Diversity and evolution of TIR-domain-containing proteins in bivalves and Metazoa: New insights from comparative genomics

Marco Gerdol [a, *], Paola Venier [b], Paolo Edomi [a], Alberto Pallavicini [a]

[a] University of Trieste, Department of Life Sciences, Via Licio Giorgieri 5, 34127 Trieste, Italy
[b] University of Padova, Department of Biology, Via Ugo Bassi 58/B, 35131 Padova, Italy

ARTICLE INFO

Keywords:
TIR
Bivalve molluscs
Toll-like receptor
MyD88
Immune signalling

ABSTRACT

The Toll/interleukin-1 receptor (TIR) domain has a fundamental role in the innate defence response of plants, vertebrate and invertebrate animals. Mostly found in the cytosolic side of membrane-bound receptor proteins, it mediates the intracellular signalling upon pathogen recognition via heterotypic interactions. Although a number of TIR-domain-containing (TIR-DC) proteins have been characterized in vertebrates, their evolutionary relationships and functional role in protostomes are still largely unknown. Due to the high abundance and diversity of TIR-DC proteins in bivalve molluscs, we investigated this class of marine invertebrates as a case study. The analysis of the available genomic and transcriptomic data allowed the identification of over 400 full-length sequences and their classification in protein families based on sequence homology and domain organization. In addition to TLRs and MyD88 adaptors, bivalves possess a surprisingly large repertoire of intracellular TIR-DC proteins, which are conserved across a broad range of metazoan taxa. Overall, we report the expansion and diversification of TIR-DC proteins in several invertebrate lineages and the identification of many novel protein families possibly involved in both immune-related signalling and embryonic development.

1. Introduction

The Toll/interleukin-1 receptor (TIR) domain has been so far identified in nearly 15,000 different proteins from animals and plants (https://www.ebi.ac.uk/interpro, December 2016). This 170 amino acids long domain displays a complex 3D structure, with a central region comprising five-stranded parallel beta-sheets surrounded on both sides by five alpha helices (Xu et al., 2000). The TIR domain is usually located in the cytosolic side of membrane-bound receptor proteins and it is deemed fundamental for immune-related signal transduction from activated Toll-like receptors (TLRs) and Interleukin-1 receptors (IL-1R) in metazoans (O'Neill and Bowie, 2007). Upon binding to extracellular ligands such as a cytokine or a bacterial, fungal or viral component, these receptors undergo conformational changes which enable the heterotypic interaction of their TIR domain with an intracellular adaptor and the recruitment of downstream kinases (Dunne et al., 2003). These events eventually lead to the activation of transcription factors and, most often, to the establishment of a pro-inflammatory response.

The repertoire of Toll-like receptors (TLRs) largely varies across animal taxa and species, possibly reflecting specific immunological adaptations. For instance, while mammals typically possess 13 TLRs, the genome of some invertebrate species encode up to a few hundred TLR genes. The molecular diversification of TLRs is particularly important for the recognition of various Pathogen Associated Molecular Pattern (PAMPs), including LPS, flagellin and exogenous nucleic acids, either in the extracellular space or inside the endosome (Uematsu and Akira, 2008). The high plasticity of TLRs is evident in the Atlantic cod, where a series of gene losses and duplications led to a unique TLR repertoire, compensating the lack of other immune receptors (Solbakken et al., 2016). Other remarkable examples are the amphioxus *Branchiostoma floridae* and the sea urchin *Strongylocentrotus purpuratus*, that possess 72 and 222 receptors, respectively (Satake and Sekiguchi, 2012). While the TLRs of these animals are not orthologous to those of vertebrates, they display similar and, to some extent, even broader immune properties (Sasaki et al., 2009). TLRs have seemingly undergone a very complex evolutionary history, with multiple events of species-specific (or phyla-specific) expansions and losses accompanied with an extensive functional diversification (Hughes

* Corresponding author.
E-mail addresses: mgerdol@units.it (M. Gerdol), paola.venier@unipd.it (P. Venier), edomi@units.it (P. Edomi), pallavic@units.it (A. Pallavicini).

and Piontkivska, 2008; Zhang et al., 2010), which might be explained as a "*response to a changing array of binding requirements*" (Buckley and Rast, 2012).

Compared to TLRs, the evolutionary history of intracellular TIR-DC proteins has been only superficially investigated. In vertebrate animals, these cytosolic proteins have been essentially categorized into five well-known families whose members can be engaged in TIR-TIR interactions to activate (MyD88, MAL, TRIF, TRAM) or to negatively regulate (SARM) the TLR/IL-1R signalling cascade (O'Neill and Bowie, 2007). While the origin of some of the five families of vertebrate adaptors can be traced back to as early as the radiation of Bilateria (Poole and Weis, 2014; Wiens et al., 2005), others represent chordate lineage-specific innovations (Yang et al., 2011).

The number of domain combinations found in amphioxus cytosolic TIR-DC proteins is much larger than in vertebrates, resulting in an astounding variety of novel TIR-DC adaptors, that have been interpreted as possible "*shortcuts between endpoints*" along intracellular signalling pathways (Zhang et al., 2008). However, such an expansion is not unique to amphioxus, as several dozen yet uncharacterized "orphan TIR-DC proteins" with little or no similarity to vertebrate intracellular adaptors have also been reported in bivalve molluscs (Gerdol and Venier, 2015; Zhang et al., 2015). Notwithstanding the fragmentary information about their spread across metazoans, some invertebrate-specific TIR domain combinations appear to be ancient and widespread, since they are present in anthozoans (Poole and Weis, 2014; van der Burg et al., 2016) as well as in amphioxus (Huang et al., 2008).

The main aim of this work is to propose an updated classification of invertebrate TIR-DC proteins based on their structural domain organization and phylogenetic relationships, with a particular emphasis on those found in bivalve molluscs, due to their high number and diversity in this taxonomic class. First, we provide an overview about the diffusion of the TIR domain across Metazoa, based on the screening of the sequenced genomes available, and we discuss specific events of gene family expansion or shrinkage which might have occurred during animal evolution. In a second part, we proceed to an in depth characterization of the TIR-DC families which appear to be conserved across bivalves and discuss their evolution along the animal tree of life. We also propose an updated nomenclature system to properly identify novel TIR-DC proteins that do not pertain to any of the previously identified families.

2. Material and methods

2.1. Identification of TIR-DC proteins from sequence data in Bivalvia

The available bivalve transcriptome data, generated either by Illumina or 454 sequencing, were downloaded from the NCBI SRA database using the SRA Toolkit fastq-dump tool (https://github.com/ncbi/sra-tools). All the reads were imported into the CLC Genomics Workbench 9.0 environment (Qiagen, Hilden, Germany), trimmed by quality (minimum quality was set at 0.05, resulting reads shorter than 75% of the original size were discarded) and assembled using the *de novo* assembly tool, as described elsewhere (Gerdol et al., 2015a). Whenever both 454 and Illumina reads were available for a single species, only the latter were used due to their higher quality. In detail, transcriptome reads were selected for the species listed in Table 1. The assembled transcriptome of *Geukensia demissa* was kindly provided by Prof. Fields (Franklin & Marshall College, Lancaster, PA) (Fields et al., 2014). The predicted proteome of the Pacific oyster *Crassostrea gigas* was downloaded from Ensembl, based on the genome assembly v9 (INSDC Assembly GCA_000297895.1) (Zhang et al., 2012). Only oyster proteins whose annotation could be fully confirmed by transcriptome data were taken into account.

Assembled transcripts were analysed with TransDecoder v.3.0 (https://transdecoder.github.io) using standard parameters to predict the encoded proteins and only the complete ones (bearing both an ATG start codon and a STOP codon) were selected for further analysis. The TIR domain matrix PS50104 was downloaded from PROSITE (Sigrist et al., 2012) and used to scan the proteomes

Table 1
List of the bivalve mollusc species whose transcriptome has been scanned for the presence of sequences encoding TIR-domain containing proteins.

Species name	Taxa	Species name	Taxa	Species name	Taxa
Alasmidonta heterodon	Palaeoheterodonta	*Geukensia demissa*	Pteriomorphia	*Neocardia* sp. *VG-2014*	Pteriomorphia
Amusium pleuronectes	Pteriomorphia	*Glossus humanus*	Imparidentia	*Neotrigonia margaritacea*	Palaeoheterodonta
Anadara trapezia	Pteriomorphia	*Hiatella arctica*	Imparidentia	*Nuculana pernula*	Protobranchia
Arctica islandica	Imparidentia	*Hyriopsis cumingii*	Palaeoheterodonta	*Ostrea chilensis*	Pteriomorphia
Argopecten irradians	Pteriomorphia	*Lampsilis cardium*	Palaeoheterodonta	*Ostrea edulis*	Pteriomorphia
Astarte sulcata	Archiheterodonta	*Lamychaena hians*	Imparidentia	*Ostrea lurida*	Pteriomorphia
Atrina rigida	Pteriomorphia	*Lasaea adansoni*	Imparidentia	*Ostreola stentina*	Pteriomorphia
Azumapecten farreri	Pteriomorphia	*Laternula elliptica*	Anomalodesmata	*Paphia textile*	Imparidentia
Bathymodiolus azoricus	Pteriomorphia	*Limnoperna fortunei*	Pteriomorphia	*Pecten maximus*	Pteriomorphia
Bathymodiolus platifrons	Pteriomorphia	*Lyonsia floridana*	Anomalodesmata	*Perna viridis*	Pteriomorphia
Cardites antiquatus	Archiheterodonta	*Macoma balthica*	Imparidentia	*Phacoides pectinatus*	Imparidentia
Cerastoderma edule	Imparidentia	*Mactra chinensis*	Imparidentia	*Phreagena okutanii*	Imparidentia
Coelomactra antiquata	Imparidentia	*Margaritifera margaritifera*	Palaeoheterodonta	*Pinctada fucata*	Pteriomorphia
Corbicula fluminea	Imparidentia	*Mercenaria campechiensis*	Imparidentia	*Pinctada margaritifera*	Pteriomorphia
Crassostrea angulata	Pteriomorphia	*Mercenaria mercenaria*	Imparidentia	*Placopecten magellanicus*	Pteriomorphia
Crassostrea corteziensis	Pteriomorphia	*Meretrix meretrix*	Imparidentia	*Polymesoda caroliniana*	Imparidentia
Crassostrea gigas	Pteriomorphia	*Mesodesma donacium*	Imparidentia	*Pyganodon grandis*	Palaeoheterodonta
Crassostrea hongkongensis	Pteriomorphia	*Mimachlamys nobilis*	Pteriomorphia	*Ruditapes decussatus*	Imparidentia
Crassostrea virginica	Pteriomorphia	*Mizuhopecten yessoensis*	Pteriomorphia	*Ruditapes philippinarum*	Imparidentia
Cycladicama cumingii	Imparidentia	*Mya arenaria*	Imparidentia	*Saccostrea glomerata*	Pteriomorphia
Cyrenoida floridana	Imparidentia	*Mya truncata*	Imparidentia	*Sinonovacula constricta*	Imparidentia
Diplodonta sp. VG-2014	Imparidentia	*Myochama anomioides*	Anomalodesmata	*Solemya velum*	Protobranchia
Donacilla cornea	Imparidentia	*Mytilus californianus*	Pteriomorphia	*Sphaerium nucleus*	Imparidentia
Elliptio complanata	Palaeoheterodonta	*Mytilus chilensis*	Pteriomorphia	*Tegillarca granosa*	Pteriomorphia
Ennucula tenuis	Protobranchia	*Mytilus coruscus*	Pteriomorphia	*Uniomerus tetralasmus*	Palaeoheterodonta
Eucrassatella cumingii	Archiheterodonta	*Mytilus edulis*	Pteriomorphia	*Villosa lienosa*	Palaeoheterodonta
Eurhomalea exalbida	Imparidentia	*Mytilus galloprovincialis*	Pteriomorphia	*Yoldia limatula*	Protobranchia
Galeomma turtoni	Imparidentia	*Mytilus trossulus*	Pteriomorphia		

obtained from each bivalve species with HMMER v. 3.1 (Eddy, 1998) using a 0.01 e-value threshold.

After the virtually translated proteins were properly classified within families (Section 2.3), they were used as queries in a BLASTp to identify fragments of other bivalve TIR-DC proteins originated from incomplete assembled nucleotide sequences. The presence of a given family in a species was determined based on e-value and sequence identity thresholds of 1×10^{-50} and 50%, respectively.

2.2. Comparative genomics analysis of TIR-DC proteins

TIR-DC proteins were also detected with HMMER in the predicted proteomes of many other metazoans. Namely, the genomes selected for this analysis are reported in Table 2.

2.3. Domain architecture of TIR-DC proteins

All the resulting full-length predicted protein sequences were analysed with Phobius (Käll et al., 2004) to detect and discriminate signal peptide and transmembrane regions. Moreover, InterProScan sequence searches were performed to annotate the presence and position of conserved protein domains (Jones et al., 2014). Due to the degeneration of some domains (e.g LRRs in some TLRs and TIR in some intracellular proteins) compared to the reference InterPro consensus, a particular attention was given to those domains displaying e-values slightly below the default threshold of significance. Regions with no obvious conserved domain were modelled with HHpred (Söding et al., 2005) to identify the possible presence of conserved three-dimensional structures. Proteins which displayed obvious premature N-terminal or C-terminal truncations (e.g. only a partial TIR domain, or TLRs lacking a signal peptide) were considered as a likely result of mis-assembly and were therefore discarded. Annotated proteins were then categorized into families based on the predicted subcellular localization (membrane-bound vs cytoplasmic) and domain combination. Proteins displaying identical domain architecture were further compared with each other by BLASTp to confirm their homology. The minimum requirement for the creation of a new family was the identification of at least 3 full-length proteins in at least two species pertaining to different taxonomic orders. Whenever, in spite of identical domain architecture, a significant homology (based on an e-value threshold of 1×10^{-10}) could not be detected, sequences were considered as the possible result of convergent evolution. Proteins which did not satisfy the above mentioned criteria were labelled as 'unclassified'.

The complete list of the full-length protein sequences considered in this study, as well as the bivalve species where they were identified is reported in Supplementary File 1.

Table 2
Summary of the metazoan genomes analysed in the present study.

Species name	Taxa	Assembly version	Reference
Intoshia linei	Mesozoa	v.1.0	unpublished
Amphimedon queenslandica	Porifera	Aqu1	(Srivastava et al., 2010)
Trichoplax adhaerens	Placozoa	ASM15027v1	(Srivastava et al., 2008)
Mnemiopsis leidyi	Ctenophora	MneLei_Aug2011	(Ryan et al., 2013)
Nematostella vectensis	Cnidaria - Anthozoa	ASM20922v1	(Putnam et al., 2007)
Acropora digitifera	Cnidaria - Anthozoa	v.1.1	(Shinzato et al., 2011)
Exaiptasia pallida	Cnidaria - Anthozoa	v.1.1	(Baumgarten et al., 2015)
Thelohanellus kitauei	Cnidaria - Myxozoa	v.1	(Yang et al., 2014)
Hydra vulgaris	Cnidaria - Hydrozoa	v.1.0	(Chapman et al., 2010)
Schistosoma mansoni	Platyhelminthes	ASM23792v2	(Berriman et al., 2009)
Gyrodactylus salaris	Platyhelminthes	v.1.0	(Hahn et al., 2014)
Echinococcus multilocularis	Platyhelminthes	EMULTI002	(Tsai et al., 2013)
Schmidtea mediterranea	Platyhelminthes	v.4.0	(Robb et al., 2015)
Caenorhabditis elegans	Nematoda	WBcel23	(C. elegans Sequencing Consortium, 1998)
Trichinella spiralis	Nematoda	Tspiralis1	(Mitreva et al., 2011)
Drosophila melanogaster	Arthropoda - Insecta	BDGP6	(Adams et al., 2000)
Acyrthosiphon pisum	Arthropoda - Insecta	v2.0	(The International Aphid Genomics Consortium, 2010)
Pediculus humanus	Arthropoda - Insecta	PhumU2	(Kirkness et al., 2010)
Daphnia pulex	Arthropoda - Crustacea	v.1.0	(Colbourne et al., 2011)
Hyalella azteca	Arthropoda - Crustacea	v.2.0	unpublished
Strigamia maritima	Arthropoda - Myriapoda	Smar1	(Chipman et al., 2014)
Limulus polyphemus	Arthropoda - Merostomata	v.2.1.2	(Nossa et al., 2014)
Ixodes scapularis	Arthropoda - Arachnida	IscaW1	(Pagel Van Zee et al., 2007)
Priapulus caudatus	Priapulida	v.5.0.1	unpublished
Hypsibius dujardini	Tardigrada	ASM145500v1	(Boothby et al., 2015)
Ramazzottius varieornatus	Tardigrada	v.1.0	(Hashimoto et al., 2016)
Capitella teleta	Annelida - Polychaeta	v.1.0	(Simakov et al., 2013)
Helobdella robusta	Annelida - Clitellata	Helro1	(Simakov et al., 2013)
Lingula anatina	Brachiopoda	LinAna1.0	(Luo et al., 2015)
Adineta vaga	Rotifera	v.2.0	(Flot et al., 2013)
Lottia gigantea	Mollusca - Gastropoda	LotGi1	(Simakov et al., 2013)
Aplysia californica	Mollusca - Gastropoda	v.3.0	unpublished
Crassostrea gigas	Mollusca - Bivalvia	v9	(Zhang et al., 2012)
Octopus bimaculoides	Mollusca - Cephalopoda	v.1.0	(Albertin et al., 2015)
Strongylocentrotus purpuratus	Echinodermata - Echinozoa	Spur_3.1	(Sea Urchin Genome Sequencing Consortium et al., 2006)
Acanthaster planci	Echinodermata - Asterozoa	v.1.0	unpublished
Saccoglossus kowalevskii	Hemichordata	v.1.0	(Simakov et al., 2015)
Ptychodera flava	Hemichordata	v.3.0	(Simakov et al., 2015)
Branchiostoma floridae	Cephalochordata	v.2.0	(Putnam et al., 2008)
Ciona intestinalis	Urochordata	v.2.0	(Dehal et al., 2002)
Oikopleura dioica	Urochordata	v.1.0	(Danks et al., 2013)
Botryllus schlosseri	Urochordata	v.1.0	(Voskoboynik et al., 2013)
Danio rerio	Vertebrata	GRCz10	(Howe et al., 2013)

2.4. Phylogenetic analyses

Based on preliminary analyses, two separate sequence sets were prepared to facilitate phylogenetic reconstruction. Both sets contained the amino acidic sequences corresponding to the TIR domain, extracted using the coordinates of the beginning and the end of the domain, as identified by HMMER. In detail: (i) TIR domains from Toll-like receptors, limited to bivalve molluscs; (ii) TIR domains from all the other membrane-bound and cytoplasmic TIR-DC proteins from bivalve molluscs and other metazoans. Unclassified proteins were disregarded. Due to their poor conservation, a few TIR domains were not considered in this analysis. Namely, the third TIR domain of ecTIR-DC family 8 proteins (section 3.14) and both TIR domains of ecTIR-DC family 14 proteins (section 3.15). In the second analysis, one representative sequence for each TIR-DC protein family and for each metazoan phyla and at least one representative sequence for each TIR-DC family and for each bivalve subclass were taken into account (Supplementary File 1).

