m9.figshare.22770722) contains data from this publication, including set of sequences and different scripts used in the methodology. Accession link is: https://figshare.com/s/1b56c8ab0896887956f4.

References

Abele D, Brey T, Philipp E. 2009. Bivalve models of aging and the determination of molluscan lifespans. *Exp Gerontol.* **44**:307–315.

Alcaide M, Edwards S V. 2011. Molecular evolution of the Toll-like receptor multigene family in birds. *Mol Biol Evol.* **28**:1703–1715.

Almeida-Silva F, Venancio TM. 2022. BioNERO: an all-in-one R/Bioconductor package for comprehensive and easy biological network reconstruction. *Funct Integr Genomics.* **22**:131–136.

Anderson KV, Jürgens G, Nüsslein-Volhard C. 1985. Establishment of dorsal-ventral polarity in the *Drosophila* embryo: genetic studies on the role of the Toll gene product. *Cell* **42**:779–789.

Anderson KV, Nüsslein-Volhard C. 1984. Information for the dorsal–ventral pattern of the *Drosophila* embryo is stored as maternal mRNA. *Nature* **311**:223–227.

Areal H, Abrantes J, Esteves PJ. 2011. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol. Biol.* **11**: 368.

Balseiro P, Moreira R, Chamorro R, Figueras A, Novoa B. 2013. Immune responses during the larval stages of *Mytilus galloprovincialis*: metamorphosis alters immunocompetence, body shape and behavior. *Fish Shellfish Immunol.* **35**:438–447.

Blier PU, Abele D, Munro D, Degletagne C, Rodriguez E, Hagen T. 2017. What modulates animal longevity? Fast and slow aging in bivalves as a model for the study of lifespan. *Semin. Cell Dev Biol.* **70**:130–140.

Brennan JJ, Gilmore TD. 2018. Evolutionary origins of Toll-like receptor signaling. *Mol Biol Evol.* **35**:1576–1587.

Brennan JJ, Messerschmidt JL, Williams LM, Matthews BJ, Reynoso M, Gilmore TD. 2017. Sea anemone model has a single Toll-like

