Università degli Studi di Padova

SCUOLA DI DOTTORATO DI RICERCA
IN BIOSCIENZE E BIOTECNOLOGIE

INDIRIZZO IN BIOTECNOLOGIE
XXIV CICLO

Transcriptional responses and AMP transcript diversity in *M. galloprovincialis*

Direttore della Scuola: Ch.mo Prof. Giuseppe Zanotti
Coordinatore d'indirizzo: Ch.mo Prof. Giorgio Valle
Supervisore: Dott.ssa Paola Venier

Dottorando: Umberto Rosani
966467-DR

Index

Riassunto

La presente tesi raccoglie il lavoro svolto durante il mio dottorato in biotecnologie (2009-2011) e, tramite tre articoli pubblicati, mostra come il gruppo di ricerca entro il quale mi sono inserito ha contribuito ad ampliare le conoscenze nel campo della genomica funzionale del mollusco bivalve *Mytilus galloprovincialis*. Utilizzando differenti approcci abbiamo cercato di comprendere aspetti di un trascrittoma solo parzialmente conosciuto, soffermandoci sulle molecole dell'immunità innata, sistema difensivo grazie al quale questo organismo invertebrato sopravvive in ambienti ricchi di patogeni.

Avendo a disposizione una collezione di sequenze espresse (EST) rappresentativa di vari tessuti di mitili nativi e di mitili sottoposti a varie condizioni di *stress*, abbiamo analizzato e descritto alcuni importanti singole sequenze e gruppi di trascritti coinvolti nelle risposte immuni del comune mitilo mediterraneo. Una selezione di questi è stata utilizzata per disegnare una piattaforma *DNA-microarray* (*Immunochip 1.0*) utile a validare sperimentalmente la grande varietà di trascritti genici potenzialmente implicati nel *sensing*, nel *signalling* e in funzioni effettrici dell'immunità innata di *M. galloprovincialis* (Venier *et al*., 2011).

Le piattaforme DNA *microarray* sono ormai largamente utilizzate per lo studio in parallello dell'espressione di migliaia di trascritti e, in generale, forniscono misure affidabili e dinamiche degli andamenti trascrizionali. Riguardo al genere *Mytilus*, sono state sviluppate piattaforme specie-specifiche e piattaforme ibride che includono sonde nucleotidiche di specie evolutivamente affini (Domeneghetti *et al.*, 2011).

Utilizzando un approccio di sequenziamento avanzato, abbiamo infine esplorato e mappato la diversità trascrizionale di nove tipici peptidi antimicrobici espressi in mitilo (varie defensine, mitiline, miticine e mitimicina) scoprendo così, pur con differenze da caso a caso, una elevata frequenza di cambiamenti a singolo nucleotide (Rosani *et al*., 2011). Sulla base di questi risultati sarà interessante studiare meccanismi genetici ed epigenetici che possano spiegare tale variabilità di sequenza, ancora indeterminati in mitilo, e caratterizzare da un punto di vista funzionale le varianti espresse in risposta a diversi stimoli antigenici.

Abstract

The present dissertation contains the work done during my PhD in Biotechnology (2009-2011) and, through three published articles, shows how the research group which I worked in has contributed to the development of functional genomics in the mollusc bivalve *Mytilus galloprovincialis*.

Using different approaches, we aimed to better understand the peculiarities of a transcriptome only partially investigated, focusing our attention on molecules possibly explaining the mussel innate immunity, defense system that allows the survival of this organism in potentially lethal habitats.

Starting from a publicly available collection of expressed sequence tags (ESTs) generated from various tissues of native mussels or from mussels subjected to various stress conditions, we have analyzed and described relevant transcript groups and singletons involved in the mussel immune responses. A selection of these sequences has been used to design a new DNA-microarray platform (*Immunochip 1.0*), a new tool for the experimental validation of the many gene transcripts possibly playing a role in the sensing, signaling and effector functions of the mussel innate immunity (Venier *et al*., 2011).

DNA microarray platforms are widely used to investigate the expression of thousands of transcripts and, in general, they provide reliable and dynamic measures of transcriptional trends. As far as the genus *Mytilus*, species-specific platforms and also platforms including probes of evolutionary close species are available (Domeneghetti *et al.*, 2011).

Using an advanced sequencing approach, we have finally explored and mapped the transcriptional diversity of nine typical antimicrobial peptides expressed in mussels (defensins, mytilins, myticins and mytimycin): although different one from the other, a high frequency of single nucleotide changes was globally evident (Rosani *et al*., 2011). These results pave the way to study genetic and epigenetic mechanisms able to explain the observed transcript diversity, yet undetermined in mussel, and to characterize, from a functional point of view, sequence variants expressed in response to different antigenic stimuli.

Introduction

Mytilus ecology

The family *Mytilidae* has originated in the Devonian era (400 million years ago, Moore, 1983). The presence of a pitted resilial ridge, several hinge teeth, an anterior adductor muscle, a continuous posterior byssus and foot retractor muscle scar differentiate *Mytilus* from other genera in the family (Gosling, 1992). Marine mussels of the genus *Mytilus* are a complex of many sibling species, *M. edulis* (L. 1758), *M. trossulus* (Gould 1850), *M. galloprovincialis* (Lamarck 1819), *M. coroscus* and *M. californianus* (Lamy, 1936 and Soot-Ryen, 1955).

The evolutionary and historical picture most consistent with genetic (Figure 1) and geological data, indicates that the genus *Mytilus* evolved in the north Pacific. Afterwards, the *M. trossulus* stock migrated (~3.5 Mya) through the Bering Strait to the north Atlantic where the *M. edulis* stock arose, and subsequently *M. galloprovincialis* diverged from *M. edulis* in the European water bodies (Seed, 1992).

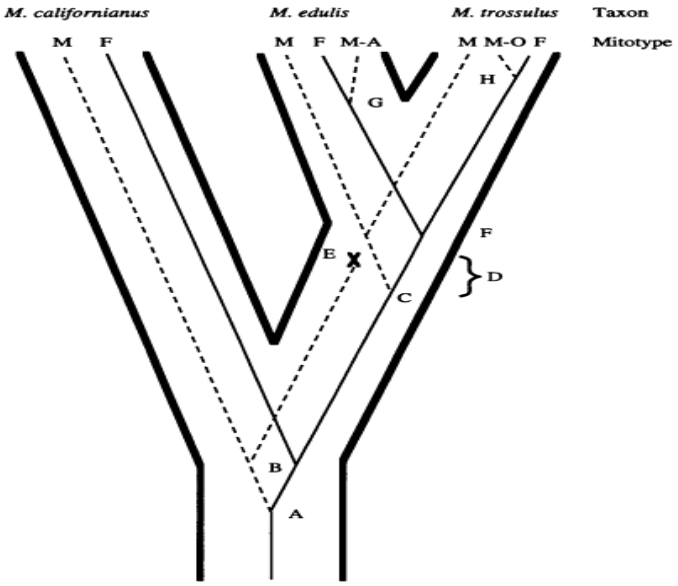


Figure 1 – Cladogram of genus *Mytilus*, according to the mitochondrial genome (Hoeh *et al*., 1997).

Since these early events, more recent translocations of both *M. galloprovincialis* and *M. edulis* have occurred to California, southern Africa, and the northwest Pacific (Hilbish *et al*. 2000, Daguin and Borsa, 2000). Because none of these recent introductions is known to have been intentional, they most likely occurred accidentally via commercial shipping transport, but this remains to be determined.

M. galloprovincialis have evolved in an environment with higher and more stable temperatures than other mussel ecotypes, and a number of physiological, biochemical and molecular studies have shown to be more warm-adapted than its north temperate congeners *M. trossulus* and *M. edulis*. For example, heart function in *M. galloprovincialis* can be sustained at higher temperatures than in the other species, whereas *M. trossulus* can maintain heart function down to lower temperatures (Braby and Somero, 2006).

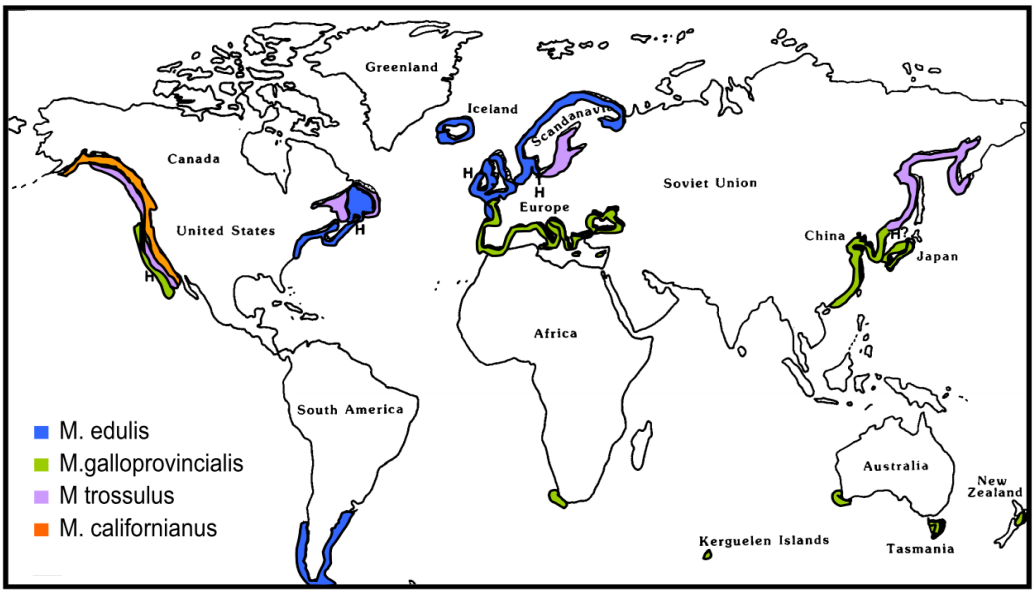


Figure 2 – Geographical distribution of the four most important *Mytilus* ecotypes (modified from Gosling, 1992).

Today, the genus *Mytilus* is widely distributed in temperate waters of the northern and southern hemispheres, with an antitropical distribution defined by the absence of these mussels in the tropics (Gerard *et al*., 2008). In certain world regions, e.g. South Brittany and Gulf of California, different *Mytilus* ecotypes cohabitate and hybridize (Braby and Somero, 2006) and current data suggest that *M. galloprovincialis* can act as an invasive species (Lockwood *et al*., 2010).

In general, mussels play an important role in the functioning of the benthic ecosystem, for example as a scaffold for rocky shore intertidal communities or as contributors to the transfer of mineral and organic matter (Gosling, 2003). Adult mussels are sessile, filter-feeding organisms able to collect microalgae, detritus and organic material and to concentrate by filtration a variety of contaminants through two pairs of gills composed by a large number of parallel filaments. Absorption of nutrients takes places in the digestion gland (brown-greenish in colour) situated in the centre of the body. The mantle, immediately in contact with the interior of the shell, secretes the shell and contains many focal points of gametogenesis. Adult mussels shed eggs and sperm directly from their

genital ducts into the open water, where fertilization takes place. Fertilized eggs usually have a diameter of 60-90 µm (Lutz and Kennish, 1992). After the initial divisions, the embryo begins to swim when cilia first appear (4-5 h).

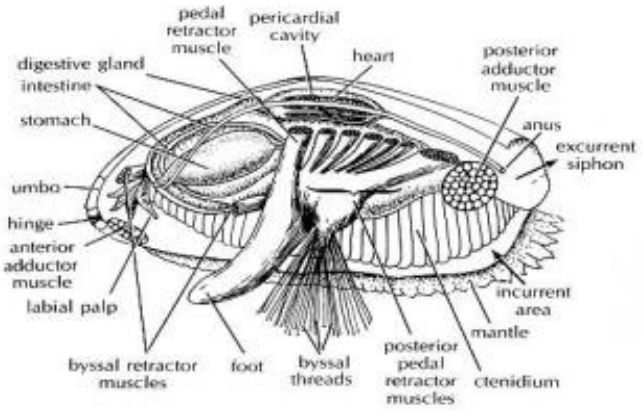


Figure 3 – Internal characteristics of *Mytilus*

But only 24-48 h after fertilization, a ciliated trochophore stage takes place and, starting from the thickened dorsal ectoderm, the larval shell begins to form. At this point the larva exhibits a “D” shape with a shell length of 100-120 µm (Sprung, 1984). Later, a second shell is secreted and begins the “veliger” stage, a pelagic larval stage with high dispersal capability. Active swimming allows the veliger *larvae* to maintain their position in the water column whereas currents increase their potential dispersal over great distances. Following metamorphosis, in which a foot organ develops, the “pediveliger” larva begins the search of a suitable substrate for settling. Duration of larval life, typically from 1 to 4 weeks, is contingent upon temperature, salinity and food availability. Byssal threads, which develop in the last larval stage, anchor the young shell to suitable substrates and announce the final metamorphosis to adult sessile mussels (Lutz and Kennish, 1992).

Growth rates in *Mytilus spp*. are highly variable. Part of this variation is explained by genotype and multilocus heterozygosity (Gosling, 1992) but the majority of variation is probably environmentally determined (Seed and Suchanek, 1992). Numerous factors may work together, depending on location and environmental conditions such as temperature, salinity, food availability, tidal exposure and parasitism or the presence of contaminants (Thompson *et al*., 2000). Several causes may contribute to mortality and to the dynamics of the *Mytilus* populations, including temperature, desiccation, storms and wave action, siltation and biodeposits, intra- and interspecific competition and predation (the most important source of mortality). Many predators target specific sizes of mussels and, therefore, influence the population size structure. The vulnerability of mussels

decreases as they grow, since they can grow larger than their predators preferred size. *Mytilus spp.* may be preyed upon by neogastropods such as *Nucella lapillus*, starfish such as *Asterias rubens*, the sea urchin, crabs such as *Carcinus maenas* and *Cancer pagurus*, fish such as *Platichthys flesus, Pleuronectes platessa* and *Limanda limanda* and birds (Seed and Suchanek, 1992). Longevity is dependent on locality and habitat; on lower seabed, few individuals probably survive more than 2-3 years due to intense predation, whereas high seabed populations are composed of numerous year classes (Seed, 1969) and specimens have been reported to reach 18-24 years of age (Thiesen, 1973).

Mytilus spp. are organisms of great relevance in aquaculture, with a world production above million tons (1.98 million tons, FAO, 2010).

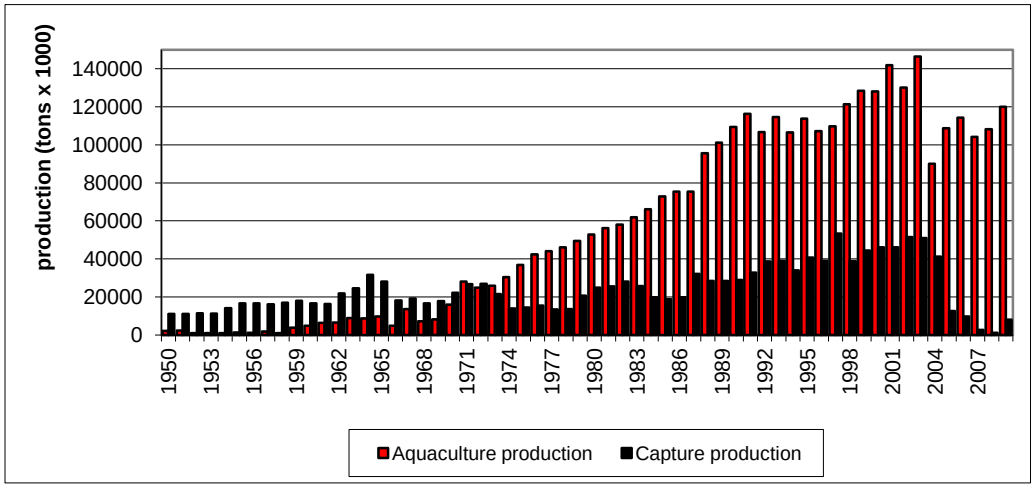


Figure 4 – Worldwide production of *Mytilus galloprovincialis* (captured or acquacultured, 1950-2010, FAO data 2010).

The rearing of *Mytilus galloprovincialis* is always extensive. Mussels can be cultured on suspended ropes covered with mussel seeds, kept in place by nylon nets and suspended in the water column, a technique which can be extended further offshore.

Even if these animals are being cultured extensively and mortalities can be frequent in similar conditions, mussels do not seem to be susceptible to the same pathogens responsible of massive deaths reported for other molluscs. Bower (1992) concluded that, although most parasites did not cause significant mortality, several species of parasites found in mussels could also infect and cause mortality in other shellfish, an indication that mussel populations may be reservoirs of disease for other species.

Mussels are commonly reported to accumulate faecal and pathogenic bacteria, viruses, and toxins from toxic algal blooms. The accumulation of toxins from toxic algal blooms

may result in paralytic shellfish poisoning (PSP), diarrhetic shellfish poisoning (DSP) or amnesic shellfish poisoning (ASP). Bacteria may be removed or significantly reduced by depuration, although outbreaks of diseases have resulted from poor depuration and viruses may not be removed by depuration. Recent improvements in waste water treatment and shellfish water quality regulations may reduce the risk of bacterial and viral contamination (Luckas *et al.*, 2005 and Shumway, 1992).

According to the Mussel Watch concept (Goldberg, 2000), ecological and physiological characteristics of mussels make these animals suitable for environmental monitoring surveys in which a small number of specific biomarkers such as "scope for growth measurements" (Widdows *et al*., 2002) and genotoxicity end-points (Mitchelmore *et al*., 1998) are used to evaluate the quality of coastal waters, management actions and long-term changes of the marine ecosystem (some recent applications are reported in Kimbrough *et al.*, Hunt and Slone, 2010; Scarpato *et al.*, 2010).

Mytilus sequence data

The haploid DNA content ranges from 0.65 pg to 5.4 pg in the bivalve species. As a whole, these values are in the middle of the range for the metazoans (Gregory, 2005). As can be seen in Figure 4, mussel genome sizes are comparable to those of known model organisms whose genome has been already sequenced, or of other important cultured marine species. *Mytilus galloprovincialis* exhibits a diploid complement of 28 chromosomes and a genome size estimated in 1.41–1.92 pg (Gregory, 2005).

The mitochondrial component of the mussel genome is especially interesting because it shows a particular type of inheritance called "doubly uniparental". Mussels have two types of mitochondrial genomes, called F and M. Female mussels are homoplasmic for the F mitotype, which is inherited maternally, while males are usually heteroplasmic, carrying a mixture of the maternal F mitotype and the paternally inherited M genome (Breton *et al*., 2005).

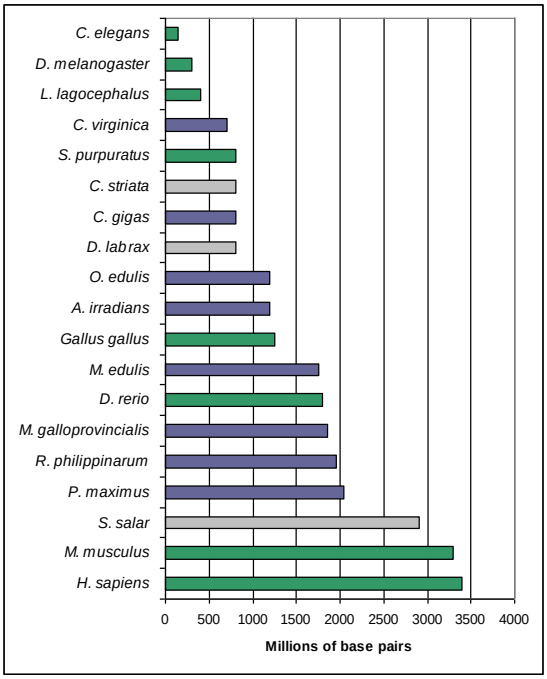


Figure 5 - Genome size of selected aquacultured bivalves (violet) as compared to other marine organisms (grey) and well-known sequenced genomes (green). Data on non-model organisms were obtained from the genome size database (www.genomebase.org, Gregory, 2005).