The two sets of sequences were separately aligned using MUSCLE (Edgar, 2004). Sequence alignments were then converted in a NEXUS format and used as an input for MrBayes v.3.2 (Huelsenbeck and Ronquist, 2001), to perform Bayesian phylogenetic inference. The LG model of molecular evolution with a Gamma distribution of rates across sites and fixed empirical priors on state frequencies (LG + G + F) was estimated to be the best-fitting model for the TLRs dataset with ProtTest (Abascal et al., 2005), based on the corrected Akaike Information Criteria (Hurvich and Tsai, 1989). The same analysis identified the WAG model (Whelan and Goldman, 2001) with a Gamma distribution of rates across sites and fixed empirical priors on state frequencies (WAG + G + F) as the best-fitting model for the second phylogenetic analysis, targeting all the remaining TIR-DC proteins.

Two independent analyses were run for each dataset until the standard deviation of observed split frequencies reached a value < 0.05 and the Effective Sample Size for each of the estimated parameters for each analysis reached a value higher than 200, disregarding the first 25% of the generated trees for the burnin procedure. This allowed the removal of trees sampled before the convergence of the runs. The remaining ones were used to build 50% consensus trees and to calculate posterior probabilities. The convergence of the runs and of the estimated parameters was inspected with Tracer v.1.6 (http://beast.bio.ed.ac.uk/Tracer). Details about convergence of the runs and estimated parameters are provided in Supplementary File 1.

2.5. Digital gene expression analysis

Raw RNA-seq reads obtained from nine tissues of adult oysters (hemocytes, digestive gland, gills, labial palp, inner mantle, mantle rim, male gonad, female gonad and adductor muscle) and 19 larval developmental stages (Zhang et al., 2012) were downloaded from the NCBI SRA database (Supplementary File 1). Sixty-one non-redundant oyster TIR-DC genes (based on a 95% sequence identity threshold), whose sequence could be fully confirmed by *de novo* assembled transcriptome and excluding those labelled as unclassified, were selected to calculate digital expression levels as described elsewhere (Gerdol et al., 2015b). Briefly, reads trimmed by quality were individually mapped with the RNA-seq mapping tool of the CLC Genomics Workbench v.9.0 (Qiagen, Hilden, Germany) on the full-length sequences of the 61 above mentioned TIR-DC mRNAs and 52 highly expressed housekeeping genes to guarantee a uniform mapping rate for all samples. Length and similarity fraction parameters were set at 0.75 and 0.98 respectively, with mismatch/insertion/deletion penalties set at 3. Read counts were used to calculate gene expression levels as TPM (Transcripts Per Million) (Wagner et al., 2012). The TPM values, transformed by log2, were used to generate a gene expression heat map.

3. Results and discussion

3.1. The TIR-DC proteins repertoire has undergone expansion and shrinkage in different taxa

We explored the full complement of TIR-DC proteins encoded by the genomes of representative metazoan species, dividing them between membrane-bound and cytoplasmic proteins (Fig. 1) and assessing the taxonomic distribution of the gene families which will be described in detail below (sections 3.2–3.20) (Fig. 2).

Overall, this analytical survey evidenced a low number of TIR-DC proteins in early-branching metazoans, with only a single gene in the mesozoan *I. linei*, 7 in the sponge *A. queenslandica*, 15 in the placozoan *T. adhaerens* and 11 in the ctenophore *M. leidyi*. Consistent with the results of previous studies (Gauthier et al., 2010), TLRs and most of the other TIR-DC proteins found in protostomes ad deuterostomes were absent in these basal metazoans, while others were likely already present in the common ancestor of all Metazoa.

A more dynamic situation was observed in Cnidaria. The myxozoan *T. kitauei* is completely devoid of TIR-DC proteins, possibly due to the genome size reduction observed in this parasitic species (Chang et al., 2015). On the other hand, the genome of the hydrozoan *H. vulgaris* possess 8 TIR-DC genes, comprising MyD88-like sequences and membrane bound receptors without extracellular LRRs (Miller et al., 2007), and anthozoan genomes have an even larger repertoire of TIR-DC genes, consisting of 14, 24 and 78 members in *N. vectensis*, *E. pallida* and *A. digitifera*, respectively. Earlier studies described the presence of TLRs and IL-1R-like receptors in anthozoans, as well as of a few previously uncharacterized cytoplasmic TIR-DC adaptors (Miller et al., 2007; Poole and Weis, 2014) which likely appeared for the first time in the common ancestor of all Eumetazoa (see details in the next section).

The most parsimonious interpretation of the data in our hands, in terms of number of gene family expansion and loss events necessary to explain current observations, implies that the TIR domain has undertaken different evolutionary routes during the course of protostome evolution. Indeed, a remarkable number of TIR-DC genes is noticeable in most Lophotrochozoa whereas both Platyhelminthes and Ecdysozoa display a greatly reduced complement of such genes, even in comparison with more basal metazoans (Fig. 1). Some TIR-DC protein families found in Parazoa and Radiata, as well as in Lophotrochozoa and basal Deuterostomia, are absent in the two aforementioned lineages, suggesting that they may have been lost along their evolution (Fig. 2). Platyhelminthes only have 1 to 8 different TIR-DC genes and lack membrane-bound TIR-DC proteins (with the exception of *S. mediterranea*, which shows a single IL-1R-like sequence) and other ancestral adaptors such as MyD88. In comparison, nematodes have two different types of TIR-DC proteins, a Toll-like receptor (sometimes missing) and SARM, but surprisingly lack MyD88, an essential mediator of Toll signalling. Likewise, the genomes of the tardigrades *H. dujardini* and *R. varieornatus* only encode two TIR-DC proteins (a Toll-like receptor and SARM). Priapulids and arthropods further confirm the limited number of TIR-DC genes in Ecdysozoa. Except from the 49 TIR-DC genes of the myriapod *S. maritima,* only 5–22 TIR-DC genes per genome could be identified across different arthropod subphyla. In these organisms, the TIR domain architectures are not highly diversified and they basically correspond to evolutionarily conserved TLRs, MyD88, SARM and ecTIR-DC families 9 and 15 (absent in Pancrustacea, see section 3.15 and section 3.20).

On the other hand, in Lophotrochozoa the high number of genes as well as the diversification of domain architectures are consistent

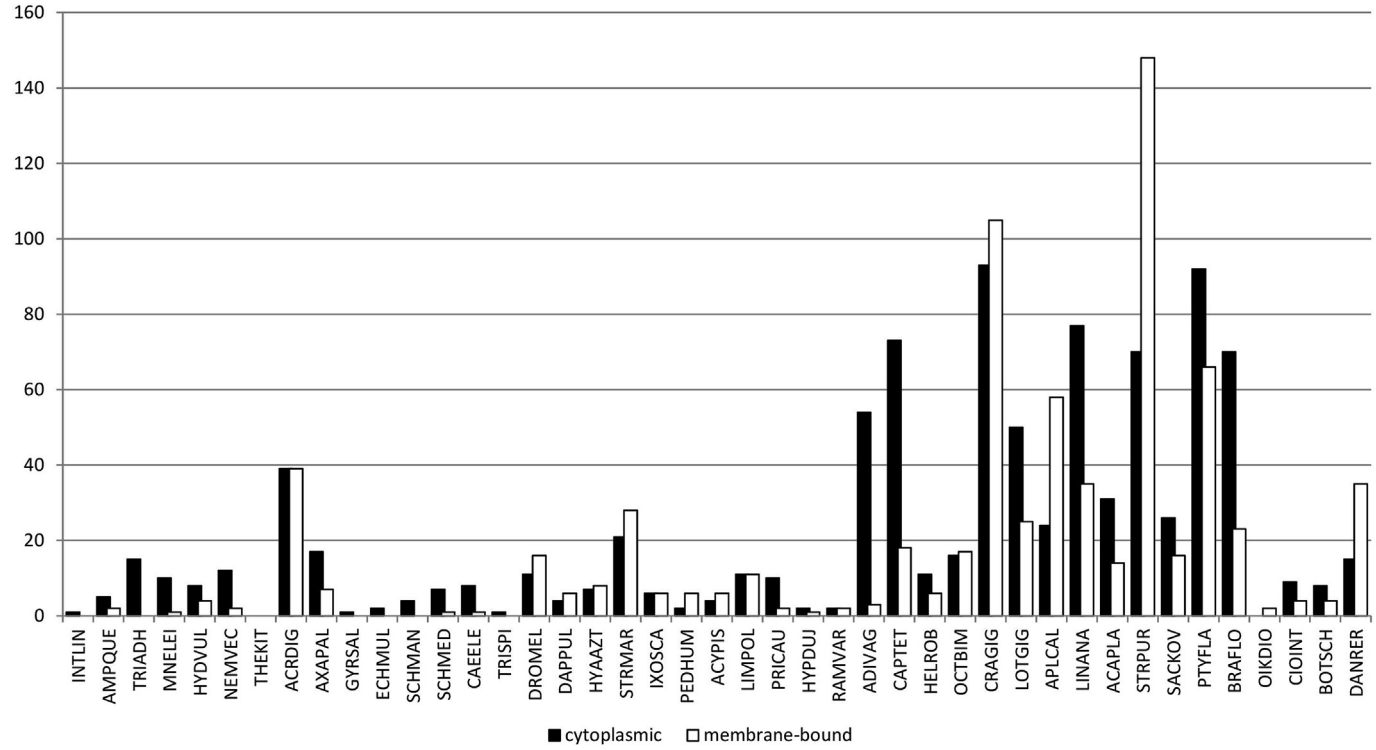

Fig. 1. Number of cytoplasmic (black bars) and membrane-bound (white bars) TIR-domain-containing proteins encoded by different metazoan genomes. ACAPLA: *Acanthaster planci*; ACRDIG: *Acropora digitifera*; ACYPIS: *Acyrthosiphon pisum*; ADIVAG: *Adineta vaga*; AMPQUE: *Amphimedon queenslandica*; APLCAL: *Aplysia californica*; AXAPAL: *Exaiptasia pallida*; BOTSCH: *Botryllus schlosseri*; BRAFLO: *Branchiostoma floridae*; CAPTET: *Capitella teleta*; CAEELE: *Caenorhabditis elegans*; CIOINT: *Ciona intestinalis*; CRAGIG: *Crassostrea gigas*; DANRER: *Danio rerio*; DAPPUL: *Daphnia pulex*; DROMEL: *Drosophila melanogaster*; ECHMUL: *Echinococcus multilocularis*; GYRSAL: *Gyrodactylus salaris*; HELROB: *Helobdella robusta*; HYAAZT: *Hyalella azteca*; HYDVUL: *Hydra vulgaris*; HYPDUJ: *Hypsibius dujardini*; INTLIN: *Intoshia linei*; IXOSCA: *Ixodes scapularis*; LIMPOL: *Limulus polyphemus*; LINANA: *Lingula anatina*; LOTGIG: *Lottia gigantea*; MNELEI: *Mnemiopsis leidyi*; NEMVEC: *Nematostella vectensis*; OCTBIM: *Octopus bimaculoides*; OIKDIO: *Oikopleura dioica*; PEDHUM: *Pediculus humanus*; PRICAU: *Priapulus caudatus*; PTYFLA: *Ptychodera flava*; RAMVAR: *Ramazzottius varieornatus*; SACKOV: *Saccoglossus kowalevskii*; SCHMAN: *Schistosoma mansoni*; SCHMED: *Schmidtea mediterranea*; STRMAR: *Strigamia maritima*; STRPUR: *Strongylocentrotus purpuratus*; THEKIT: *Thelohanellus kitauei*; TRIADH: *Trichoplax adhaerens*; TRISPI: *Trichinella spiralis*.

with a massive expansion of TIR-DC sequences, which is particularly evident in gastropod and bivalve molluscs (Figs. 1 and 2). The owl limpet *L. gigantea*, the sea slug *A. californica* and the Pacific oyster *C. gigas* bear 75, 82 and 198 different TIR-DC genes, respectively. Such gene expansion involves both TLRs, like in echinoderms, and cytoplasmic TIR-DC proteins, with at least 57 different genes labelled as "orphan TIRs" in oyster (Zhang et al., 2015). As regards cephalopods, only 33 genes could be identified in the California two-spot octopus *O. bimaculoides*, a fact that could be explained by the lineage-specific loss of many cytoplasmic TIR-DC proteins otherwise conserved in other molluscs (Fig. 2). A molecular diversification similar to that of Bivalvia and Gastropoda, with an increase in the number of both membrane-bound and cytosolic TIR-DC proteins, has apparently occurred in other Lophotrochozoa, such as the brachiopod *L. anatina* (113 TIR-DC genes) and the polychaete worm *C. teleta* (91 TIR-DC genes) but not in other segmented worms (e.g. 17 TIR-DC genes are present in the leech *H. robusta*, which overall displays a much less diversified repertoire) (Fig. 1). Despite having many TIR-DC genes (56), the rotifer *A. vaga* developed a peculiar set of intracellular proteins which combine Armadillo-type repeats to the TIR domain, and completely lacks TLRs and MyD88 (Fig. 2).

As regards basal deuterostomes, sea urchins (Echinodermata) are possibly the metazoan organisms with the highest number of TIR-DC genes, due to the expansion of TLRs (Buckley and Rast, 2012; Hibino et al., 2006). However, this situation differs significantly in other echinoderms, as for instance the crown-of-thorns sea star *A. planci* only possess six TLRs (Fig. 1). Some hemichordate species also have an expanded complement of TIR-DC proteins, like the acorn worm *P. flava* with its 158 TIR DC proteins. Similarly, 124 TIR-DC genes, including 72 TLRs have been previously reported in the amphioxus *B. floridae*, (Dishaw et al., 2012), even though we could detect a slightly lower number (93 TIR-DC proteins and 23 TLRs) based on the current genome annotation, which has been reported to be sub-optimal (Bányai and Patthy, 2016).

Based on the most parsimonious interpretation of the evolution of TIR-DC genes, the majority of the evolutionarily conserved cytoplasmic TIR-DC proteins found in invertebrates were lost somewhere along the evolution of the Chordata lineage. In detail, the TIR-DC gene family has likely undergone shrinkage in Urochordates, since *Ciona intestinalis*, *Botryllus schlosseri* and *Oikopleura dioica* only display 13, 12 and 2 TIR-DC genes, with a great reduction of both TLRs (2, 2 and just 1, respectively) and cytoplasmic proteins (Fig. 1). Finally, vertebrate animals typically possess 20–50 TIR-DC genes, including TLRs, IL-1-like receptors and five intracellular fundamental adaptors, MyD88, MAL, TRIF, TRAM and SARM (O'Neill and Bowie, 2007), but lack the vast majority of evolutionarily conserved gene families identified in this work (Fig. 2).

