Shotgun analysis of the marine mussel Mytilus edulis hemolymph proteome and mapping the innate immunity elements.
Campos A, Apraiz I, da Fonseca RR, **Cristobal S**.
Proteomics 2015, *15*, 4021–4029 DOI 10.1002/pmic.201500118

Shotgun analysis of the marine mussel *Mytilus edulis* haemolymph proteome and mapping the innate immunity elements

Alexandre Campos[1], Itxaso Apraiz[2], Rute R. da Fonseca[3] and Susana Cristobal[1,4]

[1]Department of Clinical and Experimental Medicine, Cell Biology, Faculty of Medicine and Health Sciences, Linköping University, Linköping, Sweden.

[2]Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

[3]The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark

[4]IKERBASQUE, Basque Foundation for Science. Department of Physiology, Faculty of Medicine and Dentistry, University of the Basque Country, Leioa, Spain.

Correspondence: Prof. Susana Cristobal. Department of Clinical and Experimental Medicine, Cell Biology, Faculty of Medicine and Health Sciences, Linköping University, SE-581 83 Linköping, Sweden.

E-mail: Susana.Cristobal@liu.se; Tel: +46101030881; Fax: +46 13 14 94 03

Keywords: Biomarker, haemolymph, innate immunity, *Mytilus edulis*, shotgun proteomics.

Total number of words: 5162

Abstract

The marine mussel innate immunity provides protection to pathogen invasion and inflammation. In this regard the mussel haemolymph takes a main role in the animal innate response. Despite the importance of this body fluid in determining the physiological condition of the animal, little is known about the molecular mechanisms underlying the cellular and humoral responses. In this work we have applied a mass spectrometry (nanoLC-MS/MS) strategy integrating genomic and transcriptomic data with the aim to: i) identify the main protein functional groups that characterize haemolymph and ii) to map the elements of innate immunity in the marine mussel *Mytilus edulis* haemolymph proteome. After sample analysis and first protein identification based on MS/MS data comparison, proteins with unknown functions were annotated with blast using public database (nrNCBI) information. Overall 595 haemolymph proteins were identified with high confidence and annotated. These proteins encompass primary cellular metabolic processes: energy production and metabolism of biomolecules, as well as processes related to oxidative stress defence, xenobiotic detoxification, drug metabolism, and immune response. A group of proteins was identified with putative immune effector, receptor and signalling functions in *M. edulis*. Data are available via ProteomeXchange with identifier PXD001951.

Significance of the Study:

Proteomic investigations have been hampered in *Mytilus* sp. by the scarce genomic data available. Here a high-throughput methodology was successfully employed to describe the overlooked haemolymph proteome from the marine mussel *M. edulis*, revealing a conjunct of molecular functions contributing to the immune defence. The information available contributes to enhance the

understanding of the molecular physiology and immunity in this marine invertebrate and has applicability in the assessment of animal health.

1 Introduction

Mussels are a representative and ecologically relevant group of sedentary invertebrates from coastal and estuarine habitats. *Mytilus sp.* is a cosmopolitan genus comprehending several species including *M. edulis* distributed along North Europe Coast and *M. galloprovincialis* native to the Mediterranean, the Black Sea and Western Europe. Mussels are economically exploited as a food source, representing 22% of global aquaculture production [1]. Furthermore they play an important role as sentinels of aquatic environment [2-4].

Invertebrates lack an adaptive immune system and do not have immune memory; nevertheless they rely on innate immunity to secure protection to pathogenic microorganisms [5]. Important innate immune responses are undertaken in the haemolymph, the circulatory fluid that has an analogous function to that of the blood of vertebrates. Mucosal tissues at the interface with the environment also assume a strategic importance in the defence against potential pathogens. Several genes with defence functions have been reported for instance in gills [6].

In haemolymph, the haemocytes are responsible for the main cellular defence functions consisting of pathogen phagocytosis, encapsulation, infiltration, and production of reactive oxygen and nitrogen species (ROS and RNS, respectively) [7]. Haemocytes are therefore rich in several hydrolytic enzymes (lysozymes, esterases, proteases, glucosidases), acid phosphatase, peroxidase, and phenoloxidase activities, which have already been reported for instance in oyster and mussel haemocytes [8-10]. Haemocytes also express several proteins involved in pathogen recognition and agglutination such as lectins [11] and bacteria-binding ficolin-like proteins (FLPs) as found in crayfish [12]. On the other hand, in the haemolymph serum a humoral response is carried out conducted by a conjunct of anti-microbial proteins and peptides of the family of myticins, mytilins and defensins that have been described in mussels and other bivalves [13-16].

Other haemolymph functions regard the circulation of oxygen, nutrients, hormone-like and neuropeptides [17, 18] and xenobiotic detoxification [19]. Despite these accomplishments, our understanding of the mechanisms underlying the functions of haemolymph and the immunomodulation processes in marine molluscs is still scarce and restricted to the study of individual proteins and applying principles of comparative physiology. Proteomics is helping to cover this gap enabling the identification of: i) key elements of the immune response of *Octopus vulgaris* to the parasite *Aggregata octopiana* [5], ii) markers of immune response to metal contamination [20], iii) novel elements of the innate immune response of the ascidian *Halocynthia roretzi* showing soft tunic syndrome [21], iv) 160 gene products from haemolymph of the ancient marine arthropod *Limulus polyphemus* [22], v) novel signalling mechanisms and hormone mediated regulatory processes in crustaceans and molluscs [17, 18]. Furthermore, the use of combined proteomic and transcriptomic approaches have uncovered patterns of expression of constitutive haemoglobin isoforms in model crustacean *Daphnia pulex* that are related to acute temperature-induced tissue hypoxia and to a mechanism of acclimation to moderately elevated temperatures [23].

This work aimed to gather information of the functions of marine mussel *M. edulis* haemolymph and to map the elements of innate immunity. For this purpose a shotgun proteomics approach was conducted. The majority of proteins in the haemolymph proteome are involved in the primary metabolism of haemocytes (with ion, small molecule, carbohydrate, protein, heterocyclic and organic cyclic compound biding functions). Moreover, we identified a set of proteins which link the haemolymph to xenobiotic metabolism, oxidative stress defence and cellular response to stress, humoral and innate immune functions. Within this set of proteins we find hydrolases, oxidoreductases, transferases, carbohydrate binding, and signal transducer proteins.