For the genus *Mytilus*, 67007 ESTs and 4570 protein sequences are currently available (NCBI, Dec 2011), as reported in Figure 5. Only a small number of complete gene sequences are deposited, mostly related to mitochondrial genes. Knowledge of the genome sequence, structure and related evolutionary aspects is limited and mostly based on “single gene” approaches.

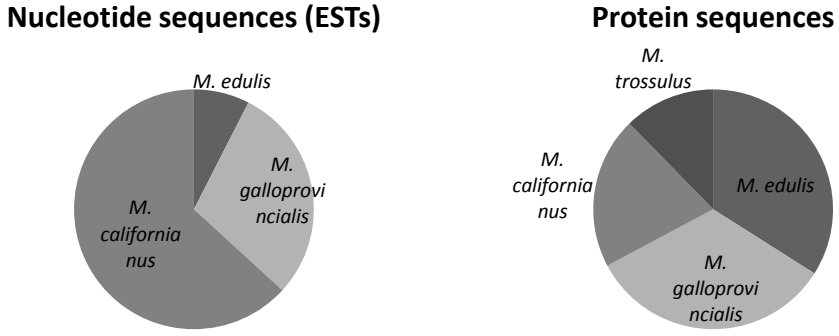


Figure 6 – Number of EST and protein sequences for the main *Mytilus* species (NCBI, Dec. 2011).

Introduction

Sequence data of “non-model” organisms are steadily increasing, and, also in mussels, they allow comprehensive molecular approaches to study the organism responses to the surrounding habitat. Firstly, gene expression profiles based on DNA microarrays (e.g.: Venier *et al*., 2006; Lockwood *et al*., 2010), single-gene studies (e.g.: Pantzartzi *et al*., 2010, Aceto *et al*., 2011) and lastly high-throughput sequencing approaches (e.g.: Craft *et al*., 2011) disclosed a high amount of data and are contributing to a progressively advanced functional genomics of the *Mytilus* spp.

In this dissertation, I pass through three articles published during my PhD (Venier *et al*. 2011, Domeneghetti *at al*. 2011, Rosani *et al*., 2011) testifying my activity and the experimental approaches used in the study of the mussel transcriptome.

BMC GENOMICS

2011

Insights into the innate immunity of the Mediterranean mussel *Mytilus galloprovincialis*

Paola Venier[1*§], Laura Varotto[1], Umberto Rosani[1], Caterina Millino[2], Barbara Celegato[2], Filippo Bernante[2],Gerolamo Lanfranchi[1,2], Beatriz Novoa[3], Philippe Roch[4], Antonio Figueras[3], Alberto Pallavicini[5*]

[1]Department of Biology, University of Padova, Via U. Bassi, 58/B, 35121, Padova, Italy.
[2]C.R.I.B.I. Biotechnology Centre, University of Padova, Via U. Bassi, 58/B, 35121, Padova, Italy.
[3]Instituto de Investigaciones Marinas, CSIC, C/Eduardo Cabello, 6, E-36208 Vigo, Spain.
[4]Ecosystèmes Lagunaires UMR CNRS-University of Montpellier 2, cc093, place E. Bataillon, F-34095 Montpellier cedex 05, France.
[5]Department of Life Sciences, University of Trieste, P.le Valmaura, 9, 34148 Trieste, Italy.

[§]Corresponding Authors

Received October 8, 2010; Accepted January 26, 2011.

BACKGROUND

Sessile bivalves of the genus *Mytilus* are suspension feeders relatively tolerant to a wide range of environmental changes, used as sentinels in ecotoxicological investigations and marketed worldwide as seafood. Mortality events caused by infective agents and parasites apparently occur less in mussels than in other bivalves but the molecular basis of such evidence is unknown. The arrangement of Mytibase, interactive catalogue of 7,112 transcripts of *M. galloprovincialis*, offered us the opportunity to look for gene sequences relevant to the host defences, in particular the innate immunity related genes.

RESULTS

We have explored and described the Mytibase sequence clusters and singletons having a putative role in recognition, intracellular signalling, and neutralization of potential pathogens in *M. galloprovincialis*. Automatically assisted searches of protein signatures and manually cured sequence analysis confirmed the molecular diversity of recognition/effector molecules such as the antimicrobial peptides and many carbohydrate binding proteins. Molecular motifs identifying complement C1q, C-type lectins and fibrinogen-like transcripts emerged as the most abundant in the Mytibase collection whereas, conversely, sequence motifs denoting the regulatory cytokine MIF and cytokine-related transcripts represent singular and unexpected findings. Using a cross-search strategy, 1,820 putatively immune-related sequences were selected to design oligonucleotide probes and define a species-specific Immunochip (DNA microarray). The Immunochip performance was tested with hemolymph RNAs from mussels injected with *Vibrio splendidus* at 3 and 48 hours post-treatment. A total of 143 and 262 differentially expressed genes exemplify the early and late hemocyte response of the Vibrio-challenged mussels, respectively, with AMP trends confirmed by qPCR and clear modulation of interrelated signalling pathways.

CONCLUSIONS

The Mytibase collection is rich in gene transcripts modulated in response to antigenic stimuli and represents an interesting window for looking at the mussel immunome (transcriptomes mediating the mussel response to non-self or abnormal antigens). On this basis, we have defined a new microarray platform, a mussel Immunochip, as a flexible tool for the experimental validation of immune-candidate sequences, and tested its performance on *Vibrio*-activated mussel hemocytes. The microarray platform and related expression data can be regarded as a step forward in the study of the adaptive response of the *Mytilus* species to an evolving microbial world.

Background

DNA sequencing, genomic and post-genomic techniques have made available long lists of partially described sequences and impose the construction of databases essential for mining very large data sets [1-2]. Whenever complete transcript sequences and gene structure information are not available, misidentification and erroneous annotation can easily occur. In fact, the greatest challenge in biology today is the precise delineation of genes and protein networks able to explain physiological and pathological phenotypes [3-5].

Besides well known model organisms, a number of invertebrate species differing in life cycles and adaptive strategies support the current understanding of the innate immunity, especially those living in fluctuating marine systems [6-9]. Filter-feeder bivalves such as mussels, oysters and clams typically harbour a community of commensal, opportunistic and pathogenic organisms composed of endoparasites such as *Mytilicola* and *Urastoma*, protozoans such as *Bonamia*, *Haplosporidium Marteilia*, *Perkinsus* spp., bacteria of the genus *Nocardia* and *Vibrio*, *Herpes* and enteric viruses. Microbial species take part in the biogeochemical cycles and some of them are expected to play a probiotic role in their typical hosts. The common rod-shaped *Vibrios* (>60 Gram negative heterotrophic species) well exemplify associations ranging from mutualistic to pathogenic in aquatic animals [10-12]. *V. cholerae*, *V. parahaemolyticus*, *V. vulnificus* and other nine *Vibrio* species cause mild or severe syndromes in humans while other halophilic *Vibrios* occurring in brackish and marine habitats can greatly affect molluscs, crustaceans and fish (e.g. *V. tapetis*, *V. alginolyticus*, *V. splendidus*, *V. pectenicida*, *V. harvey*, *V. penaeicida*, *V. anguillarum*). Often triggered by environmental factors such as temperature, salinity or pollutants, elements of such microbiota may invade and colonize the host and eventually lead to disease outbreaks and mortality, especially in larvae, spat and juveniles of natural and farmed bivalves [13-15]. Compared to oyster and clams, no apparent mortality and fewer pathologies have been reported in mussels [16-17]. It is more likely that *Mytilus* spp. are a reservoir of infective agents for aquatic organisms and humans, since, for instance, they tolerate significant amounts of *V. alginolyticus*, *V. parahemolyticus* and other vibrios [18]. In fact, comparative and advanced understanding of the early-induced host responses may sustain and improve the aquaculture production in many coastal regions worldwide [17, 19].

Immunocompetent mollusc cells, at least the circulating hemocytes, and a variety of molecular effectors provide a rapid and robust line of defence against potential pathogens. Once activated by the interaction between pathogen associated molecular patterns (PAMPs) and pathogen recognition receptors (PRR), such cells display

chemotactic and chemokinetic reactions, participate in encapsulation and melanization, carry out phagocytic or lytic killing. These events are made possible by the concerted action of transmembrane and soluble lectins, Toll-like and virus sensing receptors, hydrolytic enzymes and proteolytic reaction cascades, short-lived cytotoxic by-products and antimicrobial peptides (AMP) [20-25]. According to morphological observations and flow cytometry, bivalve hemocytes are heterogeneous and very dynamic cells of 7-10 µm size which can be classified into large granulocytes (eosinophilic) most active in phagocytosis and ROS production, large hyalinocytes with intermediate activity, small non-phagocytic semigranular cells (basophilic) and the less abundant blast-like hyalinocytes [26-28]. As *Mytilus* hemocytes respond to interleukin 1 (IL 1), tumour necrosis factor (TNF) and to opioid peptides (the endogenous messengers between the nervous and the immune system) they may be part of an ancient monokine-like network [29-30]. Also relevant to the use of mussels as biosensors of coastal pollution [31] the interdependence of cell processes modulated by chemical contaminants and infective agents requires additional study [32-33].

The sequence data available for bivalve species are slowly but steadily growing, especially through EST collections [4, 8, 34]. A set of 1,714 cDNA probes of *M. galloprovincialis* was arranged to investigate the transcriptional signatures of pollutants [35] but more work has subsequently been devoted to EST sequencing, also using technologies which provide very large amounts of short reads more difficult to annotate [36]. A double set of 5' and 3' ESTs of *M. californianus*, 42,354 in total, was used to investigate the influence of the tidal cycle on mussel physiology [7]. As a result of laboratory treatments performed with environmental pollutants, bacterial antigens and viral-like polynucleotides, 18,788 high-quality ESTs of *M. galloprovincialis* are now organized in a structured collection of 7,112 transcript sequences [37], named Mytibase and including most of the ESTs publicly available for the Mediterranean mussel (19,575 ESTs at Oct 2010).

In the absence of genomic information, this knowledge base offered us the unique opportunity to outline the available mussel immunome and develop a new microarray platform. In the following sections we present the most relevant Mytibase clusters and singletons related to mussel immunity and the validation of a species-specific Immunochip with hemocyte samples of *Vibrio*-injected mussels.

Results and Discussion

Identification of immune-related Mytibase sequences

The Mytibase descriptions report BLAST similarity searches, structured Gene Ontology vocabulary (GO) and identifiable protein features of the Interpro database (IPR) [38]. The latter, in particular, supported the characterization of unknown or poorly predicted sequences, and integrated the meaning of a substantial fraction of the mussel transcripts. Not surprisingly, the Mytibase sequences (MGCs) are often defined by multiple IPRs with the notable exception of 588 ESTs codifying the mussel AMP that could only be recognized by similarity to prototype sequences of mytilin, myticin, mytimycin and defensin. Table 1 illustrates in decreasing order of abundance the first 15 of 1753 redundant IPRs present in the MGCs and the known mussel AMP.

Table 1 - The first 15 Identifiable Protein Features and 4 Anti Microbial Peptide precursors, listed according to the EST abundance in Mytibase.

IPR or AMP	ESTs	Description
IPR008983*	581	Tumour necrosis factor-like
IPR001073*	430	Complement C1q protein
Mytilin	277	Antimicrobial peptide precursor
Myticin	267	Antimicrobial peptide precursor
IPR001304	246	C-type lectin
IPR002048	187	Calcium-binding EF-hand
IPR012677	178	Nucleotide-binding, alpha-beta plait
IPR002181**	164	Fibrinogen, alpha/beta/gamma chain, C-terminal globular
IPR014716**	159	Fibrinogen, alpha/beta/gamma chain, C-terminal globular, subdomain 1
IPR000504	146	RNA recognition motif, RNP-1
IPR011992	133	EF-Hand type
IPR002035	132	von Willebrand factor, type A
IPR001254	129	Peptidase S1 and S6, chymotrypsin/Hap
IPR009003	129	Peptidase, trypsin-like serine and cysteine
IPR001314	124	Peptidase S1A, chymotrypsin
IPR003582	119	Metridin-like ShK toxin
IPR004000	119	Actin/actin-like
…	…	…
Defensin	28	antimicrobial peptide precursor
Mytimycin	16	antimicrobial peptide precursor

*, **: the paired cases exemplify domains with overlapping architecture. See more on the protein signatures at http://www.ebi.ac.uk/interpro.

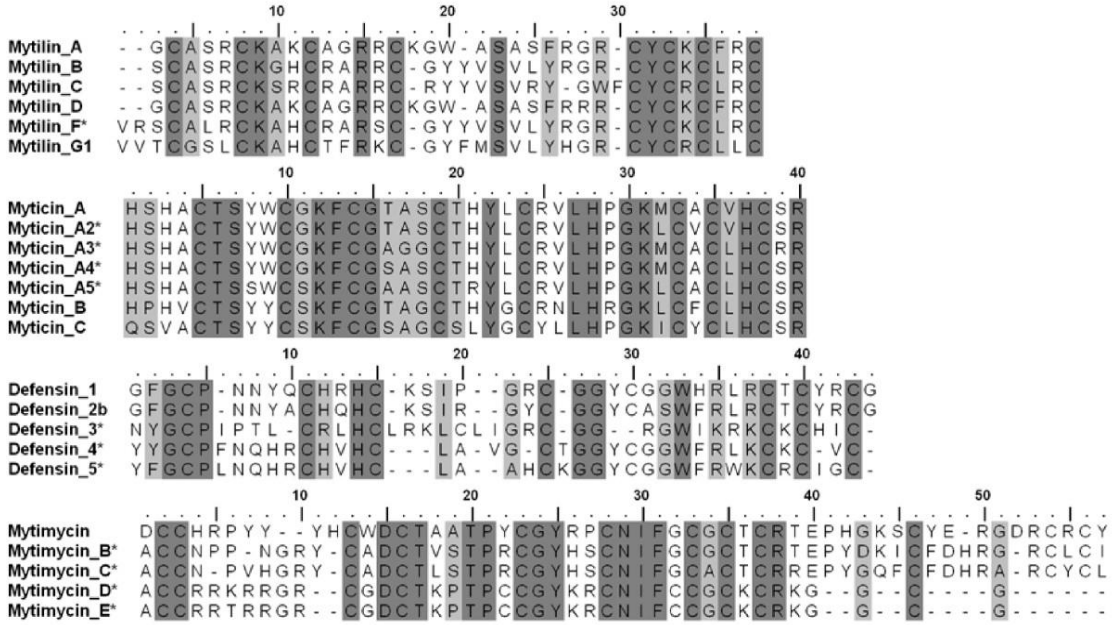
The protein motifs represented in Mytibase point to cell processes which are not restricted to the immune system as only 15% of the total IPRs directly refer to immunity. Nevertheless, the abundance of transcripts identifying AMP precursors or including domains such as Complement C1q (IPR001073) and the related Tumour Necrosis Factor-like (IPR008983), C-type lectin (IPR001304) and Fibrinogen, alpha/beta/gamma chain, C-terminal globular subdomain (IPR002181/IPR014716) definitely confirm that the Mytibase EST collection is particularly rich in immuno-related transcripts. Conversely, about 41% of the listed IPRs are exclusive of single clusters and singletons, with uncommon and intriguing protein motifs exemplified by IPR001398 (macrophage migration inhibitory factor, 4 ESTs in 3 clusters) and IPR012916 (RED-like protein, 6 EST in 1 cluster). The IPRs mentioned are easily found in Mytibase as Interpro key words.

Since the genome of *M. galloprovincialis* is not available and sequence data are still limited, we applied a multiple search strategy to identify in Mytibase a relevant set of immune-related sequences. A low-stringency tBLASTn search allowed the extraction of 309 mussel sequences related to immune system processes (GO:0002376) and 1,021 sequences similar to those indexed in the multispecies catalogue ImmunomeBase (download permitted by C. Ortutay *et al.*, Finland). Searches based on key words and manual screening yielded an additional 973 inputs and supported the final selection of 1820 mussel sequences, which can be regarded as an operational set and the starting point for the progressive authentication of immune-related candidates by transcriptional analyses and gene studies. Additional file 1 describes the selected MGCs and updates their functional annotation whereas the following paragraphs illustrate by abundance and putative function the most relevant ones to the mussel immunity.

Transcripts identifying antimicrobial peptides

Almost ubiquitous in the living species but highly diverse in structure and biological activity, host defence peptides interact with negatively charged cell membranes, lead to microbe killing and modulate both the innate and inducible antimicrobial responses in mammals [39]. Four groups of AMP are known in mussels: defensins, mytilins myticins and mytimycins [40-41]. The cationic and amphipatic structure of the mature peptides is stabilized by 4 intrachain disulphide bonds (6 in mytimycin) according to a unifying tridimensional motif [42]. Mytibase includes the full length precursor sequences of all the mussel AMP with some new variants: they are reported as mature peptide sequences in Figure 1.

Figure 1 – Multiple alignment of the four AMP families identified in Mytibase.



According to Muscle alignment and BioEdit processing, each mussel AMP family includes new sequences indicated by an asterisk (*). Identical and similar (PAM250) residues are reported in dark and light grey, respectively.

Myticins are subdivided in A, B (96 aminoacid precursor) and the polymorphic type C (100 aminoacid precursor) [43]. Searching tBLASTn similarities to prototype sequences, we identified in Mytibase many precursors of myticin C (A7DWV6, 124 ESTs), myticin A (P82103, 88 ESTs) and myticin B (P82102, 55 ESTs). Robust non-synonymous SNPs analysis allowed us to split the sequence cluster of myticin A into 5 subgroups named A, A2, A3, A4 and A5, confirmed by 23, 38, 2, 21 and 4 sequence traces of high-quality, respectively (the latter three groups present indels in the 3'-UTR region).

Mytilin precursors are more heterogeneous in length ranging between 97 and 105 residues, and can be easily differentiated from the myticin precursors due to a different cysteine pattern. Similarly, we identified mytilin A (P81612, 5 ESTs), mytilin B (Q9Y0B1, 111 ESTs), mytilin C (Q5XWD7, 127 ESTs), mytilin D (B3VT96, 9 ESTs). We could also extend the sequence of Mytilin G1 (MGC00423, 14 ESTs) [41] and we propose MGC00659 (11 ESTs) as Mytilin F, namely a new mytilin component.

The defensin precursors identified in Mytibase are MGD1 (P80571, 20 ESTs), MGD2b (Q9U6U0, 5 ESTs) and three new sequences proposed as MGD3, MGD4, and MGD5. Due to the presence of a stop codon just after the 8th conserved cystein, defensins MGD3 and MGD4 are shorter than the others whereas MGD5 is the longest with 97

aminoacid residues.

Only one Mytibase EST (Mg_Nor01_39J03) corresponds to the mytimycin described in *M. edulis* (P81614) and four other sequences grouped from 4, 4, 4 and 3 ESTs may be regarded as new mytimycin variants. Curiously, two of these ESTs (Mg_Nor01_12C12, Mg_Nor01_39B01) display a long insertion in the 5' UTR and a signal peptide with maximal cleavage probabilities between positions 18-24 from ATG. cDNAs normalization was essential to reveal the rare mytimycin ESTs whereas the other more abundant AMP sequences can be easily and mainly attributed to hemocyte libraries prepared from immunostimulated Italian and Spanish mussels, without evidence of preferential geographical distribution (see in Mytibase the EST composition of Hae01-Hae05 and NOR01 libraries). All mussel AMP (27 clusters) and one hydramacin-like transcript (MGC2887) have been included in the Immunochip.