In summary, based on this large-scale comparative survey, it appears that multiple independent and lineage-specific expansion events of TIR-DC genes have likely occurred in different taxa, including Echinodermata, the lophotrochozoan Mollusca, Polychaeta, Brachiopoda and Rotifera. TLRs were often involved in these evolutionary events, as perfectly exemplified by the case of *S. purpuratus* and, partially, by Mollusca. However, in other cases the increase and diversification also involved intracellular TIR-DC proteins, which are to some extent evolutionarily conserved

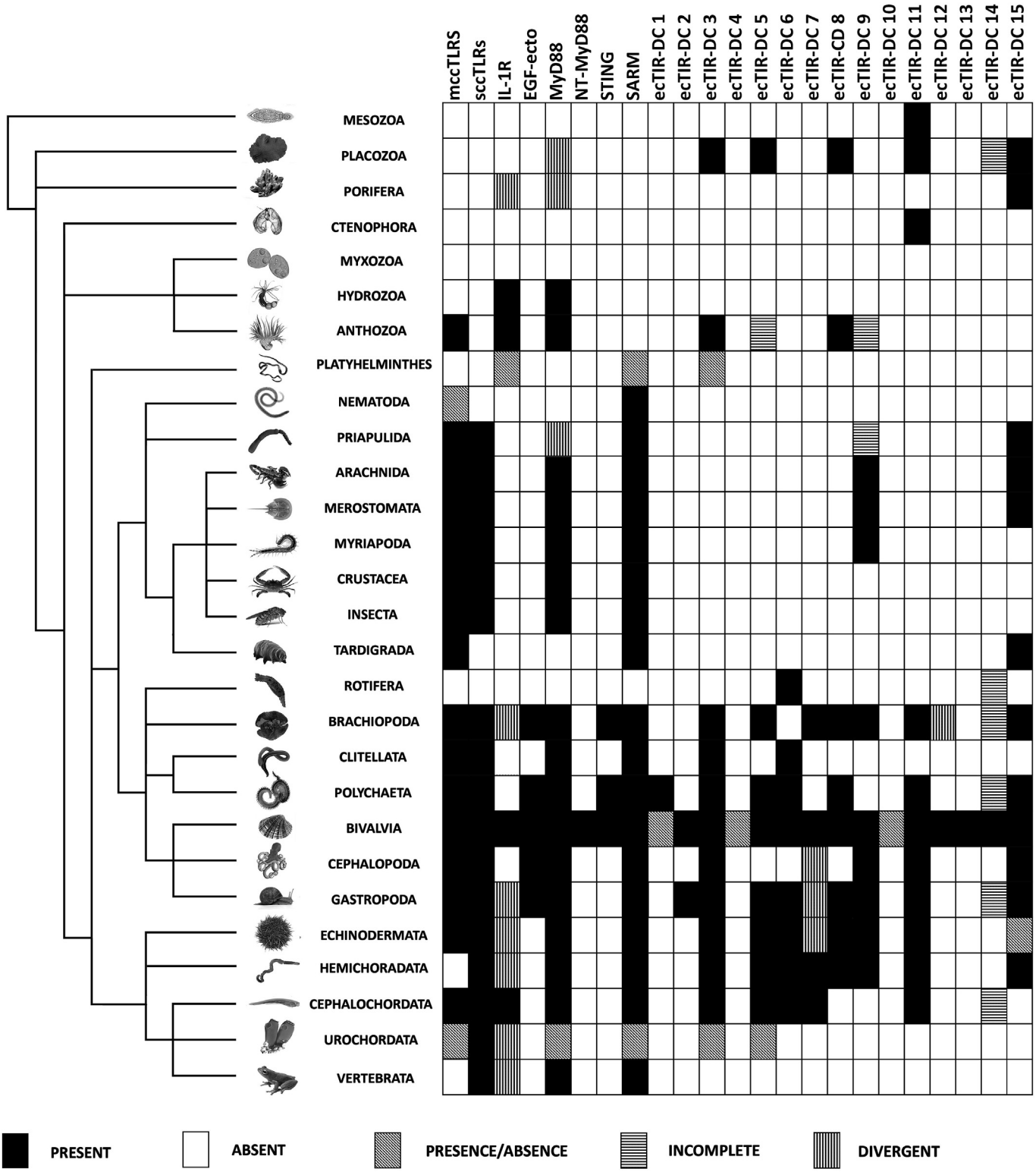

Fig. 2. Taxonomic distribution of TIR-DC protein families across metazoans. mcc: multiple cysteine cluster; scc: single cysteine cluster; TLR: Toll-like receptor; IL-1R:Interleukin-1 receptor; EGF-ecto: membrane-bound receptor with EGF ectodomains; MyD88: Myeloid differentiation primary response gene 88; STING: Stimulator of Interferon Genes; SARM: Sterile alpha and TIR motif-containing protein; ecTIR-DC: evolutionarily conserved TIR domain containing protein. The classification 'present' and 'absent' in a given taxa was based on the genome of the species considered in this study (section 2.2); 'presence/absence' defines cases where the presence of a gene family could not be ascertained in all the species pertaining to a given taxa; 'Incomplete' indicates cases where a given gene family could be identified based on high sequence homology, even though one of the characterizing domains (see Fig. 7) was missing; 'divergent' indicates the presence of proteins with domain architecture identical to that of a given family, but whose homology could not be confirmed due to high primary sequence divergence. As regards bivalves, 'presence/absence' was based on the presence of a given family in the genome of *C. gigas* and at least one additional species not pertaining to the Ostreoida order. Families recurrently found in bivalve transcriptomes but absent in the oyster genome were categorized as 'present only in some species'. The topology of the phylogenetic tree shown on the left is based on the NCBI Taxonomy classification.

across invertebrates (Fig. 2) and could therefore be involved in important, but yet uncharacterized intracellular signalling pathways.

3.2. Toll-like receptors (TLRs)

Different types of membrane-associated TIR-DC receptors are collectively known as Toll-like receptors (TLRs). Owing their name to the *Drosophila* protein Toll, implicated in the establishment of dorso-ventral embryonic polarity (Hashimoto et al., 1988), TLRs are certainly better known for their important role as Pattern Recognition Proteins (PRRs), able to recognize bacteria, fungi and exogenous nucleic acids.

All TLRs share three common features: a C-terminal and intracellular TIR domain, a transmembrane region and a series of N-terminal and extracellular Leucine Rich Repeats (LRRs), which are usually flanked by characteristic N-terminal and C-terminal LRR domains (LRRNT, LRRCT) (Fig. 3). The origins of TLRs can be traced back to the onset of Cnidaria, as the anemone *N. vectensis* is the most basal metazoan showing a single *bona fide* TLR. This protein displays a remarkable structural similarity to that of *Drosophila* Toll, with typical LRRs being interrupted by LRR-NT and LRR-CT domains, in a so-called multiple cysteine cluster TLR (mccTLR) organization. This ectodomain organization is by far the most common across ecdysozoan TLRs, and the only one present in nematodes and tardigrades. On the other hand, all vertebrates have a simpler organization, where single LRR-NT and LRR-CT domains flank typical LRRs, in a single cysteine cluster organization (sccTLRs).

Due to their structural divergence, mccTLRs and sccTLRs have long been thought to have an independent origin (Luo and Zheng, 2000). However, the remarkable conservation of downstream signalling pathways, evident from the progressive sequencing of new

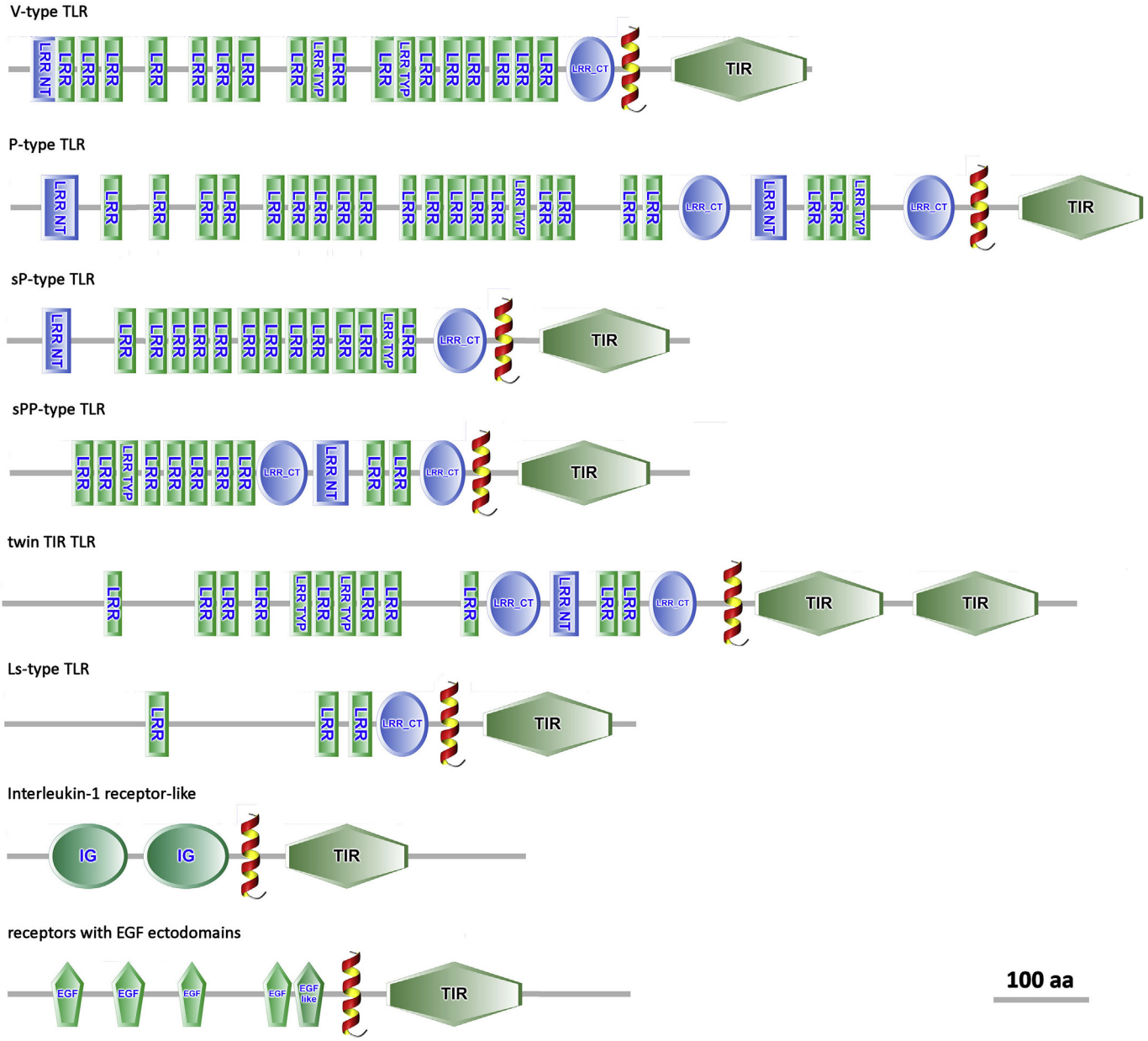

Fig. 3. Domain organization of the membrane-bound TIR-domain-containing (TIR-DC) proteins found in bivalve molluscs, including Toll-like receptors, Interleukin-1-like receptors and the newly described receptors with EGF ectodomains. Domain prediction is based on InterProscan. Helices indicate transmembrane domains.

invertebrate genomes, and the coexistence of sccTLRs and mccTLRs in a number of invertebrates (Fig. 2), challenged this view. The most ancient TLR group, mccTLRs, is present in almost all protostomes (except from flatworms and rotifers) and in some deuterostomes. In contrast, the taxonomic distribution of sccTLRs indicates a more recent origin. These receptors are not widespread in Ecdysozoa, but they underwent expansion in most deuterostomes and Lophotrochozoa (Fig. 2).

However, while this simple classification scheme based on the presence of a single or multiple cysteine clusters is sufficient to describe the general structural features of TLRs, it appears to be inadequate to fully describe the variability of the TLRs found in a given phylum from a genomic perspective (Buckley and Rast, 2012; Toubiana et al., 2013). For this reason, we used a slightly modified version of the classification scheme proposed by Zhang et al. (2015) to group bivalve TLRs within six major groups (Fig. 3). This classification system, based on the organization of LRR ectodomains, permits an easy categorization of bivalve TLRs consistent with TIR domain-based phylogeny (Fig. 4).

Among sccTLRs: (i) V-type TLRs, displaying the classical sccTLR LRR organization; (ii) sP-type TLRs, similar to V-type TLRs but somewhat shorter; (iii) Ls-type TLRs, which do not display a LRR-NT domain and often show non-canonical and degenerated LRRs.

Among mccTLRs: (iv) P-type TLRs, which resemble *Drosophila* Toll; (v) sPP-type TLRs, which are characterized by the same LRR organization but are somewhat shorter; (vi) twin-TIR TLRs, similar to the two previously mentioned groups but with two consecutive TIR domains.

Although Zhang and colleagues originally designed V-type TLRs as "Vertebrate-type" and P-, sP- and sPP-type TLRs as "Protostome-like", we might point out that such a distinction could be misleading, since V-type TLRs are present in many invertebrates and, on the other hand, protostome-like TLRs are often found in deuterostomes.

The first group of mccTLRs, P-type TLRs, share high structural conservation with insect Toll (Toubiana et al., 2013), supporting the hypothesis of a common origin for all invertebrate P-typeTLRs. Compared to the phylogenetic analysis reported by Zhang and

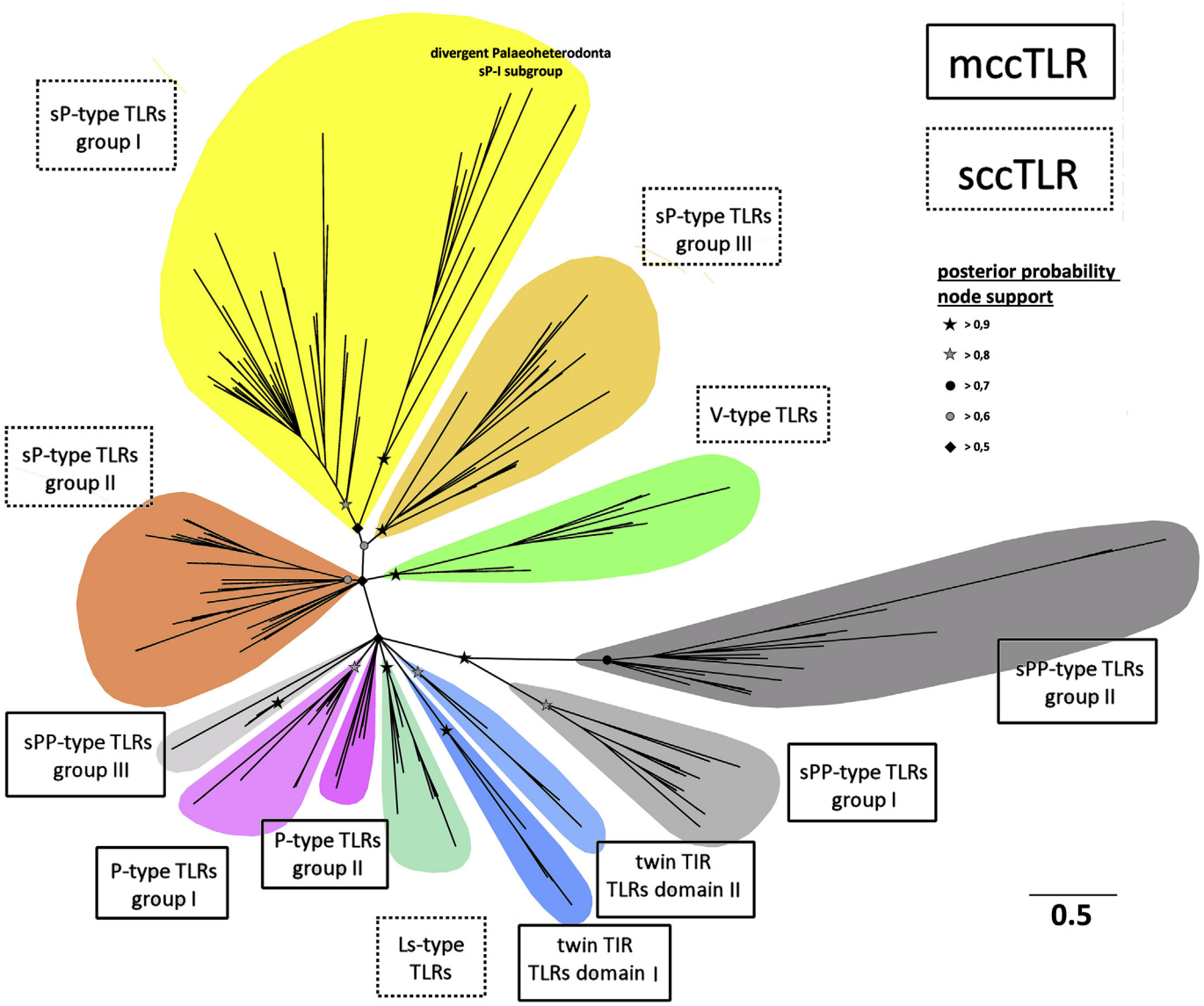

Fig. 4. Bayesian phylogeny of bivalve mollusc TLRs. Nodes with posterior probability support lower than 50% were collapsed. mcc: multiple cysteine cluster; scc: single cysteine cluster; TLR: Toll-like receptor.

colleagues we could further subdivide bivalve P-type TLRs into two subgroups (I and II), exemplified by *M. galloprovincialis* MgTLRb and MgTLRk, based on the divergence of the TIR domain (Fig. 4). P-type TLRs were found in limited number in bivalve transcriptomes (Supplementary File 1) and the presence of just two P-type TLR genes (one per subtype) in the Pacific oyster genome indicates that this family has not undergone expansion in this lineage.

The second group, sPP-type TLRs, were found in variable number in many bivalve species (Supplementary File 1). Previous analyses demonstrated that this is a relatively small gene family which only comprises a few members in *C. gigas* (Zhang et al., 2015). However, the higher number of sPP-type TLRs included in this study compared to Zhang and colleagues allowed the phylogenetic discrimination of these membrane-bound receptors in three subgroups (Fig. 4). While sPP-type TLRs I and II are broadly distributed, the small subtype III appears to be taxonomically restricted to Palaeoheterodonta.

The newly reported class of mccTLRs named twin-TIR TLRs is unique due to the presence of two consecutive TIR domains in the intracellular C-terminal region. The first member of this group ever described, Hc Toll-2, has been involved in the binding of bacterial LPS and PGN and in the regulation of AMP expression in *H. cumingii* (Ren et al., 2014b). Phylogenetic inference revealed the divergence of the two domains, suggesting their possible functional specialization. Just a low number of full-length twin-TIR TLRs was detected in our analysis, possibly due to their low expression level and relatively long size. Although only fragments of twin-TIR TLRs were found in the Pacific oyster transcriptome, a screening of its genome revealed the presence of two distinct loci encoding TLRs with this domain architecture.

In summary, despite their ancient origin and broad range of distribution across invertebrate species, none of the mccTLR groups underwent massive expansion in bivalves and in any other major invertebrate phylum.

Unlike mccTLRs, several events of gene family expansion could be inferred for sccTLRs along the evolution of metazoans. In their previous work, Zhang and colleagues defined three main groups of sccTLRs, V-, Ls- and sP-type TLRs, which were conveniently further subdivided into three clades (I, II and III) based on the divergence of the TIR domain.

Like P-type TLRs, also V-type TLRs display a remarkable similarity with sccTLRs found in other animal phyla, including vertebrates. These receptors, found in a relevant number of bivalve species (Supplementary File 1), represent a phylogenetically well-defined TLR group (Fig. 4) comprising a limited number of genes per species (5 in *C. gigas*).

On the other hand, sP-type TLRs, the largest group of bivalve sccTLRs, displayed astounding sequence variability. While the classification proposed by Zhang and colleagues for sP-type TLRs is, in line of principle, valid, the division of these receptors within the three proposed clades does not appear to be strongly supported when the entire complement of bivalve TLRs is taken into account, especially for what concerns the sP III clade, which displays the highest degree of sequence variability. Furthermore, in some cases (i.e. Palaeoheterodonda sequences within the sP clade I), divergent groups can be identified based on species phylogeny (Fig. 4).

Ls-type TLRs, which bear poorly recognizable LRR ectodomains, could not be conclusively place them neither within the sccTLR nor within the mccTLR group by the phylogenetic analysis. Ls-TLRs do not appear to be widespread (only 3 genes could be confirmed in *C. gigas*) and so far there is no indication about their function.

3.3. Interleukin-1-like and other transmembrane receptors

Interleukin-1-like receptors (IL-1RL) are the second major class of membrane-bound TIR-DC proteins of vertebrates. These receptors are characterized by three extracellular immunoglobulin (Ig) domains with ligand-binding function (IL-1 and related cytokines). The cytokine/receptor binding, accompanied by the interaction with accessory proteins, determines the activation of a downstream intracellular signalling that is basically shared with TLRs and ultimately leads to the activation of NF-κB and to the reenforcement of the inflammatory response (Subramaniam et al., 2004).