2 Materials and Methods

2.1 Collection of haemolymph

Mussels *M. edulis trossulus* 2.0-3.5 cm in length were collected from rocks and sandy substrates at 1-4 m depth in Trosa archipelago (Baltic Sea) south of the island of Askö (58°49'3''N 17°37'4''E) on March 2010. Mussels were immediately brought to the laboratory and kept in a glass aquarium with 10 L of artificial sea water (6.7-7.0 ‰ salinity, Instant Ocean Salt, Aquarium Systems, Sarrebourg, France), at a density of 3.2 mussels per liter at 6-8 °C. Mussels were fed with 5 g/L grounded food for omnivore fish (TetraMin, Tetra, Blacksburg, VA, USA) and water was exchanged and new food provided every third day.

Mussel haemolymph was collected on the third day after feeding, from the posterior adductor muscle by gentle aspiration with the aid of a 23 G x 1" syringe. A group of individual samples was used to obtain serum by allowing the haemolymph coagulate at room temperature during 1 h, after which vials were pulse-centrifuged using a board mini centrifuge for 15 min, and serum supernatants were collected into a separate vial and frozen in liquid nitrogen [24]. All the samples were subsequently stored at -80 °C until the current experiment. No permits were required for the described study, which complied with all relevant regulations. In total 3 haemocyte and serum replicate samples were considered for this study, each consisting of a pool of two individual samples.

2.2 Sample preparation: protein extraction and FASP digestion

Haemocytes were lysed directly in 2 % (v/v) SDS, 100 mM Tris, 0.1 M DTT, pH 7.6 (SDT buffer) and sonication (2 x 4 sec, 23 kHz, 105 µm). The homogenate was heated for 3 min, 95 °C and centrifuged at 16000 x *g* for 20 min. Protein samples were stored at -20 °C. Serum samples were first submitted to ultracentrifugation, 150000 x *g* for 3 h at 4 °C, to deplete haemocyanin. Serum proteins were thereafter concentrated by centrifugal filtration with membrane nominal molecular weight limit (NMWL) of 10 kDa (MRCPRT010, Millipore, Billerica, MA, USA) and denatured in SDT buffer before

trypsin digestion. Haemocyte and serum proteins were digested with trypsin following filter aided sample preparation (FASP) method essentially as described by Wiśniewski *et al.* (2009) [25]. This method consists in performing in-solution protein reduction, alkylation and digestion, using a standard filter device to enable buffer exchange. After digestion peptides are recovered by centrifugation in the filter thereby achieving a degree of purity suitable for nano-LC MS/MS. In this regard serum and haemocyte protein samples (30 µg) were alkylated and digested in solution with trypsin (recombinant, proteomics grade, Roche, Basel, Switzerland) with enzyme to protein ratio of 1:100 (w/w) for 16 h at 37 °C in centrifugal filter units with respectively 10 and 30 kDa NMWL membranes (MRCPRT010 and MRCF0R030, Millipore). Peptides were recovered, acidified with FA (10 %, v/v), desalted and concentrated by reversed phase extraction (C18 Top-Tips, TT2C18, Glygen, Columbia, MD, USA) using ACN (70 %, v/v) with FA (0.1 %, v/v) for elution. Before LC-MS/MS, peptides were recovered in 0.1 % FA and final concentration adjusted to 0.04-0.06 µg/µl.

2.3 LC-MS/MS analysis

FASP protein digests (triplicate samples) were analysed by nano-LC coupled to a hybrid Ion trap mass spectrometer (LTQ Orbitrap Velos Pro –ETD, Thermo Scientific, Waltham, MA, USA). Peptides were separated by reverse phase chromatography on a 20 mm×100 µm C18 pre column followed by a 100 mm×75 µm C18 column (particle size 5 µm, NanoSeparations, Nieuwkoop, Netherlands) using a linear gradient from 2 % buffer B (ACN + FA, 0.1 %, v/v) in buffer A (TFA, 0.1 %, v/v in water) to 30 % B in 40 min and to 95 % B (v/v) in 30 min, at a flow rate of 0.3 µl/min (total elution time 70 min). Peptides were analysed by on-line nano-electrospray ionization (easy nano-ESI) in positive mode, using Xcalibur software (version 2.6, Thermo Scientific). Full scans were performed at 30000 resolution with scan ranges of 380-2000 m/z and the top 20 most intense ions were isolated and fragmented. Collision induced fragmentation (CID) was used to fragment the precursor ions by applying normalized collision

energy of 30 % value, an isolation width of 2.0, an activation time of 10 milliseconds and a Q-value of 0.25. In total 12 nano-LC-MS/MS runs were performed, which encompassed the duplicate analysis of 3 haemocyte and 3 serum samples.

2.4 Peptide identification and quantification

LTQ raw data was searched against a protein database using SEQUEST algorithm in Proteome Discoverer software (version 1.4, Thermo Scientific). A second search was performed using X!Tandem algorithm in the Scaffold program (version Scaffold 4.3.4, Proteome Software, Portland, Oregon, USA). MS and MS/MS tolerances were set to 10 ppm and 0.6 Da. Trypsin was selected for protein cleavage allowing for 1 missed cleavage. Carbamidomethylation and oxidation were selected as static and dynamic modifications respectively. Identifications were validated by performing a decoy database search for the estimation of false Discovery rates (FDR) and peptide identifications were accepted if they could be established at greater than 95.0 % probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 99.9 % probability and contained at least 2 identified peptides. Technical duplicates were merged to produce a single output file.

The protein database utilized in this study consisted in the combination of the UniProt KnowledgeBase (KB) sequences from all organisms from the taxa Mollusca (90847 sequences released 24-10-2014), and the sequences derived from the transcriptome of gills of *M. galloprovinciallis* (46791 sequences, unpublished data) and *Bathymodiolus azoricus* (33464 sequences) [26]. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [27] via the PRIDE partner repository with the dataset identifier PXD001951 and 10.6019/PXD001951.

2.5 Data processing and quantification

Scaffold output files were used for the systematic characterization of haemolymph proteome (haemocyte and serum proteins). All proteins identified in at least two of the three replicate files were considered in the catalogue. Protein abundance was calculated using normalized spectral abundance factors (NSAFs) in Scaffold as described previously [28].

2.6 Protein homology search and gene ontology analysis

Protein sequences with unknown function were annotated with a blast search in the National Centre for Biotechnology Information database (NCBI, http://www.ncbi.nlm.nih.gov/) using blastp algorithm in Blast2Go tool (version 3.0, BioBam Bioinformatics, Valencia, Spain) employing a threshold e-value of 1E-10. Gene ontology terms were used to group proteins in the categories biological process, cellular component and molecular function.

