Bivalve transcriptomics reveal pathogen sequences and a powerful immune response of the Mediterranean mussel (*Mytilus galloprovincialis*)

Rebeca Moreira[1*], Pablo Balseiro[1,3], Gabriel Forn-Cuní[1], Massimo Milan[2], Luca Bargelloni[2], Beatriz Novoa[1], Antonio Figueras[1*]

[1] Instituto de Investigaciones Marinas (IIM). CSIC. Vigo, Spain

[2] Department of Comparative Biomedicine and Food Science. University of Padova, Legnaro (PD), Italy

[3] Present Address: Uni Research Environment, Uni Research AS, Bergen, Norway

rebecamoreira@iim.csic.es (orcid 0000-0001-7797-7221)
pablo.balseiro@uni.no
gabrielfc@iim.csic.es
massimo.milan@unipd.it
luca.bargelloni@unipd.it
beatriznovoa@iim.csic.es
antoniofigueras@iim.csic.es

* **Corresponding authors**
Rebeca Moreira
Antonio Figueras
Instituto de Investigaciones Marinas (IIM)
Consejo Superior de Investigaciones Científicas (CSIC)
Eduardo Cabello 6, 36208 Vigo – Spain
+34 986 231930 – ext. 860280
rebecamoreira@iim.csic.es
antoniofigueras@iim.csic.es

Abstract

Bivalves have colonized the interface between land and sea for the last 500 million years. Although bivalves lack an adaptive immune system, they are extraordinarily well adapted to adverse environmental conditions. Bivalves are valuable aquaculture resources worldwide and are used as sentinels for monitoring pollution.

In this work, the immune transcriptomes of mussels (*Mytilus galloprovincialis* and *edulis*) and clam (*Ruditapes decussatus*) were sequenced. For comparative purposes, an already published transcriptome dataset of *Ruditapes philippinarum* was also included in the analyses. The 454 pyrosequencing of stimulated hemocytes resulted in more than 400,000 reads for each transcriptome. The percentage of annotated sequences ranges from 50% for mussels to 30-40% for clams. Considering the 28,061 non-redundant sequences from the four transcriptomes, the four species share 785 genes.

Moreover, sequences related to different putative pathogens were found in the four bivalves. A high number of bivalve herpesvirus ORFs were found, which confirms the value of NGSs as tools to detect and quantify pathogen RNA.

Based on an examination of the immune-enriched transcriptomes of these four species, we can conclude that bivalves present an immune system that differs from its conventional characterization as a simple innate immune response against invading pathogens. Enrichment analyses showed that species in the *Mytilus* genus, especially *M. galloprovincialis*, possesses a significantly higher number of sequences related to immune processes and killing molecules than species in the *Ruditapes* genus. This could be related to the broader ecological niche occupied by mussels and the scarcity of reported mussel mass mortalities compared to the high number of mass mortalities reported for clams.

Keywords

Mytilus; *Ruditapes*; 454 pyrosequencing; transcriptome; hemocytes; *Perkinsus*; OsHV-1

Introduction

The phylum Mollusca, comprising more than 200,000 living species, is the second biggest group within invertebrates (Ponder and Lindber 2008). The majority of molluscs of the Class *Bivalvia* are present in both marine and freshwater environments, where they feed by filtering organic material from the water column. However, a huge diversity of living forms have been characterized, and bivalves can be infaunal, attached to substratum by byssus, cemented to substratum, wood-borers, rock-borers, and predators (Morton 1981). They are extraordinarily well adapted to adverse environmental conditions such as high and low temperatures, periods out of the water, salinity changes and even pollution. This diversity is mainly due to the early evolutive radiation in the late Cambrian era, which allowed bivalves to occupy different ecological niches (Plazzi and Passamonti 2010).

Bivalves are a valuable food resource, and their aquaculture is experiencing a continuous growth. The most important commercial bivalves are oysters, mussels, clams, cockles and scallops. The culturing of bivalves also has applications in ecosystem repopulation, obtaining pharmaceutical or industrial products, and ornamentation (aquaria, nacre, pearls). In addition to their economic value, bivalves have important roles in environmental energy fluxes, water quality maintenance (Dame 1993; Giles *et al.* 2006; Newel 2004) and the monitoring of environmental pollution (Milan *et al.* 2013; Zorita *et al.* 2007).

As with all invertebrates, bivalves lack antibody-based adaptive immunity. However, they possess well-adapted defense strategies to protect themselves against pathogens. Bivalve cellular defenses rely on hemocytes, circulating cells in the hemolymph with specialized immune functions and other roles in homeostasis (Fisher, 1986) or shell repair (Mount *et al*. 2004).

The natural habitat of bivalves in the oceans implies an intimate contact with some of the largest microorganism populations on the planet. In fact, the filtering bivalve lifestyle implies that these animals interact with a much higher microorganism concentration than that in the water column. Viral infections are also common in marine environments. The abundance of viruses exceeds that of bacteria and archaea by approximately 15-fold, and the estimated number of viral infections is 10^{23} per second (Suttle 2007). Still, knowledge about the antiviral response in bivalves, as in all invertebrates, is scarce. The discussion about whether invertebrates possess antiviral immunity is ongoing and further enhanced by recent works, such as the study presented by Novoa *et al.* (2016). In this study, they confirmed that constitutively expressed myticins in naïve mussels made oyster hemocytes resistant to ostreid herpesvirus 1 (OsHV-1).

Among bivalves, *Mytilus galloprovincialis* has been recognized as a very successful and robust species, able to survive pathogens that have caused massive mortalities in oysters and clams cultured in the same area (Romero *et al.* 2014) and known to be extremely resistant to environmentally adverse conditions (Kurelec and Pivčević 1991). One reason attributed for this high resistance is the presence of a rich, high expression of antimicrobial peptides and a very complex repertoire of many immune-related receptors and effectors (Costa *et al.* 2009; Gerdol and Venier 2015; Moreira *et al.* 2015; Romero *et al.* 2011; Rosani *et al.* 2015; Toubiana *et al.* 2013). However, it is often difficult to uncover the immune potential of these animals because their defense repertoire is very

105 often not expressed in normal conditions and needs to be induced (Moreira *et al.*
106 2012a). In this work, we focused on the transcriptome of bivalve hemocytes, the
107 specialized defense cells in bivalves, from three bivalve species of high commercial
108 interest, *Mytilus galloprovincialis, Mytilus edulis* and *Ruditapes decussatus.* After
109 immune stimulation, the hemocyte transcriptomes were produced by 454
110 pyrosequencing. Then, these three transcriptomes were compared during analyses that
111 included the transcriptome from *Ruditapes philippinarum* (Moreira *et al.* 2012b), which
112 was obtained following the same methodology. The objective of this work was to
113 compare the immune repertoires of four different bivalve species living in different
114 environments.

115 There are a few other recent transcriptomic studies on these bivalve species (Leite *et al.*
116 2013; Ren *et al.* 2017; Yarra *et al.* 2016), and a draft genome has been published for *M.*
117 *galloprovincialis* (Murgarella *et al.* 2016), but very little information is available on
118 hemocytes. In addition to all these efforts, this is the first time that a comparison among
119 2 species of infaunal clams and 2 species of epibenthic mussels has been carried out.

120 The raw data are accessible from the NCBI database under Bioproject PRJNA377527
121 and the Short Read Archive under SRX2600750 for *M. galloprovincialis;* SRX2600751
122 for *M. edulis* and SRX2600755 for *R. decussatus*.

123

Materials and Methods

Animals

The Mediterranean mussel, *M. galloprovincialis*, is not considered as an endangered or protected species in any international species catalogue, including the CITES list (www.cites.org), and it is not included in the list of species regulated by the EC Directive 2010/63/EU. Therefore, no specific authorization is required to work on mussel samples.

Mediterranean mussels (*M. galloprovincialis*, average size 9 cm) and carpet shell clams (*R. decussatus*, average size 4 cm) were obtained from a commercial shellfish farm (Vigo, Galicia, Spain), while blue mussels (*M. edulis*, average size 6 cm) were kindly supplied by Ainhoa Blanco (Yerseke, Netherlands). Animals were maintained in open-circuit filtered sea water tanks at 34 ppt, pH 8 and 15°C with aeration. They were fed daily with *Phaeodactylum tricornutum* and *Isochrysis galbana* (10-15 x 10^6 cells/mL). Prior to the experiments, animals were acclimatized to these conditions for one week.

Molluscs, 50 animals per species, were notched in the shell in the area adjacent to the posterior adductor muscle. The size of the notch was the minimum required for the passage of a 25G syringe. Hemolymph (1 ml) was withdrawn from the adductor muscle of the bivalves with a 25G disposable syringe, pooled and distributed in 6-well plates, with 7 ml per well. All the material used during the hemolymph extraction was placed on ice to avoid hemocyte aggregation. Hemocytes were allowed to settle for 30 min at 15°C in the darkness. After settlement, hemocytes were washed with filtered sea water (0.22 µm) and stimulated with different pathogen-associated molecular patterns (PAMPs), with 1 PAMP per well. The PAMPs were 50 µg/ml of polyinositic:polycytidylic acid (Poly I:C), CpG motifs, lipopolysaccharide (LPS), lipoteichoic acid (LTA), peptidoglycans and zymosan for 3 h at 15°C in the dark. Non-stimulated hemocytes were also included. All PAMPs were purchased from SIGMA, except for the CpG motifs, which were obtained by extracting DNA from a *Vibrio anguillarum* exponential culture. In addition, the *in vivo* stimulation of bivalves, 5 animals per species, was also carried out by injecting heat-inactivated *V. anguillarum* (10^8 CFU) in the adductor muscle of the molluscs and recovering their hemocytes after 24 h. All samples were centrifuged at 3000 g for 10 minutes at 15°C, and the pellet was resuspended in 500 µl of TRIzol (Invitrogen).

RNA isolation, cDNA production and pyrosequencing

Hemocyte RNA isolation was immediately carried out using TRIzol. Purification of RNA after DNase I treatment was performed with the RNeasy mini (Qiagen). The concentration and purity of RNA was calculated with a NanoDrop ND1000 spectrophotometer, and the quality was checked with a Bioanalyzer 2100 (Agilent Technologies) (File S1). From each challenge, *in vitro* and *in vivo*, 1 µg of RNA was pooled, and one sample from each species was used as the source of starting material for full-length enriched double stranded cDNA synthesis using the MINT cDNA synthesis kit (Evrogen, Moscow, Russia) according to manufacturer's protocol and was subsequently purified using the QIAquick PCR Purification Kit (Qiagen USA, Valencia, CA). The amplified cDNA was normalized using the Trimmer kit (Evrogen, Moscow, Russia) to minimize differences in the representation of transcripts. This method involves a denaturation-reassociation of the cDNA, followed by a digestion

with a duplex-specific nuclease (DSN) enzyme (Shagin *et al.* 2002; Zhulidov *et al.* 2004). The enzymatic degradation occurs primarily on the highly abundant cDNA fraction. The single-stranded cDNA fraction was then amplified twice by sequential PCRs according to the manufacturer's protocol. The normalized cDNA was purified using the QIAquick PCR Purification Kit (Qiagen USA, Valencia, CA).

To generate the 454 library, 500 ng of normalized cDNA was used. cDNA was fractionated into small, 300- to 800-base pair (bp) fragments, and specific A and B adaptors were ligated to both the 3' and 5' ends of the fragments. The A and B adaptors were used for the purification, amplification, and sequencing steps. The sequencing runs were performed on the GS-FLX using Titanium chemistry. All reagents and protocols used for the 454 sequencing were from Roche 454 Life Sciences, USA. The RNA was normalized, processed and sequenced by the Unitat de Genòmica (CCiT-UB, Barcelona, Spain).

Assembly and functional annotation

Raw data obtained from the 454 GS FLX pyrosequencing were processed with Cutadapt (Martin 2011) to remove adapter sequences, Prinseq (Schmieder and Edwards 2011) to filter reads based on their Phred score (Phred20: 99% base call accuracy) and FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) to check the quality of the selected reads prior to assembly. The clean, high-quality raw sequences were classified as type A or B singletons depending on their length; type A singletons were raw sequences shorter than 600 bp, while type B singletons were equal or longer than 600 bp.

Type A singletons were *de novo* assembled using the Newbler gsAssembler software (Roche). There are three levels of clustering in the Newbler pipeline: putative gene domains that are not represented in a higher level are named contigs, first level; gene splicing variants are called isotigs, second level; and putative genes including at least one isotig or contig are the isogroups. Five independent assemblies with different parameters were implemented for each species, and the best conditions were selected based on the transcriptome N50 and a BLASTx search of sequences in public databases. The parameters for the BLAST search were: seed step=21, seed length=16, seed count=1, minimum overlap length=65%, minimum overlap identity=90, alignment identity score=2 and alignment difference score=-3 for the three species. The standard Newbler approach generates three different clustering levels: raw reads are grouped in contigs; contigs are grouped into isotigs; and finally, isogroups are considered putative genes including at least one isotig or contig. Since reads longer than 600 bp are not expected using 454 technology, type B singletons were not assembled. Although long reads can introduce biases in the assembly process, they contain valid information. Therefore, Type B singletons were directly selected for annotation. Information on the assembly of the transcriptomes is available in File S2.

In addition, the previously published immune transcriptome of *R. philippinarum* (Moreira *et al*. 2012b), which was obtained following the same immune simulations and molecular sequencing methodology reported in this work, was also subjected to the described annotation pipeline.

Blast2GO software version 3.2 (Conesa *et al.* 2005) was primarily used for the transcriptome annotations. An iterative BLAST workflow was used to annotate the

sequences, using BLASTx, against the NCBI NR and NCBI Swissprot databases. Isotigs were annotated with a cut-off value threshold of 10^{-5}. Type A singletons were annotated with a cut-off e-value threshold of 10^{-10}. Type B singletons were annotated with a threshold value of 10^{-15}. This conservative e-value threshold was chosen to select only type B singletons of the highest quality (Fig S2). Note that, unlike the contigs and isotigs, only singletons successfully annotated using this threshold were added to the group of sequences selected for posterior analyses. The annotation of the transcriptomes is available in File S3.