Although proteins structurally very similar to IL-1 receptors have been described in different invertebrates (Beck et al., 2000; Poole and Weis, 2014), it is still unclear whether they are homologous to vertebrate IL-1R, due to their very scarce sequence homology and the uncertain presence of a functionally conserved IL-1-like cytokine in invertebrates (Beschin et al., 2001). Indeed, it has been previously suggested that the Ig/TIR domain combination may have evolved independently several times over the course of metazoan evolution (Zhang et al., 2010).

Our screening for IL-1RL genes across invertebrate genomes was definitely consistent with these observations. Although this domain organization could be found in several taxa (Porifera, Cnidaria, most Lophotrochozoa and all deuterostomes), the sequences appeared to be, in most cases, non-orthologous or of dubious orthology (Fig. 2). As an example, it is worth noticing that bivalve IL-1RL proteins share a relevant degree of homology with those found in cnidarians, and amphioxus, but curiously not to those found in other protostomes. Furthermore, we can report for the first time that the extracellular region of the membrane-bound TIR-DC receptors previously reported in *Hydra* (Miller et al., 2007) are likely to adopt an immunoglobulin-like fold, as predicted by HHpred analysis (Söding et al., 2005) and supported by their placement within the IL-1RL cluster by phylogenetic inference (Fig. 6). IL-1RL is present as a single copy gene in the oyster genome, and it encodes a protein with two extracellular Ig domains. We could identify full-length orthologous sequences in 4 bivalve species whereas only transcript fragments were detectable in many other bivalves (Supplementary File 1), a fact that may possibly indicate a low expression levels.

Besides TLRs and IL-1-like receptors, we can report for the first time a third group of membrane-bound TIR-DC proteins found exclusively in Lophotrochozoa (including Mollusca, Brachiopoda, Annelida but not Rotifera). These orphan receptors possess a variable number of epidermal growth factor (EGF)-like ectodomains, resulting in precursor proteins ranging in length from ~450 to ~1300 aa. Although EGF domains are extremely widespread, especially in the extracellular domain of membrane-bound animal proteins (Bork et al., 1996), their functional role in most cases is poorly understood. Due to the lack of sequence homology to proteins with known function, the biological role of these newly described receptors cannot be hypothesized at the present time. Although TIR-DC proteins with EGF ectodomains were among the TIR-DC sequences most frequently found in bivalve transcriptomes (Supplementary File 1), they are present in low number in lophotrochozoan genomes (e.g. 2 genes in the Pacific oyster).

3.4. MyD88 and its "naturally truncated" variants

The myeloid-differentiation primary response gene 88 (MyD88) is a cytoplasmic adaptor protein containing an N-terminal DEATH domain followed by a TIR domain (Fig. 5) which acts downstream Toll-like receptors and IL-1R. Upon infection, MyD88 transmits the signal from the extracellular space to the cytosol through heterotypic interaction with the intracellular TIR domain of transmembrane receptors, sometimes with the cooperation of other TIR-DC adaptors (Horng et al., 2001). This important mediator of

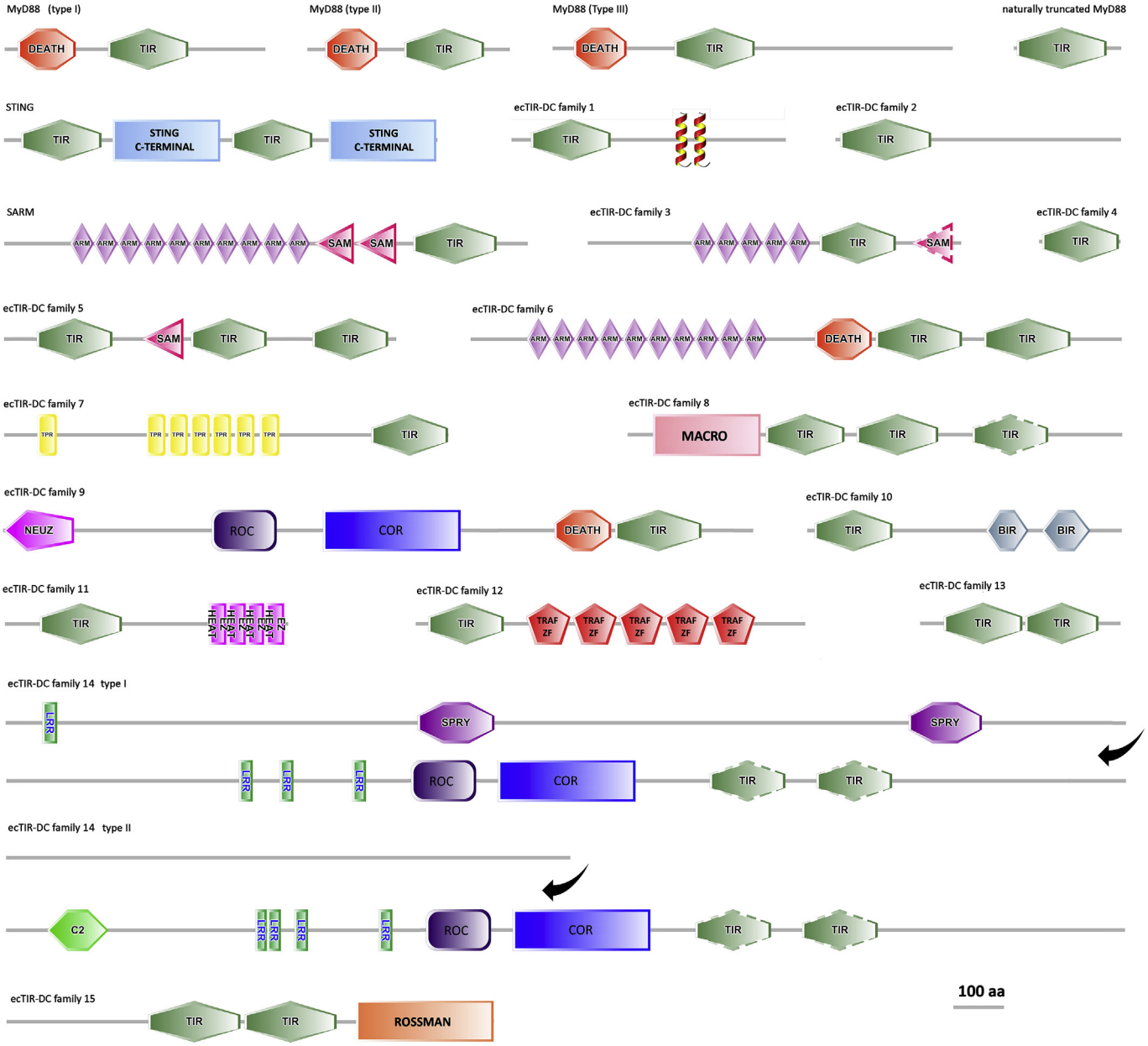

Fig. 5. Summary of the domain architecture in the evolutionary conserved TIR-DC protein families identified in bivalve molluscs. Domains were identified with InterPro. Helices indicate transmembrane domains. MyD88: Myeloid differentiation primary response gene 88; STING: Stimulator of Interferon Genes; SARM: Sterile alpha and TIR motif-containing protein; ecTIR-DC: evolutionarily conserved TIR-domain-containing protein. Due to the great protein length, the domain architecture of ecTIR-DC family 14 is represented on two consecutive lines, as indicated by a black arrow. Dashed domains indicate non canonical, poorly recognizable domains.

immune signalling activates a complex signalling cascade which ultimately leads to the transcription of pro-inflammatory genes regulated by NF-κB and AP-1 (Muzio et al., 1997; O'Neill, 2003). MyD88 is an evolutionarily conserved protein, which emerged quite early in metazoan evolution, as its homologs can be found in cnidarians and Porifera (Franzenburg et al., 2012; Wiens et al., 2005). The comparative survey of available sequence data indicates that MyD88 is missing in just a few taxonomic groups which have undergone a massive reduction of the TLR machinery, including nematodes, tardigrades and urochordates (where only one or two Toll-like protein are present), flatworms and rotifers (where TLRs are completely absent) (Fig. 2).

The first evidence of a bivalve MyD88 sequence and a MyD88-dependent immune signalling pathway were reported in *A. farreri* (Qiu et al., 2007) and *R. philippinarum* (Lee et al., 2011), respectively. More recently, multiple immune-responsive MyD88 genes have been identified in *M. galloprovincialis*, *M. yessoensis* and *C. gigas* (Ning et al., 2015; Toubiana et al., 2013; Xin et al., 2016). Functional studies further demonstrated the involvement of bivalve MyD88 adaptors in innate immune signalling, their cross-talk with Toll and NF-κB, and subsequent activation of AMP genes (Ren et al., 2014a). The description of a complete TLR signalling pathway in bivalve molluscs (Toubiana et al., 2014) further supports the key role of MyD88 in TLR signal transduction and in the activation of the downstream components of the pathway IRAK and TRAF6.

In spite of the evolutionary conservation of the TIR and DEATH domains, the variable length of the C-terminal extension of MyD88 proteins reveals a remarkable structural diversification of bivalve

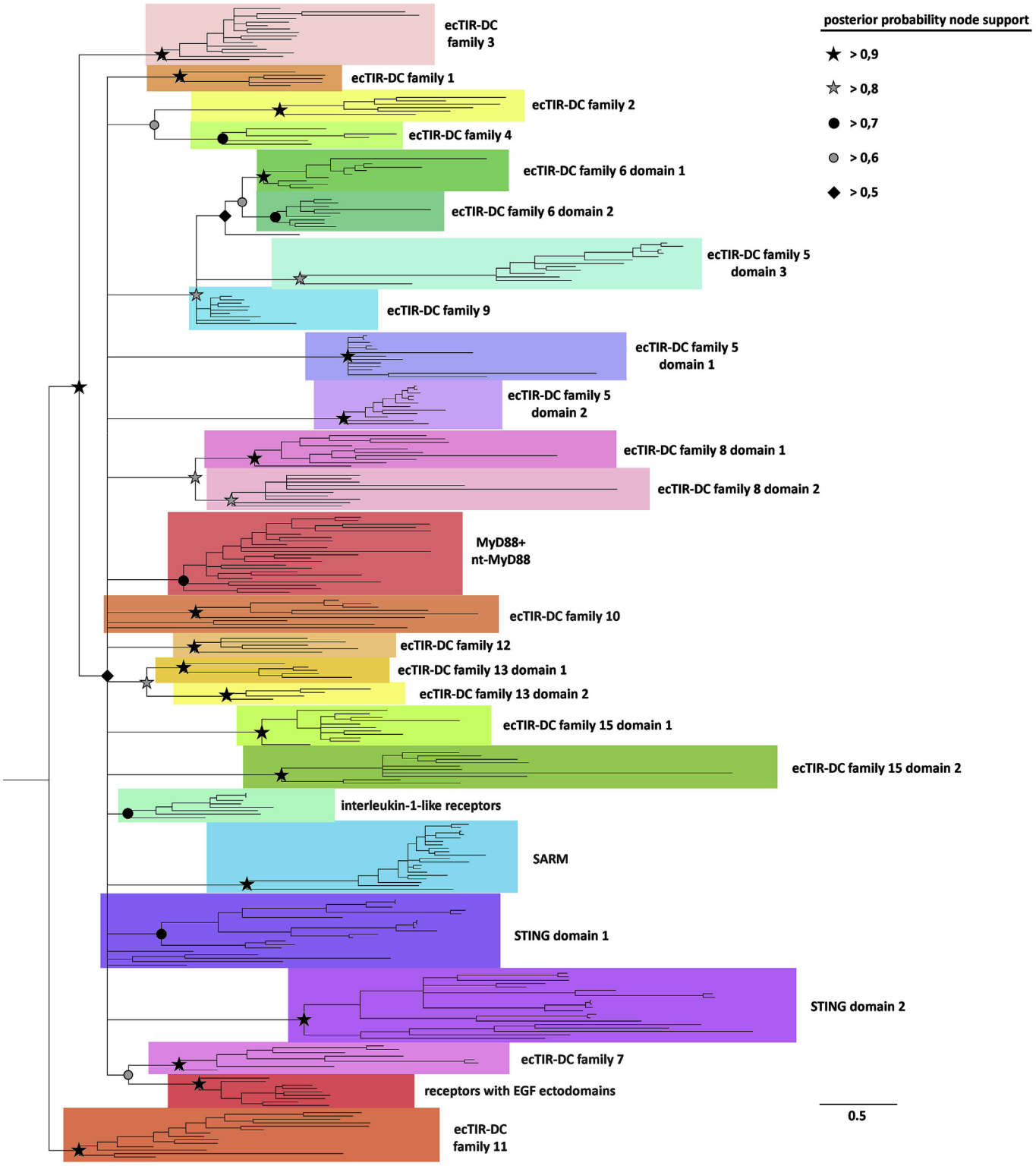

Fig. 6. Bayesian phylogeny of cytoplasmic TIR-domain containing proteins in bivalve molluscs and other metazoa, excluding those labelled as unclassified. Nodes with posterior probability support lower than 50% were collapsed. MyD88: Myeloid differentiation primary response gene 88; STING: Stimulator of Interferon Genes; SARM: Sterile alpha and TIR motif-containing protein; ecTIR-DC: evolutionarily conserved TIR-domain-containing protein. The tree was rooted on ecTIR-DC family 11, identified as the one with the most ancient inferred origin. The accession IDs of the sequences used for this analysis are reported in Supplementary File 1.

MyD88 adaptors (Fig. 5), like previously discussed by other authors (Toubiana et al., 2013). Although all MyD88 sequences clustered together in the TIR domain-based phylogeny (Fig. 6), we could define three MyD88 subtypes. Subtype I comprises proteins with a low-complexity region which extends the C-terminus to a total protein length of 420–530 aa. Such extension is shorter in subtype II proteins as their length does not exceed 380 aa. These two subtypes are widespread in bivalves and can be found together in the

same species (e.g. MgMyD88-a pertaining to subtype II, MgMyD88-b and −c pertaining to subtype I in *M. galloprovincialis*). MyD88 proteins of subtype III were found in a small number, only in a few bivalve species. These proteins have a much longer C-terminal extension and their TIR-domain is usually less recognizable. The identification of MyD88 sequences in the transcriptomes of 60 out of the 81 bivalve species analysed in this study confirms the relevance of their biological function (Supplementary File 1).

On the other hand, a family of shorter (less than 200 aa long) MyD88-like proteins, named *naturally truncated* MyD88 (NT-MyD88) was found in Ostreoida and in a few other bivalve species. The high sequence conservation of the TIR domain (see Fig. 6) leaves no doubts about the evolutionary link between these surprisingly truncated proteins and canonical MyD88 adaptors. The up-regulation of a NT-MyD88 transcript was first reported in the hemocytes of *Crassostrea ariakensis* infected by Rickettsia-like organisms, rising questions on the possible role of these unusual TIR-DC proteins in the host immune responses (Zhu and Wu, 2008). Recently, Xu and colleagues described that the two oyster variants CgMyD88-T1 and -T2, likely originated from a recent truncation of a MyD88 paralogous gene, were strongly up-regulated in hemocytes upon bacterial challenges. Consequently, the authors proposed they could act as negative regulators of the TLR pathway (Xu et al., 2015). The expression profiles of the two genes after viral infection and the lack of a DEATH domain required for signal transduction further support this hypothesis (He et al., 2015; Zhang et al., 2015).

3.5. SARM

SARM (sterile alpha- and armadillo-motif-containing protein) is one of the five typical vertebrate TIR-DC adaptors. This cytosolic protein, besides its C-terminal TIR domain, also contains two sterile alpha motifs (SAM) and Armadillo repeats towards the N-terminus (Fig. 5), i.e. two protein-interaction motifs with a broad range of functions (Kim and Bowie, 2003). In vertebrates, SARM has been implicated in the negative regulation of MyD88-independent TLR signalling (Carty et al., 2006). However, the presence of the orthologous protein Tir-1 in *C. elegans* (Belinda et al., 2008), an organism lacking the aforementioned pathway, suggests a different function for SARM in invertebrates. Indeed, TIR-1 also appears to be a key regulator of p38 MAPK, controlling both AMP production and neuronal development (Chuang and Bargmann, 2005; Couillault et al., 2004).

We could confirm the presence of SARM in many bivalve species, including the Pacific oyster, with three distinct genes (Supplementary File 1). The conservation of SARM in Lophotrochozoa and, more in general, in nearly all Bilateria (Fig. 2) implies a fundamental function for this protein. However, functional studies are needed to clarify whether bivalve SARMs mainly play a role in innate immunity, in development, or both. A role of SARM as a negative regulator of MyD88-independent TLR signalling seems unlikely, due to the absence of TRIF homologs and the consequent uncertainties about the existence of a MyD88-independent pathway itself in these organisms.

3.6. Novel intracellular evolutionarily conserved TIR-DC proteins

MyD88 and SARM are the only two intracellular TIR-DC proteins of vertebrates found in bivalves as well as in many other invertebrates. On the other hand, with the exception of a TRIF homolog in amphioxus (Yang et al., 2011), we could not find invertebrate sequences orthologous to MAL, TRIF and TRAM, and therefore these well-known vertebrate intracellular adaptors likely represent evolutionary innovations of the vertebrate lineage.

MyD88 and SARM were not the only cytosolic TIR-DC proteins conserved across many metazoan phyla (Fig. 2). In our survey we identified a high number of adaptors containing the TIR domain associated with other domains and characterized by unusual, and not yet described, domain architectures (Fig. 5). Here, we propose a novel classification scheme based on domain organization and phylogeny (Fig. 6). In detail, each newly described family of cytoplasmic TIR-DC proteins was named "ecTIR-DC" (acronym for "*evolutionarily conserved TIR-domain-containing*" family), followed by a progressive number. Overall, we describe here 15 new families, based on their recurrent presence in bivalve molluscs and, often, in other metazoan groups. Obviously, the present classification has to be considered provisional since more ecTIR-DC families absent in molluscs but conserved in other phyla might be revealed by future studies.