3 Results and discussion

3.1 Haemolymph proteome identification

A shotgun analysis was employed to investigate the proteome of *M. edulis* (Fig. 1a). In total 27754 high confident MS/MS spectra (0.01 % Decoy FDR) were obtained from the analysis of 3 biological samples enabling to identify 1121 proteins. In this set 654 proteins showed consistency in the 3 biological replicate samples and therefore were further considered to characterize haemolymph proteome. One hundred and seventy eight proteins were identified both in haemocytes and haemolymph serum (Fig. 1b), denoting a putative overlap of functions in the haemocytes and serum. One hundred and fifty five proteins were identified exclusively in the serum (24 % of total number of identifications) and 321 in haemocytes (49 % of total number of identifications) (Fig. 1b). We report

an increase of haemolymph proteome coverage in *M. edulis* in comparison to other marine species, the arthropod *L. polyphemus* and the ascidian *H. roretzi* [21, 22]. This result could be attributed in part to the FASP method employed here which, in comparison to the conventional in-gel trypsin digestion methods used in the other studies, is likely to be more efficient for protein digestion and for recovery of peptides, and thus may lead to an increased protein sequence coverage and improved protein identification [29]. Computational analysis of MS/MS spectra enabled *M. edulis* haemolymph proteins to be identified, the majority of them (89 % of the total number of identifications, Fig. 1c) showing homology to *M. galloprovincialis* sequences. A comparatively reduced number of *M. edulis* proteins showed high homology with proteins from the UniProt KB database (taxonomic group Mollusca) and *B. azoricus* proteins (sequences predicted from transcriptome). The utilisation of transcriptomic information from the congener species *M. galloprovincialis*, which is likely to share many genetic characteristics and gene similarities with *M. edulis*, enabled to increase the number of proteins identified. In fact, relying only on the UniProt KB database, will reduce the number of identified proteins by a factor of three.

3.2 Protein blast and gene ontology annotation

Functions of unknown proteins and respective gene ontology (GO) terms were assigned through searching for homologous sequences in a public database (nrNCBI) using Blast2go tool. Search results are resumed in Figure 2. In total 595 sequences matched highly homologous protein sequences from NCBI (Fig. 2a). Sequences from oyster *Crassostrea gigas* were highly represented in the blast search output, followed by sequences from two marine gastropod mollusc species *Lottia gigantea* and *Aplysia californica* (Fig. 2b). Surprisingly sequences from genus *Mytilus sp.* were less represented in the blast search which could be explained by the lack of genomic information available for this genus in the public database (Fig. 2b).

3.3 Representative molecular functions and metabolic pathways

Gene ontology terms were analysed to indicate the principal groups of proteins present in haemolymph in the categories of molecular function, biological process, and cell component (Fig. 3, Fig. S1). We observed a predominant presence in haemocytes and serum of proteins with functions associated with general cell metabolism. In this group we include the ion binding, protein binding, organic cyclic and heterocyclic compound binding proteins (Fig. 3). We also associate with cell general metabolism the groups of small molecule binding, cofactor and carbohydrate derivative binding, ligase activity. Lipid binding proteins, cytoskeleton and ribosome structural proteins were detected in haemocytes whereas enzyme inhibitor activity and signal transducer activity in serum (Fig. 3). Subsequent GO (Fig. S1a) and KEGG pathway analyses (Table 1) showed additional metabolic and biological processes associated with the identified proteins.

Other represented functional categories are constituted by hydrolases (15% and 9% of serum and haemocyte protein sequences respectively), oxidoreductases (6% and 5% of serum and haemocyte sequences), transferases (5% and 4% of serum and haemocyte sequences), transmembrane and substrate-specific transporters (3% and 4% of serum and haemocyte sequences) (Fig. 3). These proteins may assist several cellular functions and for instance link haemolymph to metabolism of xenobiotics, oxidative stress defence, and cellular response to stress in addition to main immunological functions. Specific examples of metabolic associations were retrieved with KEGG analysis (Table 1). With regard to the metabolism of xenobiotics were identified several putative enzymes, glutathione S-transferase (GSTs), trans-1,2-dihydrobenzene-1,2-diol dehydrogenase, aldehyde dehydrogenase, hypoxanthine-guanine phosphoribosyltransferase, beta-glucuronidase, inosine-5'-monophosphate dehydrogenase 2, and S-formylglutathione hydrolase (Table 1, Table S4a).

From this group only GSTs and beta-glucuronidases have been functionally characterized at the protein level.

Isomerases, lyases, ligases, lipid and carbohydrate binding proteins, and signal transducers are less represented in haemolymph. While lyases and ligases act in oxidative stress defence, signal transduction proteins could also play a role in chemotaxis, response to stimulus, and hormone activity. We also highlight the putative function of carbohydrate binding (1% of serum and haemocyte sequences) and signal transducer protein members in humoral and innate immune response and inflammation processes. The majority of the identified proteins are associated to organelles/vesicles, macromolecular complexes and membranes and only a few of them were identified as proteins of the extracellular matrix (Fig. S1b).

3.4 Elements of innate immunity

Haemolymph plays a central role in the immune response and defence against pathogen invasion and inflammation in invertebrates. In order to understand which elements complement the innate immune system in *M. edulis* haemolymph we have selected, from the catalogue, a set of proteins on the basis of GO terms and keywords that describe general functional features of proteins involved in the humoral and cellular immune responses, i.e. the extracellular proteins (GO:0005615, GO:0005576, GO:0031012) and proteins with carbohydrate binding (GO:0030246), innate immune (GO:0045087, GO:0045088, GO:0002376) and host receptor functions (Table 2).

The strategy enabled us to detect the presence of a conjunct of immune effector proteins that participate in pathogen recognition, agglutination and killing. In this group we find several hydrolytic enzymes and protease inhibitors, with a preferential accumulation in serum, and ROS production enzymes (dual oxidase and superoxide dismutase), treifol like protein involved in the production of

anti-microbial peptides such as defensins [30] and a putative chitin deacetylase with a presence in haemocytes and serum (Table 2). Chitin deacetylase participates in the metabolism of the antibacterial biopolymer chitosan in a variety of marine organisms [31]. Evidences from gene transcription in the triangle sail mussel (*Hyriopsis cumingii*) indicates that chitin deacetylase is expressed in haemolymph but also other organs such as mantle, liver, stomach, kidney, intestine, gill, and foot [32] with a possible function in the immune defence. On the other hand only genomic evidences have so far supported the treifol protein existence in bivalves.