Following the BLAST annotation, gene ontology (GO) terms were assigned. An annotation threshold of 55 and a GO weight of 5, with a default evidence code weight, was implemented. All sequences were subjected to functional annotation using the Blast2GO InterproScan analysis against Prosite, Prodom, PRINTS, SMART, PFAM, PIR, PANTHER and CATH databases, which provide information about protein families, domains and motifs, as well as cellular localization predictions. The obtained GO terms were merged into the existent annotation. Finally, Annex was run to increase the biological process and cellular component GO terms using the molecular function terms. Lastly, sequences were also annotated with the enzyme codes of the KEGG database to determine the pathways they belong to.

Comparative and enrichment analysis

Before the comparison of the four hemocyte transcriptomes, their completeness, redundancy and fragmentation were analyzed with BUSCO v2 (Simão *et al.* 2015). The application was used to compare the bivalve transcriptomes against the Metazoa odb9 database of single-copy orthologues.

To avoid redundancy of the potential discovered genes (contigs, isotigs and annotated singletons) the best representative approach was selected. Briefly, the best representative sequence was selected by considering both the e-value and similarity score using the following algorithm: $similarity^3/(e\text{-}value+1e^{-300})$. A constant 1e-300 was included to avoid indetermination when e-value=0. Only the best isoform for any particular annotated gene accession number, selected using the algorithm, was used for the following analyses.

A venn diagram based on these gene accession numbers was constructed using the software Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

A two-tailed Fisher's exact test was used for the enrichment analysis for each bivalve non-redundant transcriptome (13,245 sequences for *M. galloprovincialis*, 12,867 sequences for *M. edulis*, 6,562 sequences for *R. decussatus* and 7,206 for *R. philippinarum*), and the whole dataset of the non-redundant transcriptomes of the four species was considered. GO terms were considered significantly differentially represented at a false discovery rate (FDR) < 0.01, and a reduction was performed to find the most specific terms with the same FDR.

Results and discussion

Genomic research in non-model species has taken advantage of next-generation sequencing methods (NGSs) (Carlsson *et al.* 2013; Riesgo *et al.* 2012; Vera *et al.* 2008). Specifically, transcriptomic studies can be directed towards the discovery of genes and splice variants expressed for relevant physiological processes (Sánchez-Pla *et al.* 2012; Shi *et al.* 2013). Comparative studies among bivalve species are now gaining attention in our field. They enrich our understanding of the remarkable bivalve immune system. It is likely that other studies on different responses of bivalves to pathogens, adaptations to environment or exposures to pollution will be published in the near future. These studies will increase the knowledge on this interesting phylum of littoral communities that has a high socioeconomic impact.

In the present study, immune system triggering was carried out using PAMPs mimicking a broad range of pathogens from viruses to fungi. This approach has proven successful in the past to find new genes related to responses of interest (Moreira *et al.* 2012b).

454 pyrosequencing, assembly, and annotation

The *M. galloprovincialis*, *M. edulis*, and *R. decussatus* immune-enriched transcriptomes were sequenced using 454 GS FLX and analyzed as shown in Fig. 1. Sequencing, assembly, and annotation results are summarized in Table 1 and Fig S1. Raw read statistics and quality measures are summarized in Fig S2. Briefly, the sequencing of the stimulated hemocytes from *M. galloprovincialis*, *M. edulis*, and *R. decussatus* resulted in a combined total of more than 400,000 reads, of which 86.5% for *M. galloprovincialis*, 85.2% for *M. edulis*, and 78% for *R. decussatus* passed the first filtering and quality control test. The number of reads obtained in these transcriptomes was in agreement with numbers previously obtained for other molluscs using one-half of a picotiter plate on the 454 GS FLX system (Coppe *et al.* 2012; Milan *et al.* 2011), and it was even greater than the number obtained from high-throughput sequencing approaches performed for other mollusc species (Galindo *et al.* 2010; Shi *et al.* 2013).

In total, 14,158, 13,560 and 7,318 unique contigs and isotigs from *M. galloprovincialis*, *M. edulis*, and *R. decussatus*, were assembled, belonging to 9,256, 8,814, and 6,020 isogroups, respectively. The rest of the sequences remained as singletons: 231,672 for *M. galloprovincialis*, 230,533 for *M. edulis* and 206,785 for *R. decussatus*.

The N50 was slightly higher for isotigs than for contigs, as expected. Present sequencing works show higher N50s than those previously reported for similar species (Moreira *et al.* 2012b; Philipp *et al.* 2012). The N50s from both *Mytilus* species were higher than that obtained from *R. decussatus*. Unfortunately, although broadly used as an indicator of the quality of the assembly, the N50 is not always provided and does not directly correlate with sequence length in these two genera (Carneiro *et al.* 2012). Consequently, N50 can only be of limited use, even when comparing transcriptomes among closely related species.

As shown in Fig. 2, all the parameter distributions were consistently similar among the mussel transcriptomes. In general, the clam sequences showed smaller sizes (Fig. 2B) and less unique reads or contig clusters (Fig. 2A and 2C respectively) in the assembly step.

Previous work on *M. galloprovincialis* recovered 8,586 contigs from four different tissues of adult mussels (Craft *et al.* 2010). In contrast, the combined use of different stressors and immune stimulants allowed the detection of a higher number of sequences; for instance, almost 75,000 contigs were detected from multiple tissues in adult blue mussel (*M. edulis*) (Philipp *et al.* 2012), and more than 50,000 expressed contigs were detected in clam hemocytes after the immune stimulation of Manila clam (*R. philippinarum*) (Moreira *et al.* 2012b). The discovery of genes related to a specific physiological process is improved by using biological material that has been stimulated to trigger the physiological response of interest and by obtaining RNA from tissues related to specific physiological processes, such as immune responses, biomineralization and pearl formation, venom production and responses to variations in water pH, temperature or salinity (Bai *et al.* 2013; Hu *et al.* 2011; Kinoshita *et al.* 2011; Milan *et al.* 2011; Moreira *et al.* 2012b; Philipp *et al.* 2012; Shi *et al.* 2013; Werner *et al.* 2013; Zhao *et al.* 2012). Furthermore, new or secondary functions of the studied tissue can be detected after high-throughput sequencing. In the specific case of hemocytes, their roles in shell repair and calcification could be detected in our results (File S3) in accordance with previous works on other bivalves (Mount *et al*. 2004, Jeffroy *et al*. 2013; Sillanpää *et al*. 2016).

The annotation of the assembled sequences was performed using BLASTx, a process directly dependent on the representation of related sequences in public databases. For low-represented species in the databases, such as bivalves, only conserved sequences can be annotated (Huan *et al.* 2012). However, we successfully annotated 7,031 (49.66%) sequences from *M. galloprovincialis*, 6,683 (49.28%) sequences from *M. edulis* and 2,327 (31.8%) sequences from *R. decussatus*, which are similar values to those previously reported for bivalve transcriptomes (Bettercourt *et al.* 2010; Craft *et al.* 2010; Huang *et al.* 2013; Milan *et al.* 2011). Moreover, we also annotated singletons because they can represent low-expressed sequences and have been proven to increase the gene discovery of sequences related to immune processes in bivalves (Ewen-Campen *et al.* 2011; Meyer *et al.* 2009). Accordingly, a total of 19,431 singletons (type A and B) from *M. galloprovincialis*, 18,299 singletons from *M. edulis*, and 7,138 singletons from *R. decussatus* were successfully annotated and added to the rest of the sequences for functional annotation.

As expected, *Crassostrea gigas* was the main source of annotation for all three transcriptomes (Fig. 3). *Crassostrea gigas* is the most represented bivalve species in public databases, far from *Mytilus* genus (http://www.ncbi.nlm.nih.gov). The higher percentage of hits in both *Mytilus* species compared to *R. decussatus* could be a consequence of the ancient divergence between clams and the common antecessor of mussels and oysters (Lemer *et al.* 2016).

Even once the 454-sequenced libraries had been normalized to increase the presence of low-expression genes, the appearance frequency of the sequences could be used as a rough indication of the expression of each transcript. The top 35 BLAST hits are detailed in Fig. 3; there is a high abundance of sequences directly related to immune responses in the three studied species, including recognition proteins (e.g., lectins), stress-related proteins (e.g., HSPs), apoptosis-related proteins (e.g., caspases, IAPs), proteins of the complement cascade (e.g., C1q, C3, Factor B), and other effectors of the immune response such as lysozyme and mussel myticins.

Myticins are key antimicrobial peptides (AMPs) in the mussel immune response (Mitta *et al.* 2000a; Novoa *et al.* 2016; Pallavicini *et al.* 2008; Rosani *et al.* 2011). Myticin-related sequences have been reported as the most abundant sequences in hemocyte transcriptomes (Pallavicini *et al*. 2008) and were also found to be highly represented in our results; they are the 2nd (myticin C) and 18th (myticin A) most abundant type of sequence in the *M. galloprovincialis* transcriptome, and they are the 20th most abundant type of sequence in the *M. edulis* transcriptome. Other pore-forming molecules, such as apextrin or bactericidal permeability increasing protein (BPI), are very abundant in the mussel transcriptomes compared to the clam transcriptome. Apextrin is the most represented sequence in the *M. edulis* transcriptome, and it is the third most represented sequence in the *M. galloprovincialis* transcriptome (Fig. 3). Pore-forming molecules are important to fight against bacteria (Estévez-Calvar *et al.* 2011; Huang *et al.* 2014), and recent studies have described that BPI and some AMPs, such as myticin C, have activity against viruses as well (Novoa *et al.* 2016; Pinkenburg *et al.* 2016).

The most represented transcript in the *M. galloprovincialis* transcriptome was annotated as a cholecystokinin receptor; it was the 8th most abundant transcript in *M. edulis* (Fig. 3). This molecule, whose primary function in mammals is related to digestion, associates with G proteins to activate a phosphatidylinositol-calcium second messenger system, and it is related to cell proliferation, inflammation and cancer (Dufresne *et al.* 2006; Fino *et al.* 2012). It is not irrational to think that this protein may have a role in signal transduction in the mussel defense response.

Moreover, we found other abundant sequences for specialized immune responses in the *Mytilus* genus that are rare or absent in the clam transcriptome. A possible explanation for this, beyond a lack of expression, could be that the *R. decussatus* sequences are too divergent to be recognized by BLAST. The immune-related sequences found only in the *Mytilus* transcriptomes were predominantly FREPs, ferric-chelate reductase, GIMAP4, and TRIM proteins. FREPs are non-self-recognition molecules in invertebrates previously characterized in bivalves (Romero *et al.* 2011). Ferric-chelate reductase is part of an innate defense mechanism conserved from plants to mammals that reduces the circulating iron concentration, an essential nutrient for bacteria, therefore limiting bacterial proliferation (Cassat *et al.* 2013; Dellagi *et al.* 2009; Nairz *et al.* 2014; Sun *et al.* 2014). The GIMAP4 function in mammals is related to T-cell and B-cell development and apoptosis, and it was recently associated with the regulation of interferon-γ secretion (Heinonen *et al.* 2015). This is one of the few times that GIMAP4 has been documented in bivalves (Gerdol *et al.* 2014; McDowell *et al.* 2016). Finally, different antiviral TRIM proteins are highly represented in the mussel transcriptomes: TRIM 2, 33, 45 and 56 were found in the top 35 transcripts for both mussel species. The first mollusc orthologue of TRIM-2/3 was reported by van Diepen *et al.* (2005), but this gene family has not been further studied in molluscs. Invertebrates have 10–20 forms of TRIM proteins, while mammals have approximately 60-70 forms, and, although the functions of all specific TRIM proteins have not been completely identified, it is known that they are involved in the regulation of TLR and RIG-I-like signaling pathways (McNab *et al.* 2011; Kawai and Akira 2011).

On the other hand, we also found common immune-related sequences expressed by both mussel and clam, proteins with a growing importance in the study of innate immune responses, such as interferon-induced protein 44 (IFI44). IFI44 is one of hundreds of interferon stimulated genes (ISGs), and it is associated with viral infection. IFI44 is known to be a cytosolic protein that causes the arrest of cell division until successful

virus defense (Hallen *et al.* 2007). Recent studies have revealed new mechanisms of action; after viral infection, IFI44 spreads into the nucleus to suppress viral transcription, promoting its latency (Powe *et al.* 2015).

Put into context, all this information reveals the importance of the direct defense against pathogens for a fast and efficient immune response action in mussels compared to carpet shell clam.

Interspecies immune transcriptome comparison

To complement the interspecies immune transcriptome comparative analysis in this study, we also used the previously published *R. philippinarum* transcriptome (Moreira *et al.* 2012b). The four transcriptomes have been carried out using 454 pyrosequencing after the same hemocyte immune stimulation. To take into account the possible new information in the databases *R. philippinarum* transcriptome annotation was repeated using the same methods as the new transcriptomes.

BUSCO v2 was used to assess the completeness, redundancy and fragmentation of the transcriptomes. The results (Fig. 4a) show that the two *Mytilus* species are very similar, and *R. decussatus* has the most different profile (although the methodological approach was exactly the same as that for *Mytilus*). On the other hand, *R. philippinarum*, with its particular assembly, has a profile in-between *Mytilus* and *R. decussatus. R. philippinarum* also shows the least fragmented transcriptome, due to the singletons included in the other three species, as shown in Fig. 4b. Hemocytes are a particular tissue with a specific expression repertoire, and their transcriptional efforts are focused on the synthesis of a limited number of mRNAs from key genes. This could be the explanation of the missing orthologues in the BUSCO results, i.e., these genes are not expressed in the hemocytes. This characteristic was shared by all the studied species and has been previously reported in mussels and clams (Moreira *et al.* 2012b; Moreira *et al.* 2015). The fragmentation rates, without the singletons (Fig. 4b), are quite similar among all the species. This indicates that the assemblies nearly have the same quality. The slightly higher duplication rate for mussels could indicate the presence of more paralogs or more allelic variants compared to clams.

Even with these results, the clam immune transcriptomes are quite different. Fig. 4b shows that the completeness of the *R. philippinarum* transcriptome is higher than that of the *R. decussatus* transcriptome. We hypothesize that this could be due to biological and technical factors. Both species are known to have different immune responses to the same pathogenic stimuli; some key immune genes studied by qPCR were not detected or showed a delayed transcription pattern in *R. decussatus* compared to *R. philippinarum* (Moreira et el. 2012a). This could be indicative of a higher transcriptomic richness in *R. philippinarum*. Additionally, the sequencing depth of the 454 approach could bias the completeness of the results; the sequencing depth for *R. philippinarum* was approximately double that for *R. decussatus*, as Table 1 notes (number of reads, contigs and annotated sequences).