3.7. STING

The Stimulator of Interferon Genes (STING) is a focal cytosolic hub for signals transmitted by several intracellular foreign nucleic acids sensors and for directly sensing bacterial signalling molecules such as cyclic dinucleotides. The STING-mediated signalling then proceeds through the activation of NF-κB and IRF3, eventually triggering the production of pro-inflammatory cytokines and interferon.

Although STING is an evolutionarily ancient protein reported in nearly all metazoans, with a few exceptions such as nematodes and flatworms (Wu et al., 2014), its structure presents remarkable differences across phyla. In vertebrates, STING is a transmembrane protein which is thought to be associated with the endoplasmic reticulum membrane whereas in insects it lacks the N-terminal transmembrane domain and its subcellular localization remains unknown. In bivalve molluscs, STING has a peculiar domain organization, with the duplication of the STING globular domain and the presence of two poorly conserved TIR domains, N-terminal to STING (Fig. 5), which might be involved in downstream signalling (Gerdol and Venier, 2015). As evidenced by the phylogenetic analysis, these two domains are quite divergent between each other, with the first one being also poorly conserved across species (Fig. 6). Our comparative genomics analysis revealed that such a domain architecture is shared by two other lophotrochozoan taxa, polychaetes and brachiopods but, interestingly, not by gastropod and cephalopod molluscs (Fig. 2).

Overall, we could detect STING sequences in several bivalve species (Supplementary File 1). Furthermore, STING is present in multiple gene copies in bivalve genomes, and, based on transcriptome assembly, we could fully confirm two out the four annotated STING genes in the oyster *C. gigas* and five different STING transcripts in the mussel *M. galloprovincialis*.

3.8. ecTIR-DC family 1

The first novel TIR-DC family identified in this comparative survey comprises 400−500 aa long transmembrane proteins characterized by two C-terminal transmembrane domains and a N-terminal (intracellular) TIR domain (Fig. 6). Just a dozen amino acidic residues connect the two alpha-helical regions, leaving a very small portion of the polypeptidic chain exposed to the extracellular medium. As the members of this family completely lack ectodomains, they cannot act as receptors and they may be reasonably involved in the regulation of membrane-bound TIR-DC receptors.

This family was only found in bivalves pertaining to the Pteriomorphia subclass (Supplementary File 1), and in the polychaete worm *C. teleta* (Fig. 2). Furthermore, based on the presence of three distinct genes in *C. gigas*, it appears to be multi-genic.

3.9. ecTIR-DC families 2 and 4

These two protein families share nearly identical domain architecture, with a unique N-terminal TIR domain (Fig. 5), and are closely related based on Bayesian inference (Fig. 6). Family 4 comprises short proteins of ~130–150 aa without any C-terminal extension, whereas family 2 members display C-terminal extensions of variable length, with no recognizable conserved domain or structural fold, as long as ~800 aa in the giant floater *P. grandis*. Family 2 was found to be present as a single copy gene in oyster, as well as in gastropod molluscs (see Fig. 2). On the other hand, family 4 genes are taxonomically restricted to Bivalvia and they further appear to have a spotty distribution in this class, given the absence of an ecTIR-DC 4 gene in *C. gigas*.

3.10. ecTIR-DC family 3

This protein family is structurally characterized by a low complexity N-terminus, followed by ARM repeats and a TIR domain (Fig. 5). In addition, these proteins have a 100 aa-long C-terminal extension which contains a SAM motif, canonical in lower invertebrates and highly degenerated in Lophotrochozoa and basal Deuterostomes. EcTIR-DC 3 genes have an ancient origin, as they were found in early-branching phyla such as Placozoa and Cnidaria. This type of cytosolic TIR-DC proteins, absent in vertebrates, can be found in all protostomes and basal deuterostomes up to echinoderms, amphioxius and tunicates, with the notable exception of Ecdysozoa (Fig. 2). Curiously, they could be also identified in some flatworms, which completely lack TLRs and other membrane-bound TIR-DC proteins, a fact which suggests that their role might be not tied to TLR/IL-1R signalling. In *C. gigas*, these TIR-DC proteins are also expressed at relevant levels in larval stages (see section 3.21). In summary, ecTIR-DC 3 proteins appear to be among the most conserved and widespread intracellular TIR-DC families in invertebrates as well as in bivalves (Supplementary File 1). Four genes pertaining to this family are present in the Pacific oyster.

3.11. ecTIR-DC family 5

The structural organization of ecTIR-DC 5 proteins comprises three distinct TIR domains, with the first and the second one being separated by a SAM motif (Fig. 5). The three TIR domains cluster separately in the phylogenetic tree, even though the third one is highly similar to the TIR domains of ecTIR-DC family 6 and 9 (Fig. 6). EcTIR-DC 5 proteins represent an evolutionary conserved family with a broad taxonomic distribution that likely emerged early in metazoan evolution, as it could be detected in Placozoa, Cnidaria and in many Protostomes (with the exception of Ecdysozoa, Platyhelminthes, Rotifera, Cephalopoda and Clitellata) (Fig. 2). The function of these intracellular proteins might be not connected to TLR/IL-1R signalling, in agreement with their high expression levels during embryonic development (see section 3.21). Although this family is also found in basal deuterostomes, it was likely lost in vertebrates. No more than a single ecTIR-DC 5 gene could be identified in the genomes and transcriptomes of any of the species analysed.

3.12. ecTIR-DC family 6

Members of this family possess a long N-terminal region with a series of Arm repeats, a single central DEATH domain and two consecutive C-terminal TIR domains (Fig. 5) which appear to be duplicated and highly similar to those found in ecTIR-DC family 9 and to the third domain of ecTIR-DC family 5 proteins (Fig. 6). Owing to the relevant size of these proteins (>1000 aa) and their poor expression levels in adult individuals, the full length sequence of only three ecTIR-DC 6 proteins could be retrieved in bivalve molluscs, along with many fragments (Supplementary File 1). However, these TIR-DC proteins show a remarkable evolutionary conservation, from Lophotrochozoa (rotifers, segmented worms and gastropods) to basal deuterostomes (echinoderms, acorn worms and amphioxus) (Fig. 2).

Although their function is unknown, their presence in rotifers and their expression during embryonic development (see section 3.21) suggest that they are not involved in TLR/IL-1R signalling, like in the case of ecTIR-DC families 3 and 5.

3.13. ecTIR-DC family 7

These large proteins (700–1100 aa) display a variable number of N-terminal tetratricopeptide repeats (TPR) followed a C-terminal TIR domain (Fig. 5). EcTIR-DC 7 proteins were mostly detected in Pteriomorphia bivalves (with 2 distinct genes in *C. gigas*), brachiopods, early-branching deuterostomes and, with a very limited sequence homology, also in gastropod and cephalopod molluscs (Fig. 2). As TPRs are often used to build large scaffolds which facilitate the assembly of multiprotein complexes (D'Andrea and Regan, 2003), ecTIR-DC 7 members could be involved in the assembly of intracellular complexes to mediate signal transduction. The TIR domain of ecTIR-DC 7 proteins seems to be somewhat related to that of membrane-bound receptors with EGF ectodomains (Fig. 6).

3.14. ecTIR-DC family 8

EcTIR-DC 8 proteins possess two consecutive TIR domains, highly similar to each other (Fig. 6), and a third one which is degenerated and barely recognizable (Fig. 5). The structural modelling analysis performed with HHpred (Söding et al., 2005) permitted to identify a well-conserved Macro/A1pp domain in the N-terminal region, which might serve as an ADP-ribose binding domain (Neuvonen and Ahola, 2009). Although just a few full-length sequences could be recognized in bivalves, this TIR-DC family appears to be very ancient and conserved across several distantly related phyla, from placozoans, to anthozoans, to the large majority of lophotrochozoans, and it is retained also in hemichordates and echinoderms (Fig. 2). Interestingly, the single-copy TIR-DC 8 gene of *C. gigas* was consistently expressed at high levels since the very early stages of larval development (section 3.21).

3.15. Complex TIR-DC proteins: ecTIR-DC family 9 and 14

Most of the intracellular TIR-DCs described so far are relatively small proteins (<1000 aa). However, a few exceptions exist, as we could identify at least two evolutionarily conserved families of large TIR-DC proteins, namely ecTIR-DC 9 and ecTIR-DC 14.

The former is characterized by a N-terminal Neuralized (PF07177) domain of unknown function, common in proteins involved in neuroblast differentiation, followed by a large central region corresponding to the tandem of ROC (IPR020859) and COR (IPR032171) domains, which characterize a family of complex heterogeneous GTPases with various regulatory functions (Deng et al., 2008). In ecTIR-DC 9 proteins, encoded by a single-copy gene in *C. gigas*, the TIR domain is associated with a DEATH domain and localized at the protein C-terminus (Fig. 5). This family has a relatively wide taxonomical distribution, as it could be identified in cnidarians (*N. vectensis*, despite the absence of the Neuralized domain), molluscs, brachiopods, echinoderms, hemichordates, and in some arthropods (Fig. 2). The TIR domain of

ecTIR-DC family 9 proteins appear to be evolutionarily connected to the third domain of ecTIR-DC family 5 proteins as well as to the two domains of ec-TIR-DC family 6 proteins (Fig. 6).

The ecTIR-DC 14 family is an even more complex group, which comprises very large proteins longer than 3000 amino acids. The two *C. gigas* genes pertaining to this family encode proteins which diverge in the N-terminal region, which is quite disordered and contains different domains: in the first one (ecTIR-DC 14 type I) LRRs are associated to two SPRY domains, while in the second one (ecTIR-DC tye II) LRRs are associated to a C2 domain. Like ecTIR-DC 9, a ROC/COR tandem is also present, followed in this case by two consecutive and degenerated TIR domains. Although the N-terminal region is divergent in other metazoans, the LRR/ROC/COR/2xTIR module appears to be evolutionarily ancient, as it was detected in *T. adherens*, in most lophotrochozoans and in amphioxus, and similar sequences were also found in the flagellate unicellular eukaryotes *Thecamonas trahens* and *Guillardia theta*.

While the function of these TIR-DC proteins is unknown, they were all highly expressed in the early stages of development of oyster larvae and still produced, at lower levels, in most adult tissues (see section 3.21). Although ecTIR-DC sequences were found in many different bivalve species, due to the relevant length of the encoding mRNAs we could mostly detect sequence fragments in *de novo* assembled transcriptomes (Supplementary File 1).

3.16. ecTIR-DC family 10

The members of this TIR-DC family are characterized by a N-terminal TIR domain followed by two BIR (Baculovirus Inhibitor of apoptosis protein Repeat) domains (Fig. 5), that are commonly found in proteins related to apoptotic processes and innate immunity, such as NAIP-type NOD-like receptors. Members of this TIR-DC family were exclusively found in Bivalvia and, more in detail, just in a few distantly related species (Unionoida and Mytiloida, see Supplementary File 1), possibly reflecting a poor constitutive expression or a spotty taxonomic distribution, which is supported by their absence in the *C. gigas* genome. In addition, the TIR domain of ecTIR-DC 10 proteins resulted to be quite divergent across species (Fig. 6), possibly indicating a non-orthologous origin or relaxed evolutionary constraints.

3.17. ecTIR-DC family 11

These proteins possess an N-terminal TIR domain and a series of C-terminal HEAT repeats (Fig. 5). These repeats, which are structurally related to Armadillo repeats (also found in SARM and in the ecTIR-DC 3 and 6 families), could function as protein scaffolding units due to the flexible nature of their tandemly-organized antiparallel alpha-helices (Neuwald and Hirano, 2000).

EcTIR-DC 11 proteins are possibly the most ancient group of metazoan cytosolic TIR-DC proteins. They were indeed found in placozoans, sponges and even in the mesozoan *I. linei* (Fig. 2). Despite their absence in some lineages (cnidarians, flatworms, ecdysozoans, urochordates and vertebrates), they could be consistently observed in almost all Lophotrocozoa and basal deuterostomes. They were also among the most frequently observed TIR-DC sequences in bivalve transcriptomes, irrespective of the two only genes found in *C. gigas* (Supplementary File 1).

Considering their very early acquisition along metazoan evolution and their presence in organisms lacking membrane-bound TIR-DC proteins, ecTIR-DC 11 proteins are unlikely to be involved in TLR/IL-1R -related signalling.

3.18. ecTIR-DC family 12

This family comprises proteins with an N-terminal TIR domain followed by repeated TRAF-type zinc finger domains (Fig. 5). This small gene family was only identified in a handful of bivalve mollusc species (Supplementary File 1), including *C. gigas*, with a single-copy ecTIR-DC 12 gene. In addition, similar proteins are also encoded by the genome of the brachiopod *L. anatina*, but the TIR domain in this case appears to be quite divergent from that found in bivalve proteins. TRAF-type zinc finger domains are typically found in a family of vertebrate signal transducers associated to the TNF-alpha and TLR signalling (Bradley and Pober, 2001), including TRAF3 and TRAF6, two fundamental mediators of the activation of NF-κB also in bivalve molluscs (Toubiana et al., 2014). ecTIR-DC 12 members are therefore very likely to play a role in TLR downstream signalling consistently with the "*shortcuts between endpoints*" hypothesis (Zhang et al., 2008).

3.19. ecTIR-DC family 13

Proteins belonging to the ecTIR-DC 13 family contain two consecutive duplicated TIR domains (Figs. 5 and 6). Only two ecTIR-DC 13 genes were found in the Pacific oyster genome and members of this family could be identified only in a relatively low number of other bivalve species (Supplementary File 1). Due to the low level of homology with other vertebrate TIR-DC proteins and the absence of this domain architecture in other metazoans, the function of the ecTIR-DC family 13 is currently unknown, even though the particularly high expression from the oyster egg to blastula stages (see Section 3.21) is intriguing for what concerns a possible involvement in larval development.

3.20. ecTIR-DC family 15

The members of the ecTIR-DC family 15 have a domain architecture similar to that of ecTIR-DC family 13 but in this case the two TIR domains are phylogenetically distantly related (Fig. 6) and the proteins display longer N-terminal and C-terminal regions with no recognizable conserved domains (Fig. 5). However, structural modelling revealed that the C-terminal region is very likely to comprise a Rossman fold, which would suggest dinucleotide-binding properties (Hanukoglu, 2015). EcTIR-DC 15 genes are relatively widespread across metazoans, as they were identified in sponges, placozoans, most lophotrochozoans, some ecdysozoans (priapulids, tardigrades, arachnids and horseshoe crabs), hemichordates and echinoderms (Asterozoa, but not Echinozoa) (Fig. 2). The single copy gene identified in the Pacific oyster displayed a nearly ubiquitous expression in all tissues and developmental stages (Fig. 7).

3.21. Preliminary assessment of TIR-DC gene expression in *Crassostrea gigas*

We evaluated the expression trends of the previously identified *C. gigas* TIR-DC genes across a broad range of adult tissues in unchallenged animals and developmental stages with an RNA-seq analysis using the data generated by Zhang et al. (2012) (Supplementary File 1).

Like previously observed in mussel (Toubiana et al., 2013), most oyster TLRs appeared to be broadly expressed in all adult oyster tissues. However, significant transcripts levels of a few TLRs were found in hemocytes, including a P-type TLR (subtype I) and some sP-type TLRs, altogether bringing the cumulative expression of TLRs to their highest level in these circulating cells (Fig. 7). With the only exception of a P-type TLR pertaining to subtype 2, highly expressed

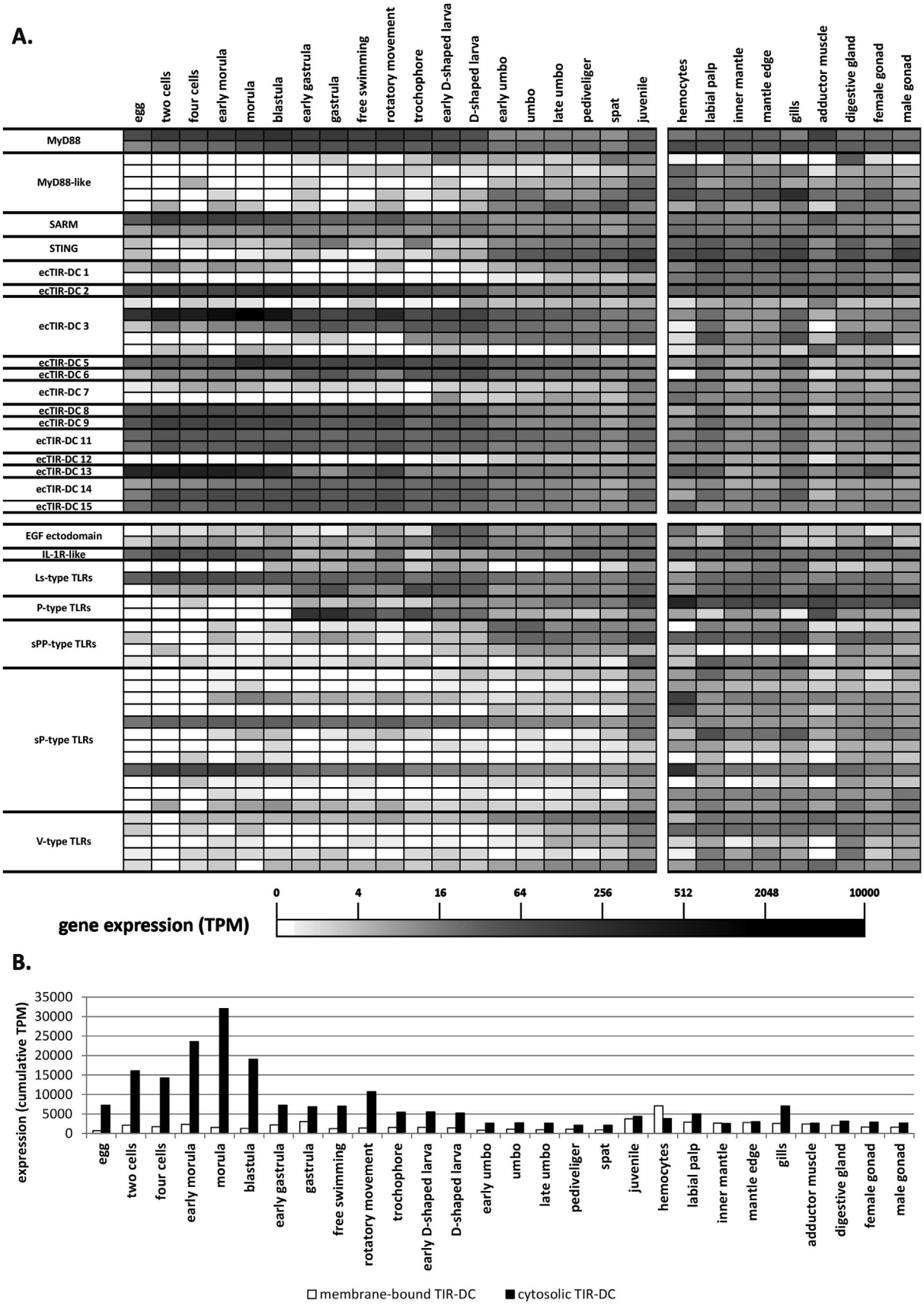

Fig. 7. Panel A: heat map summarizing the expression levels of cytosolic TIR-DCs (upper part) and membrane-bound TIR-DC proteins (lower part) throughout the larval development and in nine tissues collected from unchallenged adult *C. gigas* individuals. Panel B: cumulative expression levels (as TPM, Transcript Per Million) of cytosolic TIR-DC genes throughout larval development and in nine tissues of adult unchallenged oysters.

from the early gastrula to the D-shaped larva stage, TLRs displayed generally low expression levels throughout the different stages of embryonic development. In particular sP-type TLRs, the most expanded and diversified TLR group in bivalves (Fig. 4) were mostly undetectable up to the juvenile stage, *de facto* ruling out their possible involvement in larval development, similar to that of Toll in *Drosophila* (Hashimoto et al., 1988). Concerning the other membrane-bound TIR-DC proteins, those bearing EGF ectodomains, albeit being also expressed in adult tissues, reached their peak of expression at the D-shaped larvae stage. The only IL-1RL oyster gene was expressed at constant and relatively high levels in all developmental stages and adult tissues.