receptor that can function in pathogen detection, NF-κB signal transduction, and development. *Proc Natl Acad Sci*. **114**: E10122–E10131.
Buckley K, Rast J. 2012. Dynamic evolution of Toll-like receptor multigene families in echinoderms. *Front Immunol*. **3**:136.
Butler PG, Wanamaker AD, Scourse JD, Richardson CA, Reynolds DJ. 2013. Variability of marine climate on the North Icelandic Shelf in a 1357-year proxy archive based on growth increments in the bivalve *Arctica islandica*. *Palaeogeogr Palaeoclimatol Palaeoecol*. **373**:141–151.
Buttemer WA, Abele D, Costantini D. 2010. From bivalves to birds: oxidative stress and longevity. *Funct Ecol*. **24**:971–983.
Chen H, Cai X, Li R, Wu Y, Qiu H, Zheng J, Zhou D, Fang J, Wu X. 2022. A novel Toll-like receptor from *Crassostrea gigas* is involved in innate immune response to *Vibrio alginolyticus*. *Infect Genet Evol*. **97**:105159.
Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**:i884–i890.
Dainat J. 2019. AGAT: another Gff analysis toolkit to handle annotations in any GTF/GFF format.
Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol*. **7**:e1002195.
Emms DM, Kelly S. 2015. Orthofinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. **16**:157.
Emms DM, Kelly S. 2019. Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. **20**:238.
Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**:3047–3048.
Gauthier AE, Chandler CE, Poli V, Gardner FM, Tekiau A, Smith R, Bonham KS, Cordes EE, Shank TM, Zanoni I, *et al.* 2021. Deep-sea microbes as tools to refine the rules of innate immune pattern recognition. *Sci Immunol*. **6**:eabe0531.
Gauthier MEA, Du Pasquier L, Degnan BM. 2010. The genome of the sponge *Amphimedon queenslandica* provides new perspectives into the origin of Toll-like and interleukin 1 receptor pathways. *Evol Dev*. **12**:519–533.
Gauthier AE, Rotjan RD, Kagan JC. 2022. Lipopolysaccharide detection by the innate immune system may be an uncommon defence strategy used in nature. *Open Biol*. **12**:220146.
Gerdol M, Manfrin C, De Moro G, Figueras A, Novoa B, Venier P, Pallavicini A. 2011. The C1q domain containing proteins of the Mediterranean mussel *Mytilus galloprovincialis*: a widespread and diverse family of immune-related molecules. *Dev Comp Immunol*. **35**:635–643.
Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, Venier P, Naranjo-Ortiz MA, Murgarella M, Greco S, *et al.* 2020. Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biol*. **21**:275.
Gerdol M, Venier P. 2015. An updated molecular basis for mussel immunity. *Fish Shellfish Immunol*. **46**:17–38.
Gerdol M, Venier P, Edomi P, Pallavicini A. 2017. Diversity and evolution of TIR-domain-containing proteins in bivalves and Metazoa: new insights from comparative genomics. *Dev Comp Immunol*. **70**:145–164.
Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. **59**:307–321.
Hibino T, Loza-Coll M, Messier C, Majeske AJ, Cohen AH, Terwilliger DP, Buckley KM, Brockton V, Nair S V, Berney K, *et al.* 2006. The immune gene repertoire encoded in the purple sea urchin genome. *Dev Biol*. **300**:349–365.
Hughes AL, Piontkivska H. 2008. Functional diversification of the Toll-like receptor gene family. *Immunogenetics* **60**:249–256.
Jain C, Koren S, Dilthey A, Phillippy AM, Aluru S. 2018. A fast adaptive algorithm for computing whole-genome homology maps. *Bioinformatics* **34**:i748–i756.
Jiang F, Yue X, Wang H, Liu B. 2017. Transcriptome profiles of the clam *Meretrix petechialis* hepatopancreas in response to *Vibrio* infection. *Fish Shellfish Immunol*. **62**:175–183.
Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*. **8**:275–282.
Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol*. **338**:1027–1036.
Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform*. **20**:1160–1166.
Kawai T, Akira S. 2010. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol*. **11**:373–384.
Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res*. **19**:1639–1645.
Lefort V, Longueville J-E, Gascuel O. 2017. SMS: smart model selection in PhyML. *Mol Biol Evol*. **34**:2422–2424.
Lemaitre B, Nicolas E, Michaut L, Reichhart J-M, Hoffmann JA. 1996. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell* **86**:973–983.
Lemaitre B, Reichhart J-M, Hoffmann JA. 1997. *Drosophila* host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. *Proc Natl Acad Sci*. **94**:14614–14619.
Letunic I, Bork P. 2018. 20 Years of the SMART protein domain annotation resource. *Nucleic Acids Res*. **46**:D493–D496.
Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. **49**:W293–W296.
Letunic I, Khedkar S, Bork P. 2021. SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res*. **49**:D458–D460.
Leulier F, Lemaitre B. 2008. Toll-like receptors—taking an evolutionary approach. *Nat Rev Genet*. **9**:165–178.
Li R, Zhang R, Lu Z, Zou J, Xing Q, Dou H, Hu X, Lingling Z, Wang R, Bao Z. 2015. Characterizations and expression analyses of *NF-κB* and *Rel* genes in the Yesso scallop (*Patinopecten yessoensis*) suggest specific response patterns against Gram-negative infection in bivalves. *Fish Shellfish Immunol*. **44**:611–621.
Liu H, Huo L, Yu Q, Ge D, Chi C, Lv Z, Wang T. 2019. Molecular insights of a novel cephalopod Toll-like receptor homologue in *Sepiella japonica*, revealing its function under the stress of aquatic pathogenic bacteria. *Fish Shellfish Immunol*. **90**:297–307.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15**:550.
Luo Y-J, Kanda M, Koyanagi R, Hisata K, Akiyama T, Sakamoto H, Sakamoto T, Satoh N. 2018. Nemertean and phoronid genomes reveal lophotrochozoan evolution and the origin of bilaterian heads. *Nat Ecol Evol*. **2**:141–151.
Luo C, Zheng L. 2000. Independent evolution of Toll and related genes in insects and mammals. *Immunogenetics* **51**:92–98.
Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, Lopez R. 2022. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res*. **50**:W276–W279.
Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. **38**:4647–4654.
McGowan J, Fitzpatrick DA. 2020. Chapter five—recent advances in oomycete genomics. In: Kumar DBT, editors. Vol. 105. : *Advances in genetics*. Cambridge, USA: Academic Press. p. 175–228.
McGowan J, O'Hanlon R, Owens RA, Fitzpatrick DA. 2020. Comparative genomic and proteomic analyses of three widespread phytophthora species: *Phytophthora chlamydospora*,