Transcripts containing C1q and Tumour Necrosis Factor-like domains

The overlapping C1q (IPR001073) and TNF-like (IPR008983) domains have probably evolved by divergence from an ancient recognition molecule whose diversification could have started with urochordates and cephalocordates [44-45]. The large family of proteins with a C1q domain support many biological processes, from complement activation, modulatory immune functions, apoptotic cell clearance to coagulation, embryonic development and tissue homeostasis [46].

In mammals, the 18 polypeptide chains composing the complement subcomponent C1q are characterized by a short N-terminal region, a collagen-like Gly/Pro-rich tract and a C-terminal tulip-like structure of globular C1q domains (gC1q) also found in ficolins and other proteins. The C1q binding to immunoglobulins within immunocomplexes initiates the classic complement cascade and pathogen elimination. In the presence of Ca ions, the interaction of self and non-self ligands with charged gC1q residues causes gC1q reorientation and bending of the collagenous region. The activation signal is then transmitted to serine protease precursors (C1r, C1s) which, in turn, promote the proteolytic complement cascade and formation of a membrane attack complex [47]. Overall, the modularity and versatility of pattern recognition confirm the essential role of gC1q in both innate and acquired immune responses.

Several MGCs display sequence similarity to C1q, TNF, precerebellin, collagen and emilin proteins. Searching the TNF-like domain IPR008983 in Mytibase, we identified 146 transcripts, most of which are also characterized by the C1q domain IPR001073. Hidden Markov model analysis allowed the recognition of 22 additional C1q-containing sequences and the C1q motif was confirmed by manual validation in all 168 cases,

without evidence of a true TNF domain. To illustrate their molecular diversity, a selection of the most divergent C1q-containing MGCs is reported in Figure 2. Many of them are similar to a sequence highly expressed in the mantle of the oyster *Pinctada fucata* [48] and some are very abundant, for instance MGC0284 with 99 out of 109 ESTs originating from hemocyte cDNAs [49]. In addition to the C-terminal globular domain, most of the predicted C1q proteins of *M. galloprovincialis* (163-293 aminoacid residues) have a short N-terminal signal peptide but lack central collagen-like repeats; hence, they should represent secreted gC1q receptor proteins expected to elicit chemotaxis and pathogen lysis via more ancient complement pathways [50].

Figure 2 - Diversity of mussel transcripts containing the C1q domain.

Cladistic analysis of 65 C1q domains representing a subset of the 168 total sequences identified in Mytibase. This highly diversified non redundant subset was produced with NCBI BlastClust (-p F -L .5 -b T -S 40). The cladogram is based on the Neighbor-Joining method (1000 bootstrap replicates). The evolutionary distances were computed by using

the JTT substitution matrix (number of amino acid substitutions for each of the 406 sites).

The abundance and molecular diversity of the C1q-containing transcripts of *M. galloprovincialis* suggest pathogen-induced expansion of lectin-like PRR: the identification of related gene sequences will allow a comparison with the 32, 52 and 75 C1q gene models reported in *Homo sapiens*, *Danio rerio* and in the amphioxus *Brachiostoma floridae*, respectively [45, 51].

The new microarray platform includes 162 of these mussel transcripts and also a few mussel transcripts similar to the complement component C3 (MGC07073 and MGC05748 display the alpha-2-macroglobulin complement component domain, IPR011626, and alpha-macroglobulin receptor binding domain, IPR009048, respectively) and a Membrane attack complex/perforin/C9 (IPR001862, MGC00636) expected to be involved in the pathogen lysis.

Additional lectin-like and receptor-related transcripts

Protein-carbohydrate recognition is crucial to many cell processes and host-pathogen interactions. Lectins are membrane-associated and soluble proteins with specific carbohydrate recognition domains (CRD) which can either facilitate mutualistic interactions between host and microbiota or initiate innate and adaptive immune responses [52-54]. Acting as recognition receptors, lectins promote opsonization, phagocytosis and the activation of the complement system [25]. Structural and functional features distinguish eight to fifteen lectin groups largely related to immunity: C-type, S-type or glycan-binding galectins, I-type specific to sialic acids and glycoseaminoglycans also containing an Ig-like fold, pentraxins, fucolectins, fibrinogen-like lectins, ficolins, tachylectins and slug agglutinin, chitinase-like lectins, and orphan lectins. Transmembrane calnexins and soluble calreticulins support trafficking, sorting and maturation of glycoproteins whereas lectins localized in the plasma membrane or released into the extracellular matrix and body fluids mediate a broad range of processes including cell adhesion, cell signalling, pathogen recognition and endocytosis. Compared to more ancient lectins acting in the quality control of glycoproteins, extracellular lectins such as ficolins have evolved independently in the vertebrate and invertebrate lineages. The evolutionary radiation of these molecules emphasises the importance of the glycan code and lectin-ligand interactions in the immune responses and apoptotic cell clearance [55-56].

Table 2 summarizes in decreasing abundance the lectin-like sequences identified in Mytibase by searching archetype lectin domains (IPRs). A total of 148 MGCs (429 ESTs) share the descriptive term ‘lectin’ as Interpro key-word. The most abundant and

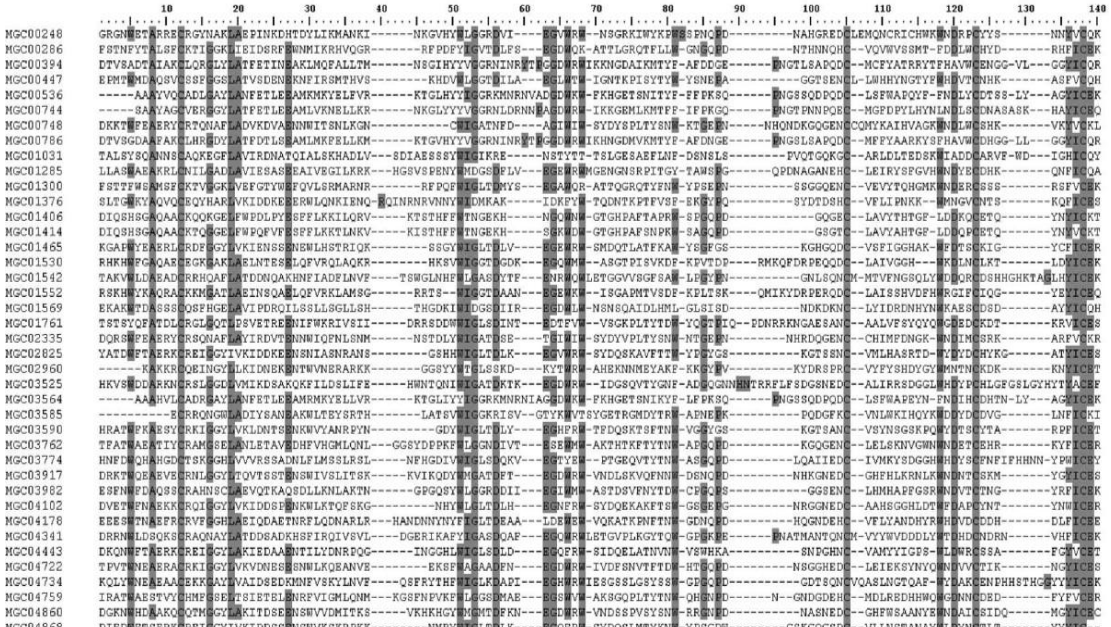
heterogeneous group refer to C-type lectins (IPR001304) originally named to reflect the importance of Ca^{+} in sugar binding. Many are similar to the nacre protein perlucin from *Haliotis laevigata* [57], while others remind of mammalian proteoglycans, type II receptors expressed particularly on macrophages and dendritic cells. For instance, among 9 MGCs the consensus MGC04167 is the most similar to the macrophage mannose receptor, protein involved in the glycoprotein endocytosis and antigen presentation, whereas 13 MGCs display similarity to the human DC-SIGN CD209 antigen [53]. Regardless of some conserved residues (e.g. cysteine and glycine) the remarkable sequence diversity of the C-type lectins expressed in mussels confirms them as candidate PRR (Figure 3). As a matter of fact, many of the *Caenorhabditis elegans* proteins containing a C-type lectin *d*omain (278 genes) support pathogen-specific responses [58].

Table 2 - Census of the Mytibase lectin-like transcripts (MGCs) by archetype IPRs

INTERPRO		MYTIBASE	
ID	Description	MGCs	ESTs
IPR001304	C-type lectin	104	246
IPR002181	Fibrinogen, alpha/beta/gamma chain, C-terminal globular	62	164
IPR013320	Concanavalin A-like lectin/glucanase, subgroup (related to galectins, S-type)	12	52
IPR006585	Fucolectin tachylectin-4 pentraxin-1 (F-type)	9	14
IPR013151	Immunoglobulin (I-type or siclecs)	7	9
IPR001223	Glycoside hydrolase, family 18, catalytic domain	4	53
IPR000922	D-galactoside/L-rhamnose binding SUEL lectin	4	17
IPR009169	Calreticulin	3	13
IPR001220	Legume lectin, beta chain	3	3
IPR009011	Mannose-6-phosphate receptor, binding (P-type)	2	4
IPR003990	Pancreatitis-associated protein	2	2
IPR008997	Ricin B-related lectin (R type)	1	13
IPR004043	LCCL	1	3
IPR005052	Legume-like lectin (L-type)	1	1

Each IPR can be referred to upstream- and downstream-related protein domain (e.g. IPR001304).

Figure 3 - Multiple alignment of 87 Mytibase sequences identified as C-lectins (IPR001304).



The SMART consensus terms for the CLECT domain (SM00034) are shown at the bottom. Four cystein residues are entirely conserved (positions 12, 105, 123 and 138 in the multiple alignment). Other conserved (75%) residues are W5, A8, L19, W51, G53, G64, W66, W68, W118, D120, I137, C138 and E139. Positions with at least 50% of conservation are also shadowed.

The second abundant lectin-like group recalls fibrinogen and fibronectin proteins and ficolins. Like the CRD of the mannose-binding lectins, the C-terminal fibrinogen-like domain (IPR002181) of ficolins has a bouquet-like structure which binds the carbohydrate residues of foreign and apoptotic cells (hence, pathogen opsonization, phagocytosis and clearance of dying cells) or in association with specific serine proteases (SP) initiates the proteolytic complement cascade and pathogen lysis [25, 59]. Species-specific expansion of fibrinogen-related proteins (FREPs) has been reported in the snail *Biomphalaria glabrata* and the mosquito *Anopheles gambiae* [5, 60]. In the crayfish *Pacifastacus leniusculus*, a protein containing the fibrinogen-like domain, but devoid of the hemagglutinating activity typical of vertebrate ficolins, acts as negative regulator of the prophenoloxidase system (proteolitic cascade similar to that of the complement system) and interferes with the transformation of quinone compounds to melanin [61].

Other MGCs point to galectins (IPR000922), I-type lectins (IPR013151) able to bind carbohydrate ligands via immunoglobulin-like domains, GH18 chitinase enzymes (IPR001223), L-type lectins (IPR005052) entailed in the intracellular protein sorting and P-type lectins (IPR009011), transmembrane proteins involved in the transport of lysosomal enzymes from the Golgi complex and the cell surface to lysosomes. For instance, chitinases are glycosyl hydrolases widely expressed from cnidarians to mammals, able to degrade the polysaccharide β-(1-4)-poly-N-acetyl D-glucosamine and confer protection against chitin-containing pathogens and parasites [62-63].

Mytibase is also rich in sequences with WD-40 repeats (IPR011046) and Leucin Rich Repeats (LRR, IPR001611). The modular organization of WD and LRR domains of vertebrate proteins sustains the diversity and plasticity of the apoptosome and inflammasome complexes in response to microbial products and metabolic stress, with the latter commonly signalled by ROS, nucleic acids, cathepsin and other molecules released by damaged cells [64-65]. In detail, the ligand binding to the carboxy-terminal LRR region of cytosolic receptors of the NOD-like family (NALPs/NLRs) can trigger receptor clustering, recruitment and activation of initiating caspases, release of IL-1R and IL18 citokines, inflammation and inflammatory cell death (pyroptosis).

Allthough many MGCs refer to nucleic acid binding proteins (IPR012677, IPR012340) or RNA/DNA binding helicases (IPR014021), further study is necessary to assign them an antiviral function typical of intracellular NOD-like and RIG-like helicase receptors or some membrane-bound TLRs [66]. With the possible exception of MGC02873, a Piwi-like singleton suggestive of silencing and regulative events in germ cells and hematopoietic stem cells, and putative RNA helicases of the DEAD-box family (IPR014014), we could not identify in Mytibase the core siRNA machinary Dcr-2, r2d2, AGO2 responsible for

antiviral responses in *Drosophila* [67].

Keeping in mind the 222 and 72 TLR gene models identified in the genome of *Strongylocentrotus purpuratus* and *Branchiostoma floridae* [51, 68], respectively, the occasional presence in Mytibase of TLR-related sequences (IPR000157, IPR004075) is disappointing. In fact, only MGC03952, MGC06978, MGC07535 and few other LRR-containing sequences display fragmentary similarity to human, fish and invertebrate TLR proteins. In the human TLRs, extracellular LRRs are arranged to recognize specific PAMPs whereas the intracellular Toll/Interleukin-1 receptor (TIR) domain activates downstream signalling pathways. According to a recent comparative overview, the identification of authentic invertebrate TLRs cannot rely on the sequence homology and requires functional studies [69].

Present in Mytibase are also putative Ig-like and MHC-related surface antigens (IPR013783, parent domain), sequences with a thyroglobulin domain (IPR000716) typical of Insulin-like Growth factor binding proteins and HLA class II invariant chain, and G-Protein-Coupled Receptors (IPR000276) involved in the transduction of various signals and accounting for about 3% of human genes [70].

Other MGCs are similar to von Willebrand Factor type C (IPR001007) found in plasma proteins promoting adhesion and thrombus formation at injured sites, Fasciclin-like proteins (IPR000782), Toll-InterLeukin Receptor (IPR000157), Speract/scavenger receptor (IPR001190), receptor-binding alpha macroglobulins (IPR009048), mannose-6-phosphate receptor (IPR009011) and TNF receptors (IPR001368: MGC05090, MGC06564).

Transcripts involved in signalling and regulative networks

Not restricted to the innate immunity, cell signalling against fungal, bacterial and viral antigens occurs in insects through the Toll, Imd, Jak-STAT and P13K/Akt/TOR pathways. The first two are similar to the vertebrate TLR/IL and TNF signalling pathways, and interact with distinct NFkB factors to induce the expression of AMP and other molecules, whereas the inhibition of the nutrient signalling P13K/Akt/TOR can restrict viral replication by cell autophagy and reallocation of the resources from growth to immune defences [67].

Related to Toll/IL and TNF signalling are MGCs putatively identifying the LPS-induced TNF alpha factor or LITAF (IPR006629, 9 MGCs), TNF receptor associated factor TRAF (IPR008974, 4 MGCs), the adapter molecule MyD88 (MGC03566, MGC07770), Pellino which is known to associate with the kinase domain of the Pelle Ser/Thr kinase (MGC02650), NF_kB_inhibitor Cactus (MGC03934), a NFkB inhibitor-interacting Ras-like protein (MGC06766) and the transcription factor NFkB/Rel/Dorsal (IPR000451,

IPR011539; MGC05614, MGC07242).

Definitely, many MGCs include the ankyrin repeat (IPR002110, 50 MGCs) typical of regulatory proteins but insufficient in itself to provide function recognition. Conversely, putative mussel kinases and phosphatases (e.g. IPR000340, IPR002290, IPR008343, IPR015731) support the existence of the mitogen-activated protein kinase (MAPK) signalling, whereas the EF-hand signature (IPR011992) and putative small G proteins (IPR003579, IPR003577, IPR005225) denote calcium regulated pathways. Putative zinc finger proteins (IPR000315, 49 MGCs), transcription factors bZIP-like (IPR004827), LIM-type (IPR001781), Jun-like (IPR002112), p53/RUNT-type (IPR012346) and repressors of transcription (IPR011991) reinforce the idea of multiple signalling pathways in mussels.

Interactions between protein kinase C, FAK and Src protein tyrosine kinases occur during the integrin-mediated spreading of *Lymnaea stagnalis* haemocytes [71] and robust intracellular signalling is essential to cytoskeleton remodelling, cell adhesion and migration of PAMP-activated haemocytes [26, 72-73]. Although more than 60 MGCs contain a DNA-binding domain and some of them include the SH2 domain (IPR000980), there is no proof in Mytibase of a mussel JAK/STAT pathway, the main signalling system for a wide array of mammalian cytokines and growth factors. Nevertheless, the remarkable presence of a mussel Macrophage Migration Inhibitory Factor (3 MGCs), transcripts recalling Platelet-Derived Growth Factor (MGC01828, MGC07226), interferon-induced proteins (IPR009311, IPR004911), an interleukin enhancer binding factor (MGC05350), an interleukin-1-receptor-associated kinase (MGC00477) and G-protein coupled chemokine-like receptors, altogether evoke a regulatory humoral network able to reinforce mussel immunity. Unquestionably, Mytibase does not contain an IL17 homologue, found instead expressed in oyster hemocytes following bacterial stimulation [74].

Finally, MGCs denoting histone proteins (IPR005819, IPR009072), deacetylase (IPR000286, IPR003000) and acetyltransferase (IPR015418) enzymes confirm the importance of chromatin remodelling and histone modifications in the regulated transcription of effector genes.

Transcripts related to oxidative stress and chaperon proteins

Scavenging and enzymatic activities protect the living cells from various stress factors, from endogenous reactive oxygen species (ROS) produced for instance by the mitochondrial respiratory chain to the oxidative burst consequent to pathogen recognition at the cell surface [75-76]. Partial or complete coding sequences of *M. galloprovincialis* super oxide dismutase (SOD), catalase, glutathione transferase, peroxisomal thiolase

and polyamine oxidase have been reported [77-80]. In Mytibase, numerous MGCs putatively identify enzymes such as amine oxidases, dehydrogenases, peroxidases, mitochondrial oxidases and reductases. In addition to SOD (IPR001424, IPR001189, 4 MGCs) and glutathione peroxidases (IPR000889, 5 MGCs) many mussel sequences are featured by the thioredoxin fold/domain (IPR012336, IPR013766), typical of proteins regulating the redox state of cellular thiol groups such as the thioredoxin-like reductases (MGC09082).

Interestingly, more than 30 MGCs indicate heat shock proteins of different sizes (HSP 20, 40, 70 and 90 Kd) and related binding factors (MGC06041, MGC03865, MGC04512), mostly known to be modulated following immunostimulation [81-83].

Transcripts identifying proteases, protease inhibitors and proteasome components

Proteases of various subfamilies and related inhibitors are essential in organism growth and development. Proteolytic reactions typically occur in the complement, coagulation and ProPO cascades, during apoptotic cell death, antimicrobial peptide synthesis and degradation of pathogen components within the lysosomal, cytosolic and extracellular compartments.

For instance, the insect clip-domain SP can act as cofactor or negatively regulate the melanization response, with a repertoire of 45 and 68 genes in *Drosophila melanogaster* and *Aedes aegypti*, respectively [5]. Cleavage of viral and host factors operated by granule-associated SP (granzymes) slows down viral replication and induces the apoptotic elimination of infected mammalian cells [70]. Caspases of the cysteine protease family also act in the proteolytic cascade of the apoptosis and, *via* NFkB signalling, regulate inflammatory responses in *Drosophila* [84].