Fig. 5 shows a Venn diagram comparing the best annotated sequences for a single gene accession code (to avoid redundancy) in the four immune transcriptomes. The number of exclusive genes for *M. galloprovincialis* was 6,006 (45.34%), for *M. edulis* was 5,694 (44.25%), for *R. decussatus* was 3,469 (52.72%) and for *R. philippinarum* was 4,461 (61.91%).

In total, we found a higher percentage of exclusive sequences in clams than mussels, although the two mussel transcriptomes share a higher number of exclusive sequences, suggesting a larger immune repertoire than clams. This could be related to the broader ecological niche of mussels and the higher mussel resistance to pathogens, with fewer reported cases of mass mortalities, compared to clams (Gómez-León *et al.* 2005).

Functional annotation and enrichment analysis

We carried out a functional annotation using Gene Ontology terms following the BLASTx annotation. Table 1 summarizes the results of the functional annotation of the two mussel transcriptomes and the clam transcriptome. To maximize the GO term assignment, a protein domain search with InterproScan (IPS) was performed. Based on the search results, 70.42% of the *M. galloprovincialis* sequences, 70.80% of the *M. edulis* sequences and 67.22% of the *R. decussatus* sequences had at least one IPS result. This additional step improved GO annotation by approximately 15%. The results of the GO annotation for the mussel transcriptomes are very similar (Table 1), showing GO terms assigned to 61% of the sequences, while only 54% of the *R. decussatus* sequences were assigned a GO term. In any case, the functional annotation results that we obtained, with annotation percentages well over 50% for the transcriptomes of all three species, is higher than other annotations previously reported in bivalves (Coppe *et al.* 2012; Bai *et al.* 2013; Bettencourt *et al.* 2010; Huang *et al.* 2013), which supports our annotation pipeline.

The differential enrichment analysis of the Gene Ontology annotations revealed that *R. philippinarum* had 221 biological processes (BPs) significantly overrepresented compared to the rest of the bivalves studied. It is noteworthy that almost 40% of the overrepresented BPs were related to immunity and inflammation. In contrast, *R. decussatus* had only 21 enriched BPs, of which none had a direct relation to the immune system. *M. edulis* presented 28 significantly overrepresented BPs, and only 2 of them were directly related to the immune defense. The *M. galloprovincialis* enrichment analysis showed 30 BPs, with key defense processes overrepresented. The high amount (almost 10 times more) of enriched GO terms in *R. philippinarum* compared to the other species, could be explained by the higher sequencing depth of this species. Most likely, *R. philippinarum* immune transcriptome is enriched for GO terms associated with genes expressed at rather low levels in hemocytes. Therefore, these genes could be only detected in the transcriptome of *R. philippinarum*.

Fig. 6 shows a summary of the 20 top BPs for the 4 bivalve species arranged by their representation in each transcriptome. Strikingly, the majority of the most enriched BPs in *R. decussatus* are metabolism-related, and none are related to immunity. However, there is a clear link between nutrient sensing and immune response processes, which co-evolved and are evolutionarily conserved. In fact, several studies have demonstrated that the balance between metabolic and inflammatory responses can be easily disrupted in highly susceptible species, as continuous contact with pathogens compromises metabolic homeostasis in a range of species from lower invertebrates to mammals (Hostamisligil and Erbay 2008; Iyer *et al.* 2015; Tornatore *et al.* 2012). This fact is exacerbated in bivalves, which live in permanent contact with pathogens as a result of their lifestyle. *R. decussatus* has been reported to be highly susceptible to diseases, which, with the immunological challenges of the experiment, may have caused the observed imbalance in the defense and metabolism processes. In contrast, *R. philippinarum* shows a strong immune profile; for example, it has an overrepresentation

of transcripts annotated with the BP "defense response to Gram-negative and Gram-positive bacterium". *M. galloprovincialis* showed a more balanced and broad-spectrum response against pathogens than the other species, e.g., AMPs clearly represented with the term "killing of cells of other organism". Indeed, if we look for the "antimicrobial" term in the NCBI database (19/10/2017), only 131 protein and nucleotide sequences are found for the *Ruditapes/Venerupis* genus, and all of them are *R. philippinarum* defensins, while the *Mytilus* genus presents 397 sequences. More than 290 of these belong to *M. galloprovincialis* and cover all the AMP spectrum.

In clams, AMPs with similarities to mussel myticin, mytilin and defensin have already been described (Gestal *et al.* 2007; Moreira *et al.* 2012b; Zhao *et al.* 2010). In this *R. decussatus* transcriptome, we found just one sequence with homology to the putative defense protein 3, and no sequence with homology to defensins, myticin or mytilin was found. Antimicrobial sequences have been deposited in the NCBI database for 6 different *Mytilus* species (*M. galloprovincialis, M. edulis, M. trossulus, M. californianus, M. chilensis* and *M. coruscus*) and a diverse range of AMPs (defensins, myticins, mytilins, mytichitin, myticusin and miticalins), some described almost two decades ago and others very recently discovered (Costa *et al.* 2009; Mitta *et al.* 2000a; Liao *et al.* 2013; Qin *et al.* 2014; Leoni *et al.* 2017). In our results, there are more than 200 and 700 AMPs in the *M. edulis* and *M. galloprovincialis* transcriptomes, respectively. Myticins are, by far, the most represented AMP, representing more than 60% of *M. edulis* AMPs and almost 80% of *M. galloprovincialis* AMPs. This is an indicator, again, of the extreme variability (Costa *et al.* 2009) and high expression (Pallavicini *et al.* 2008) of myticins in mussels, especially in *M. galloprovincialis*.

Mussels and clams had a total of 126 and 192 significantly enriched biological processes, respectively (File S4). In mussels, some overrepresented processes were cell proliferation and migration, apoptosis and the killing of cells of other organisms (due to the extremely high quantity of *M. galloprovincialis* AMPs). On the other hand, the terms "negative regulation of defense response" and "negative regulation of immune system process" were also overrepresented. These results probably mean that the efficient migration of hemocytes to the site of infection, to release AMPs and eliminate the pathogen (Balseiro *et al.* 2011; Mitta *et al.* 2000b), especially in *M. galloprovincialis*, is enough to overcome an infection without an inflammatory response. Previous studies support this theory; for example, *M. galloprovincialis* does not increment its bactericidal response after a β-glucan challenge, as observed in clams (Costa *et al.* 2008) and vertebrates (van der Meer *et al.* 2015). In addition, in all other transcriptomic studies on bivalves other than mussels, there are few AMPs compared to mussels, as previously described. In contrast, inflammatory processes are more represented in the other studied bivalves compared to *M. galloprovincialis.* Among the enriched processes in *M. edulis*, the "regulation of interferon-gamma-mediated signaling pathway" and "mast cell degranulation" processes were found; for *R. philippinarum*, many inflammatory pathways were activated, such as the TLR, NF-kB, EGF and TNF signaling pathways, the complement cascade, the acute phase response or the respiratory burst, among others.

Diverse mechanisms have evolved to prevent immune reactions and inflammation rather than accentuate the immune defense (Flajnik 2010). An example of these mechanisms is the enhancement of the mucosal response to prevent an infection, and our results (high presence of AMPs and inflammatory processes barely detected) support that *M. galloprovincialis* is highly efficient at this enhancement (Allam and Pales Espinosa

2016). Many bivalves present recognition and effector proteins (lectins, peptidoglycan-recognition proteins, Dscam proteins, thioester bearing proteins, FREPs or AMPs) in their extracellular fluids, such as hemolymph and mucus. Of note, very high amounts of FREPs and other highly diversified gene families, such as the myticin family (Costa *et al.* 2009), have been found in mussel transcriptomes. These results suggest a possible high degree of pathogen specificity and immune priming (Wang *et al.* 2013).

Taxonomic analysis of pathogen sequences

Historically, microbial culture, histology and electron microscopy were the reference methods to detect pathogens in animal tissues. These methods are especially useful for protozoan parasites. Protozoans and their associated lesions are easier to detect than bacterial or viral processes. Still, histopathology is time consuming and difficult to analyze quantitatively. In addition, the early stages of infection are difficult to detect by these methods.

The diagnosis of animal diseases has experienced dramatic advances in the last few decades. Currently, molecular techniques can rapidly detect pathogens in animal tissue samples. *In situ* hybridization or PCR can identify a known gene or DNA of pathogenic origin. This genomic information alone can give information about the presence or absence of the pathogen, but it cannot indicate its viability and infectivity. To overcome this issue, qPCR or microarrays are used to detect pathogen replication through RNA expression. With transcriptomics, pathogen presence and activity can be assessed, even in the early phase of an infection, due to the sensitivity of the actual sequencing techniques (Westermann *et al.* 2016), a goal difficult to achieve with traditional pathology approaches. Furthermore, the information obtained by this tool could be used not only in aquaculture facilities, to prevent outbreaks of disease but also in the wild, to find possible reservoirs of pathogens prior to the expansion of the disease.

We found sequences of potential putative pathogens in all the transcriptomes. The non-redundant sequences with a BLASTx match with e-value under 10^{-5} are summarized in Table 2.

<u>*Eukaryotic pathogens*</u>

Fungal sequences were highly present in all the studied species. Although it is known that bivalves are sensitive to fungal diseases (Bower *et al.* 2010; Davis *et al.* 1954), this field has barely been explored. Fungi are not important causes of bivalve mortality, and, as far as we know, there have not been many studies on fungal pathogenic species for bivalves. In fact, the most represented species in the annotation was *Capsaspora owczarzaki*, which is not pathogenic for molluscs. It was found to be a symbiont in the hemolymph of the tropical freshwater snail *Biomphalaria glabrata* (Hertel *et al.* 2002). Consequently, the sequences with a fungal origin found in the immune-stimulated transcriptomes are not necessarily from bivalve pathogens.

The fact that *Perkinsus marinus* is one of the most represented species in the *R. decussatus* annotation, with a representation level exceeding that of the *Ruditapes* genus, deserves special attention (Fig. 3). *Perkinsus* parasites are frequently detected in clams from South Galicia (Ruano *et al.*, 2015; Villalba *et al.*, 2004). Although *R. philippinarum* was obtained from the same site and in the same conditions as *R. decussatus*, the number of *Perkinsus* sequences in *R. philippinarum* (3) was much lower

than that in *R. decussatus* (131). This suggests that *Perkinsus* lives in and actively colonizes *R. decussatus* tissues. As *Perkinsus* does not easily kill its hosts (unless there are stressful environmental conditions for clams), the sampled individuals (over 50) could be heavily infected without external signs of the infection. This is in agreement with previous observations noting the higher susceptibility of *R. decussatus* than *R. philippinarum* (Moreira *et al.* 2012a).

Bacteria

As expected, the *Vibrio* genus represents the main group of pathogens detected in mussels (Romero *et al.* 2014). The genus *Vibrio* is ubiquitous in aquatic ecosystems. Mussel species are highly resistant to bacteria, but *Vibrio* species are among the main causes of disease in cultured bivalves, especially in the early life stages (Beaz-Hidalgo *et al.* 2010; Le Roux *et al.* 2015).

It is surprising that the carpet shell clam showed so few hits to *Vibrio* species compared to *R. philippinarum*, which presented the highest amount of bacterial sequences of the four studied species. The enrichment analysis showed that Manila clam had a very specialized response to bacteria, and it seems that this animal can cope with that load of bacteria, surviving and responding properly.

Viral genomes in bivalve transcriptomes

Viral diseases are a recurrent and severe problem for bivalve aquaculture. For instance, in France, *Crassostrea angulata* and *Crassostrea gigas* almost vanished from the European Atlantic coast because of serious viral outbreaks (Dègremont *et al.* 2015; Marteil 1976; Segarra *et al.* 2010).

The specific study and replication of viruses in controlled conditions requires the existence of proper cell lines. This is a huge handicap in regard to mollusc viral infections because there are no cell lines belonging to molluscs. This fact complicates the virus isolation and study with *in vitro* cultures. Consequently, NGSs have become a valuable tool to evaluate viral infections.

While few sequences of bivalve pathogenic viruses were successfully annotated in the mussels, *R. decussatus* presented 36 hits for Ostreid herpesvirus 1 (OsHV1), 11 hits for *Chlamys* acute viral necrosis virus (AVNV) and 2 hits for Abalone herpesvirus (AbHV).

We detected and analyzed in detail all the viral sequences found in the four transcriptomes. The pathogenic viruses for molluscs were OsHV1, AVNV and AbHV, as highlighted in Table 3 and Supporting File S5. OsHV1 is a well-known virus in the aquaculture industry (Batista *et al.* 2007; Green *et al.* 2015), and AVNV is a recently characterized virus (Ren *et al.* 2013) known to be a variant of OsHV-1 (the AVNV genome is 97% identical to the OsHV-1 genome). Regarding AbHV, it has been the causative agent of very high mortalities in abalone in Taiwan and Australia since the early 2000s (Chang *et al.* 2005; Corbeil *et al.* 2016). A sequence related to the white spot syndrome virus (WSSV), a virus with great importance in aquaculture, was found in the *M. galloprovincialis* transcriptome. WSSV has a remarkably broad host range, but it is known to be highly pathogenic and virulent only in shrimp (Dang *et al.* 2010).

We manually reblasted all the viral sequences in the transcriptomes and found a small percentage (less than 8% in total) that were not true viral sequences. They showed high homology to proteins with domains of viral origin, such as the baculovirus inhibitor of apoptosis protein repeat, the BIR domain of the apoptosis-related IAPs. Some other reannotated sequences, especially those with homology to AVNV, were indeed OsHV-1 sequences, as expected due to the high homology of both virus genomes.

Supporting File S5 has information about all the studied sequences, with their respective nucleotide sequences and BLASTx results (November 2016). Table 3 is a summary presenting the sequence with the best hit to each protein. These results show that, for all the 88 studied viral sequences, 49 (55.68%) belong to OsHV-1, covering 27 of the 124 ORFs in its genome. Almost all of them, 26 OsHV-1 ORFs, were found only in the *R. decussatus* transcriptome.