Moreover s-type (galectin 2) and c-type (c-type mannose receptor 2, c-type lectin domain family 4 member g, vitelline coat lysin m7) lectins and C1q-domain containing proteins were identified. Of these immune effector groups c-type lectins appear to be the most represented and is preferentially distributed in the haemocytes, whereas C1q-domain containing proteins may be found in haemocytes and serum and galectin 2 (s-type) was detected only in the serum (Table 2). The immune effector proteins are responsible for pathogen recognition, binding, and trigger subsequent defence reactions. Lectins and C1q-domain proteins are viewed as key elements of the innate immunity, being common to invertebrates, vertebrates and mammals [33, 34]. Extracellular as well as intracellular functions have been attributed to this super-family of proteins that are expressed in a variety of organs [35]. Bivalves display a remarkable diversification of C1q gene family with several hundreds of genes being reported in *M. galloprovincialis* and *C. gigas* [33, 34], which may represent an evolutionary adaptation to provide protection in environments highly rich in microorganisms.

Regarding cell membrane receptor functions we have detected the presence of beta-1,3-glucan binding (bGRP) and integrin homologous proteins. In this regard the exclusive detection of these proteins in the haemocytes and not in the serum (Table 2) is in agreement with the attributed functions. bGRP binds specifically to beta-1,3-glucan polysaccharides of fungal cell walls and is known to activate the haemolymph specific phenoloxidase cascade [36]. A gene transcription study provided

evidences for the presence of this protein in bivalves [37]. Integrins, are evolutionarily conserved and their functions in immune response are linked with leukocyte recruitment to the site of infection and complement receptor-dependent phagocytosis [38]. In bivalves however there is a lack of evidences of the expression of these proteins.

A conserved allograft inflammatory factor 1-like signal protein (AIF-1) was detected in *M. edulis* serum; this protein plays a regulatory role in inflammation and immune response by augmenting cytokine production in macrophage cells [39]. Genome evidences indicate that this protein may also be present in lower metazoans [38].

Concluding remarks

A comprehensive proteomic analysis is a fundamental key in the research towards the understanding of the physiological processes in any organism. In many marine organisms, including mussels, this goal has been hampered by the lack of genomic information. In an attempt to circumvent this limitation we utilized complementary transcriptomic resources to characterize mussel haemolymph proteome. The present catalogue, with 595 annotated gene products, illustrates the various molecular functions performed by the circulatory system where the innate immune response is highlighted. This strategy also proved to be useful in the sense that enabled to confirm the presence of several gene products that have been ascribed previously from genomic and transcriptomic data, and enabled to discover novel putative protein functions. This set of gene products can open new explanatory paths for regulation of immunity in bivalves and allow for a reflection on the development of the immune system from an evolutionary perspective. These findings could facilitate the implementation of proteomic-based methods in sectors such as aquaculture, where the analysis of immunological markers could support the assessment of mussel health status and disease surveillance.

Acknowledgements

This work was supported by grants from the Swedish Research Council-Natural Science, VR-NT, Carl Trygger Foundation, VINNOVA-Vinnmer program, Oscar and Lilli Lamms Minne Foundation, Ångpanneförening Research foundation, IKERBASQUE, Basque Foundation for science, ALF and Linköping University. The authors would like to thank Raul Bettencourt, Dept. of Oceanography and Fisheries, University of the Azores, for kindly providing the RNAseq data for this study, and to Drs. Hanna Oscarsson and Linda Kumblad from Stockholm University for the collaborative effort in the mussel sampling and exposure experiments.

Conflict of interest

The authors have declared no conflict of interest.

Shotgun analysis of the marine mussel Mytilus edulis hemolymph proteome and mapping the innate immunity elements.
Campos A, Apraiz I, da Fonseca RR, **Cristobal S**.
Proteomics 2015, *15*, 4021–4029 DOI 10.1002/pmic.201500118

References

[1] FAO, The State of World Fisheries and Aquaculture Opportunities and challenges. *FAO* 2014.
[2] Bayne, B., Watch on mussels. *Mar. Poll. Bull.* 1976, *7*, 217-218.
[3] Sparks, C., Odendaal, J., Snyman, R., An analysis of historical Mussel Watch Programme data from the west coast of the Cape Peninsula, Cape Town. *Marine pollution bulletin* 2014, *87*, 374-380.
[4] Melwani, A. R., Gregorio, D., Jin, Y., Stephenson, M.*, et al.*, Mussel Watch update: long-term trends in selected contaminants from coastal California, 1977-2010. *Marine pollution bulletin* 2014, *81*, 291-302.
[5] Castellanos-Martinez, S., Diz, A. P., Alvarez-Chaver, P., Gestal, C., Proteomic characterization of the hemolymph of Octopus vulgaris infected by the protozoan parasite Aggregata octopiana. *Journal of proteomics* 2014, *105*, 151-163.
[6] Martins, E., Figueras, A., Novoa, B., Santos, R. S.*, et al.*, Comparative study of immune responses in the deep-sea hydrothermal vent mussel Bathymodiolus azoricus and the shallow-water mussel Mytilus galloprovincialis challenged with Vibrio bacteria. *Fish & shellfish immunology* 2014, *40*, 485-499.
[7] Rodríguez-Domínguez, H., Soto-Búa, M., Iglesias-Blanco, R., Crespo-González, C., Arias-Fernández, C., et al., , Preliminary study on the phagocytic ability of *Octopus vulgaris Cuvier*, 1797 (Mollusca: Cephalopoda) haemocytes in vitro. *Aquaculture* 2006, 563–570.
[8] Carballal, M. J., Lopez, C., Azevedo, C., Villalba, A., Enzymes Involved in Defense Functions of Hemocytes of Mussel Mytilus galloprovincialis. *Journal of invertebrate pathology* 1997, *70*, 96-105.
[9] Aladaileh, S., Nair, S. V., Birch, D., Raftos, D. A., Sydney rock oyster (Saccostrea glomerata) hemocytes: morphology and function. *Journal of invertebrate pathology* 2007, *96*, 48-63.
[10] Luna-Acosta, A., Saulnier, D., Pommier, M., Haffner, P.*, et al.*, First evidence of a potential antibacterial activity involving a laccase-type enzyme of the phenoloxidase system in Pacific oyster Crassostrea gigas haemocytes. *Fish & shellfish immunology* 2011, *31*, 795-800.
[11] Tunkijjanukij, S., Giaever, H., Chin, C. C., Olafsen, J. A., Sialic acid in hemolymph and affinity purified lectins from two marine bivalves. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology* 1998, *119*, 705-713.
[12] Wu, C., Soderhall, K., Soderhall, I., Two novel ficolin-like proteins act as pattern recognition receptors for invading pathogens in the freshwater crayfish Pacifastacus leniusculus. *Proteomics* 2011, *11*, 2249-2264.
[13] Charlet, M., Chernysh, S., Philippe, H., Hetru, C.*, et al.*, Innate immunity. Isolation of several cysteine-rich antimicrobial peptides from the blood of a mollusc, Mytilus edulis. *The Journal of biological chemistry* 1996, *271*, 21808-21813.
[14] Mitta, G., Hubert, F., Noel, T., Roch, P., Myticin, a novel cysteine-rich antimicrobial peptide isolated from haemocytes and plasma of the mussel Mytilus galloprovincialis. *European journal of biochemistry / FEBS* 1999, *265*, 71-78.
[15] Mitta, G., Vandenbulcke, F., Hubert, F., Roch, P., Mussel defensins are synthesised and processed in granulocytes then released into the plasma after bacterial challenge. *Journal of cell science* 1999, *112 (Pt 23)*, 4233-4242.
[16] Gonzalez, M., Gueguen, Y., Desserre, G., de Lorgeril, J.*, et al.*, Molecular characterization of two isoforms of defensin from hemocytes of the oyster Crassostrea gigas. *Developmental and comparative immunology* 2007, *31*, 332-339.
[17] Hui, L., Cunningham, R., Zhang, Z., Cao, W.*, et al.*, Discovery and characterization of the Crustacean hyperglycemic hormone precursor related peptides (CPRP) and orcokinin neuropeptides in the sinus glands of the blue crab Callinectes sapidus using multiple tandem mass spectrometry techniques. *Journal of proteome research* 2011, *10*, 4219-4229.