Phytophthora gonapodyides and *Phytophthora pseudosyringae*. *Microorganisms* **8**:653.
Medzhitov R, Preston-Hurlburt P, Janeway CA. 1997. A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**:394–397.
Messier-Solek C, Buckley KM, Rast JP. 2010. Highly diversified innate receptor systems and new forms of animal immunity. *Semin Immunol.* **22**:39–47.
Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, Funayama N, Agata K, Bosch TCG. 2007. The innate immune repertoire in Cnidaria—ancestral complexity and stochastic gene loss. *Genome Biol*. **8**:R59.
Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, *et al.* 2021. Sustainable data analysis with Snakemake [version 2; peer review: 2 approved]. *F1000Res.* **10**:33.
Moss DK, Ivany LC, Judd EJ, Cummings PW, Bearden CE, Kim W-J, Artruc EG, Driscoll JR. 2016. Lifespan, growth rate, and body size across latitude in marine Bivalvia, with implications for Phanerozoic evolution. *Proc R Soc B Biol Sci*. **283**:20161364.
Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. *Mol Biol Evol.* **30**:1196–1205.
Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. *PLOS Genet*. **8**:e1002764.
Narbonne-Reveau K, Charroux B, Royet J. 2011. Lack of an antibacterial response defect in *Drosophila* Toll-9 mutant. *PLoS One* **6**: e17470.
Nie L, Cai S-Y, Shao J-Z, Chen J. 2018. Toll-like receptors, associated biological roles, and signaling networks in non-mammals. *Front Immunol.* **9**:1523.
O'Connor EA, Cornwallis CK. 2022. Immunity and lifespan: answering long-standing questions with comparative genomics. *Trends Genet*. **38**:650–661.
Orús-Alcalde A, Lu T-M, Børve A, Hejnol A. 2021. The evolution of the metazoan Toll receptor family and its expression during protostome development. *BMC Ecol Evol*. **21**:208.
Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* **14**:417–419.
Poltorak A, He X, Smirnova I, Liu MY, Van Huffel C, Du X, Birdwell D, Alejos E, Silva M, Galanos C, *et al.* 1998. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**:2085–2088.
Priyathilaka TT, Bathige SDNK, Lee S, Nam B-H, Lee J. 2019. Transcriptome-wide identification, functional characterization, and expression analysis of two novel invertebrate-type Toll-like receptors from disk abalone (*Haliotis discus discus*). *Fish Shellfish Immunol*. **84**:802–815.
Prochazkova P, Roubalova R, Skanta F, Dvorak J, Pacheco NIN, Kolarik M, Bilej M. 2019. Developmental and immune role of a novel multiple cysteine cluster TLR from *Eisenia andrei* earthworms. *Front Immunol.* **10**:1277.
Prussin AJII, Garcia EB, Marr LC. 2015. Total concentrations of virus and bacteria in indoor and outdoor air. *Environ Sci Technol Lett.* **2**:84–88.
Regan T, Stevens L, Peñaloza C, Houston RD, Robledo D, Bean TP. 2021. Ancestral physical stress and later immune gene family expansions shaped bivalve mollusc evolution. *Genome Biol Evol*. **13**: evab177.
Ren Y, Ding D, Pan B, Bu W. 2017. The TLR13-MyD88-NF-κB signalling pathway of *Cyclina sinensis* plays vital roles in innate immune responses. *Fish Shellfish Immunol*. **70**:720–730.
Rey-Campos M, Moreira R, Gerdol M, Pallavicini A, Novoa B, Figueras A. 2019. Immune tolerance in *Mytilus galloprovincialis* hemocytes after repeated contact with *Vibrio splendidus*. *Front Immunol.* **10**:1894.
Rey-Campos M, Moreira R, Valenzuela-Muñoz V, Gallardo-Escárate C, Novoa B, Figueras A. 2019. High individual variability in the transcriptomic response of Mediterranean mussels to *Vibrio* reveals the involvement of myticins in tissue injury. *Sci Rep.* **9**:3569.
Richter DJ, Fozouni P, Eisen MB, King N. 2018. Gene family innovation, conservation and loss on the animal stem lineage. *eLife*. **7**:e34226.
Ridgway ID, Richardson CA, Austad SN. 2011. Maximum shell size, growth rate, and maturation age correlate with longevity in bivalve molluscs. *J Gerontol A Biol Sci Med Sci*. **66A**: 183–190.
Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A. 2005. The evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci*. **102**:9577–9582.
Saco A, Rey-Campos M, Novoa B, Figueras A. 2020. Transcriptomic response of mussel gills after a *Vibrio splendidus* infection demonstrates their role in the immune response. *Front Immunol.* **11**:3273.
Saco A, Rey-Campos M, Novoa B, Figueras A. 2023. Mussel antiviral transcriptome response and elimination of viral haemorrhagic septicaemia virus (VHSV). *Fish Shellfish Immunol*. **136**:108735.
Saco A, Rey-Campos M, Rosani U, Novoa B, Figueras A. 2021. The evolution and diversity of interleukin-17 highlight an expansion in marine invertebrates and its conserved role in mucosal immunity. *Front Immunol.* **12**:692997.
Sasaki N, Ogasawara M, Sekiguchi T, Kusumoto S, Satake H. 2009. Toll-like receptors of the ascidian *Ciona intestinalis*: prototypes with hybrid functionalities of vertebrate Toll-like receptors. *J Biol Chem.* **284**:27336–27343.
Satake H, Sekiguchi T. 2012. Toll-like receptors of deuterostome invertebrates. *Front Immunol.* **3**:34.
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. **13**:2498–2504.
Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere.". *Proc Natl Acad Sci*. **103**:12115–12120.
Sollitto M, Kenny NJ, Greco S, Tucci CF, Calcino AD, Gerdol M. 2022. Detecting structural variants and associated gene presence–absence variation phenomena in the genomes of marine organisms. In: Verde C and Giordano D, editors. *Marine genomics: methods and protocols*. New York, NY: Springer US. p. 53–76.
Soneson C, Love MI, Robinson MD. 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; peer review: 2 approved]. *F1000Res.* **4**:1521.
Tassia MG, Whelan N V, Halanych KM. 2017. Toll-like receptor pathway evolution in deuterostomes. *Proc Natl Acad Sci*. **114**: 7055–7060.
Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. 2022. Signalp 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol.* **40**:1023–1025.
Tirapé A, Bacque C, Brizard R, Vandenbulcke F, Boulo V. 2007. Expression of immune-related genes in the oyster *Crassostrea gigas* during ontogenesis. *Dev Comp Immunol.* **31**:859–873.
Valanne S, Wang J-H, Rämet M. 2011. The *Drosophila* Toll signaling pathway. *J Immunol.* **186**:649 LP–649656. Available from: http://www.jimmunol.org/content/186/2/649.abstract
Wang P, Zhang Z, Xu Z, Guo B, Liao Z, Qi P. 2019. A novel invertebrate Toll-like receptor with broad recognition spectrum from thick shell mussel *Mytilus coruscus*. *Fish Shellfish Immunol*. **89**: 132–140.
Weaver S, Shank SD, Spielman SJ, Li M, Muse S V, Kosakovsky Pond SL. 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol.* **35**:773–777.