As it occurs with protozoan infections, we detected more viral agents in *R. decussatus* than in the other species. Similar to *Perkinsus*, OsHV-1 may be present in *R. decussatus* tissues without killing its host. These results perfectly fit with the comparison of the four transcriptomes and the enrichment analysis. Our results show that *Mytilus* is the genus that presents the fewest pathogen-associated sequences. *R. philippinarum* showed the highest amount of sequences of bacterial origin. Last, *R. decussatus* had the highest amount of sequences belonging to important pathogens in the aquaculture industry, OsHV-1 and *Perkinsus.*

Conclusions

In this study, we complement the previously published immune-enriched transcriptome of the clam *R. philippinarum* (Moreira *et al.* 2012b) with three additional immune-enriched mollusc transcriptomes from the carpet-shell clam *R. decussatus,* and the mussels *M. galloprovincialis* and *M. edulis*.

Our study increases the knowledge in the field of molluscan transcriptomics by NGSs, in particular transcriptomics using 454 sequencing (Bettencourt *et al.* 2010; Craft *et al.* 2010; Hu *et al.* 2011; Huan *et al.* 2012; Jiang *et al.* 2011). Thus far, one of the most interesting results obtained by bivalve transcriptomics and further confirmed in the present work is that almost half of the genes expressed in mussel hemocytes are AMPs and that myticin-related sequences are among most abundant genes (Costa *et al.* 2009; Pallavicini *et al.* 2008).

After examining the hemocyte transcriptomes of these four bivalve species, we can conclude that bivalves present a more complex immune system than previously known, with an efficient innate immune response against invading pathogens.

Of the four studied species, *R. decussatus* was found to have the most pathogen-associated sequences in its transcriptome, with an immune system that seems to be easily imbalanced by stress and pathogens, as no putative defense genes or functions were found in the enrichment analysis. Furthermore, the carpet shell clam presented a very high amount of pathogen sequences, especially from *Perkinsus* and the virus OsHV-1. Indeed, we found a great variety of ORFs of viral origin, covering approximately 1/4 of the OsHV-1 genome. This finding highlights the usefulness of NGSs to detect pathogen sequences in animal tissues and their particular importance for identifying non-culturable microorganisms and studying the very early stages of an infection.

M. galloprovincialis, when compared with the other species included in this study, showed an immune system that was the most competent at overcoming infections. Among all the studied species, it had the most powerful transcriptomic immune response, with the highest amount and diversity of AMPs and the broadest gene spectrum against pathogens. These results might help explain why there is little literature reporting mass mortalities in mussels caused by pathogens or physiological disorders, such as cancer.

Author Contributions

BN, LB and AF conceived and designed the project. PB and MM performed the annotation step. RM and PB made functional annotation analyses. GFC did the BUSCO analyses. BN and RM made the analysis of pathogen sequences. RM wrote the manuscript. All listed authors revised, edited, read and approved the manuscript.

Acknowledgments

This work has been funded by the EU Project REPROSEED (245119) and 10 PXIB 402 096 PR from Xunta de Galicia; and partially supported by Ministerio de Economía y Competitividad through Intramural 201640E024 and MYTIPEP (AGL2015-65705-R). We also acknowledge the support of Xunta de Galicia to our group (IN607B 2016/12).

RM wishes to acknowledge the Spanish MICINN for her FPI Spanish research grant (BES-2009-029765) and the EU H2020 funded project VIVALDI (678589).

We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI). We would also like to thank Ainhoa Blanco for the supply of the blue mussel, *Mytilus edulis.*

Conflict of interest

The authors declare that they have no conflict of interest.

694 **References**

695 Allam B, Pales Espinosa E (2016) Bivalve immunity and response to infections: Are
696 we looking at the right place? Fish Shellfish Immunol. 53:4-12.

697 Bai Z, Zheng H, Lin J, Wang G, Li J (2013) Comparative Analysis of the
698 Transcriptome in Tissues Secreting Purple and White Nacre in the Pearl Mussel
699 *Hyriopsis cumingii*. PLoS One. 8:e53617.

700 Balseiro P, Falcó A, Romero A, Dios S, Martínez-López A, Figueras A, Estepa A,
701 Novoa B (2011) *Mytilus galloprovincialis* myticin C: a chemotactic molecule with
702 antiviral activity and immunoregulatory properties. PLoS One. 6:e23140.

703 Batista FM, Arzul I, Pepin JF, Ruano F, Friedman CS, Boudry P, Renault T (2007)
704 Detection of ostreid herpesvirus 1 DNA by PCR in bivalve molluscs: a critical review.
705 J Virol Methods. 139:1-11.

706 Beaz-Hidalgo R, Balboa S, Romalde JL, Figueras MJ (2010) Diversity and
707 pathogenecity of *Vibrio* species in cultured bivalve molluscs. Environ Microbiol Rep.
708 2:34-43.

709 Bettencourt R, Pinheiro M, Egas C, Gomes P, Afonso M, Shank T, Santos RS (2010)
710 High-throughput sequencing and analysis of the gill tissue transcriptome from the
711 deep-sea hydrothermal vent mussel *Bathymodiolus azoricus*. BMC Genomics. 11:559.

712 Bower SM (2010) Synopsis of infectious diseases and parasites of commercially
713 exploited shellfish. Fisheries and Oceans Canada. Available from: http://www.dfo-
714 mpo.gc.ca/science/aah-saa/diseases-maladies/index-eng.html

715 Carlsson J, Gauthier DT, Carlsson JEL, Coughlan JP, Dillane E, Fitzgerald RD,
716 Keating U, McGinnity P, Mirimin L, Cross TF (2013) Rapid, economical single-
717 nucleotide polymorphism and microsatellite discovery based on de novo assembly of a
718 reduced representation genome in a non-model organism: a case study of Atlantic cod
719 *Gadus morhua*. J Fish Biol. 82:944-958.

720 Carneiro AR, Ramos RT, Barbosa HP, Schneider MP, Barh D, Azevedo V, Silva A
721 (2012) Quality of prokaryote genome assembly: Indispensable issues of factors
722 affecting prokaryote genome assembly quality. Gene. 505:365-367.

723 Cassat JE, Skaar EP (2013) Iron in infection and immunity. Cell Host Microbe. 13:509-
724 519.

725 Chang PH, Kuo ST, Lai SH, Yang HS, Ting YY, Hsu CL, Chen HC (2005) Herpes-like
726 virus infection causing mortality of cultured abalone *Haliotis diversicolor supertexta* in
727 Taiwan. Dis Aquat Organ. 65:23-27.

728 Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a
729 universal tool for annotation, visualization and analysis in functional genomics
730 research. Bioinformatics. 21:3674-3676.

731 Coppe A, Bortoluzzi S, Murari G, Marino IA, Zane L, Papetti C (2012) Sequencing
732 and Characterization of Striped Venus Transcriptome Expand Resources for Clam
733 Fishery Genetics. PLoS One. 7:e44185.

734 Corbeil S, Williams LM, McColl KA, Crane MS (2016) Australian abalone (*Haliotis*
735 *laevigata, H. rubra* and *H. conicopora*) are susceptible to infection by multiple abalone
736 herpesvirus genotypes. Dis Aquat Organ. 119:101-106.

737 Costa MM, Dios S, Alonso-Gutierrez J, Romero A, Novoa B, Figueras A (2009)
738 Evidence of high individual diversity on myticin C in mussel (*Mytilus*
739 *galloprovincialis*). Dev Comp Immunol. 33:162-170.

740 Costa MM, Novoa B, Figueras A (2008) Influence of beta-glucans on the immune
741 responses of carpet shell clam (*Ruditapes decussatus*) and Mediterranean mussel
742 (*Mytilus galloprovincialis*). Fish Shellfish Immunol. 24:498-505.

743 Craft JA, Gilbert JA, Temperton B, Dempsey KE, Ashelford K, Tiwari B, Hutchinson
744 TH, Chipman JK (2010) Pyrosequencing of *Mytilus galloprovincialis* cDNAs: Tissue-
745 Specific Expression Patterns. PLoS One. 5:e8875.

746 Dame RF (1993) Bivalve Filter Feeders. Ed. R.F. Dame. Berlin Heidelberg: Springer-
747 Verlag. pp. 245-269.

748 Dang LT, Koyama T, Shitara A, Kondo H, Aoki T, Hirono I (2010) Involvement of
749 WSSV-shrimp homologs in WSSV infectivity in kuruma shrimp: *Marsupenaeus*
750 *japonicus*. Antiviral Res. 88:217-226.

751 Davis HC, Loosanoff VL, Weston WH, Martin C (1954) A fungus disease in clam and
752 oyster larvae. Science. 120:36-38.

753 Dégremont L, Lamy JB, Pépin JF, Travers MA, Renault T (2015) New Insight for the
754 Genetic Evaluation of Resistance to Ostreid Herpesvirus Infection, a Worldwide
755 Disease, in *Crassostrea gigas*. PLoS One. 10:e0127917.

756 Dellagi A, Segond D, Rigault M, Fagard M, Simon C, Saindrenan P, Expert D (2009)
757 Microbial siderophores exert a subtle role in *Arabidopsis* during infection by
758 manipulating the immune response and the iron status. Plant Physiol. 150:1687-1696.

759 Dufresne M, Seva C, Fourmy D (2006) Cholecystokinin and gastrin receptors. Physiol
760 Rev. 86:805-847.

761 Estévez-Calvar N, Romero A, Figueras A, Novoa B (2011) Involvement of pore-
762 forming molecules in immune defense and development of the Mediterranean mussel
763 (*Mytilus galloprovincialis*). Dev Comp Immunol. 35:1017-1031.

764 Ewen-Campen B, Shaner N, Panfilio K, Suzuki Y, Roth S, Extavour CG (2011) The
765 maternal and early embryonic transcriptome of the milkweed bug *Oncopeltus fasciatus*.
766 BMC Genomics. 12:61.

767 Fino KK, Matters GL, McGovern CO, Gilius EL, Smith JP (2012) Downregulation of
768 the CCK-B receptor in pancreatic cancer cells blocks proliferation and promotes
769 apoptosis. Am J Physiol Gastrointest Liver Physiol. 302:G1244-52.

Fisher WS (1986) Structure and Functions of Oyster Hemocytes. In: Brehélin M. (eds) Immunity in Invertebrates. Proceedings in Life Sciences. Springer, Berlin, Heidelberg.

Flajnik MF (2010) All GOD's creatures got dedicated mucosal immunity. Nat Immunol. 11:777-779.

Galindo J, Grahame JW, Butlin RK (2010) An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. J. Evol. Biol. 23:2004-2016.

Gerdol M, De Moro G, Manfrin C, Milandri A, Riccardi E, Beran A, Venier P, Pallavicini A (2014) RNA sequencing and de novo assembly of the digestive gland transcriptome in *Mytilus galloprovincialis* fed with toxinogenic and non-toxic strains of *Alexandrium minutum*. BMC Res Notes. 7:722.

Gerdol M, Venier P (2015) An updated molecular basis for mussel immunity. Fish Shellfish Immunol. 46:17-38.

Gestal C, Costa M, Figueras A, Novoa B (2007) Analysis of differentially expressed genes in response to bacterial stimulation in hemocytes of the carpetshell clam *Ruditapes decussatus*: identification of new antimicrobial peptides. Gene. 406:134-143.

Giles H, Pilditch CA, Bell DG (2006) Sedimentation from mussel (*Perna canaliculus*) culture in the Firth of Thames, New Zealand: Impacts on sediment oxygen and nutrient fluxes. Aquaculture. 261:125-140.

Gómez-León J, Villamil L, Lemos ML, Novoa B, Figueras A (2005) Isolation of *V. alginolyticus* and *Vibrio splendidus* from aquacultured carpet shell clam (*Ruditapes decussatus*) larvae associated with mass mortalities. Appl Environ Microb. 71:98-104.

Green TJ, Raftos D, Speck P, Montagnani C (2015) Antiviral immunity in marine molluscs. J Gen Virol. 96:2471-2482.

Hallen LC, Burki Y, Ebeling M, Broger C, Siegrist F, Oroszlan-Szovik K, Bohrmann B, Certa U, Foser S (2007) Antiproliferative activity of the human IFN-alpha-inducible protein IFI44. J Interferon Cytokine Res. 27:675-680.

Heinonen MT, Kanduri K, Lähdesmäki HJ, Lahesmaa R, Henttinen TA (2015) Tubulin- and actin-associating GIMAP4 is required for IFN-γ secretion during Th cell differentiation. Immunol Cell Biol. 93:158-166.

Hertel LA, Bayne CJ, Loker ES (2002) The symbiont *Capsaspora owczarzaki*, nov. gen. nov. sp., isolated from three strains of the pulmonate snail *Biomphalaria glabrata* is related to members of the Mesomycetozoea. Int J Parasitol. 32:1183-1191.

Hotamisligil GS and Erbay E (2008) Nutrient sensing and inflammation in metabolic diseases. Nat Rev Immunol. 8:923-934.

Hu H, Bandyopadhyay PK, Olivera BM, Yandell M (2011) Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. BMC Genomics. 12:60.

Huan P, Wang H, Liu B (2012) Transcriptomic analysis of the clam *Meretrix meretrix* on different larval stages. Mar Biotechnol. 14:69-78.

Huang XD, Zhao M, Liu WG, Guan YY, Shi Y, Wang Q, Wu SZ, He MX (2013) Gigabase-Scale Transcriptome Analysis on Four Species of Pearl Oysters. Mar Biotechnol. 5:253-264.

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

Iyer A, Brown L, Whitehead JP, Prins JB, Fairlie DP (2015) Nutrient and immune sensing are obligate pathways in metabolism, immunity, and disease. FASEB J. 29:3612-3625.

Jeffroy F, Brulle F, Paillard C (2013) Differential expression of genes involved in immunity and biomineralization during Brown Ring Disease development and shell repair in the Manila clam, *Ruditapes philippinarum*. J Invertebr Pathol. 113:129-136.