The most striking result of this preliminary assessment of gene expression was the unexpectedly high expression of many cytoplasmic TIR-DC genes during embryonic development, starting from the very early stages (two-cells stage) up to the blastula. This was mostly ascribable to the ecTIR-DC protein families 2, 3, 5, 6, 8, 9, 11, 13, 14 and 15 (Fig. 7), but also to MyD88 and SARM sequences. Such a strong expression was not coupled with an increase in the expression of TLRs, suggesting functions independent from those of membrane-bound TIR-DC proteins. However, while members of the above mentioned protein families were mostly expressed during the early stages of embryonic development, they also retained a significant level of expression in most adult tissues (Fig. 7), thus indicating a probable involvement in the functionality of adult cells in addition to development.

4. Conclusions

Our large scale analysis confirmed previous indications about the expansion of the TLR gene family in bivalve molluscs (Zhang et al., 2015). At the present time there is no indication about whether the expansion and structural diversification among different TLRs subfamilies is paired by functional diversification, potentially resulting in a broader spectrum ligands. Although the main function of these rapidly evolving and highly diversified receptors has not yet been unequivocally associated to immune recognition in molluscs and in other invertebrates, a growing body of evidence suggests that at least some of them are directly involved in PAMP binding and related immune signalling (Lu et al., 2016; Pila et al., 2016; Toubiana et al., 2014; Zhang et al., 2015, 2013). Moreover, additional studies are needed to clarify whether any of the bivalve TLRs are localized in the endosome membrane, which might indicate a role in the recognition of foreign nucleic acids (Gerdol and Venier, 2015). The impressive expansion of certain TLR subfamilies (i.e. sP-type TLRs) certainly suggests that selective forces are acting to enhance the diversification of these receptors, in a similar fashion to other bivalve immune-related gene families (Gerdol et al., 2015b).

Besides the massive expansion of TLRs, another fascinating aspect is represented by the presence of several previously uncharacterized cytoplasmic TIR-DC families which are conserved, to some extent, across several metazoan phyla. We have proposed a novel classification of such previously defined “orphan TIR” proteins, eventually describing 15 novel families of TIR-DC cytosolic proteins sharing common domain architecture and displaying, in some cases, a very ancient origin. The broad taxonomical distribution of some of these families further suggests that they might be part of important and well conserved signalling pathways common to most metazoan phyla, that were likely secondarily lost in the vertebrate lineage. While some of these protein families do not appear to be linked with TLR/IL-1R signalling, others could cover a role similar to that of vertebrate MAL, TRIF and TRAM, i.e. intracellular adaptors conveying the signal transmitted by membrane-bound receptors to cytosolic protein scaffolds involved in the activation of the proinflammatory signalling cascade. Equally important was the observation of the unexpected high expression of cytosolic TIR-DC transcripts throughout oyster embryonic development, which leaves many open questions about their possible involvement in the determination of embryonic polarity and in the development of adult organs. Given the nearly complete absence of functional data for most of these protein families, well designed experiments should be planned in the future to investigate in detail the expression patterns of TIR-DC genes in challenged animals.

Considering that most of the bivalve TIR-DC proteins reported in this study are derived from the virtual translation of assembled transcriptome data, mostly obtained from adult naïve specimens, the data presented in this work is far from being complete, but they aim to be a comparative basis of knowledge for future studies. We only took into consideration full-length proteins (according to TransDecoder predictions), disregarding those which were incomplete due to the fragmentation of the corresponding transcript, which is a common issue in the assembly of RNA-seq data. Overall, this procedure might have introduced a bias of detection towards relatively short proteins, since long transcripts can suffer from increased fragmentation due to a drop of sequencing coverage towards the 3' region. However, we implemented a protocol to identify such cases, which contributed to obtain a more comprehensive overview in bivalve molluscs (Supplementary File 1). While a complete census of all the TIR-DC genes present in all animal species in not feasible at the moment, due to the limited genomic resources available, we believe that our conservative investigation strategy, applied to the transcriptome of over 80 bivalve species and to many reference metazoan genomes, was optimal to highlight evolutionary conserved TIR-DC proteins and we offer the present work as a contribution in the study of the evolution and classification of TIR domain-containing proteins in metazoans for future studies.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 678589.

Appendix A. Supplementary data

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinforma. Oxf. Engl. 21, 2104–2105. http://dx.doi.org/ 10.1093/bioinformatics/bti263.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B.,

McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Sidén-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., WoodageT, null, Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., Venter, J.C., 2000. The genome sequence of Drosophila melanogaster. Science 287, 2185–2195.

Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., Rokhsar, D.S., 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature 524, 220–224. http://dx.doi.org/10.1038/nature14668.

Bányai, L., Patthy, L., 2016. Putative extremely high rate of proteome innovation in lancelets might be explained by high rate of gene prediction errors. Sci. Rep. 6, 30700. http://dx.doi.org/10.1038/srep30700.

Baumgarten, S., Simakov, O., Esherick, L.Y., Liew, Y.J., Lehnert, E.M., Michell, C.T., Li, Y., Hambleton, E.A., Guse, A., Oates, M.E., Gough, J., Weis, V.M., Aranda, M., Pringle, J.R., Voolstra, C.R., 2015. The genome of Aiptasia, a sea anemone model for coral symbiosis. Proc. Natl. Acad. Sci. 112, 11893–11898. http://dx.doi.org/10.1073/pnas.1513318112.

Beck, G., Ellis, T.W., Truong, N., 2000. Characterization of an IL-1 receptor from Asterias forbesi coelomocytes. Cell. Immunol. 203, 66–73. http://dx.doi.org/10.1006/cimm.2000.1674.

Belinda, L.W.-C., Wei, W.X., Hanh, B.T.H., Lei, L.X., Bow, H., Ling, D.J., 2008. SARM: a novel Toll-like receptor adaptor, is functionally conserved from arthropod to human. Mol. Immunol. 45, 1732–1742. http://dx.doi.org/10.1016/j.molimm.2007.09.030.

Berriman, M., Haas, B.J., LoVerde, P.T., Wilson, R.A., Dillon, G.P., Cerqueira, G.C., Mashiyama, S.T., Al-Lazikani, B., Andrade, L.F., Ashton, P.D., Aslett, M.A., Bartholomeu, D.C., Blandin, G., Caffrey, C.R., Coghlan, A., Coulson, R., Day, T.A., Delcher, A., DeMarco, R., Djikeng, A., Eyre, T., Gamble, J.A., Ghedin, E., Gu, Y., Hertz-Fowler, C., Hirai, H., Hirai, Y., Houston, R., Ivens, A., Johnston, D.A., Lacerda, D., Macedo, C.D., McVeigh, P., Ning, Z., Oliveira, G., Overington, J.P., Parkhill, J., Pertea, M., Pierce, R.J., Protasio, A.V., Quail, M.A., Rajandream, M.-A., Rogers, J., Sajid, M., Salzberg, S.L., Stanke, M., Tivey, A.R., White, O., Williams, D.L., Wortman, J., Wu, W., Zamanian, M., Zerlotini, A., Fraser-Liggett, C.M., Barrell, B.G., El-Sayed, N.M., 2009. The genome of the blood fluke Schistosoma mansoni. Nature 460, 352–358. http://dx.doi.org/10.1038/nature08160.

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

Boothby, T.C., Tenlen, J.R., Smith, F.W., Wang, J.R., Patanella, K.A., Nishimura, E.O., Tintori, S.C., Li, Q., Jones, C.D., Yandell, M., Messina, D.N., Glasscock, J., Goldstein, B., 2015. Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade, 201510461 Proc. Natl. Acad. Sci.. http://dx.doi.org/10.1073/pnas.1510461112.

Bork, P., Downing, A.K., Kieffer, B., Campbell, I.D., 1996. Structure and distribution of modules in extracellular proteins. Q. Rev. Biophys. 29, 119–167. http://dx.doi.org/10.1017/S0033583500005783.

Bradley, J.R., Pober, J.S., 2001. Tumor necrosis factor receptor-associated factors (TRAFs). Oncogene 20, 6482–6491. http://dx.doi.org/10.1038/sj.onc.1204788.

Buckley, K.M., Rast, J.P., 2012. Dynamic evolution of toll-like receptor multigene families in echinoderms. Front. Immunol. 3, 136. http://dx.doi.org/10.3389/fimmu.2012.00136.

C. elegans Sequencing Consortium, 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282, 2012–2018.

Carty, M., Goodbody, R., Schröder, M., Stack, J., Moynagh, P.N., Bowie, A.G., 2006. The human adaptor SARM negatively regulates adaptor protein TRIF-dependent Toll-like receptor signaling. Nat. Immunol 7, 1074–1081. http://dx.doi.org/10.1038/ni1382.

Chang, E.S., Neuhof, M., Rubinstein, N.D., Diamant, A., Philippe, H., Huchon, D., Cartwright, P., 2015. Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. Proc. Natl. Acad. Sci. 112, 14912–14917. http://dx.doi.org/10.1073/pnas.1511468112.

Chapman, J.A., Kirkness, E.F., Simakov, O., Hampson, S.E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P.G., Borman, J., Busam, D., Disbennett, K., Pfannkoch, C., Sumin, N., Sutton, G.G., Viswanathan, L.D., Walenz, B., Goodstein, D.M., Hellsten, U., Kawashima, T., Prochnik, S.E., Putnam, N.H., Shu, S., Blumberg, B., Dana, C.E., Gee, L., Kibler, D.F., Law, L., Lindgens, D., Martinez, D.E., Peng, J., Wigge, P.A., Bertulat, B., Guder, C., Nakamura, Y., Ozbek, S., Watanabe, H., Khalturin, K., Hemmrich, G., Franke, A., Augustin, R., Fraune, S., Hayakawa, E., Hayakawa, S., Hirose, M., Hwang, J.S., Ikeo, K., Nishimiya-Fujisawa, C., Ogura, A., Takahashi, T., Steinmetz, P.R.H., Zhang, X., Aufschnaiter, R., Eder, M.-K., Gorny, A.-K., Salvenmoser, W., Heimberg, A.M., Wheeler, B.M., Peterson, K.J., Böttger, A., Tischler, P., Wolf, A., Gojobori, T., Remington, K.A., Strausberg, R.L., Venter, J.C., Technau, U., Hobmayer, B., Bosch, T.C.G., Holstein, T.W., Fujisawa, T., Bode, H.R., David, C.N., Rokhsar, D.S., Steele, R.E., 2010. The dynamic genome of Hydra. Nature 464, 592–596. http://dx.doi.org/10.1038/nature08830.

Chipman, A.D., Ferrier, D.E.K., Brena, C., Qu, J., Hughes, D.S.T., Schröder, R., Torres-Oliva, M., Znassi, N., Jiang, H., Almeida, F.C., Alonso, C.R., Apostolou, Z., Aqrawi, P., Arthur, W., Barna, J.C.J., Blankenburg, K.P., Brites, D., Capella-Gutiérrez, S., Coyle, M., Dearden, P.K., Du Pasquier, L., Duncan, E.J., Ebert, D., Eibner, C., Erikson, G., Evans, P.D., Extavour, C.G., Francisco, L., Gabaldón, T., Gillis, W.J., Goodwin-Horn, E.A., Green, J.E., Griffiths-Jones, S., Grimmelikhuijzen, C.J.P., Gubbala, S., Guigó, R., Han, Y., Hauser, F., Havlak, P., Hayden, L., Helbing, S., Holder, M., Hui, J.H.L., Hunn, J.P., Hunnekuhl, V.S., Jackson, L., Javaid, M., Jhangiani, S.N., Jiggins, F.M., Jones, T.E., Kaiser, T.S., Kalra, D., Kenny, N.J., Korchina, V., Kovar, C.L., Kraus, F.B., Lapraz, F., Lee, S.L., Lv, J., Mandapat, C., Manning, G., Mariotti, M., Mata, R., Mathew, T., Neumann, T., Newsham, I., Ngo, D.N., Ninova, M., Okwuonu, G., Ongeri, F., Palmer, W.J., Patil, S., Patraquim, P., Pham, C., Pu, L.-L., Putman, N.H., Rabouille, C., Ramos, O.M., Rhodes, A.C., Robertson, H.E., Robertson, H.M., Ronshaugen, M., Rozas, J., Saada, N., Sánchez-Gracia, A., Scherer, S.E., Schurko, A.M., Siggens, K.W., Simmons, D., Stief, A., Stolle, E., Telford, M.J., Tessmar-Raible, K., Thornton, R., van der Zee, M., von Haeseler, A., Williams, J.M., Willis, J.H., Wu, Y., Zou, X., Lawson, D., Muzny, D.M., Worley, K.C., Gibbs, R.A., Akam, M., Richards, S., 2014. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede Strigamia maritima. PLoS Biol. 12, e1002005. http://dx.doi.org/10.1371/journal.pbio.1002005.

Chuang, C.-F., Bargmann, C.I., 2005. A Toll-interleukin 1 repeat protein at the synapse specifies asymmetric odorant receptor expression via ASK1 MAPKKK signaling. Genes Dev. 19, 270–281. http://dx.doi.org/10.1101/gad.1276505.

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., Tokishita, S., Aerts, A., Arnold, G.J., Basu, M.K., Bauer, D.J., Caceres, C.E., Carmel, L., Casola, C., Choi, J.-H., Detter, J.C., Dong, Q., Dusheyko, S., Eads, B.D., Frohlich, T., Geiler-Samerotte, K.A., Gerlach, D., Hatcher, P., Jogdeo, S., Krijgsveld, J., Kriventseva, E.V., Kultz, D., Laforsch, C., Lindquist, E., Lopez, J., Manak, J.R., Muller, J., Pangilinan, J., Patwardhan, R.P., Pitluck, S., Pritham, E.J., Rechtsteiner, A., Rho, M., Rogozin, I.B., Sakarya, O., Salamov, A., Schaack, S., Shapiro, H., Shiga, Y., Skalitzky, C., Smith, Z., Souvorov, A., Sung, W., Tang, Z., Tsuchiya, D., Tu, H., Vos, H., Wang, M., Wolf, Y.I., Yamagata, H., Yamada, T., Ye, Y., Shaw, J.R., Andrews, J., Crease, T.J., Tang, H., Lucas, S.M., Robertson, H.M., Bork, P., Koonin, E.V., Zdobnov, E.M., Grigoriev, I.V., Lynch, M., Boore, J.L., 2011. The ecoresponsive genome of Daphnia pulex. Science 331, 555–561. http://dx.doi.org/10.1126/science.1197761.

Couillault, C., Pujol, N., Reboul, J., Sabatier, L., Guichou, J.-F., Kohara, Y., Ewbank, J.J., 2004. TLR-independent control of innate immunity in Caenorhabditis elegans by the TIR domain adaptor protein TIR-1, an ortholog of human SARM. Nat. Immunol. 5, 488–494. http://dx.doi.org/10.1038/ni1060.

Danks, G., Campsteijn, C., Parida, M., Butcher, S., Doddapaneni, H., Fu, B., Petrin, R., Metpally, R., Lenhard, B., Wincker, P., Chourrout, D., Thompson, E.M., Manak, J.R., 2013. OikoBase: a genomics and developmental transcriptomics resource for the urochordate Oikopleura dioica. Nucleic Acids Res. 41, D845–D853. http://dx.doi.org/10.1093/nar/gks1159.

Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., Tomaso, A.D., Davidson, B., Gregorio, A.D., Gelpke, M., Goodstein, D.M., Harafuji, N., Hastings, K.E.M., Ho, I., Hotta, K., Huang, W., Kawashima, T., Lemaire, P., Martinez, D., Meinertzhagen, I.A., Necula, S., Nonaka, M., Putnam, N., Rash, S., Saiga, H., Satake, M., Terry, A., Yamada, L., Wang, H.-G., Awazu, S., Azumi, K., Boore, J., Branno, M., Chin-bow, S., DeSantis, R., Doyle, S., Francino, P., Keys, D.N., Haga, S., Hayashi, H., Hino, K., Imai, K.S., Inaba, K., Kano, S., Kobayashi, K., Kobayashi, M., Lee, B.-I., Makabe, K.W., Manohar, C., Matassi, G., Medina, M., Mochizuki, Y., Mount, S., Morishita, T., Miura, S., Nakayama, A., Nishizaka, S., Nomoto, H., Ohta, F., Oishi, K., Rigoutsos, I., Sano, M., Sasaki, A., Sasakura, Y., Shoguchi, E., Shin-i, T., Spagnuolo, A., Stainier, D., Suzuki, M.M., Tassy, O., Takatori, N., Tokuoka, M., Yagi, K., Yoshizaki, F., Wada, S., Zhang, C., Hyatt, P.D., Larimer, F., Detter, C., Doggett, N., Glavina, T., Hawkins, T., Richardson, P., Lucas, S., Kohara, Y., Levine, M., Satoh, N., Rokhsar, D.S., 2002. The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. Science 298, 2157–2167. http://dx.doi.org/10.1126/science.1080049.