Wei X, Zhao T, Ai K, Li H, Jiang X, Li C, Wang Q, Jianmin Y, Zhang R, Jialong Y. 2018. Role of scavenger receptor from *Octopus ocellatus* as a co-receptor of Toll-like receptor in initiation of TLR-NF-κB signaling during anti-bacterial response. *Dev Comp Immunol.* **84**:14–27.

Wlasiuk G, Nachman MW. 2010. Adaptation and constraint at Toll-like receptors in primates. *Mol Biol Evol.* **27**:2172–2186.

Xu K, Zhang Z, Xu Z, Tang Z, Liu L, Lu Z, Qi P. 2019. A novel invertebrate Toll-like receptor is involved in TLR mediated signal pathway of thick shell mussel *Mytilus coruscus*. *Dev Comp Immunol.* **97**:11–19.

Zhang Y, He X, Yu F, Xiang Z, Li J, Thorpe KL, Yu Z. 2013. Characteristic and functional analysis of Toll-like receptors (TLRs) in the lophotrocozoan, *Crassostrea gigas*, reveals ancient origin of TLR-mediated innate immunity. *PLoS One* **8**: e76464.

Zhang L, Li L, Guo X, Litman GW, Dishaw LJ, Zhang G. 2015. Massive expansion and functional divergence of innate immune genes in a protostome. *Sci Rep.* **5**:8693.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**:153.

Zhao X, Duan X, Wang Z, Zhang W, Li Y, Jin C, Xiong J, Li C. 2017. Comparative transcriptome analysis of *Sinonovacula constricta* in gills and hepatopancreas in response to *Vibrio parahaemolyticus* infection. *Fish Shellfish Immunol.* **67**:523–535.