Jiang JZ, Zhang W, Guo ZX, Cai CC, Su YL, Wang RX, Wang JY (2011) Functional annotation of an expressed sequence tag library from *Haliotis diversicolor* and analysis of its plant-like sequences. Mar Genomics. 4:189-196. doi: 10.1016/j.margen.2011.05.001

Kawai T, Akira S (2011) Regulation of innate immune signalling pathways by the tripartite motif (TRIM) family proteins. EMBO Mol Med. 3:513-527.

Kinoshita S, Wang N, Inoue H, Maeyama K, Okamoto K, Nagai K, Kondo H, Hirono I, Asakawa S, Watabe S (2011) Deep Sequencing of ESTs from Nacreous and Prismatic Layer Producing Tissues and a Screen for Novel Shell Formation-Related Genes in the Pearl Oyster. PLoS One. 6:e21238.

Kurelec B, Pivčević B (1991) Evidence for a multixenobiotic resistance mechanism in the mussel *Mytilus galloprovincialis*. Aquat Toxicol. 19:291-301.

Le Roux F, Wegner KM, Baker-Austin C, Vezzulli L, Osorio CR, Amaro C, Ritchie JM, Defoirdt T, Destoumieux-Garzón D, Blokesch M, Mazel D, Jacq A, Cava F, Gram L, Wendling CC, Strauch E, Kirschner A, Huehn S (2015) The emergence of *Vibrio* pathogens in Europe: ecology, evolution, and pathogenesis. Front Microbiol. 6:830.

Leite RB, Milan M, Coppe A, Bortoluzzi S, dos Anjos A, Reinhardt R, Saavedra C, Patarnello T, Cancela ML, Bargelloni L (2013) mRNA-Seq and microarray development for the Grooved Carpet shell clam, *Ruditapes decussatus*: a functional approach to unravel host-parasite interaction. BMC Genomics, 14:741.

Lemer S, González VL, Bieler R, Giribet G (2016) Cementing mussels to oysters in the pteriomorphian tree: a phylogenomic approach. Proc Biol Sci; 283:pii: 20160857.

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

847 Liao Z, Wang XC, Liu HH, Fan MH, Sun JJ, Shen W (2013) Molecular
848 characterization of a novel antimicrobial peptide from *Mytilus coruscus*. Fish Shellfish
849 Immunol. 34:610-616.

850 Marteil L (1976) Shellfish culture in France. Part 2. Oyster and mussel biology.
851 Original title: La conchyliculture française. Biologie de l'huître et de la moule. Revue
852 des Travaux de l´institut des Pêches maritimes. 40, 149-345. France: IFREMER.
853 Available from: http://archimer.ifremer.fr/doc/00000/1796/

854 Martin M (2011) Cutadapt removes adapter sequences from high-throughput
855 sequencing reads. EMBnet.journal. 17:1.

856 McDowell IC, Modak TH, Lane CE, Gomez-Chiarri M (2016) Multi-species protein
857 similarity clustering reveals novel expanded immune gene families in the eastern oyster
858 *Crassostrea virginica*. Fish Shellfish Immunol. 53:13-23.

859 McNab FW, Rajsbaum R, Stoye JP, O'Garra A (2011) Tripartite-motif proteins and
860 innate immune regulation. Curr Opin Immunol. 23:46-56.

861 Meyer E, Aglyamova G, Wang S, Buchanan-Carter J, Abrego D, Colbourne JK, Willis
862 BL, Matz MV (2009) Sequencing and de novo analysis of a coral larval transcriptome
863 using 454 GSFlx. BMC Genomics. 10:219.

864 Milan M, Coppe A, Reinhardt R, Cancela LM, Leite RB, Saavedra C, Ciofi C, Chelazzi
865 G, Patarnello T, Bortoluzzi S, Bargelloni L (2011) Transcriptome sequencing and
866 microarray development for the Manila clam, *Ruditapes philippinarum*: genomic tools
867 for environmental monitoring. BMC Genomics. 12:234.

868 Milan M, Ferraresso S, Ciofi C, Chelazzi G, Carrer C, Ferrari G, Pavan L, Patarnello T,
869 Bargelloni L (2013) Exploring the effects of seasonality and chemical pollution on the
870 hepatopancreas transcriptome of the Manila clam. Mol Ecol. 22:2157-2172.

871 Mitta G, Vandenbulcke F, Noël T, Romestand B, Beauvillain JC, Salzet M, Roch P
872 (2000a) Differential distribution and defence involvement of antimicrobial peptides in
873 mussel. J Cell Sci. 113:2759-2769.

874 Mitta G, Vandenbulcke F, Roch P (2000b). Original involvement of antimicrobial
875 peptides in mussel innate immunity. FEBS Lett. 486:185-190.

876 Moreira R, Balseiro P, Romero A, Dios S, Posada D, Novoa B, Figueras A (2012a)
877 Gene expression analysis of clams *Ruditapes philippinarum* and *Ruditapes decussatus*
878 following bacterial infection yields molecular insights into pathogen resistance and
879 immunity. Dev Comp Immunol. 36:140-149.

880 Moreira R, Balseiro P, Planas JV, Fuste B, Beltran S, Novoa B, Figueras A (2012b)
881 Transcriptomics of in vitro immune-stimulated hemocytes from the Manila clam
882 *Ruditapes philippinarum* using high-throughput sequencing. PLoS One. 7:e35009.

883 Moreira R, Pereiro P, Canchaya C, Posada D, Figueras A, Novoa B (2015) RNA-Seq in
884 *Mytilus galloprovincialis*: comparative transcriptomics and expression profiles among
885 different tissues. BMC Genomics. 16:728.

886 Morton B (1981) The Anomalodesmata. Malacologia. 21:35-60.

887 Mount AS, Wheeler AP, Paradkar RP, Snider D (2004) Hemocyte-mediated shell
888 mineralization in the eastern oyster. Science. 304:297-300.

889 Murgarella M, Puiu D, Novoa B, Figueras A, Posada D, Canchaya C (2016) A First
890 Insight into the Genome of the Filter-Feeder Mussel *Mytilus galloprovincialis*. PLoS
891 One. 11:e0151561.

892 Nairz M, Haschka D, Demetz E, Weiss G (2014) Iron at the interface of immunity and
893 infection. Front Pharmacol. 5:152.

894 Newell RIE (2004) Ecosystem influences of natural and cultivated populations of
895 suspension-feeding bivalve molluscs: a review. J shellfish res. 23:51-61.

896 Novoa B, Romero A, Álvarez ÁL, Moreira R, Pereiro P, Costa MM, Dios S, Estepa A,
897 Parra F, Figueras A (2016) Antiviral activity of myticin C peptide from mussel: an
898 ancient defence against herpesviruses. J Virol. 90:7692:7702.

899 Pallavicini A, Costa M del M, Gestal C, Dreos R, Figueras A, Venier P, Novoa B
900 (2008) High sequence variability of myticin transcripts in hemocytes of immune-
901 stimulated mussels suggests ancient host-pathogen interactions. Dev Comp Immunol.
902 32:213-226.

903 Philipp EE, Kraemer L, Melzner F, Poustka AJ, Thieme S, Findeisen U, Schreiber S,
904 Rosenstiel P (2012) Massively parallel RNA sequencing identifies a complex immune
905 gene repertoire in the lophotrochozoan *Mytilus edulis*. PLoS One. 7: e33091.

906 Pinkenburg O, Meyer T, Bannert N, Norley S, Bolte K, Czudai-Matwich V Herold S,
907 Gessner A, Schnare M (2016) The Human Antimicrobial Protein
908 Bactericidal/Permeability-Increasing Protein (BPI) Inhibits the Infectivity of Influenza
909 A Virus. PLoS One. 11:e0156929.

910 Plazzi F, Passamonti M (2010) Towards a molecular phylogeny of Molluscs: Bivalves'
911 early evolution as revealed by mitochondrial genes. Mol Phylogenet Evol. 57:641-657.

912 Ponder WF, Lindberg DR (2008) Phylogeny and Evolution of the Mollusca. Oakland,
913 CA: University of California Press.

914 Power D, Santoso N, Dieringer M, Yu J, Huang H, Simpson S, Seth I, Miao H, Zhu J
915 (2015) IFI44 suppresses HIV-1 LTR promoter activity and facilitates its latency.
916 Virology. 481:142-150

917 Qin CL, Huang W, Zhou SQ, Wang XC, Liu HH, Fan MH, Wang RX, Gao P, Liao Z
918 (2014) Characterization of a novel antimicrobial peptide with chitin-biding domain
919 from *Mytilus coruscus*. Fish Shellfish Immunol. 41:362-370.

920 Ren W, Chen H, Renault T, Cai Y, Bai C, Wang C, Huang J (2013) Complete genome
921 sequence of acute viral necrosis virus associated with massive mortality outbreaks in
922 the Chinese scallop, *Chlamys farreri*. Virol J. 10:110.

Ren Y, Xue J, Yang H, Pan B, Bu W (2017) Transcriptome analysis of *Ruditapes philippinarum* hepatopancreas provides insights into immune signaling pathways under *Vibrio anguillarum* infection. Fish Shellfish Immunol. 64:14-23.

Riesgo A, Andrade SC, Sharma P, Novo M, Pérez-Porro AR, Vahtera V, González VL, Kawauchi GY, Giribet G (2012) Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. Front Zool. 9:33.

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

Romero A, Costa Md, Forn-Cuni G, Balseiro P, Chamorro R, Dios S, Figueras A, Novoa B (2014) Occurrence, seasonality and infectivity of *Vibrio* strains in natural populations of mussels *Mytilus galloprovincialis*. Dis Aquat Organ. 108:149-163.

Rosani U, Varotto L, Rossi A, Roch P, Novoa B, Figueras A, Pallavicini A, Venier P (2011) Massively parallel amplicon sequencing reveals isotype-specific variability of antimicrobial peptide transcripts in *Mytilus galloprovincialis*. PLoS One. 6:e26680.

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

Ruano F, Batista FM, Arcangeli G (2015) Perkinsosis in the clams *Ruditapes decussatus* and *R. philippinarum* in the Northeastern Atlantic and Mediterranean Sea: A review. J Invertebr Pathol. 131:58-67.

Sánchez-Pla A, Reverter F, Ruíz de Villa MC, Comabella M (2012) Transcriptomics: mRNA and alternative splicing. J Neuroimmunol. 248:23-31.

Schmieder R and Edwards R (2011) Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. PLoS One. 6:e17288.

Segarra A, Pepin JF, Arzul I, Morga B, Faury N, Renault T (2010) Detection and description of a particular Ostreid herpesvirus 1 genotype associated with massive mortality outbreaks of Pacific oysters, *Crassostrea gigas*, in France in 2008. Virus Res. 153:92-99.

Shagin DA, Rebrikov DV, Kozhemyako VB, Altshuler IM, Shcheglov AS, Zhulidov PA, Bogdanova EA, Staroverov DB, Rasskazov VA, Lukyanov S (2002) A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. Genome Res. 12:1935-1942.

Shi Y, Yu C, Gu Z, Zhan X, Wang Y, Wang A (2013) Characterization of the Pearl Oyster (*Pinctada martensii*) Mantle Transcriptome Unravels Biomineralization Genes. Mar Biotechnol. 15:175-187.

962 Sillanpää JK, Ramesh K, Melzner F, Sundh H, Sundell K (2016) Calcium mobilisation
963 following shell damage in the Pacific oyster, *Crassostrea gigas*. Mar Genomics. 27:75-
964 83.

965 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015)
966 BUSCO: assessing genome assembly and annotation completeness with single-copy
967 orthologs. Bioinformatics. 31:3210-3212.

968 Sun Y, Zhang Y, Fu X, Zhang R, Zou J, Wang S, Hu X, Zhang L, Bao Z (2014)
969 Identification of two secreted ferritin subunits involved in immune defense of Yesso
970 scallop *Patinopecten yessoensis*. Fish Shellfish Immunol. 37:53-59.

971 Suttle CA (2007) Marine viruses – major players in the global ecosystem. Nat Rev
972 Microbiol. 5:801-812.

973 Tornatore L, Thotakura AK, Bennett J, Moretti M, Franzoso G (2012) The nuclear
974 factor kappa B signaling pathway: integrating metabolism with inflammation. Trends
975 Cell Biol. 22:557-566.

976 Toubiana M, Gerdol M, Rosani U, Pallavicini A, Venier P, Roch P (2013)Toll-like
977 receptors and MyD88 adaptors in *Mytilus*: complete cds and gene expression levels.
978 Dev Comp Immunol. 40:158-166.

979 van der Meer JW, Joosten LA, Riksen N, Netea MG (2015) Trained immunity: A smart
980 way to enhance innate immune defence. Mol Immunol. 68:40-44.

981 van Diepen MT, Spencer GE, van Minnen J, Gouwenberg Y, Bouwman J, Smit AB,
982 van Kesteren RE (2005) The molluscan RING-finger protein L-TRIM is essential for
983 neuronal outgrowth. Mol Cell Neurosci. 29:74-81.

984 Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden
985 JH (2008) Rapid transcriptome characterization for a nonmodel organism using 454
986 pyrosequencing. Mol Ecol. 17:1636-1647.

987 Villalba A, Reece KS, Ordas MC, Casas SM, Figueras A (2004) Perkinsosis in
988 molluscs: a review. Aquat Living Res. 17:411e32.

989 Wang J, Wang L, Yang C, Jiang Q, Zhang H, Yue F, Huang M, Sun Z, Song L (2013)
990 The response of mRNA expression upon secondary challenge with *Vibrio anguillarum*
991 suggests the involvement of C-lectins in the immune priming of scallop *Chlamys*
992 *farreri*. Dev Comp Immunol. 40:142-147.

993 Werner GD, Gemmell P, Grosser S, Hamer R, Shimeld SM (2013) Analysis of a deep
994 transcriptome from the mantle tissue of *Patella vulgata* Linnaeus (Mollusca:
995 Gastropoda: Patellidae) reveals candidate biomineralising genes. Mar Biotechnol.
996 15:230-243

997 Westermann AJ, Förstner KU, Amman F, Barquist L, Chao Y, Schulte LN, Müller L,
998 Reinhardt R, Stadler PF, Vogel J (2016) Dual RNA-seq unveils noncoding RNA
999 functions in host-pathogen interactions. Nature. 529:496-501.