Deng, J., Lewis, P.A., Greggio, E., Sluch, E., Beilina, A., Cookson, M.R., 2008. Structure of the ROC domain from the Parkinson's disease-associated leucine-rich repeat kinase 2 reveals a dimeric GTPase. Proc. Natl. Acad. Sci. U. S. A. 105, 1499–1504. http://dx.doi.org/10.1073/pnas.0709098105.

Dishaw, L.J., Haire, R.N., Litman, G.W., 2012. The amphioxus genome provides unique insight into the evolution of immunity. Brief. Funct. Genom. 11, 167–176. http://dx.doi.org/10.1093/bfgp/els007.

Dunne, A., Ejdebäck, M., Ludidi, P.L., O'Neill, L.A.J., Gay, N.J., 2003. Structural complementarity of toll/interleukin-1 receptor domains in toll-like receptors and the adaptors mal and MyD88. J. Biol. Chem. 278, 41443–41451. http://dx.doi.org/10.1074/jbc.M301742200.

D'Andrea, L.D., Regan, L., 2003. TPR proteins: the versatile helix. Trends Biochem. Sci. 28, 655–662. http://dx.doi.org/10.1016/j.tibs.2003.10.007.

Eddy, S.R., 1998. Profile hidden Markov models. Bioinformatics 14, 755–763. http://dx.doi.org/10.1093/bioinformatics/14.9.755.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. http://dx.doi.org/10.1093/nar/gkh340.

Fields, P.A., Eurich, C., Gao, W.L., Cela, B., 2014. Changes in protein expression in the salt marsh mussel Geukensia demissa: evidence for a shift from anaerobic to aerobic metabolism during prolonged aerial exposure. J. Exp. Biol. 217, 1601–1612. http://dx.doi.org/10.1242/jeb.101758.

Flot, J.-F., Hespeels, B., Li, X., Noel, B., Arkhipova, I., Danchin, E.G.J., Hejnol, A.,

Henrissat, B., Koszul, R., Aury, J.-M., Barbe, V., Barthélémy, R.-M., Bast, J., Bazykin, G.A., Chabrol, O., Couloux, A., Da Rocha, M., Da Silva, C., Gladyshev, E., Gouret, P., Hallatschek, O., Hecox-Lea, B., Labadie, K., Lejeune, B., Piskurek, O., Poulain, J., Rodriguez, F., Ryan, J.F., Vakhrusheva, O.A., Wajnberg, E., Wirth, B., Yushenova, I., Kellis, M., Kondrashov, A.S., Mark Welch, D.B., Pontarotti, P., Weissenbach, J., Wincker, P., Jaillon, O., Van Doninck, K., 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga. Nature 500, 453–457. http://dx.doi.org/10.1038/nature12326.

Franzenburg, S., Fraune, S., Künzel, S., Baines, J.F., Domazet-Loso, T., Bosch, T.C.G., 2012. MyD88-deficient Hydra reveal an ancient function of TLR signaling in sensing bacterial colonizers. Proc. Natl. Acad. Sci. U. S. A. 109, 19374–19379. http://dx.doi.org/10.1073/pnas.1213110109.

Gauthier, M.E.A., Du Pasquier, L., Degnan, B.M., 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. http://dx.doi.org/10.1111/j.1525-142X.2010.00436.x.

Gerdol, M., Venier, P., 2015. An updated molecular basis for mussel immunity. Fish. Shellfish Immunol. 46, 17–38. http://dx.doi.org/10.1016/j.fsi.2015.02.013.

Gerdol, M., De Moro, G., Venier, P., Pallavicini, A., 2015a. Analysis of synonymous codon usage patterns in sixty-four different bivalve species. PeerJ 3, e1520. http://dx.doi.org/10.7717/peerj.1520.

Gerdol, M., Venier, P., Pallavicini, A., 2015b. 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. 49, 59–71. http://dx.doi.org/10.1016/j.dci.2014.11.007.

Hahn, C., Fromm, B., Bachmann, L., 2014. Comparative genomics of flatworms (platyhelminthes) reveals shared genomic features of ecto- and endoparasitic neodermata. Genome Biol. Evol. 6, 1105–1117. http://dx.doi.org/10.1093/gbe/evu078.

Hanukoglu, I., 2015. Proteopedia: Rossmann fold: a beta-alpha-beta fold at dinucleotide binding sites. Biochem. Mol. Biol. Educ. Bimon. Publ. Int. Union Biochem. Mol. Biol. 43, 206–209. http://dx.doi.org/10.1002/bmb.20849.

Hashimoto, C., Hudson, K.L., Anderson, K.V., 1988. The Toll gene of Drosophila, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. Cell 52, 269–279.

Hashimoto, T., Horikawa, D.D., Saito, Y., Kuwahara, H., Kozuka-Hata, H., Shin-I, T., Minakuchi, Y., Ohishi, K., Motoyama, A., Aizu, T., Enomoto, A., Kondo, K., Tanaka, S., Hara, Y., Koshikawa, S., Sagara, H., Miura, T., Yokobori, S., Miyagawa, K., Suzuki, Y., Kubo, T., Oyama, M., Kohara, Y., Fujiyama, A., Arakawa, K., Katayama, T., Toyoda, A., Kunieda, T., 2016. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. Nat. Commun. 7, 12808. http://dx.doi.org/10.1038/ncomms12808.

He, Y., Jouaux, A., Ford, S.E., Lelong, C., Sourdaine, P., Mathieu, M., Guo, X., 2015. Transcriptome analysis reveals strong and complex antiviral response in a mollusc. Fish. Shellfish Immunol. http://dx.doi.org/10.1016/j.fsi.2015.05.023.

Hibino, T., Loza-Coll, M., Messier, C., Majeske, A.J., Cohen, A.H., Terwilliger, D.P., Buckley, K.M., Brockton, V., Nair, S.V., Berney, K., Fugmann, S.D., Anderson, M.K., Pancer, Z., Cameron, R.A., Smith, L.C., Rast, J.P., 2006. The immune gene repertoire encoded in the purple sea urchin genome. Dev. Biol. 300, 349–365. http://dx.doi.org/10.1016/j.ydbio.2006.08.065.

Horng, T., Barton, G.M., Medzhitov, R., 2001. TIRAP: an adapter molecule in the Toll signaling pathway. Nat. Immunol. 2, 835–841. http://dx.doi.org/10.1038/ni0901-835.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J.C., Koch, R., Rauch, G.-J., White, S., Chow, W., Kilian, B., Quintais, L.T., Guerra-Assunção, J.A., Zhou, Y., Gu, Y., Yen, J., Vogel, J.-H., Eyre, T., Redmond, S., Banerjee, R., Chi, J., Fu, B., Langley, E., Maguire, S.F., Laird, G.K., Lloyd, D., Kenyon, E., Donaldson, S., Sehra, H., Almeida-King, J., Loveland, J., Trevanion, S., Jones, M., Quail, M., Willey, D., Hunt, A., Burton, J., Sims, S., McLay, K., Plumb, B., Davis, J., Clee, C., Oliver, K., Clark, R., Riddle, C., Elliott, D., Threadgold, G., Harden, G., Ware, D., Begum, S., Beverley, Mortimore, Kerry, G., Heath, P., Phillimore, B., Tracey, A., Corby, N., Dunn, M., Johnson, C., Wood, J., Clark, S., Pelan, S., Griffiths, G., Smith, M., Glithero, R., Howden, P., Barker, N., Christine, Lloyd, Stevens, C., Harley, J., Holt, K., Panagiotidis, G., Lovell, J., Beasley, H., Henderson, C., Gordon, D., Auger, K., Wright, D., Collins, J., Raisen, C., Dyer, L., Leung, K., Robertson, L., Ambridge, K., Leongamornlert, D., McGuire, S., Gilderthorp, R., Griffiths, C., Manthravadi, D., Nichol, S., Barker, G., Whitehead, S., Kay, M., Brown, J., Murnane, C., Gray, E., Humphries, M., Sycamore, N., Barker, D., Saunders, D., Wallis, J., Babbage, A., Hammond, S., Mashreghi-Mohammadi, M., Barr, L., Martin, S., Wray, P., Ellington, A., Matthews, N., Ellwood, M., Woodmansey, R., Clark, G., Cooper, J.D., Tromans, A., Grafham, D., Skuce, C., Pandian, R., Andrews, R., Harrison, E., Kimberley, A., Garnett, J., Fosker, N., Hall, R., Garner, P., Kelly, D., Bird, C., Palmer, S., Gehring, I., Berger, A., Dooley, C.M., Ersan-Ürün, Z., Eser, C., Geiger, H., Geisler, M., Karotki, L., Kirn, A., Konantz, J., Konantz, M., Oberländer, M., Rudolph-Geiger, S., Teucke, M., Lanz, C., Raddatz, G., Osoegawa, K., Zhu, B., Rapp, A., Widaa, S., Langford, C., Yang, F., Schuster, S.C., Carter, N.P., Harrow, J., Ning, Z., Herrero, J., Searle, S.M.J., Enright, A., Geisler, R., Plasterk, R.H.A., Lee, C., Westerfield, M., Jong, P.J., de, Zon, L.I., Postlethwait, J.H., Nüsslein-Volhard, C., Hubbard, T.J.P., Crollius, H.R., Rogers, J., Stemple, D.L., 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature 496, 498–503. http://dx.doi.org/10.1038/nature12111.

Huang, S., Yuan, S., Guo, L., Yu, Y., Li, J., Wu, T., Liu, T., Yang, M., Wu, K., Liu, H., Ge, J., Yu, Y., Huang, H., Dong, M., Yu, C., Chen, S., Xu, A., 2008. Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. Genome Res. 18, 1112–1126. http://dx.doi.org/10.1101/gr.069674.107.

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

Hughes, A.L., Piontkivska, H., 2008. Functional diversification of the toll-like receptor gene family. Immunogenetics 60, 249–256. http://dx.doi.org/10.1007/s00251-008-0283-5.

Hurvich, C.M., Tsai, C.-L., 1989. Regression and time series model selection in small samples. Biometrika 76, 297–307. http://dx.doi.org/10.1093/biomet/76.2.297.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinforma. Oxf. Engl. 30, 1236–1240. http://dx.doi.org/10.1093/bioinformatics/btu031.

Käll, L., Krogh, A., Sonnhammer, E.L.L., 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338, 1027–1036. http://dx.doi.org/10.1016/j.jmb.2004.03.016.

Kim, C.A., Bowie, J.U., 2003. SAM domains: uniform structure, diversity of function. Trends Biochem. Sci. 28, 625–628. http://dx.doi.org/10.1016/j.tibs.2003.11.001.

Kirkness, E.F., Haas, B.J., Sun, W., Braig, H.R., Perotti, M.A., Clark, J.M., Lee, S.H., Robertson, H.M., Kennedy, R.C., Elhaik, E., Gerlach, D., Kriventseva, E.V., Elsik, C.G., Graur, D., Hill, C.A., Veenstra, J.A., Walenz, B., Tubío, J.M.C., Ribeiro, J.M.C., Rozas, J., Johnston, J.S., Reese, J.T., Popadic, A., Tojo, M., Raoult, D., Reed, D.L., Tomoyasu, Y., Kraus, E., Mittapalli, O., Margam, V.M., Li, H.-M., Meyer, J.M., Johnson, R.M., Romero-Severson, J., VanZee, J.P., Alvarez-Ponce, D., Vieira, F.G., Aguadé, M., Guirao-Rico, S., Anzola, J.M., Yoon, K.S., Strycharz, J.P., Unger, M.F., Christley, S., Lobo, N.F., Seufferheld, M.J., Wang, N., Dasch, G.A., Struchiner, C.J., Madey, G., Hannick, L.I., Bidwell, S., Joardar, V., Caler, E., Shao, R., Barker, S.C., Cameron, S., Bruggner, R.V., Regier, A., Johnson, J., Viswanathan, L., Utterback, T.R., Sutton, G.G., Lawson, D., Waterhouse, R.M., Venter, J.C., Strausberg, R.L., Berenbaum, M.R., Collins, F.H., Zdobnov, E.M., Pittendrigh, B.R., 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc. Natl. Acad. Sci. 107, 12168–12173. http://dx.doi.org/10.1073/pnas.1003379107.

Lee, Y., Whang, I., Umasuthan, N., De Zoysa, M., Oh, C., Kang, D.-H., Choi, C.Y., Park, C.-J., Lee, J., 2011. Characterization of a novel molluscan MyD88 family protein from manila clam, Ruditapes philippinarum. Fish. Shellfish Immunol. 31, 887–893. http://dx.doi.org/10.1016/j.fsi.2011.08.003.

Lu, Y., Zheng, H., Zhang, H., Yang, J., Wang, Q., 2016. Cloning and differential expression of a novel toll-like receptor gene in noble scallop Chlamys nobilis with different total carotenoid content. Fish. Shellfish Immunol. 56, 229–238. http://dx.doi.org/10.1016/j.fsi.2016.07.007.

Luo, C., Zheng, L., 2000. Independent evolution of Toll and related genes in insects and mammals. Immunogenetics 51, 92–98.

Luo, Y.-J., Takeuchi, T., Koyanagi, R., Yamada, L., Kanda, M., Khalturina, M., Fujie, M., Yamasaki, S., Endo, K., Satoh, N., 2015. The Lingula genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. Nat. Commun. 6, 8301. http://dx.doi.org/10.1038/ncomms9301.

Miller, D.J., Hemmrich, G., Ball, E.E., Hayward, D.C., Khalturin, K., Funayama, N., Agata, K., Bosch, T.C.G., 2007. The innate immune repertoire in cnidaria–ancestral complexity and stochastic gene loss. Genome Biol. 8, R59. http://dx.doi.org/10.1186/gb-2007-8-4-r59.

Mitreva, M., Jasmer, D.P., Zarlenga, D.S., Wang, Z., Abubucker, S., Martin, J., Taylor, C.M., Yin, Y., Fulton, L., Minx, P., Yang, S.-P., Warren, W.C., Fulton, R.S., Bhonagiri, V., Zhang, X., Hallsworth-Pepin, K., Clifton, S.W., McCarter, J.P., Appleton, J., Mardis, E.R., Wilson, R.K., 2011. The draft genome of the parasitic nematode Trichinella spiralis. Nat. Genet. 43, 228–235. http://dx.doi.org/10.1038/ng.769.

Muzio, M., Ni, J., Feng, P., Dixit, V.M., 1997. IRAK (Pelle) family member IRAK-2 and MyD88 as proximal mediators of IL-1 signaling. Science 278, 1612–1615.

Neuvonen, M., Ahola, T., 2009. Differential activities of cellular and viral macro domain proteins in binding of ADP-ribose metabolites. J. Mol. Biol. 385, 212–225. http://dx.doi.org/10.1016/j.jmb.2008.10.045.

Neuwald, A.F., Hirano, T., 2000. HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. Genome Res. 10, 1445–1452. http://dx.doi.org/10.1101/gr.147400.

Ning, X., Wang, R., Li, X., Wang, S., Zhang, M., Xing, Q., Sun, Y., Wang, S., Zhang, L., Hu, X., Bao, Z., 2015. Genome-wide identification and characterization of five MyD88 duplication genes in Yesso scallop (Patinopecten yessoensis) and expression changes in response to bacterial challenge. Fish. Shellfish Immunol. 46, 181–191. http://dx.doi.org/10.1016/j.fsi.2015.06.028.

Nossa, C.W., Havlak, P., Yue, J.-X., Lv, J., Vincent, K.Y., Brockmann, H.J., Putnam, N.H., 2014. Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. GigaScience 3, 9. http://dx.doi.org/10.1186/2047-217X-3-9.

O'Neill, L. a. J., 2003. The role of MyD88-like adapters in Toll-like receptor signal transduction. Biochem. Soc. Trans. 31, 643–647. http://dx.doi.org/10.1042/.

O'Neill, L.A.J., Bowie, A.G., 2007. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. Nat. Rev. Immunol. 7, 353–364. http://dx.doi.org/10.1038/nri2079.