Specific enzyme inhibitors are expected to modulate the same biological processes but also inhibit pathogen growth and invasive behaviour. In fact, trypsin and chymotrypsin inhibitor levels correlate with the plant resistance to pathogens, and in the basal metazoan *Hydra magnipapillata* the bactericidal activity of a kazal-type SP inhibitor possibly compensates the absence of migratory phagocytic cells [85-86].

In Mytibase, as much as 57 and 14 domains denote proteases/proteinases/peptidases and their inhibitors, respectively. Many MGCs indicate inherently secreted serine-type endopeptidases of the chymotrypsin/Hap family (IPR001254, 18 MGCs), SP inhibitors with Kazal-like repeats (IPR002350, 14 MGCs) or BIR repeats (IPR001370, 12 MGCs), with the latter belonging to the Inhibitor of Apoptosis (IAP) family [84]. Other MGCs point to cysteine caspase-like peptidases (IPR002398, 6 MGCs), astacin-like zinc metallopeptidases (IPR001506, 8 MGCs) and related inhibitors (I2, I8, I14/I15

hirudin/antistatin, I19, I29, I32, serpin and Tissue Inhibitors of MetalloProteases among others).

More than 60 MGCs denote ubiquitin, ubiquitin-related and proteasome-related components and give emphasis to intracellular processes oriented towards the pathogen elimination [87].

Transcripts identifying lysozyme

Recognized in 1922 as an antibacterial molecule and abundant in various animal secretions, lyzozyme hydrolyzes 1,4-beta-linkages in peptidoglycan and chitodextrin structures. In flies and other invertebrates, lysozyme expression and activity increase after exposure to bacteria, and the species-specific gene number partly depends on the use of bacteria as food resource [5]. Up-regulation of the mussel lysozyme, with increased percentage of hemocytes expressing lysozyme mRNA, was observed at 2-3 days post-injection of *Vibrio anguillarum* or *Micrococcus lysodeikticus* [88] whereas maximum expression occurred after 3 hours in hemocytes immunostimulated *in vitro* [75]. In Mytibase, the 8 MGCs denoting lysozymes can mainly be classified in types C and G: among them, MGC02986 is similar to a C-type lysozyme described in insects but not yet reported in molluscs.

Definition and validation of a *M. galloprovincialis* Immunochip

Owing to the continuous growth of the GenBank/UniProtKB/SwissProt databases, recurrent similarity searches and manual validation of the emerging similarities guided the progressive selection of 1,820 MGCs to be validated as components of the mussel immunome. Probes of 54-57 nucleotide length (plus 652 unrelated human probes) have been designed using the 3' end transcript region templates and spotted in four replicates to prepare a new DNA microarray platform, namely a *M. galloprovincialis* Immunochip.

Taking advantage of a large immunostimulation trial conducted *in vivo* on mussels from three different European regions [89] we selected and processed hemolymph samples collected at 3 and 48 hours after the injection of 10 million exponentially growing *Vibrio splendidus* cells into the adductor muscle. Total RNA was purified from two hemolymph pools (N=10) per time point, and from paired saline-injected control mussels sampled at 3 and 48 h (N=40, one unique reference pool). As the amplified Cy3/Cy5-labeled targets were competitively dye-swap tested on the mussel Immunochip, the reciprocal hybridizations of a target pair on quadruplicated probes yielded 8 fluorescence signals per probe (16 values per time point).

Looking at the total hybridization data set, 21.8% of the mussel probes gave significant

fluorescence (median fluorescence values averaged per probe resulting above the background, i.e. average BG*2St.Dev.) with a range of 13.5-27.7% per individual array and average values of 17.2% and 26.4 % lighted spots at 3 and 48 hours, respectively. These percentages reasonably relate to the number of differentially expressed genes (143 and 262, respectively) estimated by permutation from the absolute level and standard deviation of the replicates (SAM analysis, Table 3 and Additional files 2, 3 and 4). Soon after the immune stimulation, the over-expressed genes are consistently more numerous than the under-expressed (106 vs. 37 at 3 h), whereas later in time their proportion roughly equals (124 and 138 at 48 h). Converting the $\log_2$ (test/control) values by the relative fold change of expression, they range over two orders of magnitudes from +7.3 to -8.9 (at 3 h) and from + 7.6 to -9.6 (at 48 h).

Table 3 - Total number of differentially expressed genes (over↑ under↓) in hemocytes sampled at 3 and 48 hours from mussels injected with live *V. splendidus*

Sampling time (hours post-injection)	Number of differentially expressed genes	FDR	Δ
3	143 (106↑ 37↓)	1.30%	0.457
48	262 (124↑ 138↓)	1.20%	0.534

False discovery rate (FDR) and delta value are also indicated (one class SAM, 16 replicates per time point).

Hierarchical clustering of the Immunochip profiles clearly shows the resemblance between biological replicates (3.1 - 3.2 and 48.1 - 48.2), greater differences between the time points of 3 h and 48 h, and interesting discriminant signals such as those related to LITAF and IAP-like apoptosis inhibitors (Figure 4).

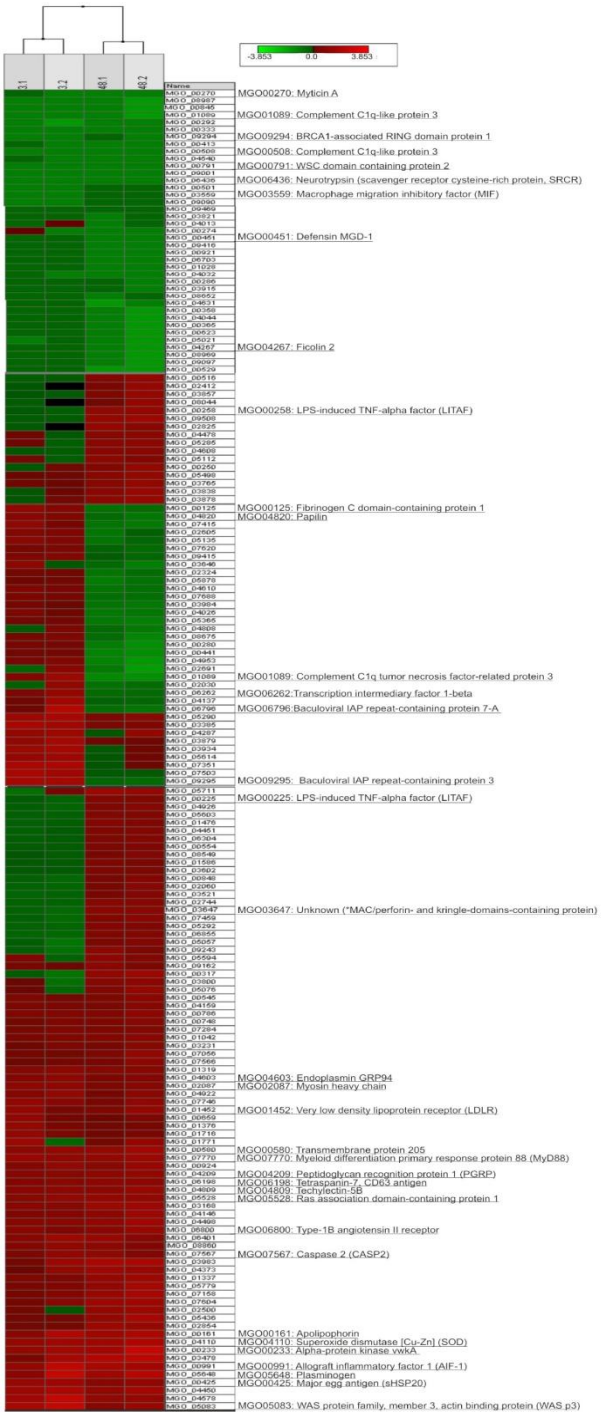
The general AMP down-regulation detected in the hemocytes of *Vibrio*-injected mussels (for instance 4.7-6 fold down-regulation of defensin MGD1b at 3-48 h after challenge) confirms previous qPCR data [89]. Similarly, putative acute phase response proteins (apextrin) and the macrophage Migration Inhibitory Factor (MIF cytokine) were under-expressed. Conversely, probes pointing to Allograft inflammatory factor 1, SOD, small HSP20, plasminogen as well as various recognition receptors and molecules supporting intracellular signalling (e.g. MyD88) or cytoskeleton remodelling/motility (e.g. actin, myosin) were commonly up-regulated (Additional file 4). Compared to the early response, after 2 days we detected a significant expression of proteases and protease inhibitors, LITAF (TNF signalling) and sequences suggesting various cell functions (Additional file

3).

In general, no consistent trends could be defined for the C1q-like and lectin-like molecules. Due to their abundance and high sequence diversity, further study is necessary to understand their constitutive and PAMP-induced expression in mussel hemocytes.

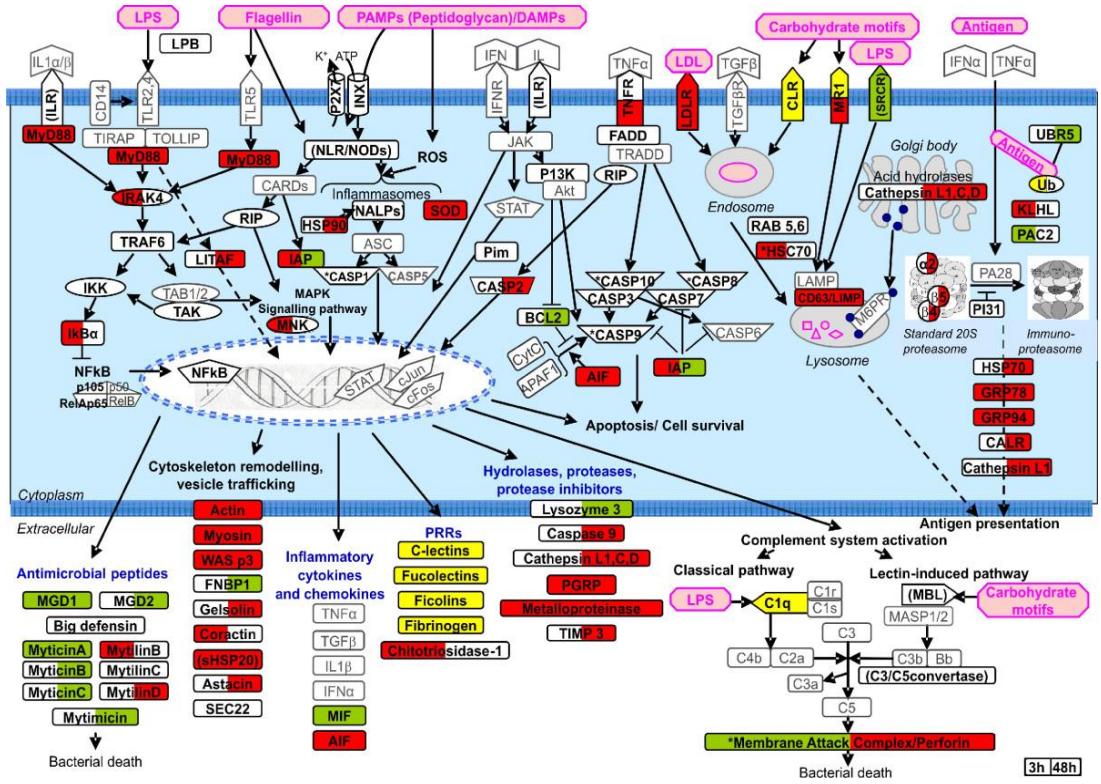
Based on the Immunochip hybridization data, the molecular pathways and gene functions mapping out the mussel hemocyte response to the *Vibrio* injection are modelled in Figure 5. Functionally similar to dendritic cells or macrophages, the mussel hemocytes display a pleiotropic response to the bacterial attack. Interacting with bacterial PAMPs, versatile and redundant recognition receptors undergo conformational changes, oligomerization or clustering. The subsequent activation of cross-talking signal transduction pathways adjusts the cell biochemical machinery towards the expression of specific gene sets and key effector molecules (in bold, on coloured field). Pathogen-induced oxidative burst and damage-associated molecular patterns (DAMPs) also sustain the inflammasome activation and intracellular signalling. Eventually, the endolysosomal and proteasome systems, secretory pathways and whole cell behaviour are recruited to achieve the pathogen killing.

Figure 4 - Hierarchical clustering of the Immunochip profiles referring to 3 and 48 h post-injection of live *V. splendidus*



Two mussel groups with 8 replicates per time point have been analyzed (R-package software). Scale of the expression values and probe ID are indicated. For space reasons, only instructive parts of the clustered data are reported.

Figure 5 – Modelling the putative transcriptional hemocyte response to the Vibrio injection.



The model outlines PRR, interrelated signalling pathways and cell processes significantly modulated at 3 and 48 h post-treatment. Various bacterial PAMPs and DAMPs are indicated in pink. The gene functions represented on the Immunochip are indicated in bold according to sequence similarities/identities: framed bold characters on red, green and yellow fields indicate over-expression, under-expression and heterogeneous trends, respectively, of specific Immunochip probes at 3 h (left half) and 48 h (right half). Other gene functions not represented in Mytibase have been introduced to recall specific molecular pathways. Similarities resulting from InterproScan Analysis are reported in brackets whereas an asterisk indicates annotations based on manual inspection of relevant similarities. All the acronyms with related extended names are reported in the list of abbreviation.

Overview of the mussel response to live *Vibrio splendidus*

The most ancient defences of the living organisms are based on neuropeptide and protein hormone receptors, receptor kinases and PRR (mainly TLRs, lectin-like, NLRs and RIG-like families) able to signal the danger and increase the expression of various inflammatory and effector molecules [90-92]. In view of the most recently sequenced invertebrate genomes, the pleiotropic innate immune responses could be described as a coordinated system of elements rapidly evolving and expanding the ability for pathogen sensing/targeting, and evolutionary conserved regulatory factors which finely adjust basic

cell processes and direct the development and performance of the immune cells [5, 51, 93].

Ancient signalling pathways like those of MAPKs and NFkB are not exclusive of the immune responses and, not solved by standard sequence searching, the identification of invertebrate interleukine homologues (or molecules with functions analogous to the specialized signalling provided in vertebrates by interleukins and cytokines) makes new exploratory approaches necessary [9, 74, 94].

Although the hemolymph cells are fundamental in the mussel immunity, it is not clear if cells other than hemocytes contribute to the complex spectrum of rapid innate responses to potential pathogens. Consistent with a more general view linking immunity to metabolism and other body processes, typical immune genes and proteins should also be expressed in 'non-immune' cells, tissues and organs [2, 41, 95-97]. For instance, the expression of C1q/TNF-like molecules (MGC0284) has been detected in various tissues, with hemocytes showing the greatest levels, and throughout the development of *M. galloprovincialis* [49].

Similar to cells of the vertebrate monocyte/macrophage lineage, PAMP-activated immunocytes achieve pathogen elimination essentially through chemotaxis, phagocytosis, and cytotoxic processes [30]. In the Mediterranean mussel, agranular hemocytes are cells able to divide as they show replication-dependent chromosomal damage [98] whereas the heterogeneous and abundant granulocytes can be regarded as differentiated cells, mostly phagocytic and able to release antimicrobial peptides [41]. Accordingly, distinct hemocyte subpopulations appear to respond to potential pathogens with specific patterns of gene expression [88, 99].

In addition to the host response, pathogen-related and physico-chemical factors are other main determinants of disease onset and mortality in aquacultured bivalves. The survival and niche occupation of *Vibrio* cells in changeable habitats (including the infected hosts) depend on the overall nutritional versatility of these bacteria, chemico-physical conditions for growth but also on the expression of hemagglutinins or lectins mediating the interaction with host cells and active secretions able to inhibit or disrupt the host defence reactions such as proteases, pore-forming hemolysins, ciliostatic and hemocyte-killer toxins [100-102]. As suggested for *V. harvey*, the modulation of signalling pathways essential to the antimicrobial immune response is an additional way to attack and escape the host response [103].

Testing the Immunochip performance with hemocytes sampled at 3 and 48 h from *Vibrio*-injected mussels revealed a general AMP downregulation, possibly related to the toxicity of live bacteria and contrasting the enhanced response to the stimulus obtained with

heat-killed bacteria [43, 75]. According to quantitative real time PCR assays performed on the hemolymph cells, the injection of control mussels with saline solution did not affect the expression of immune-relevant genes, namely mytilin B, myticin B, defensin, lysozyme and HSP70 [89]. The increase in transcriptional changes from 3 to 48 h and the slight prevalence of down-regulation signals at 48 h in the hemocytes of mussels injected with 10 million potentially infective *V.splendidus* cells mark an incoming functional decline. Indeed, a not negligible fraction of the *Vibrio*-injected mussels (22%) showed very slow or unapparent reactivity at 48 h (hence, considered close to death and discarded) whereas no mortality was observed at 3 h or in the control mussels injected with the saline solution only. As Spanish and French mussels injected in parallel with equal doses of live *Vibrio* cells in their respective locations did not show signs of distress [89] we suppose that season-related life history factors may underlie the overall reaction of these mussels to the injected bacteria. The delayed (48 h) over-expression of a number of proteases and stress proteins supports the functional hypothesis. Timing and complexity of the mussel immune response as well as the immunostimulation protocol (following shell notching, small mantle wounds may have attracted the hemocytes and induced their degranulation far from the adductor muscle) could also explain the progressive AMP down-regulation observed in the hemocytes of the *Vibrio*-challenged mussels. The HSPs (HSPs 20.6, 70 and 90 in particular) showed instead opposing expression trends with only a couple of probes for small HSPs down-regulated at 48 h post-challenge. These stress-inducible protein chaperons probably support pro-survival pathways but their multiple roles and complex expression patterns suggest further study [99, 104]. In the same hemocyte samples, lectin-like and fibrinogen-like adhesion/recognition molecules showed heterogeneous expression trends whereas the frequent up-regulation of mussel genes relating to the cell shape and motility points to chemotactic and phagocytic hemocyte behaviour. The enhanced expression of LITAF and persistent MIF down-regulation in response to the injected bacteria encourage us to search regulatory mussel monokines with new immunostimulation trials and approaches other than DNA microarray testing.

The samples tested on the Immunochip exemplify only two temporal stages of the multi-step response to a reference dose of live *V. splendidus* cells. The observed transcriptional changes apparently mark the hemocyte activity against the *Vibrio* cells with a mounting inflammatory response (3 h) and a shift towards a more general stress condition (48 h). A previous equal treatment of *M. galloprovincialis* with live *V. splendidus*, caused a dramatic increase in living intra-hemocyte bacteria in less than an hour, suggestive of intense phagocytosis, and a subsequent gradual decrease with only a few viable bacteria at 24 h post-injection [105]. Recruited against active bacteria, the total

counts of three distinct hemolymph cells (ialinocytes, small and large granulocytes) almost halved at 3 h post-injection and, after 48 h were still below the normal levels. Full understanding of the complex and dynamic response of *M. galloprovincialis* to the bacterial attack requires further study.