[18] Liang, Z., Schmerberg, C. M., Li, L., Mass spectrometric measurement of neuropeptide secretion in the crab, Cancer borealis, by in vivo microdialysis. *The Analyst* 2014.
[19] Mello, D. F., de Oliveira, E. S., Vieira, R. C., Simoes, E.*, et al.*, Cellular and transcriptional responses of Crassostrea gigas hemocytes exposed in vitro to brevetoxin (PbTx-2). *Marine drugs* 2012, *10*, 583-597.
[20] Thompson, E. L., Taylor, D. A., Nair, S. V., Birch, G.*, et al.*, Proteomic discovery of biomarkers of metal contamination in Sydney Rock oysters (Saccostrea glomerata). *Aquatic toxicology* 2012, *109*, 202-212.
[21] Cha, I. S., Castillo, C. S., Nho, S. W., Hikima, J.*, et al.*, Innate immune response in the hemolymph of an ascidian, Halocynthia roretzi, showing soft tunic syndrome, using label-free quantitative proteomics. *Developmental and comparative immunology* 2011, *35*, 809-816.
[22] D'Amato, A., Cereda, A., Bachi, A., Pierce, J. C., Righetti, P. G., In depth exploration of the hemolymph of Limulus polyphemus via combinatorial peptide ligand libraries. *Journal of proteome research* 2010, *9*, 3260-3269.
[23] Zeis, B., Becker, D., Gerke, P., Koch, M., Paul, R. J., Hypoxia-inducible haemoglobins of Daphnia pulex and their role in the response to acute and chronic temperature increase. *Biochimica et biophysica acta* 2013, *1834*, 1704-1710.
[24] Luengen, A. C., Friedman, C. S., Raimondi, P. T., Flegal, A. R., Evaluation of mussel immune responses as indicators of contamination in San Francisco Bay. *Marine environmental research* 2004, *57*, 197-212.
[25] Wisniewski, J. R., Zougman, A., Nagaraj, N., Mann, M., Universal sample preparation method for proteome analysis. *Nature methods* 2009, *6*, 359-362.
[26] Bettencourt, R., Pinheiro, M., Egas, C., Gomes, P.*, et al.*, High-throughput sequencing and analysis of the gill tissue transcriptome from the deep-sea hydrothermal vent mussel Bathymodiolus azoricus. *BMC genomics* 2010, *11*, 559.
[27] Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A.*, et al.*, ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature biotechnology* 2014, *32*, 223-226.
[28] Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K.*, et al.*, Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *Journal of proteome research* 2006, *5*, 2339-2347.
[29] Sun, X., Jiang, X., Combination of FASP and fully automated 2D-LC-MS/MS allows in-depth proteomic characterization of mouse zymogen granules. *Biomedical chromatography : BMC* 2013, *27*, 407-408.
[30] Barrera, G. J., Sanchez, G., Gonzalez, J. E., Trefoil factor 3 isolated from human breast milk downregulates cytokines (IL8 and IL6) and promotes human beta defensin (hBD2 and hBD4) expression in intestinal epithelial cells HT-29. *Bosnian journal of basic medical sciences / Udruzenje basicnih mediciniskih znanosti = Association of Basic Medical Sciences* 2012, *12*, 256-264.
[31] Zhao, Y., Park, R. D., Muzzarelli, R. A., Chitin deacetylases: properties and applications. *Marine drugs* 2010, *8*, 24-46.
[32] Wang, G. L., Xu, B., Bai, Z. Y., Li, J. L., Two chitin metabolic enzyme genes from Hyriopsis cumingii: cloning, characterization, and potential functions. *Genetics and molecular research : GMR* 2012, *11*, 4539-4551.
[33] Gerdol, M., Manfrin, C., De Moro, G., Figueras, A.*, et al.*, The C1q domain containing proteins of the Mediterranean mussel Mytilus galloprovincialis: a widespread and diverse family of immune-related molecules. *Developmental and comparative immunology* 2011, *35*, 635-643.
[34] Gerdol, M., Venier, P., Pallavicini, A., The genome of the Pacific oyster Crassostrea gigas brings new insights on the massive expansion of the C1q gene family in Bivalvia. *Developmental and comparative immunology* 2015, *49*, 59-71.

[35] Zick, Y., Eisenstein, M., Goren, R. A., Hadari, Y. R.*, et al.*, Role of galectin-8 as a modulator of cell adhesion and cell growth. *Glycoconjugate journal* 2004, *19*, 517-526.
[36] Takahasi, K., Ochiai, M., Horiuchi, M., Kumeta, H.*, et al.*, Solution structure of the silkworm betaGRP/GNBP3 N-terminal domain reveals the mechanism for beta-1,3-glucan-specific recognition. *Proceedings of the National Academy of Sciences of the United States of America* 2009, *106*, 11679-11684.
[37] Garcia-Lagunas, N., Romero-Geraldo, R., Hernandez-Saavedra, N. Y., Genomics study of the exposure effect of Gymnodinium catenatum, a paralyzing toxin producer, on Crassostrea gigas' defense system and detoxification genes. *PloS one* 2013, *8*, e72323.
[38] Zhang, G., Fang, X., Guo, X., Li, L.*, et al.*, The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 2012, *490*, 49-54.
[39] Watano, K., Iwabuchi, K., Fujii, S., Ishimori, N.*, et al.*, Allograft inflammatory factor-1 augments production of interleukin-6, -10 and -12 by a mouse macrophage line. *Immunology* 2001, *104*, 307-316.