1000 Yarra T, Gharbi K, Blaxter M, Peck LS, Clark MS (2016) Characterization of the
1001 mantle transcriptome in bivalves: *Pecten maximus, Mytilus edulis* and *Crassostrea*
1002 *gigas*. Mar Genomics. 27:9-15.

1003 Zhao J, Li C, Chen A, Li L, Su X, Li T (2010) Molecular characterization of a novel
1004 big defensin from clam *Venerupis philippinarum*. PLoS One. 5:e13480.

1005 Zhao X, Wang Q, Jiao Y, Huang R, Deng Y, Wang H, Du X (2012) Identification of
1006 Genes Potentially Related to Biomineralization and Immunity by Transcriptome
1007 Analysis of Pearl Sac in Pearl Oyster *Pinctada martensii*. Mar Biotechnol. 14:730-739.

1008 Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL,
1009 Kozhemyako VB, Matz MV, Meleshkevitch E, Moroz LL, Lukyanov SA, Shagin DA
1010 (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease.
1011 Nucleic Acids Res. 32:e37.

1012 Zorita I, Apraiz I, Ortiz-Zarragoitia M, Orbea A, Cancio I, Soto M, Marigómez I,
1013 Cajaraville MP (2007) Assessment of biological effects of environmental pollution
1014 along the NW Mediterranean Sea using mussels as sentinel organisms. Environ Pollut.
1015 148:236-250.

1016

1017

Figure legends

Figure 1. Schematic flow-through of the work tasks and data processing pipeline.

Figure 2. Assembly statistics per species. A. Contigs arranged by read mapping. B. Contigs arranged by length in bp. C. Isotigs arranged by contig clustering. D. Number of sequences that belong to a unique gene or isogroup.

Figure 3. Top 35 hit sequences identified by BLASTx. Graphs represent the top 35 hits against non-redundant and SwissProt databases. Pie charts represent the top hit species in the BLASTx results.

Figure 4. BUSCO analysis per species. A. with singletons; B. without singletons.

Figure 5. Comparison among unique Gene Accession numbers of the immune transcriptomes per species. Numbers refer to the number of sequences belonging to each group: *M. galloprovincialis, M. edulis, R. decussatus* and *R. philippinarum*.

Figure 6. Blast2GO enrichment analysis showing the comparison of the four bivalve species considered. Only the top 20 overrepresented GO terms for each species are shown. Black bars (all) represent the whole dataset of the four transcriptomes.

Tables

Table 1. Summary of sequencing data and assembly process.

	M.galloprovincialis	*M. edulis*	*R. decussatus*	*R. philippinarum*
Sequencing statistics				Moreira *et al.*, 2012b
Total number of reads	447,336	421,821	457,559	975,190
Filtered	386,838 (86.5%)	359,392 (85.2%)	356,896 (78%)	974,976
≥ 600 bp	33,648	39,822	47,682	-
Assembly statistics				Moreira *et al.*, 2012b
Reads assembled	215,664 (48.21%)	191,288 (45.35%)	250,774 (54.80%)	842,957 (86.44%)
Bases assembled (ba)	178,431,264	166,441,769	101,270,122	216,454,498
Contigs	9,273	9,016	4,280	51,265
Contig N50	1,034	1,034	946	677
Isotigs	14,098	13,285	7,305	-
Isotig N50	1,242	1,263	981	-
Estimated transcriptome size (eTs)	24,684,842	23,834,877	9,894,038	29,856,736
Isogroups	9,256	8,814	6,020	29,679 clusters
Estimated coverage (eC) eC=ba/eTs	7x	7x	10x	7x
Annotation statistics				
BLASTx				Present work
Total sequences* to be annotated	14,158	13,560	7,318	44,925
Annotated contigs and isotigs	7,031 (49.66%)	6,683 (49.28%)	2,327 (31.8%)	17,565 (39.09%)
Annotated singletons	19,431	18,299	7,138	-
Total annotated sequences	26,462	24,982	9,465	-
Blast2GO				Present work
Total sequences** to be assigned GO terms	33,589	31,859	14,456	44,925
sequences with GO terms	20,382 (60.68%)	19,483 (61.15%)	7,815 (54.06%)	16,568 (36.87%)
sequences without GO terms	13,206 (39.32%)	12,376 (38.85%)	6,641 (45.94%)	28,357 (63.13%)

* Contigs + isotigs

** Total contigs + isotigs and only annotated singletons

1040 **Table 2. Putative pathogen sequences found in the four studied species.**

Pathogen species	***M.galloprovincialis*** **nº sequences**	***M. edulis*** **nº sequences**	***R. decussatus*** **nº sequences**	***R. philippinarum*** **nº sequences**
Virus				
Ostreid herpesvirus 1	1	1	36	-
Chlamys acute necrosis virus	-	1	11	2
Abalone herpesvirus	-	-	2	-
White spot syndrome virus	1	1	-	-
Other viruses	9	11	4	8
Bacteria				
Vibrio alginolyticus	2	1	-	1
Vibrio anguillarum	18	10	5	18
Vibrio harveyi	-	-	1	2
Vibrio parahemolyticus	2	1	-	9
Vibrio splendidus	16	16	-	10
Other *Vibrio* species	42	42	8	154
Genus *Aeromonas*	1	-	-	3
Genus *Pseudomomas*	-	2	3	10
Order *Ricketsiales*	1	-	-	1
Protozoa				
Perkinsus marinus	-	4	131	3
Fungi				
-	63	69	31	155

1041

1042

1043 **Table 3. Analysis of viral sequences found in the four bivalves transcriptomes.**
1044 Cells highlighted in grey show viruses of special interest in aquaculture.

	BLASTx	Virus	e-value	Acc nº
M. galloprovincialis				
HJVT2KG01DOKNY	ubiquitin/ribosomal protein	*Amsacta moorei* entomopoxvirus	9.00E-25	NP_064949.1
HJVT2KG01AHE8T	inhibitor of apoptosis protein 3	*Anticarsia gemmatalis* multiple nucleopolyhedrovirus	2.00E-12	ALR70562.1
HJVT2KG01BZJ42	prostaglandin G/H synthase 2	Cercopithecine herpesvirus 5	7.00E-17	YP_004935976.1
HJVT2KG01B3SFE	replication-associated protein	Columbid circovirus	2.00E-20	AID18515.1
HJVT2KG01D3YL8	replication associated protein	Dragonfly larvae associated circular virus-10	2.00E-38	ALE29830.1
HJVT2KG01CP5K3	JP-B_gp2	Marine RNA virus JP-B	4.00E-10	YP_001429584.1
isotig07745	ORF MSV248	*Melanoplus sanguinipes* entomopoxvirus	8.00E-07	NP_048319.1
isotig09592	hypothetical protein MpV1_050	*Micromonas sp.* RCC1109 virus MpV1	2.00E-10	YP_004061933.1
isotig09267	ORF87	Ostreid herpesvirus 1	7.00E-24	YP_024626.1
isotig05220	FIC protein	White spot syndrome virus	3.00E-17	ABM92267.1
M. edulis				
HJVT2KG02H0VP6	inhibitor of apoptosis protein 3	*Anticarsia gemmatalis* nucleopolyhedrovirus	9.00E-11	YP_803428.1
HJVT2KG02IXCNJ	poryprotein	Bovine viral diarrhea virus 1	2,77E-16	BAD04937.1
HJVT2KG02HBT0R	replication associated protein	Dragonfly larvae associated circular virus-10	1.00E-37	ALE29830.1
HJVT2KG02J2K4D	ORF99	Ostreid herpesvirus 1	1.00E-17	AKM21032.1
R. decussatus				
HIUQ1KI01BT4EW	putative methyltransferase	Abalone herpesvirus Victoria/AUS/2009	8.00E-26	YP_006908738.1
HIUQ1KI01EGWYH	19.5g1 protein	*Chelonus inanitus* bracovirus	2.00E-12	CAO98966.1
HIUQ1KI01D3L3C	putative replicase protein	Dromedary stool-associated circular ssDNA virus	5.00E-18	AIY31261.1
HIUQ1KI01DCN0E	RNA-dependent DNA polymerase	Lymphocystis disease virus Sa	5.00E-07	AOC55195.1
HIUQ1KI01BSHPU	putative serine_threonine protein kinase receptor	Moumouvirus Monve	8.00E-11	AEX62336.1
isotig04097	ORF22	Ostreid herpesvirus 1	3.00E-30	YP_024567.1
HIUQ1KI01BI46K	ORF24	Ostreid herpesvirus 1	1.00E-11	YP_024569.1
isotig06796	ORF25	Ostreid herpesvirus 1	2.00E-09	AKM20963.1
HIUQ1KI01DLMDF	ORF44	Ostreid herpesvirus 1	2.00E-22	AKM20983.1
HIUQ1KI01AM8YO	ORF47	Ostreid herpesvirus 1	2.00E-30	YP_024591.1
HIUQ1KI01AXRA8	ORF54	Ostreid herpesvirus 1	3.00E-40	AKM20991.1
isotig02001	ORF57	Ostreid herpesvirus 1	8.00E-25	YP_024600.1
HIUQ1KI01CNG7E	ORF61	Ostreid herpesvirus 1	9.00E-08	YP_024604.1
HIUQ1KI01BUFVO	ORF66	Ostreid herpesvirus 1	5.00E-10	AKM21000.1
isotig03837	ORF67	Ostreid herpesvirus 1	2.00E-59	AKM21001.1
HIUQ1KI01ETI0R	ORF68	Ostreid herpesvirus 1	2.00E-52	AKM21002.1
HIUQ1KI01EH4XM	ORF69	Ostreid herpesvirus 1	7.00E-31	AKM21003.1
isotig04787	ORF71	Ostreid herpesvirus 1	2.00E-05	AKM21005.1
isotig04807	ORF82	Ostreid herpesvirus 1	6.00E-19	AKM21015.1
HIUQ1KI01EMT37	ORF87	Ostreid herpesvirus 1	5.00E-11	YP_024626.1
isotig06164	ORF88	Ostreid herpesvirus 1	3.00E-12	AKM21021.1
isotig05588	ORF89	Ostreid herpesvirus 1	5.00E-23	YP_024628.2
HIUQ1KI01C3VNN	ORF95	Ostreid herpesvirus 1	4.00E-25	YP_024634.1
HIUQ1KI01DIBSQ	ORF102	Ostreid herpesvirus 1	4.00E-14	YP_024641.1

HIUQ1KI01CZSCN	ORF103	Ostreid herpesvirus 1	1.00E-11	AKM21036.1
isotig01947	ORF104	Ostreid herpesvirus 1	2.00E-25	YP_024643.1
HIUQ1KI01DWDXO	ORF104	Ostreid herpesvirus 1	1.00E-07	AKM21037.1
HIUQ1KI01AXS3R	ORF111	Ostreid herpesvirus 1	7.00E-31	YP_024649.1
HIUQ1KI01BQLZY	ORF112	Ostreid herpesvirus 1	9.00E-18	YP_024650.1
HIUQ1KI01BM59F	ORF115	Ostreid herpesvirus 1	3.00E-16	YP_024657.1
HIUQ1KI01ECIBG	ORF117	Ostreid herpesvirus 1	2.00E-18	YP_024659.1
R. philippinarum				
c28532	large T and small t antigens	Aves polyomavirus 1	2.00E-06	AAC33626.1
rep_c37545	putative helicase/proteas	Bovine viral diarrhea virus 1	3.00E-26	AAA42855.1
rep_c36945	p125 protein	Bovine viral diarrhea virus 1	3.00E-40	AAA53067.1
rep_c34991	membrane glycoprotein UL144	Cynomolgus macaque cytomegalovirus strain Ottawa	8.00E-06	AEQ32275.1
c13859	JM145	*Macaca fuscata* rhadinovirus	2.00E-11	AAT00122.1
c22684	ORF99	Ostreid herpesvirus 1	8.00E-16	AKM21032.1
c18828	polyprotein	Posavirus 1	1.00E-11	BAV31554.1

1045

1046

Supporting information

Figure S1. Read distribution per species. Raw data before (blue) and after (red and green) filtering obtained for the 454 GS FLX platform for the three species.

Figure S2. Quality control per species. Quality score per base position in the filtered reads under 600 bp (A, C and E) and over 600 bp (B, D, F).

File S1. Bioanalyzer results for the sequenced samples.

File S2 Nucleotidic sequences in fasta format of contigs and singletons of *M. edulis*, *M. galloprovincialis* and *R. decussatus.*

File S3 List of annotated contigs of *M. edulis*, *M. galloprovincialis* and *R. decussatus.* The file includes among other relevant information: sequence name, length, subject mapping, e-value, GO.

File S4. Enrichment analysis between *Mytilus* and *Ruditapes* genus.

File S5. Viral sequences found in the four transcriptomes.