The great number of deep sea vent mussel transcripts made available during manuscript submission [106] and the launch of a new InterProScan Sequence Search interface (http://www.ebi.ac.uk/Tools/pfa/iprscan) will probably speed up the cross-species identification and validation of immune-related genes of marine bivalves. A partial comparison between Mytibase and the DeepSeaVent database (http://transcriptomics.biocant.pt:8080/deepSeaVent) rescued 5,261 annotated protein sequences expressed in both *M. galloprovincialis* and *Bathymodiolus azoricus* [106]. New BLASTN queries performed with the MGC transcript sequences significantly modulated at 3 and 48 h in the *Vibrio*-injected mussels against the 75,407 transcript sequences of *Bathymodiolus azoricus* confirmed the robustness of the Mytibase annotations (see Additional files 2 and 3). Such similarity searches also ascertained a not negligible sequence diversity of putative homologues (only 75 to 91% identical nucleotide positions in *M. galloprovincialis* and *B. azoricus* transcripts coupled by BLASTN e-value equal to 0.0) and the absence of typical AMP (the e-values obtained by querying the Mytibase AMP ranged from 0.97 to 5e-04). These findings, as well as previous comparative analysis of large EST sets from *M. californianus* and *M. galloprovincialis* [7, 37], support the use of species–specific DNA microarrays.

Conclusions

The great molecular diversification of pathogen binding molecules such as the insect Down syndrome cell adhesion molecule [107], snail FREPs [60], sea urchin TLRs [68] as well as the individual variant patterns reported for sea urchin 185/333 molecules [108] and mussel myticins [43, 109] emphasize the emerging complexity and divergent evolution of the invertebrate immune systems. Filter-feeding bivalves such as the *Mytilus* species commonly interact with a 'sea' of microscopic living forms, and can reveal interesting adaptations to co-evolving invaders and environmental changes. As many proteins involved in the immune responses also participate in basic cell processes, evolutionary adaptations differ between and within taxa and the *Mytilus* genomes are not yet available, the use of species-specific DNA microarrays represent a rational choice for studying transcriptional profiles and co-expression landscapes, and to validate many immune-related candidate molecules.

In fact, Mytibase includes almost all the domains featuring the innate PRR, i.e. C-type lectin and Ig-like domains, LRRs (and pyrin) domain, nucleotide-binding and Toll-Interleukin receptor domains, caspase recruitment and helicase domains [110], and reports abundance and diversity of the C1q/TNF-like, lectin-like and AMP mussel transcripts. Using the protein domains as instructive identifiers of sequence homology and other bioinformatics tools, we have designed 1,820 immune-candidate probes, organized them into a *M. galloprovincialis* Immunochip and tested this new DNA microarray with haemolymph samples exemplifying the early and late response to live *V. splendidus* cells. From one fifth to one fourth of the ImmunoChip probes gave significant fluorescence signals, respectively, and indicated both the modulation of various cell processes and a very specialized hemocyte transcriptome. Accordingly, the Immunochip could be confidently used to expand the validation of candidate probes on hemocytes and also in other mussel tissues. The putative relational map resulting from the Immunochip data certainly requires further study. In the meantime, a good number of Mytibase sequences relevant to the mussel immunity such as for instance the fibrinogen-like peptides are the object of new studies [111, 112].

Methods

Identification of immune related mussel sequences in Mytibase

A multiple search strategy guided the extraction of putative immune-related sequences from Mytibase, the mussel transcript database [37]. We used 2,915 Gene Ontology (GO) sequences associated with UniProt Knowledgebase (UniProtKB) below the node GO:0002376-Immune system processes [113] and 4,216 sequences downloaded from the multispecies ImmunomeBase [2] to seek related mussel transcripts by tBLASTn similarity search (cut-off 1.0E-4). A working list of 1,233 keywords relating to mussels and innate immunity also supported the extraction of MytiBase sequences. Finally, BLAST similarities, gene ontologies and protein features reported in Mytibase were manually screened to confirm the core set of immune-related mussel transcripts.

Descriptive analysis of selected sequence clusters

Selected immune sequence groups, mainly identified in Mytibase by textual search of Interpro domains [114] and/or BLAST similarity searches [115] were evaluated in more detail. The raw sequence traces identifying AMP and those containing the molecular signature of C-type lectin (IPR001304) and C1q (TNF-like) (IPR001073) were manually cleaned to perform multiple sequence alignment and compute phylogenetic trees by the Neighbour Joining with Bootstrap test. To multialign and validate the identification of AMP precursors and C1q domain containing sequences, we used different editors: Muscle [116], BioLign/BioEdit [117] and Jalview [118]. The C1q signature was confirmed by sequence homology search based on profile hidden Markov models (HMMER3) [119] whereas SignalP was used for prediction of signal peptide cleavage sites [120].

Probe design and Immunochip preparation

One thousand and 820 oligonucleotide probes were designed with OligoArray 2.1 [121] on the selected MGCs according to the following requirements: 56.7 average length (54-57 range), 300 bases of distance between the oligo 5' end and transcript 3' end, 10-80% CG content, 70-92°C melting temperature with 65°C and 60°C as thresholds for cross-hybridization and hairpin formation, respectively. Additional 38 oligonucleotides (56-mers) with no virtual hybridization against the whole mussel EST collection were similarly designed using unrelated human sequences as templates. The designed probes were custom synthesized (Europe Services Invitrogen), arranged and deposited (BioRobotics MicroGrid II, Digilab, Genomic Solutions) on derivatized glass slides (MICROMAX Glass Slides: SuperChipTM I, Perkin-Elmer) at 50% relative humidity. The resulting species-

specific Immunochip includes two equal arrays, each one organized in 16 subarrays and containing 4x1,820 mussel probes, 652 unrelated probes in multiple replicates and 112 alignment spots. Probe fixation on the slide was performed by UV cross linker (Stratagene) at a total power of 300 mJ. Slides were rinsed once in 1% SDS, 3× SSC for 1 min at room temperature, twice in distilled water for 5 min at room temperature, dried in laminar flux chamber and stored at room temperature under vacuum.

Mussel challenge with *Vibrio splendidus*

Native mussels of commercial size (6.8±0.7 cm lenght) from one outlet of the Venice lagoon (Italy, water temperature of 24-25°C) were acclimatized for one week in sea water collected at flood tide (32 psu, 22°C) and fed with *Isochrisis galbana*. Following careful shell notching, 0.1 ml of exponentially growing bacteria (10^7 Colony Forming Units of *V. splendidus* LGP32 in Trypsin-Casein-Soy medium) were injected into the posterior adductor muscle. One ml of hemolymph was withdrawn from individual mussels at 3 and 48 h post-injection and 10 hemolymph/group were pooled [89]. Hemolymph samples were similarly collected from paired control mussels injected with NaCl-enriched PBS (PBS-NaCl). Following centrifugation at 800xg, 4°C for 15 min, the pelleted hemocytes were re-suspended in 1 ml Trizol reagent (Invitrogen) and immediately stored at -80°C. Basically, two biological replicates per time point (3.1 and 3.2; 48.1 and 48.2) each one representing 10 individual hemolymphs were processed for hybridisation on the Immunochip in dye-swap combinations with a unique reference composed by all the hemolymphs sampled in parallel at 3 and 48 h from the control mussels (N=40).

RNA sample processing and microarray analysis

Total RNA from pooled hemolymph of treated and control mussels was extracted and additionally purified with high molar LiCl. RNA concentration and quality were ascertained by using the NanoDrop® ND-1000UV spectrophotometer and Agilent 2100 Bioanalyzer (microcapillary electrophoresis on RNA 6000 Nano LabChips, Agilent Technologies). Equal amounts of 4 pooled hemolymph samples, representing 40 mussels injected with PBS-NaCl, were mixed to define one unique reference sample to be competitively hybridized on the Immunochip.

Hemolymph mRNA was linearly amplified from total RNA with the Message-Amp™ II aRNA Amplification kit (Ambion): 5-(3-aminoallyl)-UTP modified nucleotides were incorporated into the aRNA during the *in vitro* transcription reaction, then mono-functional NHS-esters of Cy3 or -Cy5 dyes (CyDye Post-Labeling Reactive Dye Pack, Amersham GE Healthcare) were resuspended in DMSO and covalently coupled to the aminoallyl-

aRNA probes for 1 h at room temperature in the dark [122]. Following purification (Gene Elute PCR Clean-up kit, Sigma-Aldrich) and UV-quantification, 500 ng of both reference and test aaRNAs were combined and ethanol-precipitated. Cy3/Cy5-coupled samples were re-suspended in 18 µl of hybridization buffer (5x SSC, 50% formamide, 0.1% SDS), denaturated for 3 min at 70°C and competitively hybridised to the Immunochip for 24 h at 48°C in humidified dual-slide chamber (HybChamber, GeneMachines). Slides were first conditioned for 12 h at 48°C in a solution of 5x SSC, 100 ng/µl salmon sperm ssDNA, 5x Denhardt's solution and 0.1% SDS). Reference and test samples were then simultaneously hybridised in dye-swap crossed combinations on the 2 identical arrays of the same slide. The slides were sequentially washed at room temperature with mild shaking in buffer: 1x SSC, 0.2% SDS; 0.1x SSC, 0.2% SDS; 0.2x SSC (4 min each) and 0.1x SSC (3 min), with final drying by air flow.

Microarray data analysis

Immunochip fluorescence signals were scanned using two lasers (633 nm and 543 nm) at 5 µm resolution with a GSI Lumonics LITE dual confocal laser scanner. Image processing and signal quantification were performed with the software ScanArray Express® (PerkinElmer). Normalisation of the fluorescence signals was performed by using the total and LOWESS (Logfit) algorithm with MIDAS (MIcroarray Data Analysis System, http://www.tigr.org/software) [123]. The log2 test/reference ratio of all the normalised fluorescence values was computed and the genes differentially expressed in the test sample *versus* control sample were identified by means of the Significance Analysis of Microarrays available from the Stanford University, CA (SAM software package v3.0, One-class analysis with 200 minimal permutations and FDR<2%) [124-125]. Similarities among the Immunochip profiles were assessed by hierarchical clustering of the Pearson correlation similarity matrix (J-Express v2.1) [126].

Mytibase home page and GEO accession numbers

The mussel knowledgebase is available following registration at http://mussel.cribi.unipd.it. The *Mytilus galloprovincialis* Immunochip (GPL10758) and related expression data (GSM575753, GSM575790, GSM577075-80) have been recorded at http://www.ncbi.nlm.nih.gov/geo.

Authors' Contributions

LV and UR performed RNA purification and processing, microarray experiments and data analysis, LV also supported the overall interpretation. FB (former Mytibase curator) designed the oligonucleotides and performed the AMP analysis. GL assured the equipment and supervised the microarray work (http://microcribi.cribi.unipd.it). BC and CM prepared the Immunochip slides using custom-synthesized oligonucleotides. AP guided the Mytibase annotations, carried out specific sequence analysis and supported manuscript development. FR, AF and BN provided valuable inputs for Immunochip definition and text improvements (AF partially screened the 1820 Mytibase transcripts). PV performed the mussel treatment, evaluated the Mytibase sequences for the Immunochip definition, drafted and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors would like to thank C. Ortutay, K.Rannikko and M. Vihinen for making available the sequences indexed in the Immunomebase (http://bioinf.uta.fi) and Marco Gerdol for a preliminary evaluation of the C1q/TNF-like MGCs. This work is supported by the European Integrated Project FOOD-CT-2005-007103 (http://imaquanim.dfvf.dk/info) and in part by Regione Friuli Venezia Giulia, Direzione Centrale Risorse Agricole, Naturali, Forestali e Montagna, L.R. 26/2005 prot. RAF/9/7.15/47174.

INVERTEBRATE SURVIVAL JOURNAL

REVIEW, 2011

How gene expression profiles disclose vital processes and immune responses in *Mytilus spp.*

S Domeneghetti[1*], C Manfrin[2*], L Varotto[1], U Rosani[1], M Gerdol[2], G De Moro[2], A Pallavicini[2], P Venier[1]

[1]Department of Biology, University of Padua, Padua, Italy
[2]Department of Life Sciences, University of Trieste, Trieste, Italy
*Equal contribution

Abstract

Gene expression studies largely support the understanding of gene-environment interactions in humans and other living organisms but the lack of genomic and genetic information often complicates the analysis of functional responses in non-traditional model species. Nevertheless, the fast advancement of DNA microarray and sequencing technologies now makes global gene expression analysis possible in virtually any species of interest. As regards the *Mytilus* genus, tens of thousands Expressed Sequence Tags (ESTs) are currently available for *M. californianus* and *M. galloprovincialis*, and DNA microarrays have been developed. Among them, Immunochip 1.0 specifically includes 1,820 probes of genes centrally involved or modulated in the innate immune responses of the Mediterranean mussel. This review recalls peculiarities and applications of the existing mussel DNA microarrays and finally summarizes facts concerning a variety of transcript sequences likely involved in the mussel immunity. Beside DNA microarrays, Next Generation Sequencing (NGS) technologies now offer new and broader research perspectives, from the whole transcriptome coverage to the *Mytilus* genome sequencing.

Key Words: *Mytilus*; DNA microarray; innate immunity; ESTs; antimicrobial peptides; C1q

Introduction

Global gene expression analyses in organisms selected to represent a given ecosystem currently support ecotoxicological investigations and create a conceptual bridge between the early organism responses and late population changes (Steinberg *et al*., 2008). The animal response to a variety of detrimental conditions usually starts with alarm signals followed by adjustment reactions aimed to neutralize the physiological unbalance, and may end up in a general decline of vital processes ultimately marked by disease and death. Depending on the stress type and exposure intensity, the expression of definite sets of genes makes available specific proteins and other molecules in cells and tissues.

Appeared in the 1990s, the DNA microarray technology enables the simultaneous expression measure of thousands of genes represented in the microarray platform by unambiguous polynucleotide probes (Schena *et al*., 1995; Lockhart *et al*., 1996). The gene expression profiles emerging from suitable sampled cells or tissues can provide a dynamic view of biological processes and allow the correct sorting of different functional states. Based on the availability of sequence data, DNA microarrays can be used to solve a variety of biological questions: from the identification of molecular markers pathognomonic of disease and transcriptional signatures of various stress factors to the understanding of complex phenomena such as the epigenome in normality and disease (Martín-Subero and Esteller, 2011).

Specific microarray platforms and advanced deep sequencing technologies now support studies on the cellular functions of microRNAs and their role in human diseases (Thomas *et al*., 2010). Leading research institutions are currently using both the mRNA and miRNA expression profiling to examine the genomic responses to environmental stresses (NCT). Central to the toxicogenomics studies is the concept of '*phenotypic anchoring*' which recalls the importance to correlate the observed gene expression changes to adverse effects defined by conventional parameters of toxicity and pathology.

In the controlled vocabulary of the Natl. Library of Medicine, the term 'DNA microarray' is indexed under the following category which indicates the large application range of such innovative technology (MESH):

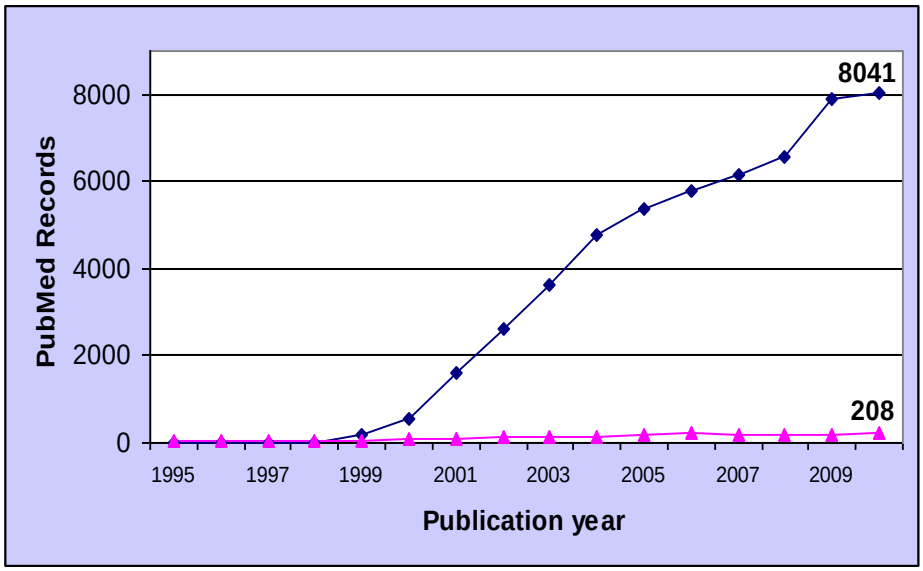
Oligonucleotide Array Sequence Analysis - the hybridization of a nucleic acid sample to a very large set of oligonucleotide probes, which are attached to a solid support, to determine sequence or to detect variations in a gene sequence or expression or for gene mapping.

Relevant to the gene expression profiling research area is Gene Expression Omnibus, a public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomics data submitted by

the scientific community (GEO). To fulfil the current standards (Minimum Information About a Microarray Experiment) the contents submitted to GEO should include the following: raw hybridization data; normalized data from which the main experimental findings can be outlined; description of the tested samples and whole experimental design, with details on the biological and technical replicates; identity and location of all probes and controls of the microarray platform, with external reference in the case of commercial arrays; concise but precise description of laboratory and data processing protocols related to the experiment under submission. According to the aims of the Microarray Gene Expression Data Society, dating back to the late '90s, the compliance to the MIAME standards should assure the data comparability among different platforms and testing protocols while supporting common work criteria and the reduction of random data variation (Rogers and Cambrosio, 2007). Based on the comparative data analysis, the guidelines for standardization and reporting have been further refined (Chen *et al*., 2007; Shi *et al*., 2008). At present, GEO contains as much as 9,000 platform records which can be accessed and browsed in full detail.

Figure 1 illustrates the annual increase of PubMed records including the term "DNA-microarray" or "*Mytilus*" (subject heading or title/abstract) and suggests that pioneering technologies open the way to new ideas more than an unconventional model organisms.

Figure 1 - Number of PubMed publications including the terms "DNA-microarray" (blue line) or "Mytilus" (purple line) from 1995 to 2010. Has the DNA microarray revolution reached its peak?



In fact, the gene expression profiling field has substantially diversified: specialized equipments and various related software make today the DNA microarrays powerful tools for the study of gene sequence, structure and expression, particularly for the best known model organisms. Nonetheless, one must remember that transcription is just one step in gene expression, and post transcriptional events referred to maturation of the primary transcript, RNA editing and RNA silencing as well as various modifications of the translation products overall influence the final amounts and activity of cellular proteins.

Mytilus DNA microarrays: preparation strategy and applications

Six GEO records refer to mussel DNA microarrays at July 2011.