Figure Legends

Figure 1: Proteomics workflow employed to characterize haemolymph proteome in *M. edulis* (a). Venn diagram depicting the total number of proteins in haemocyte (H) and haemolymph serum (S) as well as the number of overlapped proteins (b); and the percentage of proteins matching sequences from the three complementary protein databases (c).

Figure 2: Statistical information regarding the blast search and gene ontology annotation using Blast2GO tool. Data distribution (a), and top-hit species distribution (b). The full list of annotated sequences is presented in supporting information (Table S3).

Figure 3: Top molecular functional groups of identified haemocyte (dark blue bars) and serum (light grey bars) protein sequences, obtained from the analysis of GO terms.

Table 1: Top KEGG pathways in haemolymph. The full KEGG pathway list is shown in supporting information (Table S4). Prokaryote metabolic pathways lacking unequivocal evidences (exclusive prokaryote sequences) were excluded from this list.

Table 2: Proteins with suggested functions in the innate immune response in *M. edulis.*

Supplementary information on-line available:

Table 1 SuppInfo: Protein identification report table.

Table 2 SuppInfo: Peptide identification report table.

Table 3 SuppInfo: complete list of annotated protein sequences retrieved by Blast2go.

Table 4 SuppInfo: complete list of KEGG pathways retrieved by Blast2go.

Figure S1: Classification based on GO terms of haemocyte (dark blue bars) and serum (light grey bars) protein sequences in the categories of biological processes (a) and cellular components (b).

Table 1

Pathway	Seqs in Pathway	Enzyme
Purine metabolism	55	12
Thiamine metabolism	47	2
Glycolysis / Gluconeogenesis	20	15
Pyruvate metabolism	15	12
Citrate cycle (TCA cycle)	15	11
Other glycan degradation	15	5
Phenylalanine metabolism	15	9
Arginine and proline metabolism	13	10
Valine, leucine and isoleucine degradation	10	7
Glutathione metabolism	9	5
Tryptophan metabolism	9	7
Pentose phosphate pathway	8	6
Lysine degradation	8	6
Alanine, aspartate and glutamate metabolism	8	7
Drug metabolism - other enzymes	7	4
beta-Alanine metabolism	7	5
Starch and sucrose metabolism	7	7
Fatty acid degradation	7	6
Phenylpropanoid biosynthesis	7	1
Metabolism of xenobiotics by cytochrome P450	6	3
Cysteine and methionine metabolism	6	7
Butanoate metabolism	6	5
Tyrosine metabolism	6	7
Glyoxylate and dicarboxylate metabolism	6	5
Pentose and glucuronate interconversions	6	7

Table 2

Seq. Name [1]	Accession [2]	Seq. Description	Seq. Length	min. eValue	mean Similarity	H [3]	S [3]
Hydrolases and inhibitors							
comp70715_c0_seq1	gi|405974410	zinc carboxypeptidase a 1-like	418	1.45E-151	60.65%	-	0.035
comp78560_c0_seq1	gi|47117147	ribonuclease oy	258	4.25E-55	55.7%	-	0.029
comp80382_c1_seq1	gi|405970893	plasma alpha-l-fucosidase	254	1.88E-109	72.95%	-	0.013
comp73791_c0_seq1	gi|322802735	tissue inhibitor of metalloproteases	185	4.54E-11	47.0%	-	0.029
comp81888_c0_seq4	gi|405969441	plasma alpha-l-fucosidase	454	0.0	72.55%	-	0.005
Q8ITU2_MYTED	gi|23263574	lysozyme	176	4.04E-127	73.45%	0.060	0.166
comp78211_c0_seq2	gi|405971836	Serpin B9	386	1.08E-79	56.95%	-	0.006
comp83852_c1_seq7	gi|40074252	alpha-2-macroglobulin-like protein 1	656	5.97E-104	54.6%	-	0.007
comp61482_c0_seq1	gi|114158675	leukocyte elastase inhibitor (Serpin)	180	2.06E-22	58.15%		0.032
comp81152_c0_seq1	gi|612568509	cathepsin c	455	0.0	71.45%	0.0047	-
Lectins							
F0V443_MYTGA	gi|46395578	c1q domain containing protein 1q6	231	9.47E-144	77.36%	0.016	0.054
comp65559_c0_seq1	gi|46395578	c1q domain containing protein 1q6	231	1.03E-134	67.57%	0.012	0.014
comp74298_c0_seq1	gi|112181167	complement component 1 q subcomponent-binding mitochondrial	285	4.72E-35	55.45%	0.006	-
comp76240_c0_seq1	gi|405954560	c-type lectin domain family 4 member g	208	9.63E-50	66.67%	-	0.036
comp76848_c0_seq1	gi|241913786	galectin 2	556	0.0	62.75%	0.0156	-

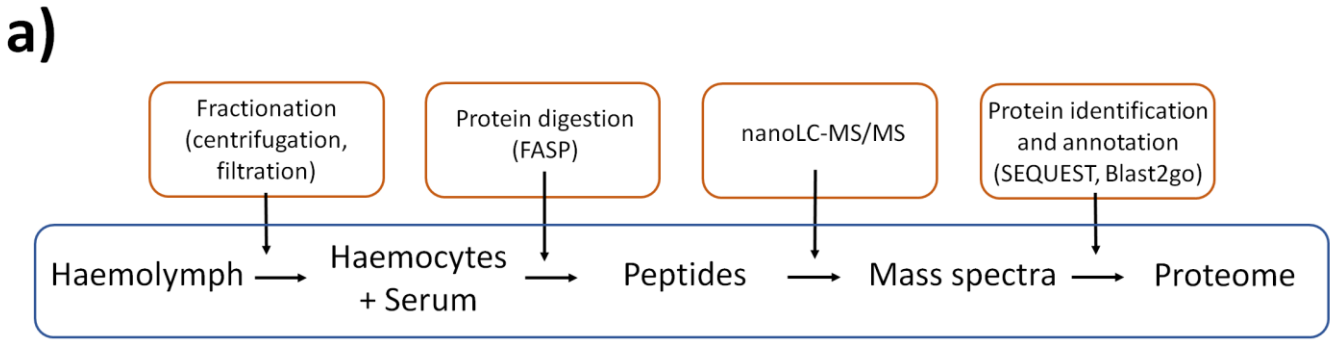
comp66285_c0_seq2	gi|405978398	c-type mannose receptor 2	186	8.57E-17	46.17%	0.007	-
A4UDS9_MYTGA	gi|222531725	vitelline coat lysin m7	180	1.16E-130	98.7%	0.015	-
A4UDX1_MYTTR	gi|126216491	vitelline coat lysin m7	180	5.57E-116	97.1%	0.020	-
receptors							
comp81033_c0_seq1	gi|405952567	beta-1,3-glucan-binding protein	1458	0.0	47.05%	0.026	-
comp75625_c3_seq10	gi|15004986	integrin beta-ps	801	0.0	64.85%	0.011	-
signal molecules							
comp78503_c1_seq1	gi|405969953	allograft inflammatory factor 1-like	153	8.27E-64	77.95%	-	0.036
Other immune effector related molecules							
comp33963_c0_seq1	gi|405973191	chitin deacetylase 3 precursor	273	2.40E-69	48.55%	0.022	0.039
comp74040_c0_seq1	gi|148223694	trefoil factor	209	3.27E-39	55.15%	0.044	0.014
comp75300_c0_seq1	gi|402122771	superoxide dismutase	349	8.86E-149	63.55%	0.022	0.050
comp83316_c0_seq4	gi|405978205	dual oxidase 2-like	1594	0.0	61.9%	0.007	-