Figure 1

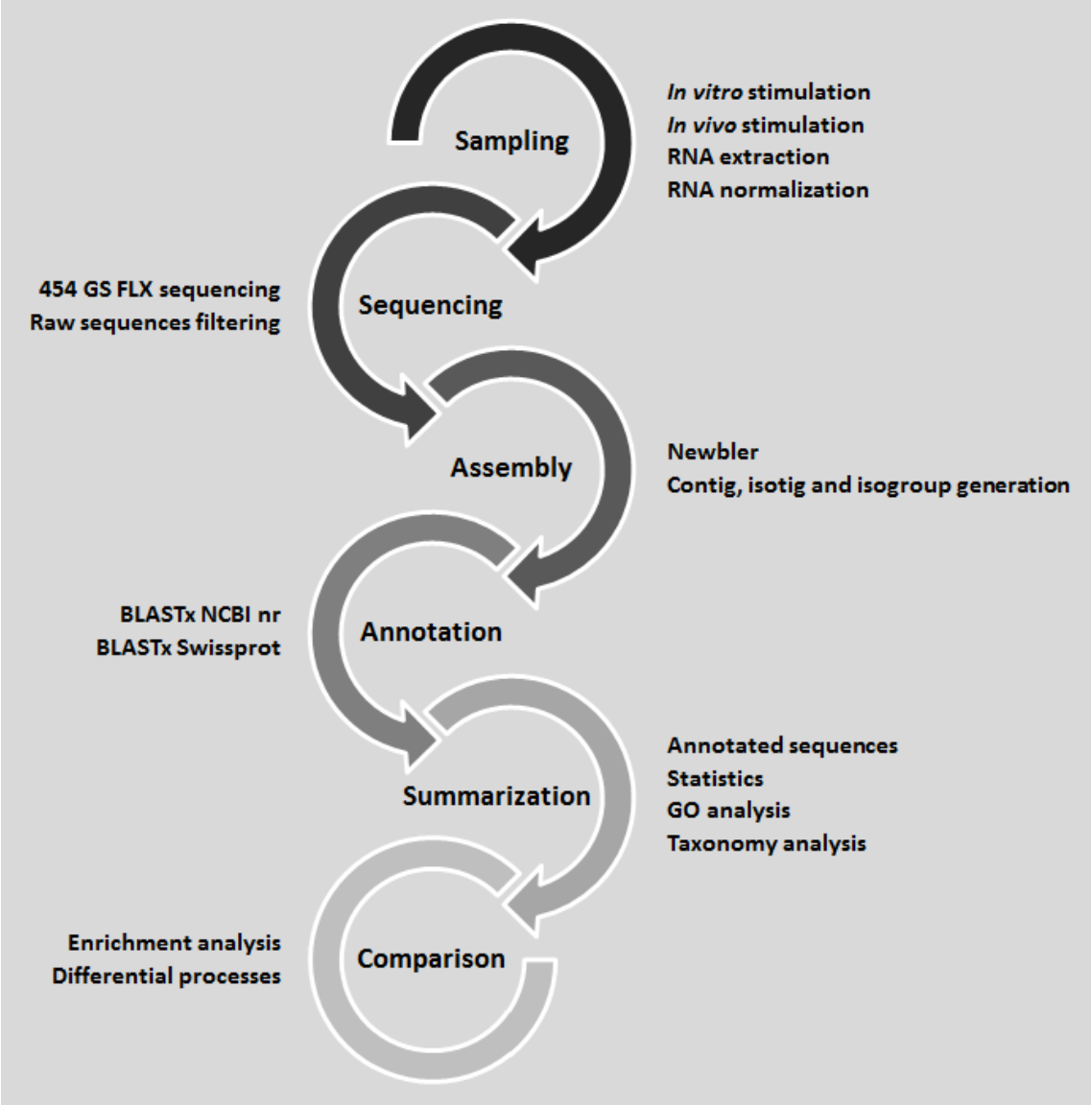

Sampling
In vitro stimulation
In vivo stimulation
RNA extraction
RNA normalization
Sequencing
454 GS FLX sequencing
Raw sequences filtering
Assembly
Newbler
Contig, isotig and isogroup generation
Annotation
BLASTx NCBI nr
BLASTx Swissprot
Summarization
Annotated sequences
Statistics
GO analysis
Taxonomy analysis
Comparison
Enrichment analysis
Differential processes

Figure 2

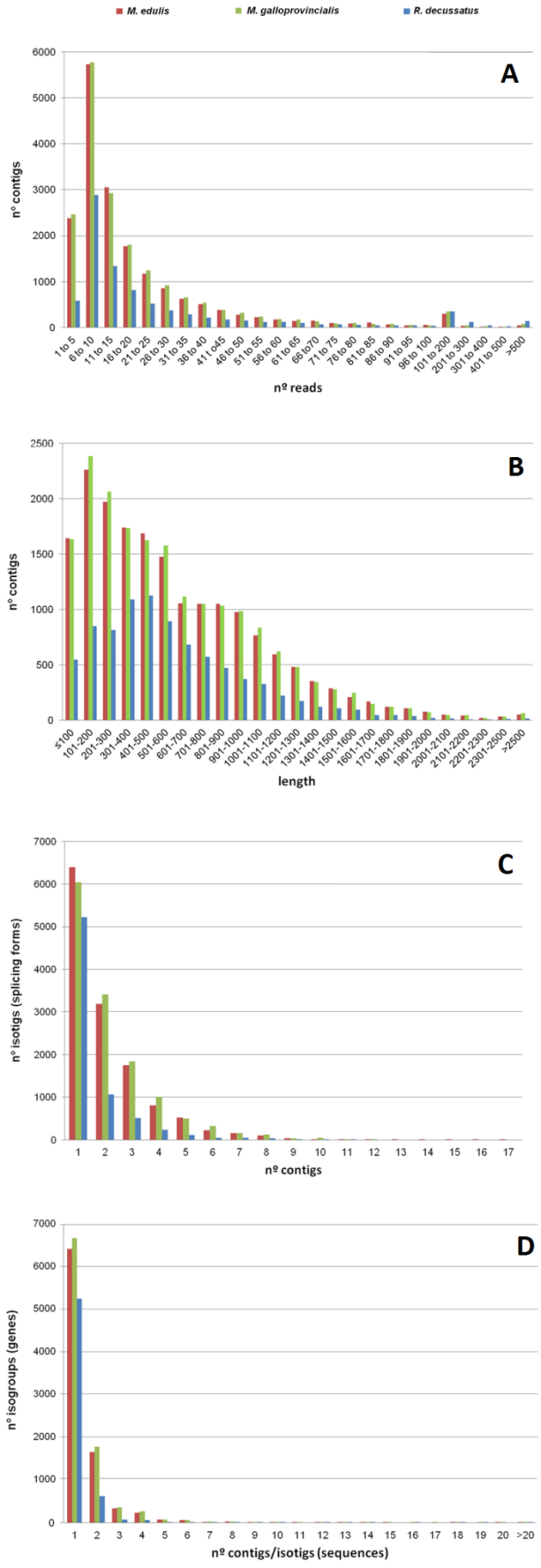

M. edulis
M. galloprovincialis
R. decussatus
A
nº contigs
nº reads
B
nº contigs
length
C
nº isotigs (splicing forms)
nº contigs
D
nº isogroups (genes)
nº contigs/isotigs (sequences)

Figure 3

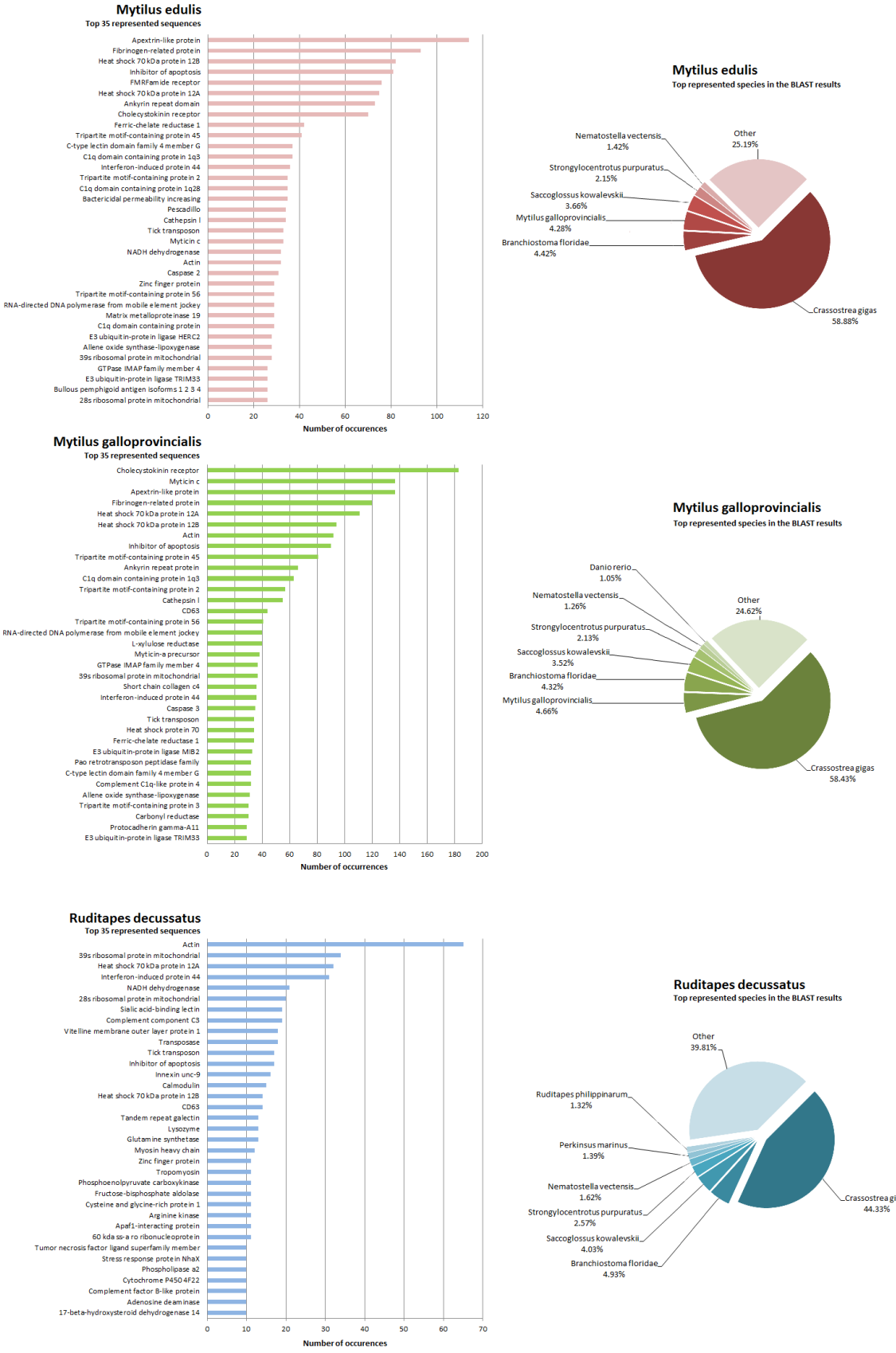

Mytilus edulis
Top 35 represented sequences
Apextrin-like protein
Fibrinogen-related protein
Heat shock 70 kDa protein 12B
Inhibitor of apoptosis
FMRFamide receptor
Heat shock 70 kDa protein 12A
Ankyrin repeat domain
Cholecystokinin receptor
Ferric-chelate reductase 1
Tripartite motif-containing protein 45
C-type lectin domain family 4 member G
C1q domain containing protein 1q3
Interferon-induced protein 44
Tripartite motif-containing protein 2
C1q domain containing protein 1q28
Bactericidal permeability increasing
Pescadillo
Cathepsin l
Tick transposon
Myticin c
NADH dehydrogenase
Actin
Caspase 2
Zinc finger protein
Tripartite motif-containing protein 56
RNA-directed DNA polymerase from mobile element jockey
Matrix metalloproteinase 19
C1q domain containing protein
E3 ubiquitin-protein ligase HERC2
Allene oxide synthase-lipoxygenase
39s ribosomal protein mitochondrial
GTPase IMAP family member 4
E3 ubiquitin-protein ligase TRIM33
Bullous pemphigoid antigen isoforms 1 2 3 4
28s ribosomal protein mitochondrial
0 20 40 60 80 100 120
Number of occurences
Mytilus edulis
Top represented species in the BLAST results
Other 25.19%
Nematostella vectensis 1.42%
Strongylocentrotus purpuratus 2.15%
Saccoglossus kowalevskii 3.66%
Mytilus galloprovincialis 4.28%
Branchiostoma floridae 4.42%
Crassostrea gigas 58.88%
Mytilus galloprovincialis
Top 35 represented sequences
Cholecystokinin receptor
Myticin c
Apextrin-like protein
Fibrinogen-related protein
Heat shock 70 kDa protein 12A
Heat shock 70 kDa protein 12B
Actin
Inhibitor of apoptosis
Tripartite motif-containing protein 45
Ankyrin repeat protein
C1q domain containing protein 1q3
Tripartite motif-containing protein 2
Cathepsin l
CD63
Tripartite motif-containing protein 56
RNA-directed DNA polymerase from mobile element jockey
L-xylulose reductase
Myticin-a precursor
GTPase IMAP family member 4
39s ribosomal protein mitochondrial
Short chain collagen c4
Interferon-induced protein 44
Caspase 3
Tick transposon
Heat shock protein 70
Ferric-chelate reductase 1
E3 ubiquitin-protein ligase MIB2
Pao retrotransposon peptidase family
C-type lectin domain family 4 member G
Complement C1q-like protein 4
Allene oxide synthase-lipoxygenase
Tripartite motif-containing protein 3
Carbonyl reductase
Protocadherin gamma-A11
E3 ubiquitin-protein ligase TRIM33
0 20 40 60 80 100 120 140 160 180 200
Number of occurrences
Mytilus galloprovincialis
Top represented species in the BLAST results
Danio rerio 1.05%
Nematostella vectensis 1.26%
Other 24.62%
Strongylocentrotus purpuratus 2.13%
Saccoglossus kowalevskii 3.52%
Branchiostoma floridae 4.32%
Mytilus galloprovincialis 4.66%
Crassostrea gigas 58.43%
Ruditapes decussatus
Top 35 represented sequences
Actin
39s ribosomal protein mitochondrial
Heat shock 70 kDa protein 12A
Interferon-induced protein 44
NADH dehydrogenase
28s ribosomal protein mitochondrial
Sialic acid-binding lectin
Complement component C3
Vitelline membrane outer layer protein 1
Transposase
Tick transposon
Inhibitor of apoptosis
Innexin unc-9
Calmodulin
Heat shock 70 kDa protein 12B
CD63
Tandem repeat galectin
Lysozyme
Glutamine synthetase
Myosin heavy chain
Zinc finger protein
Tropomyosin
Phosphoenolpyruvate carboxykinase
Fructose-bisphosphate aldolase
Cysteine and glycine-rich protein 1
Arginine kinase
Apaf1-interacting protein
60 kda ss-a ro ribonucleoprotein
Tumor necrosis factor ligand superfamily member
Stress response protein NhaX
Phospholipase a2
Cytochrome P450 4F22
Complement factor B-like protein
Adenosine deaminase
17-beta-hydroxysteroid dehydrogenase 14
0 10 20 30 40 50 60 70
Number of occurences
Ruditapes decussatus
Top represented species in the BLAST results
Other 39.81%
Ruditapes philippinarum 1.32%
Perkinsus marinus 1.39%
Nematostella vectensis 1.62%
Strongylocentrotus purpuratus 2.57%
Saccoglossus kowalevskii 4.03%
Branchiostoma floridae 4.93%
Crassostrea gigas 44.33%