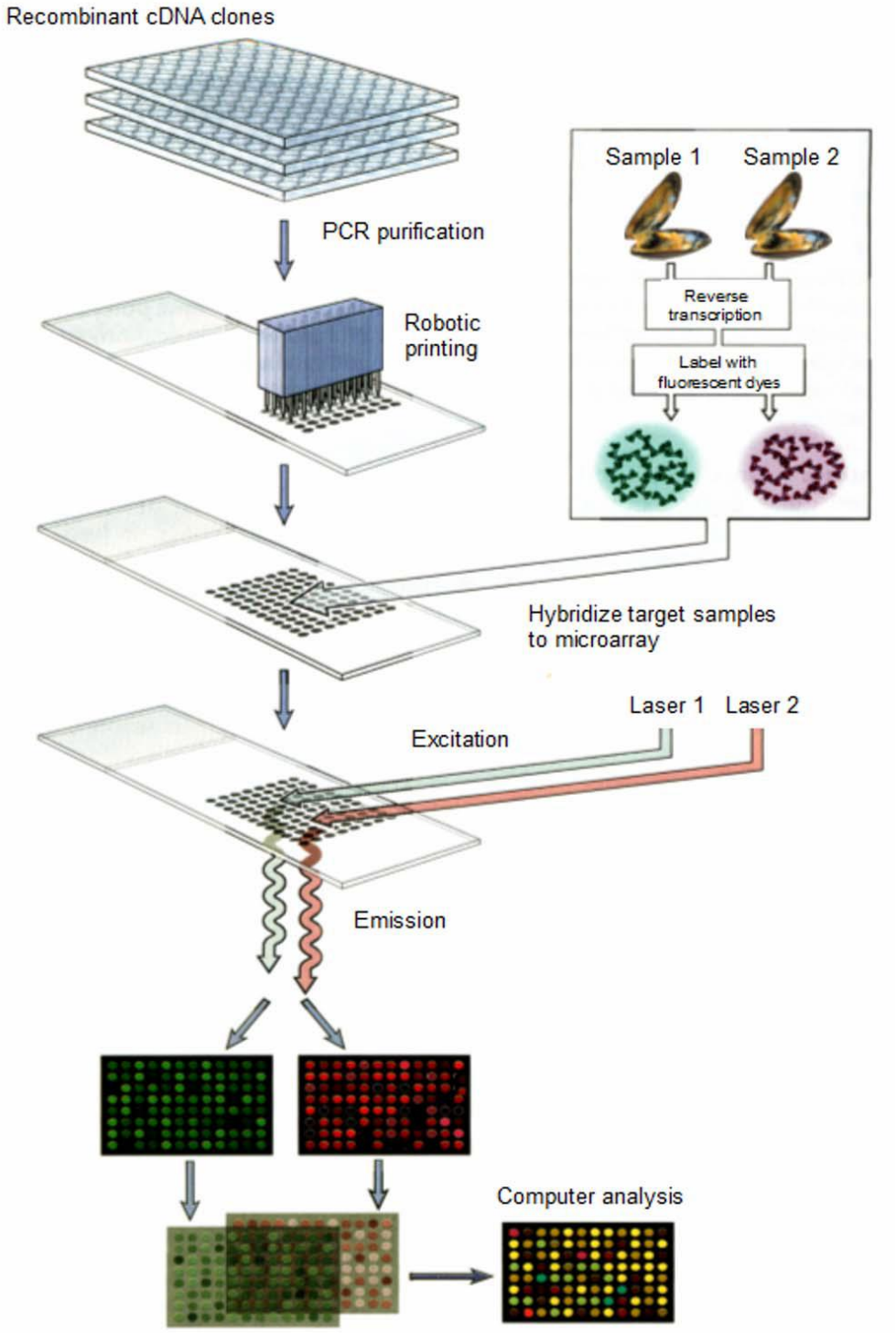
MytArray 1.0 (GEO platform GPL1799, Oct 2006) is composed by 1,712 cDNA probes, univocally tagging the 3'-end region of transcripts from the main tissues of adult mussels (*Mytilus galloprovincialis*) and 46 unrelated cDNA control probes, all printed in duplicate and twice per slide (1.7 k mussel probes per array, 7.0 k total probes per slide). The probes were designed in the 3'-UTR, one among the least conserved gene regions, so that competition of different mRNAs from genes with similar coding sequence and cross-hybridization to the same microarray probe should be minimal. Also, the probe size of 400 - 800 bp is expected to ensure comparable efficiency in the amplification and spotting of the cDNA inserts as well as uniform hybridization kinetics (Venier *et al*., 2006).

MytArray 1.0 was first used to investigate the specificity of gene transcription in mussel tissues with different functional role and the transcriptional profiles of mussels treated with chemical mixtures or living wild in different sites of the Venice lagoon (Venier *et al.*, 2006; GEO series GSE2176, GSE2183 and GSE2184). Sample pairs combined according to dye-swap labelling (reference and test samples labelled with Cy3/Cy5 cyanine dyes in alternate combinations) were competitively hybridized on the two equal arrays of cDNAs spotted on the same slide (Figure 2).

Gills, digestive gland, tissues involved in contraction/motility (foot, adductor muscles, ligaments) and reproduction (gonads and mantle) displayed specific transcriptional footprints, as expected. The results obtained in mussels treated with mixtures of inorganic metal salts or persistent organic chemicals guided the interpretation of the gene expression profiles of mussels living in the inner industrial canals or at the lagoon border open to the sea (this exercise yielded a provisional list of contamination marker probes).

In this study, the evident transcriptional down-regulation detected in the reproductive tissues was consistent with the depleted status of the mussel gonads whereas the greatest variety and abundance of transcripts was found in the digestive gland.

Figure 2 - Work diagramm referred to the competitive hybridization of two dye-swap-labelled samples on a cDNA microarray with two-channel detection of the fluorescence signals (modified from Gibson and Muse, 2004).



Additional analysis of these expression data is reported elsewhere (Pantzartzi *et al*., 2010).

The same platform was then used to evaluate in a time-course study the gene expression changes in the digestive gland of mussels exposed to okadaic acid (OA) via food contamination for five weeks (Manfrin *et al*., 2010; GEO series GSE14885). One relevant purpose of the study was the identification of molecular biomarkers which could enable an easy and rapid detection of the Diarrhoeic Shellfish Poisoning biotoxins in marketable mussel stocks, *i.e.*, novel reliable assays complementing the existing diagnostic methods. An unsaturated loop design, combining control and treated samples with different dye-labelling for the competitive hybridization on Mytarray 1.0, was adopted to take into account all the time points and the biological replicates, with some combinations only inferred (Kerr and Churchill, 2001). A considerable number of transcriptional changes was detected in the OA-exposed mussels, with a prevalence of up-regulated probes at 3 days and a subsequent progressive increase of down-regulated probes (from 58 % over-expressed to 76 % under-expressed genes, respectively detected at day 3 and day 35). The biphasic time-related trend of response observed in this study recalls the changes occurring in the mussel digestive gland along different phases of the mussel reaction to the experimental stimulus, from the early acute response to the late overall unbalance of the functional processes. Many candidate markers are now under study to evaluate their predictive value in the diagnosis of biotoxin-contaminated mussels.

MytArray 1.1 (GPL102699, March 2010) contains the same cDNA probes of MytArray 1.0 in a slightly modified platform geometry. It has been used to study the gene expression profiles of *M. galloprovincialis* with monthly samplings for one year, hence taking into account seasonal differences which are known to influence metabolism rates and gonad development among other vital functions (Banni *et al*., 2011; GEO series GSE22915, GSE23049- GSE23051). Mussels were collected from an anthropized and industrialized lagoon of the Southern Mediterranean Sea (Ben Said *et al*., 2009) and competitive hybridizations were performed with dye-swap-labelled samples (dual colour analysis). Following a loop design with 3-4 biological replicates and parallel histological evaluation of the gonad status, the authors could analyze the transcriptional profiles of digestive gland tissue of female mussels collected during 12 months, and those of digestive gland and mantle tissues from male and female individuals representing all four gonad maturation stages. In the examined annual period, the transcriptional profiles globally highlighted the higher expression of genes associated to mussel nutrition and digestion in May-August compared to the other months, and trends for gonad transcripts consistent with the reproductive mussel status.

The same cDNA platform contributed to the toxicological evaluation of a neonicotinoid insecticide mixture (Dondero *et al*., 2010), an organophosphate compound (Canesi *et al*., 2011) and to the integrated measure of the functional mussel responses in the estuarine Tamar region in UK (Shaw *et al*., 2011).

The Hofmann_UCSB_Mytilus_2.5K_v1.0 record (GPL5795, Mar 2008) describes a platform of nearly 2500 spotted cDNAs of *Mytilus californianus* consisting of both unsequenced and sequenced clones referring to gill and muscle of environmentally challenged mussels. The related GEO series GSE8935 include data on latitudinal gene expression changes. Five biological replicates from four populations of Californian mussels were compared to a common reference sample in dual colour analysis (dye-swap labelling).

The HMS/SomeroLab-Mytilus-105K array-v1.0 (GPL9676, Jun 2010) and HMS/Somero-Mytilus-105K Agilent-v1.0 salinity stress (GPL11156, Jan 2011) are two successive versions of a platform composed by oligomer probes *in-situ* synthesized by Agilent Technologies (Santa Clara, CA, U.S.A)). These microarrays include probes of both *M. californianus* and *M. galloprovincialis,* and are intended for homologous and heterologous gene expression profiling. The processing and assembling of about 26,000 ESTs from *M. californianus* (Gracey *et al*., 2008) and 3,984 ESTs from *M. galloprovincialis* (Venier *et al*., 2003) resulted in a total of 12,961 and 1,688 transcript clusters or singletons, respectively. Long (60-mer) oligoprobes were designed against the *M. californianus* series and the resulting 43,969 total unique probes (2.6 probes per transcript sequence) were analyzed through BLAST searches against the *M. galloprovincialis* series to support selection and design of related probes (556 probe pairs matching transcripts of both species, with a mean number of 4.6 divergent nucleotide bases per probe). A total of 44,524 unique probes were duplicated or triplicated randomly to fill a microarray of 105,000 elements (105 k probes).

These two platforms have been used to investigate the transcriptional responses to thermal and osmotic stresses in *M. californianus*, *M. trossulus* and *M. galloprovincialis* (Evans and Somero, 2010; Lockwood *et al*., 2010; Lockwood and Somero, 2011). To control the effects of sequence mismatches in the case of *M. galloprovincialis* probes included in the GPL9676 platform, only probes experimentally confirmed in the hybridization of 84 samples of both *M. galloprovincialis* and *M. trossulus* were used in the related data analysis. Following a large set of hybridization experiments and stringent quality control, misleading probes were removed from the dataset and the second platform version (GPL11156/Agilent 019153) was generated.

In the central and southern coasts of California, *M. galloprovincialis* has largely displaced

the native congener, *M. trossulus*, and such evidence could be explained by species differences in physiological traits related to the adaptation to warm habitats. To investigate the hypothesis, gene expression profiling was performed on gill RNA from mussels subjected to acute heat-stress (GEO series GSE19031). A total of 1,531 probes, out of 4,488 different genes represented on the microarray and recognizing mRNAs of both species, showed temperature-dependent expression changes highly similar in the two congeners whereas 96 probes denoting oxidative stress, proteolysis, energy metabolism, ion transport, cell signalling, and cytoskeleton reorganization outlined species-specific responses to the heat-stress. Among them, the one encoding the small heat shock protein 24 was highly induced in the Mediterranean mussel and showed only a small change in *M. trossulus*. Six biological replicates per mussel group were included in this study which exemplifies the use of a cross-species microarray as well as heterologous and homologous hybridization. According to the authors and published literature*, M. trossulus* and *M. galloprovincialis* are approximately 7.6 million years divergent from *M. californianus*, and only 3.5 million years divergent from each other: in other words, the heterologous hybridization of target sequences from *M. trossulus* should occur on microarray probes from *M. galloprovincialis* without inherent sequence bias and should provide a reliable comparison of their transcriptional responses. Though debated, prudent evaluations of the sequence divergence by *in silico* approaches and phylogenetic data could expand the use of cross-species hybridization as a compromise solution for investigating gene expression in species with unsequenced genomes (Costa *et al*., 2010; Nazar *et al*., 2010; Ptitsyn *et al*., 2010).

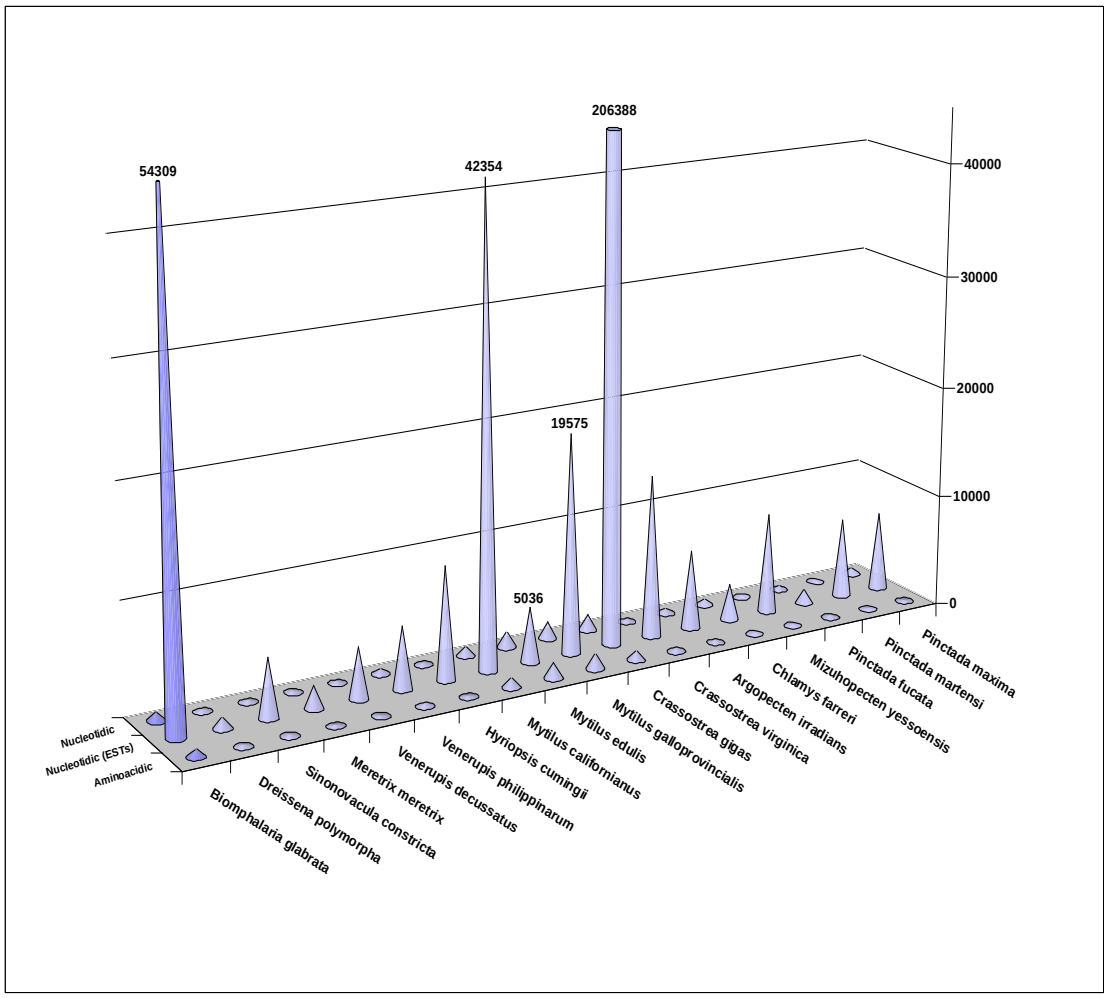
Gene expression profiling was also performed on gill RNA from mussels subjected to salinity stress (GEO series GSE25111). A total of 117 probes, out of 6,777 genes represented on the microarray, showed significant changes similar between *M. californianus* and *M. galloprovincialis* whereas 12 probes, denoting mRNA splicing, polyamine synthesis, exocytosis, translation, cell adhesion, and cell signaling, outlined species-specific responses. The study was based on AlexaFluor-labelling (555 and 647 fluorescence dyes) of amplified RNA, pooled reference samples, six biological replicates, and competitive hybridization in agreement to the recommended Agilent protocols. In addition to the overall stringent processing of the fluorescence signals, the heterologous hybridization design suggested the elimination of data from probes with low signal intensity (signal intensity < 150 % of the local background and hybridized spot diameter <30 % of the nominal spot diameter).

The work performed at the A. Gracey's and G.N. Somero's laboratories (University of Southern California -Los Angeles, CA, U.S.A. and Stanford University -Palo Alto, CA,

U.S.A., respectively) on *Mytilus* (GEO series GSE19031 and GSE25111) and other species is facing the fundamental aspects of the organism adaptation to fluctuating environments and global climate changes, and gene expression profiling has been essential to their findings. For instance, the study of gene-expression changes in the Californian mussels at different phases in the tidal cycle revealed at least four distinct physiological states, corresponding to metabolism and respiration phase, cell-division phase, and two stress-response signatures linked to moderate and severe heat-stress events. The metabolism and cell-division phases appeared to be functionally linked and anti-correlated in time whereas magnitude and timing of the above states resulted to be influenced by the microhabitat conditions according to the vertical position on the shore (Gracey *et al*., 2008). Based on comparative physiology, a recent paper offers an overview on the expected consequences of global climate changes (Somero, 2011).
Finally, the Mussel Immunochip 1.0 (GPL10758, April 2011) is a spotted oligonucleotide platform consisting of four-replicated 1820 oligomer probes plus unrelated controls prepared at CRIBI for the purposes of a recent European project (IMAQUANIM). Oligomers of 57 bases average length were designed at short distance from the 3' end of transcript sequences selected previously in Mytibase, the interactive knowledgebase of *M. galloprovincialis* which includes most of the ESTs publicly available for this species (Venier *et al*., 2009). Based on multiple criteria, the subset of transcripts selected from Mytibase as putatively immune-related molecules should denote central "players" of the mussel innate immunity or genes whose expression is modulated during the mussel responses to immunostimulation (Venier *et al*., 2011). In the platform description, the probe ID is hyperlinked to the relative Mytibase record: for instance the probe MGO_07346 relates to MGC07346, a mussel transcript featured by the protein domain IPR000098-Interleukin 10 and yet functionally unknown. The performance of Immunochip 1.0 was tested with hemolymph samples collected at 3 and 48 h from *Vibrio*-challenged mussels (GEO series GSE23535) according to competitive hybridization of dye-swap labelled amplified RNA samples.

In agreement with the above descriptions, Figure 3 provides an updated summary of the nucleotide and protein sequences publicly available at July 2011 and highlights the importance of EST sequencing for the preparation of new DNA microarrays. More about the molecular "players" of the innate immunity and the immune responses of *M. galloprovincialis* is reported in the following paragraph.

Figure 3 - Number of sequence records available for selected mollusc species at the Natl Center for Biotechnology Information at July 2011.



DNA, RNA and protein sequences refer to *Biomphalaria glabrata* (Gastropoda) and bivalves belonging to the *Veneroida, Unionoida, Mytiloida, Ostreoida, Pectinoida* and *Pterioida* orders.

How much can simple sequences tell us about the mussel immune responses?

Taking advantage of the continuous increase of the nucleotide and amino acid sequences in the public databases, the current methods of bioinformatics can extract instructive data from simple sequences: from the analysis of various gene/transcript regions to the evaluation of protein/peptide structure and to the comparative analysis of evolutionary differences across the tree of life. This procedural approach complements and integrates the data derived from long-standing disciplines such as measures of structural changes and protein amounts/activity, among others.

The overall analysis of 18,788 high-quality ESTs rationally organized in 7,112

independent clusters or singletons (Mytibase transcript collection) highlighted some particularly abundant transcript groups: namely, transcripts featured by a complement component C1q-like domain, antimicrobial peptide (AMP) precursors of all four families known in the Mediterranean mussel and many heterogeneous lectins including fibrinogen-related molecules (Venier *et al*., 2011). To explain the abundance of immune-related molecules in Mytibase it is important to remember that such collection has been prepared by 16 primary (5 from hemocytes) and 1 normalized cDNA libraries from mussels subjected to various challenges, for instance mussels immune stimulated with preparations of Gram positive and Gram negative cells and viral-like molecules.

Searches by protein domain revealed a total of 168 different Mytibase transcripts containing the C1q signature IPR001073, almost invariably associated with the overlapping TNF-like IPR008983 motif. Curiously, the C1q domain-containing proteins predicted from the transcript sequences, display a short N-terminal signal peptide and a C-terminal globular domain but no central collagen-like repeats which are instead typical of vertebrate C1q domain-containing proteins. According to the current literature, these mussel proteins could represent secreted globular receptors, components of ancient complement pathways expected to mediate pathogen recognition and lysis (Dodds and Matsushita, 2007). The modularity and versatility of binding mediated by the globular C1q domain explain the variety of roles currently attributed to this still expanding family of proteins, and also supports their involvement in pathogen pattern recognition (Carland and Gerwick, 2010). The abundance and variety of mussel C1q domain-containing transcripts are consistent with this view.