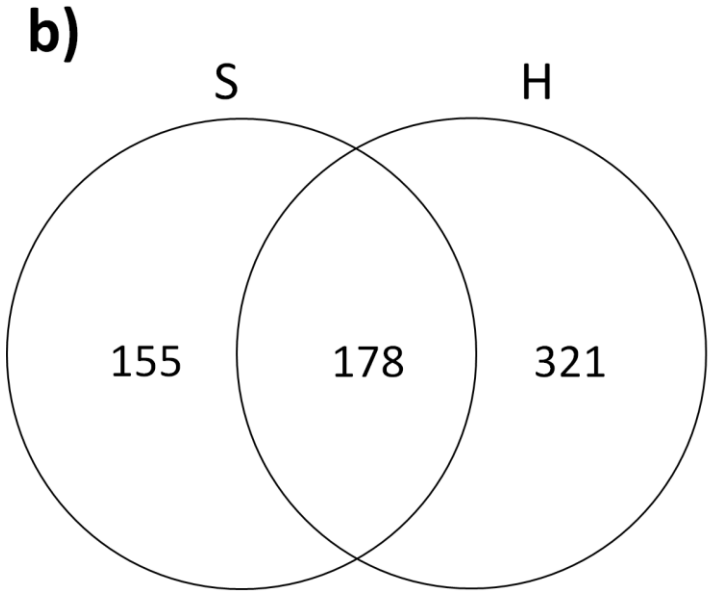
1) Reference of the sequences identified by mass spectrometry

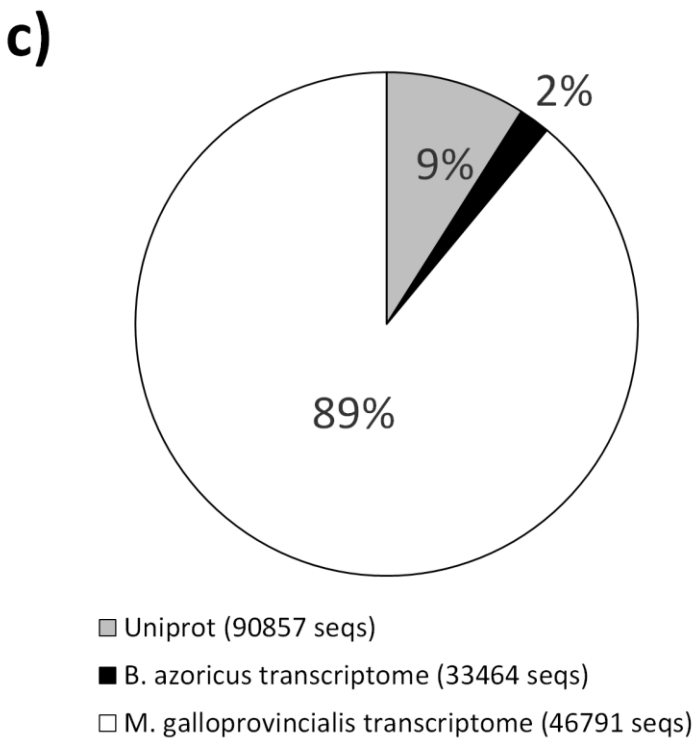
2) Accession numbers (NCBI) of the homologous sequences retrieved with Blast

3) Relative abundances (NSFA values) of proteins in haemocytes (H) and serum (S). Not detected (-).

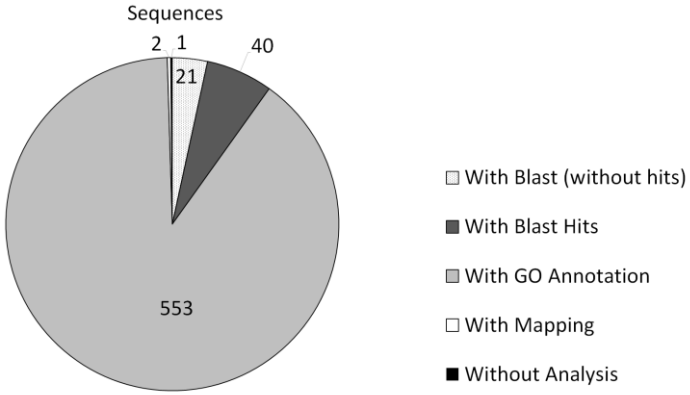
Shotgun analysis of the marine mussel Mytilus edulis hemolymph proteome and mapping the innate immunity elements.
Campos A, Apraiz I, da Fonseca RR, **Cristobal S**.
Proteomics 2015, *15*, 4021–4029 DOI 10.1002/pmic.201500118


a)
Fractionation (centrifugation, filtration)
Protein digestion (FASP)
nanoLC-MS/MS
Protein identification and annotation (SEQUEST, Blast2go)
Haemolymph
Haemocytes + Serum
Peptides
Mass spectra
Proteome