Pila, E.A., Tarrabain, M., Kabore, A.L., Hanington, P.C., 2016. A novel toll-like receptor (TLR) influences compatibility between the gastropod Biomphalaria glabrata, and the Digenean Trematode Schistosoma mansoni. PLoS Pathog. 12, e1005513.

http://dx.doi.org/10.1371/journal.ppat.1005513.
Poole, A.Z., Weis, V.M., 2014. TIR-domain-containing protein repertoire of nine anthozoan species reveals coral-specific expansions and uncharacterized proteins. Dev. Comp. Immunol. 46, 480–488. http://dx.doi.org/10.1016/j.dci.2014.06.002.
Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V.V., Jurka, J., Genikhovich, G., Grigoriev, I.V., Lucas, S.M., Steele, R.E., Finnerty, J.R., Technau, U., Martindale, M.Q., Rokhsar, D.S., 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317, 86–94. http://dx.doi.org/10.1126/science.1139158.
Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., Benito-Gutiérrez, E., Dubchak, I., Garcia-Fernàndez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W.H., Satoh, N., Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064–1071. http://dx.doi.org/10.1038/nature06967.
Qiu, L., Song, L., Yu, Y., Xu, W., Ni, D., Zhang, Q., 2007. Identification and characterization of a myeloid differentiation factor 88 (MyD88) cDNA from Zhikong scallop Chlamys farreri. Fish. Shellfish Immunol. 23, 614–623. http://dx.doi.org/10.1016/j.fsi.2007.01.012.
Ren, Q., Chen, Y.-H., Ding, Z.-F., Huang, Y., Shi, Y.-R., 2014a. Identification and function of two myeloid differentiation factor 88 variants in triangle-shell pearl mussel (Hyriopsis cumingii). Dev. Comp. Immunol. 42, 286–293. http://dx.doi.org/10.1016/j.dci.2013.09.012.
Ren, Q., Lan, J.-F., Zhong, X., Song, X.-J., Ma, F., Hui, K.-M., Wang, W., Yu, X.-Q., Wang, J.-X., 2014b. A novel Toll like receptor with two TIR domains (HcToll-2) is involved in regulation of antimicrobial peptide gene expression of Hyriopsis cumingii. Dev. Comp. Immunol. 45, 198–208. http://dx.doi.org/10.1016/j.dci.2014.02.020.
Robb, S.M.C., Gotting, K., Ross, E., Sánchez Alvarado, A., 2015. SmedGD 2.0: the Schmidtea mediterranea genome database. Genes. N. Y. N. 2000 (53), 535–546. http://dx.doi.org/10.1002/dvg.22872.
Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A.-D., Moreland, R.T., Simmons, D.K., Koch, B.J., Francis, W.R., Havlak, P., Smith, S.A., Putnam, N.H., Haddock, S.H.D., Dunn, C.W., Wolfsberg, T.G., Mullikin, J.C., Martindale, M.Q., Baxevanis, A.D., 2013. The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution. Science 342, 1242592. http://dx.doi.org/10.1126/science.1242592.
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. http://dx.doi.org/10.1074/jbc.M109.032433.
Satake, H., Sekiguchi, T., 2012. Toll-like receptors of deuterostome invertebrates. Front. Immunol. 3, 34. http://dx.doi.org/10.3389/fimmu.2012.00034.
Sea Urchin Genome Sequencing Consortium, Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., Burke, R.D., Coffman, J.A., Dean, M., Elphick, M.R., Ettensohn, C.A., Foltz, K.R., Hamdoun, A., Hynes, R.O., Klein, W.H., Marzluff, W., McClay, D.R., Morris, R.L., Mushegian, A., Rast, J.P., Smith, L.C., Thorndyke, M.C., Vacquier, V.D., Wessel, G.M., Wray, G., Zhang, L., Elsik, C.G., Ermolaeva, O., Hlavina, W., Hofmann, G., Kitts, P., Landrum, M.J., Mackey, A.J., Maglott, D., Panopoulou, G., Poustka, A.J., Pruitt, K., Sapojnikov, V., Song, X., Souvorov, A., Solovyev, V., Wei, Z., Whittaker, C.A., Worley, K., Durbin, K.J., Shen, Y., Fedrigo, O., Garfield, D., Haygood, R., Primus, A., Satija, R., Severson, T., Gonzalez-Garay, M.L., Jackson, A.R., Milosavljevic, A., Tong, M., Killian, C.E., Livingston, B.T., Wilt, F.H., Adams, N., Belle, R., Carbonneau, S., Cheung, R., Cormier, P., Cosson, B., Croce, J., Fernandez-Guerra, A., Geneviere, A.-M., Goel, M., Kelkar, H., Morales, J., Mulner-Lorillon, O., Robertson, A.J., Goldstone, J.V., Cole, B., Epel, D., Gold, B., Hahn, M.E., Howard-Ashby, M., Scally, M., Stegeman, J.J., Allgood, E.L., Cool, J., Judkins, K.M., McCafferty, S.S., Musante, A.M., Obar, R.A., Rawson, A.P., Rossetti, B.J., Gibbons, I.R., Hoffman, M.P., Leone, A., Istrail, S., Materna, S.C., Samanta, M.P., Stolc, V., Tongprasit, W., Tu, Q., Bergeron, K.-F., Brandhorst, B.P., Whittle, J., Berney, K., Bottjer, D.J., Calestani, C., Peterson, K., Chow, E., Yuan, Q.A., Elhaik, E., Graur, D., Reese, J.T., Bosdet, I., Heesun, S., Marra, M.A., Schein, J., Anderson, M.K., Brockton, V., Buckley, K.M., Cohen, A.H., Fugmann, S.D., Hibino, T., Loza-Coll, M., Majeske, A.J., Messier, C., Nair, S.V., Pancer, Z., Terwilliger, D.P., Agca, C., Arboleda, E., Chen, N., Churcher, A.M., Hallbook, F., Humphrey, G.W., Idris, M.M., Kiyama, T., Liang, S., Mellott, D., Mu, X., Murray, G., Olinski, R.P., Raible, F., Rowe, M., Taylor, J.S., Tessmar-Raible, K., Wang, D., Wilson, K.H., Yaguchi, S., Gaasterland, T., Galindo, B.E., Gunaratne, H.J., Juliano, C., Kinukawa, M., Moy, G.W., Neill, A.T., Nomura, M., Raisch, M., Reade, A., Roux, M.M., Song, J.L., Su, Y.-H., Townley, I.K., Voronina, E., Wong, J.L., Amore, G., Branno, M., Brown, E.R., Cavalieri, V., Duboc, V., Duloquin, L., Flytzanis, C., Gache, C., Lapraz, F., Lepage, T., Locascio, A., Martinez, P., Matassi, G., Matranga, V., Range, R., Rizzo, F., Rottinger, E., Beane, W., Bradham, C., Byrum, C., Glenn, T., Hussain, S., Manning, F.G., Miranda, E., Thomason, R., Walton, K., Wikramanayke, A., Wu, S.-Y., Xu, R., Brown, C.T., Chen, L., Gray, R.F., Lee, P.Y., Nam, J., Oliveri, P., Smith, J., Muzny, D., Bell, S., Chacko, J., Cree, A., Curry, S., Davis, C., Dinh, H., Dugan-Rocha, S., Fowler, J., Gill, R., Hamilton, C., Hernandez, J., Hines, S., Hume, J., Jackson, L., Jolivet, A., Kovar, C., Lee, S., Lewis, L., Miner, G., Morgan, M., Nazareth, L.V., Okwuonu, G., Parker, D., Pu, L.-L., Thorn, R., Wright, R., 2006. The genome of the sea urchin Strongylocentrotus purpuratus. Science 314, 941–952. http://dx.doi.org/10.1126/science.1133609.
Shinzato, C., Shoguchi, E., Kawashima, T., Hamada, M., Hisata, K., Tanaka, M., Fujie, M., Fujiwara, M., Koyanagi, R., Ikuta, T., Fujiyama, A., Miller, D.J., Satoh, N., 2011. Using the Acropora digitifera genome to understand coral responses to environmental change. Nature 476, 320–323. http://dx.doi.org/10.1038/nature10249.
Sigrist, C.J.A., Castro, E., de, Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L., Xenarios, I., 2012. New and continuing developments at PROSITE. gks1067 Nucleic Acids Res.. http://dx.doi.org/10.1093/nar/gks1067.
Simakov, O., Marletaz, F., Cho, S.-J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D.-H., Larsson, T., Lv, J., Arendt, D., Savage, R., Osoegawa, K., de Jong, P., Grimwood, J., Chapman, J.A., Shapiro, H., Aerts, A., Otillar, R.P., Terry, A.Y., Boore, J.L., Grigoriev, I.V., Lindberg, D.R., Seaver, E.C., Weisblat, D.A., Putnam, N.H., Rokhsar, D.S., 2013. Insights into bilaterian evolution from three spiralian genomes. Nature 493, 526–531. http://dx.doi.org/10.1038/nature11696.
Simakov, O., Kawashima, T., Marlétaz, F., Jenkins, J., Koyanagi, R., Mitros, T., Hisata, K., Bredeson, J., Shoguchi, E., Gyoja, F., Yue, J.-X., Chen, Y.-C., Freeman, R., Sasaki, A., Hikosaka-Katayama, T., Sato, A., Fujie, M., Baughman, K., Levine, J., Gonzalez, P., Cameron, C., Fritzenwanker, J., Pani, A., Goto, H., Kanda, M., Arakaki, N., Yamasaki, S., Qu, J., Cree, A., Ding, Y., Dinh, H., Dugan, S., Holder, M., Jhangiani, S., Kovar, C., Lee, S., Lewis, L., Morton, D., Nazareth, L., Okwuonu, G., Santibanez, J., Chen, R., Richards, S., Muzny, D., Gillis, A., Peshkin, L., Wu, M., Humphreys, T., Su, Y.-H., Putnam, N., Schmutz, J., Fujiyama, A., Yu, J.-K., Tagawa, K., Worley, K., Gibbs, R., Kirschner, M., Lowe, C., Satoh, N., Rokhsar, D., Gerhart, J., 2015. Hemichordate genomes and deuterostome origins. Nature 527. http://dx.doi.org/10.1038/nature16150.
Söding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244–W248. http://dx.doi.org/10.1093/nar/gki408.
Solbakken, M.H., Tørresen, O.K., Nederbragt, A.J., Seppola, M., Gregers, T.F., Jakobsen, K.S., Jentoft, S., 2016. Evolutionary redesign of the Atlantic cod (Gadus morhua L.) Toll-like receptor repertoire by gene losses and expansions. Sci. Rep. 6, 25211. http://dx.doi.org/10.1038/srep25211.
Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M.L., Signorovitch, A.Y., Moreno, M.A., Kamm, K., Grimwood, J., Schmutz, J., Shapiro, H., Grigoriev, I.V., Buss, L.W., Schierwater, B., Dellaporta, S.L., Rokhsar, D.S., 2008. The Trichoplax genome and the nature of placozoans. Nature 454, 955–960. http://dx.doi.org/10.1038/nature07191.
Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E.A., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N.H., Stanke, M., Adamska, M., Darling, A., Degnan, S.M., Oakley, T.H., Plachetzki, D.C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S.F., Goodstein, D.M., Harris, C., Jackson, D.J., Leys, S.P., Shu, S., Woodcroft, B.J., Vervoort, M., Kosik, K.S., Manning, G., Degnan, B.M., Rokhsar, D.S., 2010. The Amphimedon queenslandica genome and the evolution of animal complexity. Nature 466, 720–726. http://dx.doi.org/10.1038/nature09201.
Subramaniam, S., Stansberg, C., Cunningham, C., 2004. The interleukin 1 receptor family. Dev. Comp. Immunol., Cytokines- Evol. Perspect. 28, 415–428. http://dx.doi.org/10.1016/j.dci.2003.09.016.
The International Aphid Genomics Consortium, 2010. Genome sequence of the Pea Aphid Acyrthosiphon pisum. PLoS Biol. 8, e1000313. http://dx.doi.org/10.1371/journal.pbio.1000313.
Toubiana, M., Gerdol, M., Rosani, U., Pallavicini, A., Venier, P., Roch, P., 2013. Toll-like receptors and MyD88 adaptors in Mytilus: complete cds and gene expression levels. Dev. Comp. Immunol. 40, 158–166. http://dx.doi.org/10.1016/j.dci.2013.02.006.
Toubiana, M., Rosani, U., Giambelluca, S., Cammarata, M., Gerdol, M., Pallavicini, A., Venier, P., Roch, P., 2014. Toll signal transduction pathway in bivalves: complete cds of intermediate elements and related gene transcription levels in hemocytes of immune stimulated Mytilus galloprovincialis. Dev. Comp. Immunol. 45, 300–312. http://dx.doi.org/10.1016/j.dci.2014.03.021.
Tsai, I.J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K.L., Tracey, A., Bobes, R.J., Fragoso, G., Sciutto, E., Aslett, M., Beasley, H., Bennett, H.M., Cai, J., Camicia, F., Clark, R., Cucher, M., De Silva, N., Day, T.A., Deplazes, P., Estrada, K., Fernández, C., Holland, P.W.H., Hou, J., Hu, S., Huckvale, T., Hung, S.S., Kamenetzky, L., Keane, J.A., Kiss, F., Koziol, U., Lambert, O., Liu, K., Luo, X., Luo, Y., Macchiaroli, N., Nichol, S., Paps, J., Parkinson, J., Pouchkina-Stantcheva, N., Riddiford, N., Rosenzvit, M., Salinas, G., Wasmuth, J.D., Zamanian, M., Zheng, Y., Taenia solium Genome Consortium, Cai, X., Soberón, X., Olson, P.D., Laclette, J.P., Brehm, K., Berriman, M., 2013. The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496, 57–63. http://dx.doi.org/10.1038/nature12031.
Uematsu, S., Akira, S., 2008. Toll-like receptors (TLRs) and their ligands. In: Bauer, P.D.S., Hartmann, P.D.G. (Eds.), Toll-like Receptors (TLRs) and Innate Immunity, Handbook of Experimental Pharmacology. Springer, Berlin Heidelberg, pp. 1–20.
van der Burg, C.A., Prentis, P.J., Surm, J.M., Pavasovic, A., 2016. Insights into the innate immunome of actiniarians using a comparative genomic approach. BMC Genomics 17, 850. http://dx.doi.org/10.1186/s12864-016-3204-2.
Pagel Van Zee, J., Geraci, N.S., Guerrero, F.D., Wikel, S.K., Stuart, J.J., Nene, V.M., Hill, C.A., 2007. Tick genomics: the Ixodes genome project and beyond. Int. J.

Parasitol. 37, 1297–1305. http://dx.doi.org/10.1016/j.ijpara.2007.05.011.
Voskoboynik, A., Neff, N.F., Sahoo, D., Newman, A.M., Pushkarev, D., Koh, W., Passarelli, B., Fan, H.C., Mantalas, G.L., Palmeri, K.J., Ishizuka, K.J., Gissi, C., Griggio, F., Ben-Shlomo, R., Corey, D.M., Penland, L., White, R.A., Weissman, I.L., Quake, S.R., 2013. The genome sequence of the colonial chordate, Botryllus schlosseri. eLife 2, e00569. http://dx.doi.org/10.7554/eLife.00569.
Wagner, G.P., Kin, K., Lynch, V.J., 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. Theor. Den. Biowiss. 131, 281–285. http://dx.doi.org/10.1007/s12064-012-0162-3.
Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18, 691–699.
Wiens, M., Korzhev, M., Krasko, A., Thakur, N.L., Perović-Ottstadt, S., Breter, H.J., Ushijima, H., Diehl-Seifert, B., Müller, I.M., Müller, W.E.G., 2005. Innate immune defense of the sponge Suberites domuncula against bacteria involves a MyD88-dependent signaling pathway. Induction of a perforin-like molecule. J. Biol. Chem. 280, 27949–27959. http://dx.doi.org/10.1074/jbc.M504049200.
Wu, X., Wu, F.-H., Wang, X., Wang, L., Siedow, J.N., Zhang, W., Pei, Z.-M., 2014. Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. Nucleic Acids Res. 42, 8243–8257. http://dx.doi.org/10.1093/nar/gku569.
Xin, L., Wang, M., Zhang, H., Li, M., Wang, H., Wang, L., Song, L., 2016. The categorization and mutual modulation of expanded MyD88s in Crassostrea gigas. Fish. Shellfish Immunol. 54, 118–127. http://dx.doi.org/10.1016/j.fsi.2016.04.014.
Xu, Y., Tao, X., Shen, B., Horng, T., Medzhitov, R., Manley, J.L., Tong, L., 2000. Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. Nature 408, 111–115. http://dx.doi.org/10.1038/35040600.
Xu, F., Zhang, Y., Li, J., Zhang, Y., Xiang, Z., Yu, Z., 2015. Expression and function analysis of two naturally truncated MyD88 variants in the Pacific oyster Crassostrea gigas. Fish. Shellfish Immunol. 45, 510–516. http://dx.doi.org/10.1016/j.fsi.2015.04.034.
Yang, M., Yuan, S., Huang, S., Li, J., Xu, L., Huang, H., Tao, X., Peng, J., Xu, A., 2011. Characterization of bbtTICAM from amphioxus suggests the emergence of a MyD88-independent pathway in basal chordates. Cell Res. 21, 1410–1423. http://dx.doi.org/10.1038/cr.2011.156.
Yang, Y., Xiong, J., Zhou, Z., Huo, F., Miao, W., Ran, C., Liu, Y., Zhang, J., Feng, J., Wang, M., Wang, M., Wang, L., Yao, B., 2014. The genome of the Myxosporean Thelohanellus kitauei shows adaptations to nutrient acquisition within its fish host. Genome Biol. Evol. 6, 3182–3198. http://dx.doi.org/10.1093/gbe/evu247.
Zhang, Q., Zmasek, C.M., Dishaw, L.J., Mueller, M.G., Ye, Y., Litman, G.W., Godzik, A., 2008. Novel genes dramatically alter regulatory network topology in amphioxus. Genome Biol. 9, R123. http://dx.doi.org/10.1186/gb-2008-9-8-r123.
Zhang, Q., Zmasek, C.M., Godzik, A., 2010. Domain architecture evolution of pattern-recognition receptors. Immunogenetics 62, 263–272. http://dx.doi.org/10.1007/s00251-010-0428-1.
Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., Xiong, Z., Que, H., Xie, Y., Holland, P.W.H., Paps, J., Zhu, Y., Wu, F., Chen, Y., Wang, J., Peng, C., Meng, J., Yang, L., Liu, J., Wen, B., Zhang, N., Huang, Z., Zhu, Q., Feng, Y., Mount, A., Hedgecock, D., Xu, Z., Liu, Y., Domazet-Lošo, T., Du, Y., Sun, X., Zhang, S., Liu, B., Cheng, P., Jiang, X., Li, J., Fan, D., Wang, W., Fu, W., Wang, T., Wang, B., Zhang, J., Peng, Z., Li, Y., Li, N., Wang, J., Chen, M., He, Y., Tan, F., Song, X., Zheng, Q., Huang, R., Yang, H., Du, X., Chen, L., Yang, M., Gaffney, P.M., Wang, S., Luo, L., She, Z., Ming, Y., Huang, W., Zhang, S., Huang, B., Zhang, Y., Qu, T., Ni, P., Miao, G., Wang, J., Wang, Q., Steinberg, C.E.W., Wang, H., Li, N., Qian, L., Zhang, G., Li, Y., Yang, H., Liu, X., Wang, J., Yin, Y., Wang, J., 2012. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490, 49–54. http://dx.doi.org/10.1038/nature11413.
Zhang, Y., He, X., Yu, F., Xiang, Z., Li, J., Thorpe, K.L., 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. http://dx.doi.org/10.1371/journal.pone.0076464.
Zhang, L., Li, L., Guo, X., Litman, G.W., Dishaw, L.J., Zhang, G., 2015. Massive expansion and functional divergence of innate immune genes in a protostome. Sci. Rep. 5 http://dx.doi.org/10.1038/srep08693.
Zhu, B., Wu, X., 2008. Identification of outer membrane protein ompR from rickettsia-like organism and induction of immune response in Crassostrea ariakensis. Mol. Immunol. 45, 3198–3204. http://dx.doi.org/10.1016/j.molimm.2008.02.026.