One among these transcripts, named MgC1q, resulted to be expressed at detectable levels in the main tissues of naïve adult mussels, with the hemocytes showing the highest expression levels, and from 2 h post-fertilization up to 3 months later. The MgC1q expression was significantly modulated after mussel infection with Gram positive or Gram negative bacteria, data which confirm MgC1q as an immune-related gene. The striking molecular diversity of MgC1q was confirmed at both the DNA and cDNA levels, hence posing mechanistic questions on the origin of such variation (Gestal *et al*., 2010). Experimental findings and sequence analyses support the hypothesis of gene duplication, functional diversification and positive selection of many C1qDC variants in selected taxa, including the mussel lineage (Gerdol *et al*., 2011).

Defensins, mytilins, myticins and mytimycins are cationic antimicrobial peptides stabilized by 4 intrachain disulphide bonds (6 in mytimycin) in a typical 3-D motif (Yeaman and Yount, 2007). A remarkable diversity of a new group of myticins, with specific variant profiles detectable in single mussels, was reported in *M. galloprovincialis* (Pallavicini *et*

al., 2008; Costa *et al*., 2009). Following the discovery of the myticin-C variants, their molecular diversity and evolution has been further discussed (Padhi and Verghese, 2008) and the most recent findings indicate myticin C as a chemotactic molecule with antiviral activity and immunoregulatory properties (Balseiro *et al*., 2011). Just one singleton and other four similar sequences denote the antifungal AMP mytimycin in Mytibase (rare transcript). Mytimycin is composed by 54 aminoacids (6.2–6.3 kDa, 12 cysteines) and two main precursor variants, both featured by a signal peptide and a C-terminal extension, are expressed in mussels from different European regions (Sonthi et al., 2011). The presence of a calcium binding (EF hand) motif in the C-terminal extension suggests further characterization of such unusual AMP.

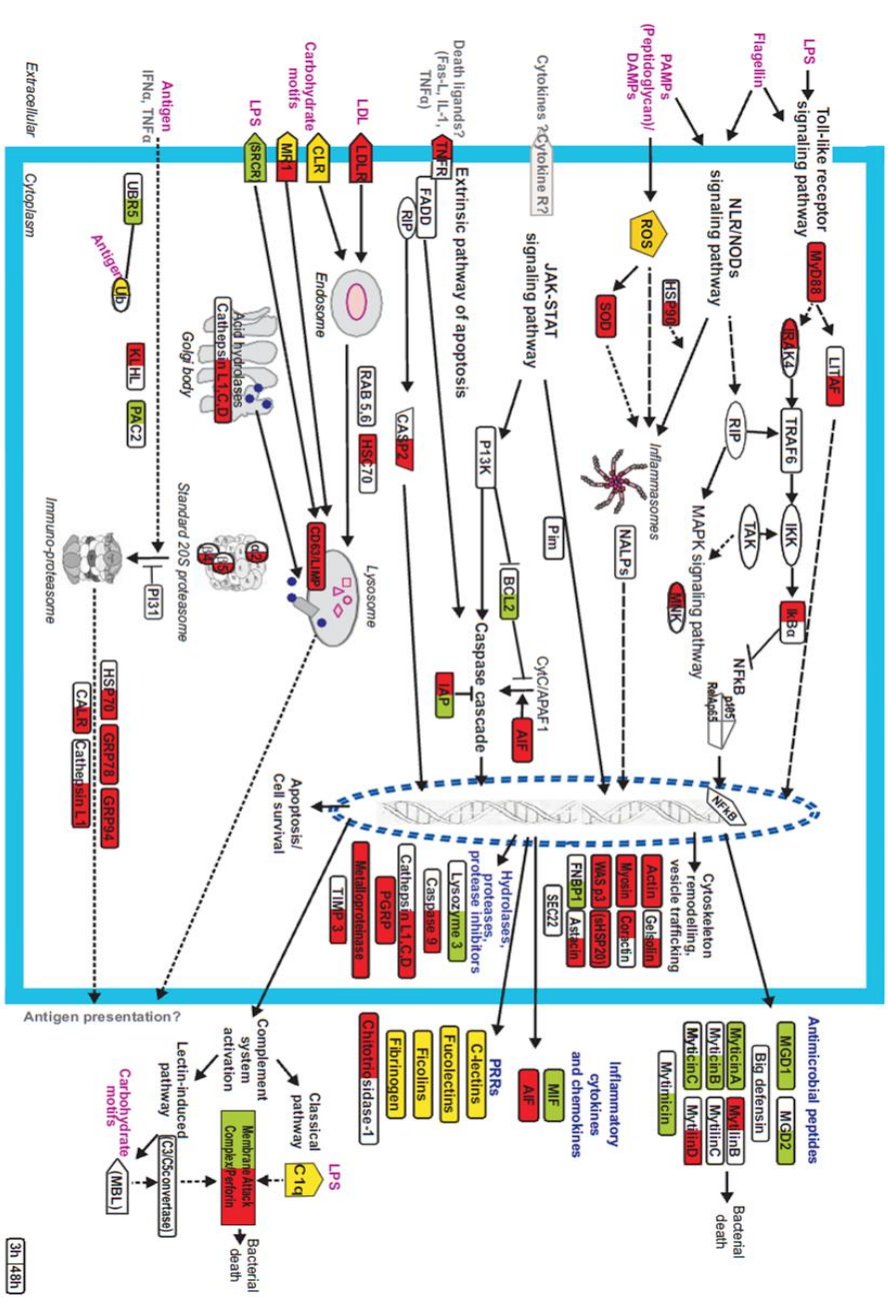
The "effector" role of the mussel antimicrobial peptides (AMPs) is confirmed in many experimental studies and a comprehensive review have been recently provided (Li *et al.*, 2011). Whether these effectors can modulate the mussel immune responses with mechanisms other than membrane disruption, as reported for mammalian AMPs, it is not clear. Based on deep amplicon sequencing, the sequence diversity of mussel AMPs is now under study in natural mussel populations from different geographical regions and in mussels challenged with bacterial cells.

Lectins are a rather heterogeneous protein family comprising 8 to 15 subgroups, depending on the scientist's view (Dodd and Drickamer, 2001). Lectins typically possess carbohydrate binding domains and participate in many cell processes. Similarly to the mammalian C1q, the C-terminal fibrinogen-like domain IPR002181 of ficolins forms a tulip-like structure able to bind the carbohydrate residues of foreign and apoptotic cells (with consequent opsonization, phagocytosis and cell clearance) or triggering the proteolytic complement cascade and pathogen lysis. Fibrinogen-related lectin proteins (FREPs) are expressed also in mussels (Venier *et al*., 2011) and are codified by at least 2 (M. edulis) 4 (M. californianus) and 7 genes (*M. galloprovincialis)* (Gorbushin and Iakovleva, 2011). These molecules can be regarded as immune pattern-recognition receptors and their involvement in the native immunity is supported by the evidence of species-specific expansion of FREPs in the snail *Biomphalaria glabrata* and the mosquito *Anopheles gambiae* (Waterhouse *et al*., 2007; Zhang *et al*., 2008). In mussel, FREPs are significantly up-regulated after bacterial infection or PAMP treatment, and display opsonizing activity similar to that of mammalian ficolins; moreover, the different sets of FREP sequences detected among and within individuals further emphasize the great complexity of the invertebrate immune systems (Romero *et al.*, 2011). Other lectin-like sequences expressed in mussels are commented in Venier *et al*. (2011).

The cases reported above are a few examples of the many classes of transcripts

specifically expressed or modulated during the mussel response to potential pathogens.

Figure 4 - Main transcriptional changes detected in mussels at 3 h and 48 h from the injection of live *Vibrio* cells (modified from Venier *et al*., 2011).



Only relevant molecular "players" represented in the Immunochip of *M. galloprovincialis* are reported (framed). In each frame, the detected expression trends are indicated in red, green and yellow (up- and down-regulation and not homogeneous trends, respectively). Annotations based only on protein domains are reported in brackets. Overall, the figure draws the attention to a number of mussel genes, still not characterized, whose expression is modulated in response to immune stimulation.

Considering in a dynamic view the behaviour of one cell population only, the versatile mussel hemocytes, one can imagine that almost all cellular processes could be influenced by the contact with pathogen-associated molecular patterns: from the cytoskeleton remodelling supportive of chemotaxis, migration and phagocytosis to the intracellular signalling possibly shaping the inflammatory response and finely tuned expression of many regulatory and effector genes. Cross-talking signalling pathways have been traced in mussel and the Mytibase collection includes transcripts denoting the regulatory cytokine MIF (migration inhibiting factor) and cytokine-related molecules, consistent with the idea of an invertebrate cytokine network (Malagoli, 2010). The recent definition of a species-specific Immunochip aims to the experimental validation of a selected subset of transcripts: a synopsis of the main gene expression changes detected in mussels at 3 and 48 h after challenge with live bacterial cells is reported in Figure 4. The general AMP down-regulation observed in this particular laboratory treatment was confirmed by quantitative PCR data and is discussed also in Li *et al*. (2010).

Concluding remarks

EST sequencing and DNA microarrays have substantially improved the identification of genes expressed in the *Mytilus* species. Compared to the first EST collection and the related cDNA microarray, Mytibase includes an interesting variety of immune-related molecules which can be further characterized with traditional and innovative approaches as exemplified by Romero *et al.* (2011). Nonetheless, in the Mytibase collection about half of the mussel transcripts are still unknown, devoid of functional annotation. Hence, much work remains to be done both *in silico* and in laboratory to provide a comprehensive view of the global gene transcription in mussels, particularly the part of the transcriptome mediating the response to potential invaders (immunome).

Undoubtedly, the application of the available mussel DNA microarray platforms can further reveal expression trends of different gene categories and identify useful markers of functional state, if not global molecular signatures useful to disentangle the complex mussel physiology. Depending on the study design and on the type of microarray platform, independent validation of the expression data can be accomplished by quantitative PCR or with other experimental measures. All the steps of the DNA microarray testing could be used to strengthen the final data interpretation, from the microarray preparation strategy to the stringency of the hybridization reaction to the algorithms applied to data processing.

The maintenance of the physical collection of the cDNAs, *i.e.*, recombinant bacterial

clones, is a prerequisite for the use of spotted cDNA microarrays (for instance, the current use of Mytarray 1.0 slides, printed at the CRIBI facility depends on long work performed at the Dept of Biology - University of Padua). Such work is not more affordable as long as the clustered ESTs increase in number, and external commercial services or deep sequencing become an attractive alternative.

As a matter of fact, next-generation sequencing (NGS) technologies are now complementing and challenging the DNA microarrays as alternative tools for genome analysis and transcriptome sequencing (Hurd and Nelson, 2009; Morozova *et al*., 2009). For instance, the so called 454 pyrosequencing has been already applied to the study of tissue-specific expression patterns in *M. galloprovincialis* (Craft *et al*., 2010) and many laboratories in the world are now investing in this kind of work.

Acknowledgments

This work has been sequentially supported by FOOD-CT-2005-007103 (IMAQUANIM), FP7-KBBE-2010-4-266157 (BIVALIFE) and, partially, 20084BEJ9F (PRIN08).

PLOSONE

2011

Massively parallel amplicon sequencing reveals isotype-specific variability of antimicrobial peptide transcripts in *Mytilus galloprovincialis*

Umberto Rosani[1]*, Laura Varotto[1], Alberta Rossi[1], Philippe Roch[2], Beatriz Novoa[3], Antonio Figueras[3], Alberto Pallavicini[4], Paola Venier[1]*

[1]Department of Biology, University of Padua, Padova, Italy.
[2]Ecologie des Systèmes Marins et Côtiers, CNRS-IRD-University of Montpellier 2, Montpellier, France.
[3]Instituto de Investigaciones Marinas, CSIC, Vigo, Spain.
[4]Department of Life Sciences, University of Trieste, Trieste, Italy.

*Corresponding Authors

Editor: Christos A. Ouzounis, The Centre for Research and Technology, Greece

Received July 2, 2011; Accepted October 2, 2011; Published November 7, 2011

Funding: This study was funded by the European project FOOD-CT-2005-007103 (http://imaquanim.dfvf.dk/info) and, in part, by the Italian Project 20084BEJ9FPRIN08 (http://prin.miur.it). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Background: Effective innate responses against potential pathogens are essential in the living world and possibly contributed to the evolutionary success of invertebrates. Taken together, antimicrobial peptide (AMP) precursors of defensin, mytilin, myticin and mytimycin can represent about 40% of the hemocyte transcriptome in mussels injected with viral-like and bacterial preparations, and unique profiles of myticin C variants are expressed in single mussels. Based on amplicon pyrosequencing, we have ascertained and compared the natural and *Vibrio*-induced diversity of AMP transcripts in mussel hemocytes from three European regions.

Methodology/Principal Findings: Hemolymph was collected from mussels farmed in the coastal regions of Palavas (France), Vigo (Spain) and Venice (Italy). To represent the AMP families known in *M. galloprovincialis*, nine transcript sequences have been selected, amplified from hemocyte RNA and subjected to pyrosequencing. Hemolymph from farmed (offshore) and wild (lagoon) Venice mussels, both injected with 10^7 *Vibrio* cells, were similarly processed. Amplicon pyrosequencing emphasized the AMP transcript diversity, with Single Nucleotide Changes (SNC) minimal for mytilin B/C and maximal for arthropod-like defensin and myticin C. Ratio of non-synonymous vs. synonymous changes also greatly differed between AMP isotypes. Overall, each amplicon revealed similar levels of nucleotidic variation across geographical regions, with two main sequence patterns confirmed for mytimycin and no substantial changes after immunostimulation.

Conclusions/Significance: Barcoding and bidirectional pyrosequencing allowed us to map and compare the transcript diversity of known mussel AMPs. Though most of the genuine cds variation was common to the analyzed samples we could estimate from 9 to 106 peptide variants in hemolymph pools representing 100 mussels, depending on the AMP isoform and sampling site. In this study, no prevailing SNC patterns related to geographical origin or *Vibrio* injection emerged. Whether or not the contact with potential pathogens can increase the amount of AMP transcript variants in mussels requires additional study.

PlosOne, 2011

Introduction

Mytilus species (Phylum Mollusca, Class Bivalvia) are intertidal filter-feeders distributed worldwide, anchored to hard substrates in dense communities and widely used as bio-sensors of coastal pollution. Mussel populations of the northern and southern hemisphere probably separated 0.54-1.31 million years ago, far after the trans-Arctic expansion towards North America, and before the divergence between the Atlantic and Mediterranean ecotypes [1]. *M. galloprovincialis* hybridizes with *M. edulis* in southwest England and the Mediterranean mussel is now reported in Eastern Asia, California, Chile and Western Australia [1-2].

At different latitudes, mussels face tidal and seasonal fluctuations, changeable pollutant loads and also the surrounding bioma with behavioral changes [3], metabolic adjustments [4] and a variety of defense reactions [5-7]. With the exception of a few metazoan parasites which also somewhat affect the *Mytilus* species [8-9], mussels seem refractory to diseases and could instead influence the prevalence of pathogens such as *Perkinsus* spp. and *Betanodavirus* (*Nodaviridae*) in other bivalves and fishes, respectively [10]. Like other invertebrates, bivalve molluscs rely on ancient and rapid defenses to fight potential pathogens, and gene-encoded antimicrobial peptides (AMPs) are major humoral components of their immune system.

Host defense peptides are present in virtually all living organisms, with more than 30 AMPs expressed in humans and about 200 peptides identified in insects (approximately 1500 molecules very diverse in sequence and secondary structures are reported in specific databases).

[11-13]. Among other structural features, a conserved γ-core motif originated from the bidirectional orientation of specific aminoacid residues including an invariant cysteine array has indicated the evolutionary relatedness of cysteine-stabilized α-β (CS-αβ) AMPs, kinocidins, invertebrate toxins and snake venoms: such unifying structure provides an interesting hypothesis for context-specific action modes, from the perturbation of negatively charged cell membranes and ion channels to the immunoregulatory functions [14].

In the continuous fight with competitors, predators and pathogens, the evolutionary diversification of AMP types and gene families likely occurred through events of gene duplication, shuffling of functional elements and selection for variation at positions adjacent, or integral to, the conserved structural motifs [13, 15]. In *Crassostrea gigas*, combined mechanisms of sequence diversification (e.g. recombination, parallel homoplasic mutations, indel events) and directional selection have been suggested to explain the remarkable gene multiplicity and variable copy number of defensins and

proline-rich peptides, whereas the marked transcript diversity of Cg-bpi, a bactericidal permeability protein, has been mainly referred to the allelic polymorphism of one single gene [16]. Microsatellite-mediated mosaics of sequence elements, low-transcription fidelity and transcript editing support the evidence of about 50 polymorphic genes, and an extraordinary diverse set of Sp185/333 proteins expressed in response to pathogens by the purple sea urchin [17-18]. Worthy of note, the copy number polymorphism of α and β defensin genes with proportional peptide levels in neutrophylic granulocytes, has been related to the individual risk of infection in humans [19-20].

Tens of different AMPs or AMP families have been discovered in marine invertebrates [21]. In the mussels *M. galloprovincialis* and *M. edulis*, four different groups of CS-αβ AMPs with multiple isoforms have been discovered and classified according to their primary sequence and secondary structure: defensins reported as MGD1 and MGD2, mytilin A, B, C, D and E, myticin A, B and C, mytimycin, the only strictly antifungal peptide with an EF-hand like domain [22-25]. These AMPs share small size (3.7-4.5 kDa, except mytimycin of 6.2 kDa), positive charge and amphiphilic behavior. Their precursors (pre-pro-peptides) consist of an N-terminal signal peptide, a central mature peptide and a C-terminal extension. Each family is characterized by a cysteine array of 8 (12 in mytimycin) cysteines engaged in intramolecular disulfide bonds.

A broad spectrum of activity, often complementary and not strictly antibacterial, was reported for the mussel defensins, mytilins and myticins [26-27] whereas mytimycin, a 6.2 kDa peptide isolated from normal and immunostimulated mussels, selectively inhibited *Neurospora* and *Fusarium* growth [22]. Whether purified in sufficient amounts from cellular fractions or obtained in stable conformations by chemical synthesis or recombinant system, pure peptides are essential to investigate the antibiotic power of the different mussel AMPs.

In situ hybridization and immunolocalization assays performed on mussel hemocytes demonstrated a partially overlapping expression of defensins and mytilins [28]. AMP expression and stored peptides have been observed in several tissues and developmental stages [28-29] to indicate that cells other than hemocytes can produce and release AMPs, a phenomenon well known from frog skin [30] and the male reproductive system of rats [31]. Overall, mussel AMPs display rather complex expression patterns, dependent on developmental stage, seasonality and immunostimulation [32-34].

In *M. galloprovincialis*, massive EST sequencing confirmed the abundance and transcript diversity of AMPs and other key players of the innate immunity [35]. The AMP precursors represented 26-43% of the hemocyte transcripts in mussels injected with viral-like and bacterial preparations; in particular, 74 precursor and 25 mature peptide variants of

myticin C were detected in a sample of only 100 mussels, with unique profiles of transcript variants in single mussels and less common alleles differing at single nucleotide positions from the two most common ones [29, 36].