Figure 4

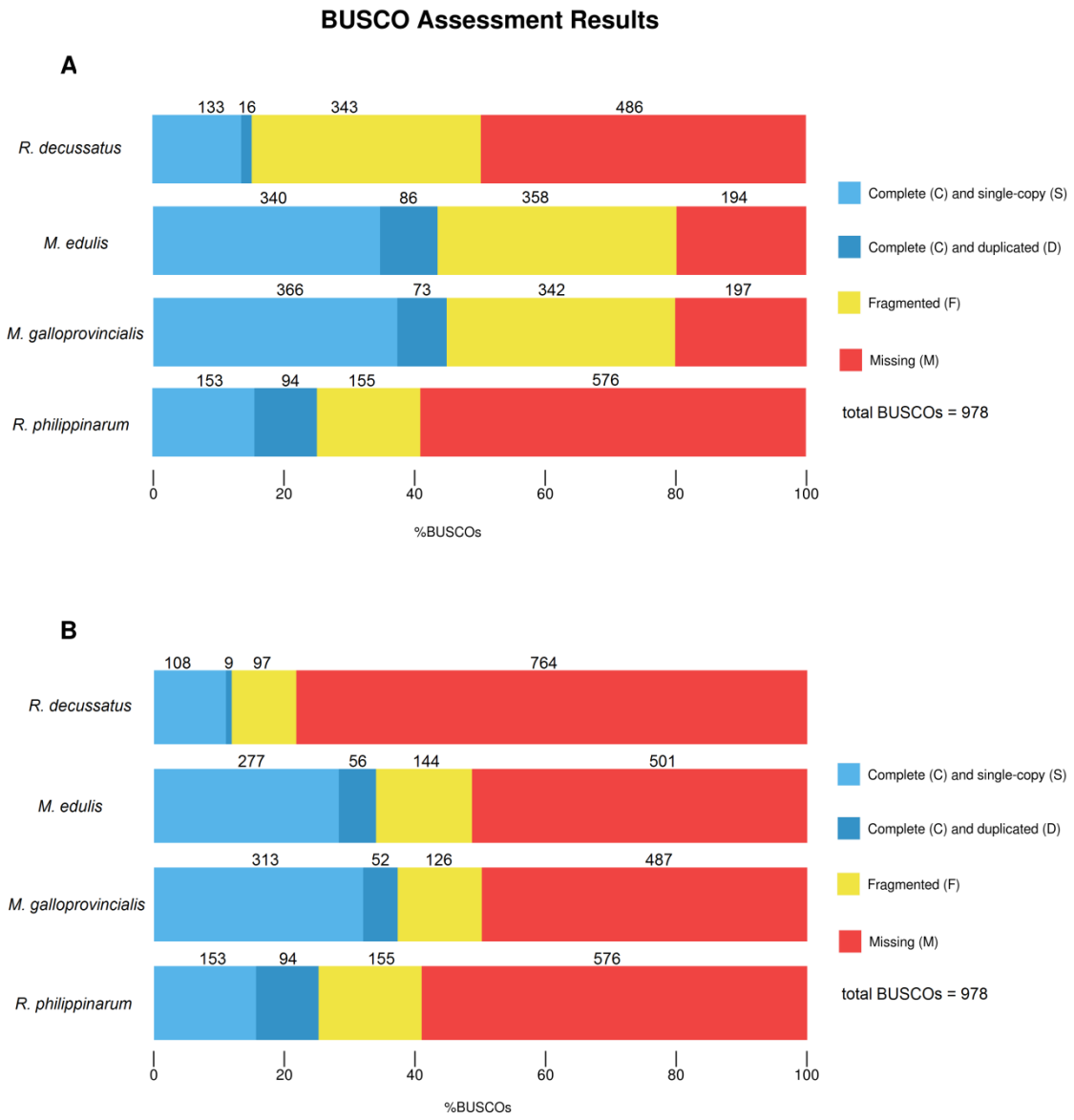

BUSCO Assessment Results
A
R. decussatus
133
16
343
486
M. edulis
340
86
358
194
M. galloprovincialis
366
73
342
197
R. philippinarum
153
94
155
576
0
20
40
60
80
100
%BUSCOs
Complete (C) and single-copy (S)
Complete (C) and duplicated (D)
Fragmented (F)
Missing (M)
total BUSCOs = 978
B
R. decussatus
108
9
97
764
M. edulis
277
56
144
501
M. galloprovincialis
313
52
126
487
R. philippinarum
153
94
155
576
0
20
40
60
80
100
%BUSCOs
Complete (C) and single-copy (S)
Complete (C) and duplicated (D)
Fragmented (F)
Missing (M)
total BUSCOs = 978

Figure 5

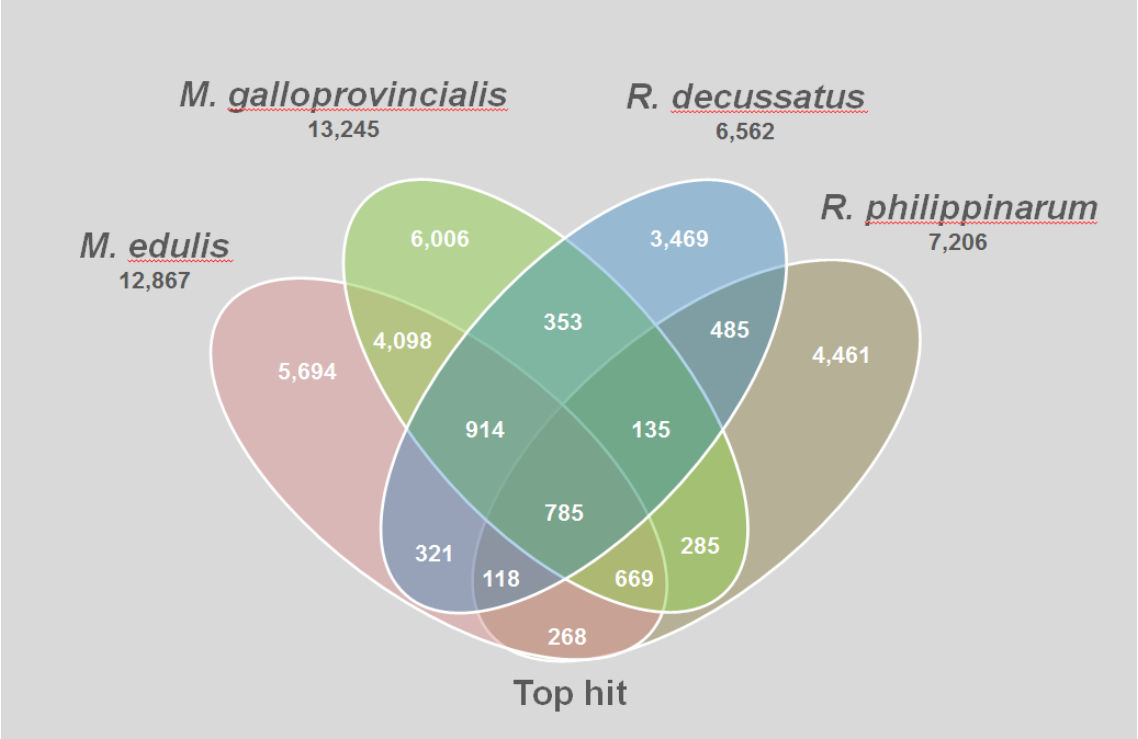
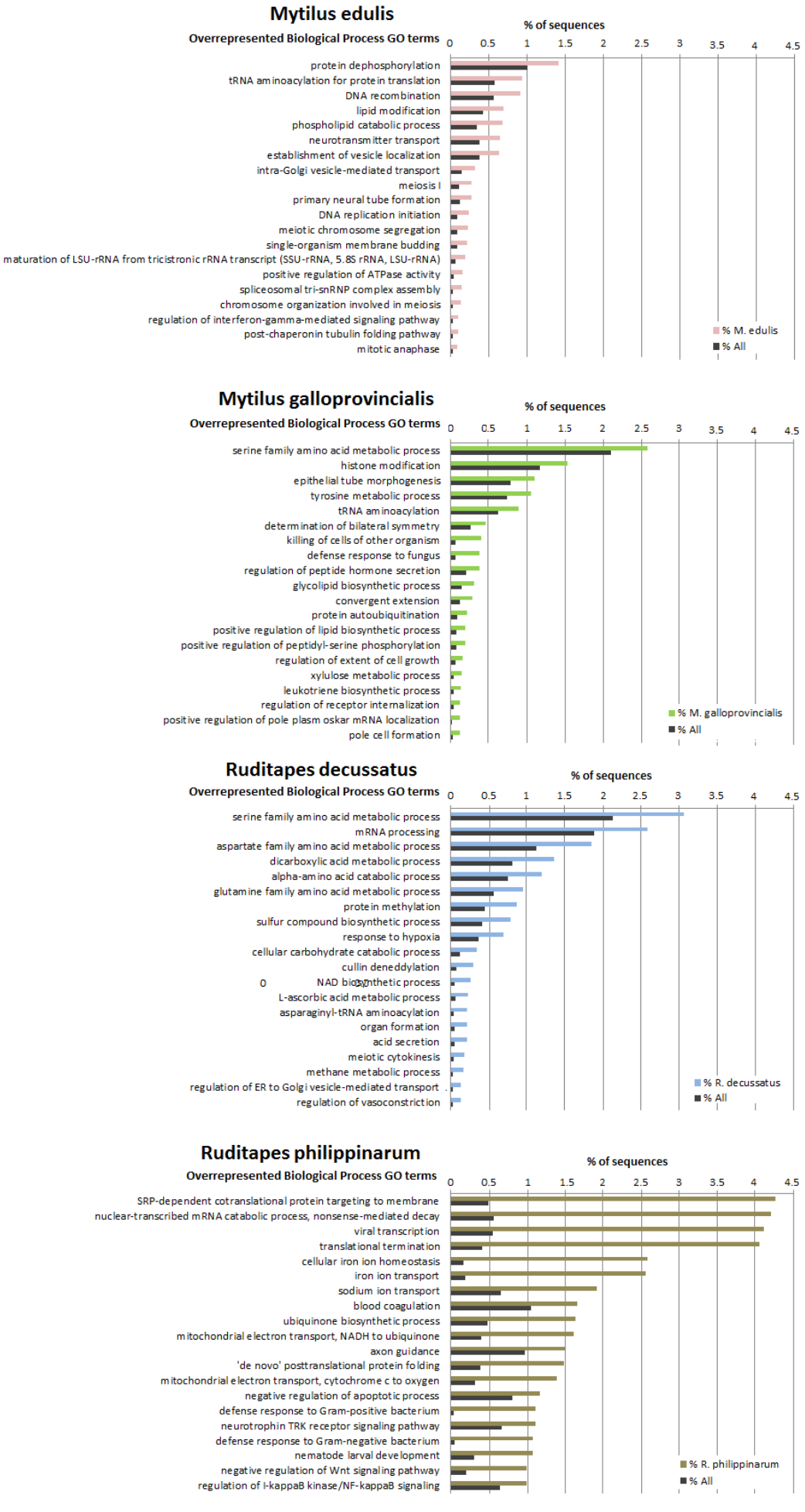

M. galloprovincialis
13,245
R. decussatus
6,562
M. edulis
12,867
R. philippinarum
7,206
6,006
3,469
5,694
4,098
353
485
4,461
914
135
785
321
285
118
669
268
Top hit

Figure 6


Mytilus edulis
Overrepresented Biological Process GO terms
% of sequences
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5
protein dephosphorylation
tRNA aminoacylation for protein translation
DNA recombination
lipid modification
phospholipid catabolic process
neurotransmitter transport
establishment of vesicle localization
intra-Golgi vesicle-mediated transport
meiosis I
primary neural tube formation
DNA replication initiation
meiotic chromosome segregation
single-organism membrane budding
maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)
positive regulation of ATPase activity
spliceosomal tri-snRNP complex assembly
chromosome organization involved in meiosis
regulation of interferon-gamma-mediated signaling pathway
post-chaperonin tubulin folding pathway
mitotic anaphase
% M. edulis
% All
Mytilus galloprovincialis
Overrepresented Biological Process GO terms
% of sequences
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5
serine family amino acid metabolic process
histone modification
epithelial tube morphogenesis
tyrosine metabolic process
tRNA aminoacylation
determination of bilateral symmetry
killing of cells of other organism
defense response to fungus
regulation of peptide hormone secretion
glycolipid biosynthetic process
convergent extension
protein autoubiquitination
positive regulation of lipid biosynthetic process
positive regulation of peptidyl-serine phosphorylation
regulation of extent of cell growth
xylulose metabolic process
leukotriene biosynthetic process
regulation of receptor internalization
positive regulation of pole plasm oskar mRNA localization
pole cell formation
% M. galloprovincialis
% All
Ruditapes decussatus
Overrepresented Biological Process GO terms
% of sequences
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5
serine family amino acid metabolic process
mRNA processing
aspartate family amino acid metabolic process
dicarboxylic acid metabolic process
alpha-amino acid catabolic process
glutamine family amino acid metabolic process
protein methylation
sulfur compound biosynthetic process
response to hypoxia
cellular carbohydrate catabolic process
cullin deneddylation
NAD biosynthetic process
L-ascorbic acid metabolic process
asparaginyl-tRNA aminoacylation
organ formation
acid secretion
meiotic cytokinesis
methane metabolic process
regulation of ER to Golgi vesicle-mediated transport
regulation of vasoconstriction
% R. decussatus
% All
Ruditapes philippinarum
Overrepresented Biological Process GO terms
% of sequences
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5
SRP-dependent cotranslational protein targeting to membrane
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
viral transcription
translational termination
cellular iron ion homeostasis
iron ion transport
sodium ion transport
blood coagulation
ubiquinone biosynthetic process
mitochondrial electron transport, NADH to ubiquinone
axon guidance
'de novo' posttranslational protein folding
mitochondrial electron transport, cytochrome c to oxygen
negative regulation of apoptotic process
defense response to Gram-positive bacterium
neurotrophin TRK receptor signaling pathway
defense response to Gram-negative bacterium
nematode larval development
negative regulation of Wnt signaling pathway
regulation of I-kappaB kinase/NF-kappaB signaling
% R. philippinarum
% All