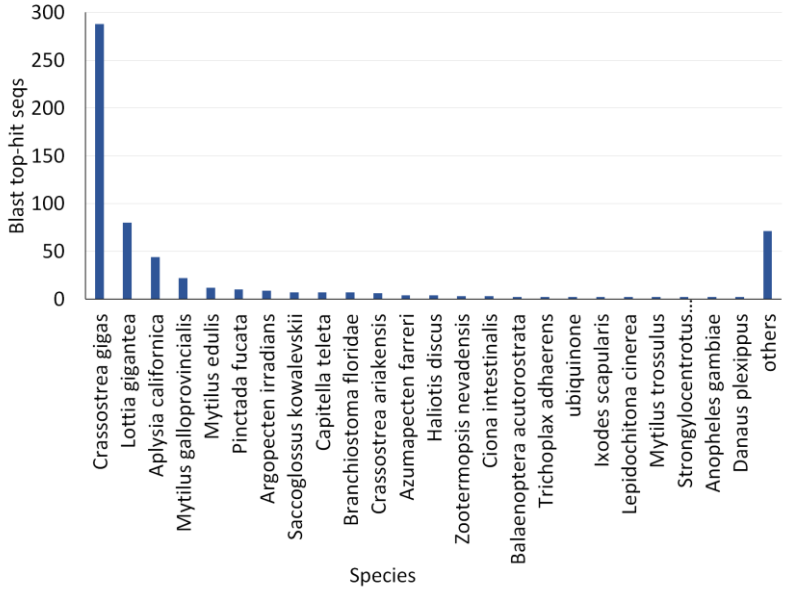

b)
S
H
155
178
321

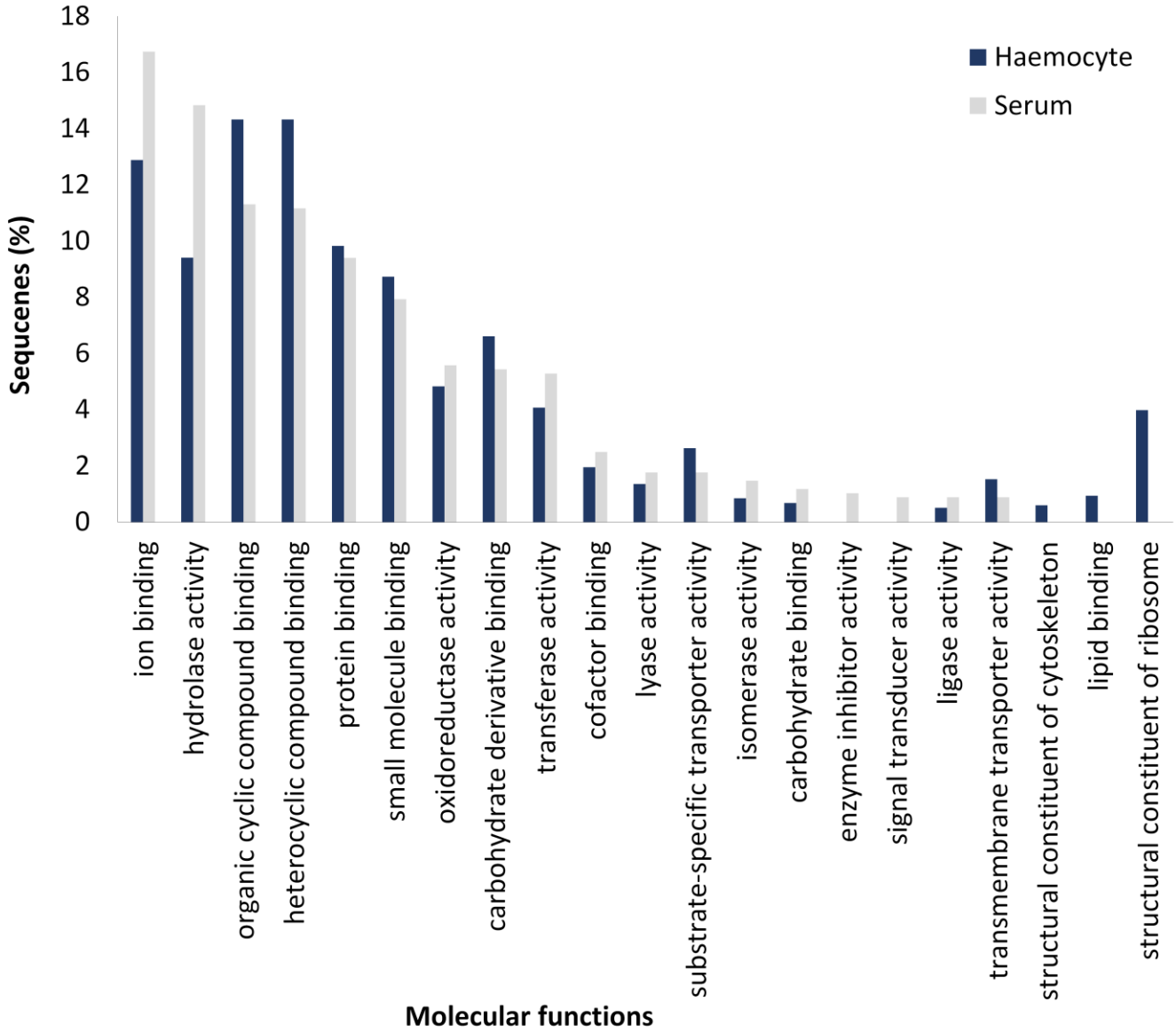

c)
2%
9%
89%
Uniprot (90857 seqs)
B. azoricus transcriptome (33464 seqs)
M. galloprovincialis transcriptome (46791 seqs)

a)


Sequences
2
1
40
21
553
With Blast (without hits)
With Blast Hits
With GO Annotation
With Mapping
Without Analysis

b)


Blast top-hit seqs
300
250
200
150
100
50
0
Crassostrea gigas
Lottia gigantea
Aplysia californica
Mytilus galloprovincialis
Mytilus edulis
Pinctada fucata
Argopecten irradians
Saccoglossus kowalevskii
Capitella teleta
Branchiostoma floridae
Crassostrea ariakensis
Azumapecten farreri
Haliotis discus
Zootermopsis nevadensis
Ciona intestinalis
Balaenoptera acutorostrata
Trichoplax adhaerens
ubiquinone
Ixodes scapularis
Lepidochitona cinerea
Mytilus trossulus
Strongylocentrotus...
Anopheles gambiae
Danaus plexippus
others
Species


Haemocyte
Serum
Sequcenes (%)
0
2
4
6
8
10
12
14
16
18
ion binding
hydrolase activity
organic cyclic compound binding
heterocyclic compound binding
protein binding
small molecule binding
oxidoreductase activity
carbohydrate derivative binding
transferase activity
cofactor binding
lyase activity
substrate-specific transporter activity
isomerase activity
carbohydrate binding
enzyme inhibitor activity
signal transducer activity
ligase activity
transmembrane transporter activity
structural constituent of cytoskeleton
lipid binding
structural constituent of ribosome
Molecular functions