The myticin C variation is also remarkable compared to mytilin B, as one mussel can produce 2-10 different mytilin B transcripts but silent substitutions restrict the peptide variants to only a few [37]. In spite of the abundance of other AMPs, just one singleton plus 4 similar sequences denote mytimycin in Mytibase, interactive catalogue including 18788 expressed sequence tags (ESTs) of *M. galloprovincialis* [25]. Sequencing and Southern blot data indicate one gene copy per genome for defensin MGD2, mytilin B and myticin C [38, 36]. Two gene copies or allelic polymorphism could explain the simultaneous presence of two length variants of the mytimycin gene per mussel [39]. The gene copy number of the mussel AMPs need verification since partial gene sequences covering the coding sequence (cds) are only available for MGD1, mytilin B, myticin C and mytimycin [29, 35-39]. Two 3D structures have been established by NMR spectrometry, defensin MGD1 [40] and mytilin B [41].

Thus far no *Mytilus* genome has been sequenced and, compared to better known model organisms, a limited number of genes have been investigated: for instance those concerning defensin and mytilins [38], heat shock proteins [42], metallothioneins [43] and apoptotic caspases [44]. Also gene-centered studies take advantage of the massive production of ESTs which currently contributes to the identification of molecules and pathways underlying the mussel response to various natural and experimental conditions [45-48]. Among the 67,726 ESTs and 4680 aminoacid sequences publicly available for the *Mytilus* genus, about 29 and 32%, respectively, refer to *M. galloprovincialis* (www.ncbi.nlm.nih.gov, June 2011). Recently, new ESTs have been produced from the digestive gland, foot, gill and mantle of *M. galloprovincialis* by advanced sequencing [49]. The so called ‘pyrosequencing’ was the first alternative to the use of chain-terminating inhibitors [50] and it has radically increased the sequencing power as well as the resolution of low-abundance variants [51-53]. Adequate read coverage can assure reliable quantification of single nucleotide changes (SNC) when seeking critical mutations or sequence polymorphisms.

Based on 454 pyrosequencing, we have thoroughly studied the sequence diversity of 9 different AMP precursors expressed in hemocytes of mussels (*M. galloprovincialis*) farmed in three European regions. Similarly, we have analyzed and compared mussels farmed offshore or living inside the Venice Lagoon (Italy), before and after injection with live *Vibrio* cells.

Results

In Table 1 we summarized the main features of nine CS-αβ AMPs expressed in *M. galloprovincialis*.

Table 1 - Features of the selected mussel AMPs.

AMP	Precursor length (aa)	Mature peptide length (aa)	Cysteine array	Weight (kDa)	Isoelectric point (pH)	Hydrophobicity ratio (%)
MytA	96	40	8	4.5	9.1	35
MytB	96	40	8	4.6	9.2	30
MytC	100	40	8	4.4	8.8	35
MytIB	103	34	8	4.0	9.7	27
MytIC	100	34	8	4.2	9.9	24
MytID	97	34	8	3.9	10.4	32
MytM	152	54	12	6.4	8.4	26
MGD1	82	39	8	4.4	9.4	36
MGDt	61	38	8	4.4	9.3	45

MGDt lacks of C-terminal extension. Table data are calculated from Mytibase ESTs. See also AF162334.1, AF162335.1, AF162336.1, AF162337.1, EU810204.1 and EU927448.1 at NCBI.

Following appropriate primer design, we amplified the related transcript sequences from hemolymph pools representing groups of 100 mussels farmed in south France (Pa), northwest Spain (Vi) and northeast Italy (Ve) or native from the industrial canals of the Lagoon of Venice (Ve nc). In addition, groups of 40 offshore-farmed (Ve ft) and lagoon-native (Ve nt) mussels were injected with 10^7 live *Vibrio splendidus* cells and similarly processed (Table 2).

Table 2 - Description of the mussel haemolymph samples processed for 454 pyrosequencing.

Sample ID	Origin	No. of mussels	Status	Treatment
Pa	Palavas, F	100	farmed	untreated
Vi	Ria de Vigo, S	100	farmed	untreated
Ve	Offshore Venice, I	100	farmed	untreated
Ve nc	Venice Lagoon, I	100	native	untreated
Ve ft	Offshore Venice, I	40	farmed	*Vibrio* injection
Ve nt	Venice Lagoon, I	40	native	*Vibrio* injection

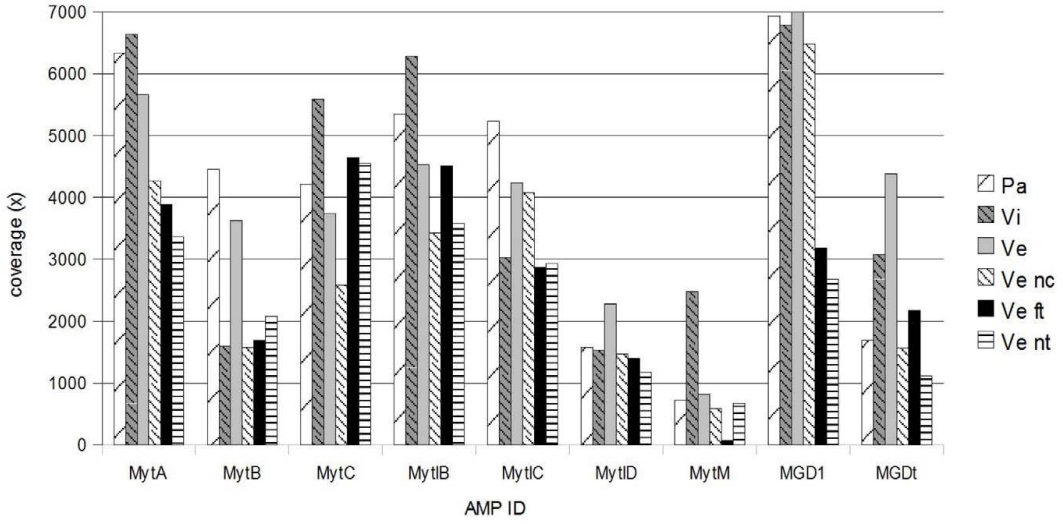
The resulting 78 PCR products (13 amplicons x 6 samples) were then purified, quantified, diluted to the appropriate concentration (~$7*10^9$ molecules/µl) to compose two equimolecular pools for the emulsion PCR (cDNA concentration was 2.3 and 2.5 ng/µl,

respectively) and bidirectional sequencing.

Overall, massively parallel sequencing produced 359,867 output reads and more than 73 Mbases with good quality scores, for a total of 304,621 trimmed reads with 226 bp average length. We discarded about 15% of reads per sample (88% of them shorter than 70 bp) which possibly originated in the PCR amplification or sequencing reaction (short sequences are not expected to bias the amplicon coverage nor the accuracy of SNCs detection).

Total or partial overlapping of the forward and reverse reads allowed the complete coverage of the 13 reference sequences. Hence, 97.5% of the good quality reads correctly mapped against the 9 selected AMPs and could be attributed to the 6 original samples. With the exception of MytM in the sample Ve ft, at least 1034 reads mapped on each AMP precursor transcript (range 1034-10814, File S1). Figure 1 shows relevant differences in the average base coverage calculated per AMP precursor transcript in each sample (3378x, the average read depth per AMP).

Figure 1 - Average base coverage (x) per AMP and sample.



Coverage is calculated as total sequenced base divided by the length of the amplified transcript.

Separately for each AMP and sample, we subsequently grouped the reads having the same length and 100% identity in two cluster types: (1) equal to the original transcript and (2) with at least 1 SNC. The last ones were employed for SNC detection and related analysis. Only the SNCs covered 30x, and representing at least 3% of the reads mapping a given AMP, were considered genuine and counted per AMP precursor (Table 3). Based on the SNC counts, the average value of SNC per base calculated for the coding

sequence was 0.18. Considering all AMPs together, as much as 134 SNCs were common to the 6 samples and represented 86% of the genuine cds variation (File S2). The SNC frequency values (cds) indicated MytC and MGD1 as the most polymorphic AMP transcripts, opposite to those of mytilins B and C. The fraction of non-synonymous changes ranged from null (MytlC) to 25 (MytC) whereas the related ω values highlighted the opposite cases of MytlC (null) and MytlD (9-15).

All SNC frequency values were used to test statistically the sequence diversity of the 9 selected AMPs (one-way ANOVA, α=0.001, followed by Tukey's Honestly Significant Difference test, α=0.05). The null hypothesis (equal SNC frequency between AMPs) was rejected and, according to the Tukey's HSD test we classified the 9 cases from the least changeable mytilins, to the most polymorphic myticins and defensins (Figure 2).

Common and exclusive SNCs are reported in Figure 3 per geographical region: the common changes represent the majority, with 68, 74 and 78% in the Vi, Ve and Pa samples, respectively. Similar percentages of common SNCs were found in mussels farmed offshore or living wild in the industrial canals of the Venice lagoon (Ve versus Ve nc), injected or not with 10^7 live *V. splendidus* (Ve versus Ve ft, Ve nc versus Ve nt). Details on the immune stimulation and related host response are reported elsewhere [34, 35].

Table 3 - Number of SNCs detected in the whole AMP transcript precursor and cds, related frequencies, non-synonymous changes and ω values.

AMP precursor	Sample	SNCs (total)	SNCs (cds)	Frequency (SNC / base)	Freq / AMP (Mean ± SD)	ns SNCs*	ω**
MytA	Pa	38	25	0.26	0.25 ± 0.04	10	0.7
	Vi	36	20	0.21		8	0.7
	Ve	41	28	0.29		12	0.8
	Ve nc	38	23	0.24		10	0.8
	Ve ft	30	19	0.20		10	1.1
	Ve nt	39	26	0.27		12	0.9
MytB	Pa	21	17	0.18	0.20 ± 0.05	12	2.4
	Vi	40	29	0.30		15	1.1
	Ve	22	17	0.18		11	1.8
	Ve nc	30	19	0.20		11	1.4
	Ve ft	27	18	0.19		12	2.0
	Ve nt	19	14	0.14		7	1.0
MytC	Pa	33	27	0.27	0.26 ± 0.04	19	2.4
	Vi	41	33	0.33		25	3.1
	Ve	31	23	0.23		17	2.8
	Ve nc	34	26	0.26		17	1.9
	Ve ft	32	24	0.24		19	3.8
	Ve nt	27	22	0.22		15	2.1
MytlB	Pa	11	5	0.05	0.09 ± 0.03	1	0.3
	Vi	17	10	0.10		5	1.0
	Ve	16	9	0.09		3	0.5

	Ve nc	23	13	0.13		9	2.3
	Ve ft	15	10	0.10		5	1.0
	Ve nt	16	9	0.09		4	0.8
MytlC	Pa	8	1	0.01	0.01 ± 0.01	0	0.0
	Vi	8	1	0.01		0	0.0
	Ve	11	2	0.02		0	0.0
	Ve nc	7	1	0.01		0	0.0
	Ve ft	6	1	0.01		0	0.0
	Ve nt	8	2	0.02		0	0.0
MytlD	Pa	23	16	0.17	0.15 ± 0.04	15	15.0
	Vi	25	20	0.22		18	9.0
	Ve	21	12	0.13		11	11.0
	Ve nc	19	11	0.12		11	[11]
	Ve ft	18	13	0.14		13	[13]
	Ve nt	19	10	0.11		9	9.0
MytM	Pa	28	28	0.20	0.21 ± 0.01	19	2.1
	Vi	29	29	0.21		21	2.6
	Ve	31	31	0.22		23	2.9
	Ve nc	30	30	0.21		22	2.8
	Ve ft	29	29	0.21		21	2.6
	Ve nt	28	28	0.2		20	2.5
MGD1	Pa	28	21	0.26	0.28 ± 0.02	14	2.0
	Vi	29	23	0.28		15	1.9
	Ve	28	21	0.26		14	2.0
	Ve nc	31	24	0.30		15	1.7
	Ve ft	28	22	0.27		15	02.0 1.00
	Ve nt	30	23	0.28		15	1.9
MGDt	Pa	43	12	0.19	0.18 ± 0.02	11	11.0
	Vi	39	13	0.21		10	3.3
	Ve	40	12	0.19		10	5.0
	Ve nc	41	10	0.16		9	9.0
	Ve ft	35	12	0.19		10	5.0
	Ve nt	36	10	0.16		9	9.0

Frequency values, ns SNCs and ω values refer to the cds; * ns SNCs, non-synonymous changes;
** ω, ratio between non-synonymous and synonymous SNCs (in brackets, cases without synonymous substitutions).

Figure 2 - AMP precursors grouped on the basis of their variability.

Grouping based on HSD test, α =0.05; sample-specific and averaged diversity levels are reported in Table 3.

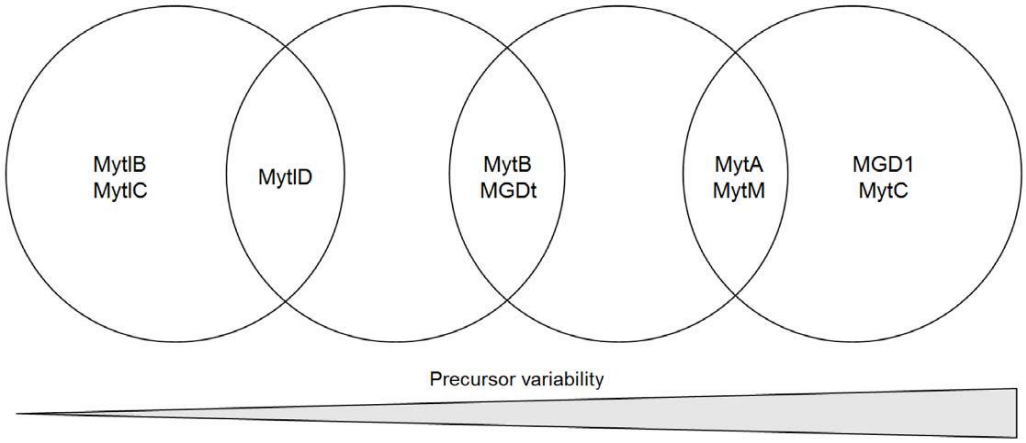
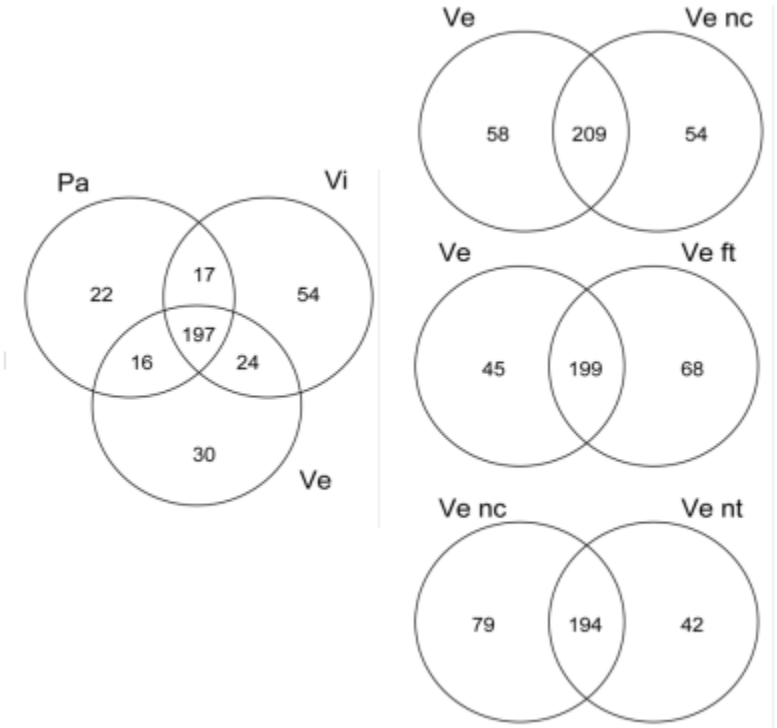


Figure 3 - Common and exclusive SNCs in the Pa, Vi and Ve samples and paired samples from native and immunostimulated mussels.



Finally, we analyzed the nucleotidic and aminoacidic substitution patterns of each AMP precursor, separately in the samples from Palavas, Vigo and Venice. The percentage of read clusters differing at least 1 SNC ranged from 56% (MytlD) to 100% (MytB and MGDt). The read clusters covered at least 3x were virtually translated into amino acids, and redundancy due to silent substitutions was removed. As reported in Table 4 we could estimate a number of AMP transcript variants ranging from 9 (Pa MytM) to nearly 100 (Ve MytB, Vi MGD1, Ve MGD1 and Ve MGDt).

Table 4 - AMP transcript variants and expected number of non-redundant peptides per geographical region.

AMP ID	Sample	Total read clusters*	Fraction with at least 1 SNC (%)	Non-redundant peptides**
	Pa	300	83	50
MytA	Vi	684	90	68
	Ve	486	86	51
	Pa	763	100	95
MytB	Vi	280	100	63
	Ve	654	100	106
	Pa	844	97	41
MytC	Vi	777	100	93
	Ve	484	99	59
	Pa	206	70	18
MytlB	Vi	338	90	40
	Ve	190	88	28
	Pa	240	68	37
MytlC	Vi	138	83	23
	Ve	177	79	32
	Pa	61	64	17
MytlD	Vi	59	69	17
	Ve	61	56	14
	Pa	31	94	9
MytM	Vi	206	97	49
	Ve	32	94	12
	Pa	410	99	85
MGD1	Vi	557	99	100
	Ve	686	89	101
	Pa	204	100	45
MGDt	Vi	729	100	84
	Ve	793	100	103

*Total cluster number refers to reads differing in length and/or sequence
** Virtual number of peptides differing at least for one residue

The facts emerging from 454 pyrosequencing of the Mytimycin and Myticin C amplicons are reported here below in more detail as instructive examples.

The two gene sequences publicly available for the MytM of *M. galloprovincialis* (FJ804479.1, FJ804478.1) denote 3 exons and 2 introns, and differ only in the length of intron 2; the short and long version of it occurring simultaneously in single mussels [39]. The amplicon designed in the present work covered 429/456 bp, i.e. 94% of the cds.

The MytM pyrosequencing yielded 6645 aligned reads (33, 30 and 35% of them differing from the reference sequence in the Pa, Vi and Ve samples, respectively).

A nucleotide change in position 58 (Thymine in the place of Cytosine) was detected in all reads and was not considered as SNC because it could represent an error occurred during Sanger sequencing of the original Mytibase singleton (MGC05878, File S3 in Supplementary Materials). The AVA software (Roche Life Sciences) grouped the Pa, Vi and Ve MytM reads in 31, 206 and 32 clusters or singletons (27, 199 and 31 of high quality), respectively. In total, 5690 high quality sequences (86%) were translated in amino acids and produced 9-49 expected peptide variants (Table 4).

Irrespective of the geographical origin, we could consistently identify the two most abundant MytM types: a consensus very similar to the original sequence (MytM_1, MGC05878) and a second one (MytM_2) similar to the sequence MytM-P recently described [39]. Jointly, MytM_1 and MytM_2 represent the 87, 78 and 76% of the MytM reads in the samples Pa, Vi and Ve, respectively (Vi MytM display higher sequence variability than the other two samples).

MytM_2 represents 20, 8 and 11% of the MytM reads in samples Pa, Vi and Ve, respectively, and displays 25 SNCs. Eighteen out of 25 changes are non-synonymous (7 in the signal peptide, 8 in the mature peptide and 3 in the C-terminal extension) and confirm the substitutions detected in MytM-P by Sonthi *et al.* using a Sanger approach. Figure 4 locates all genuine SNCs along the original MytM sequence (cds).

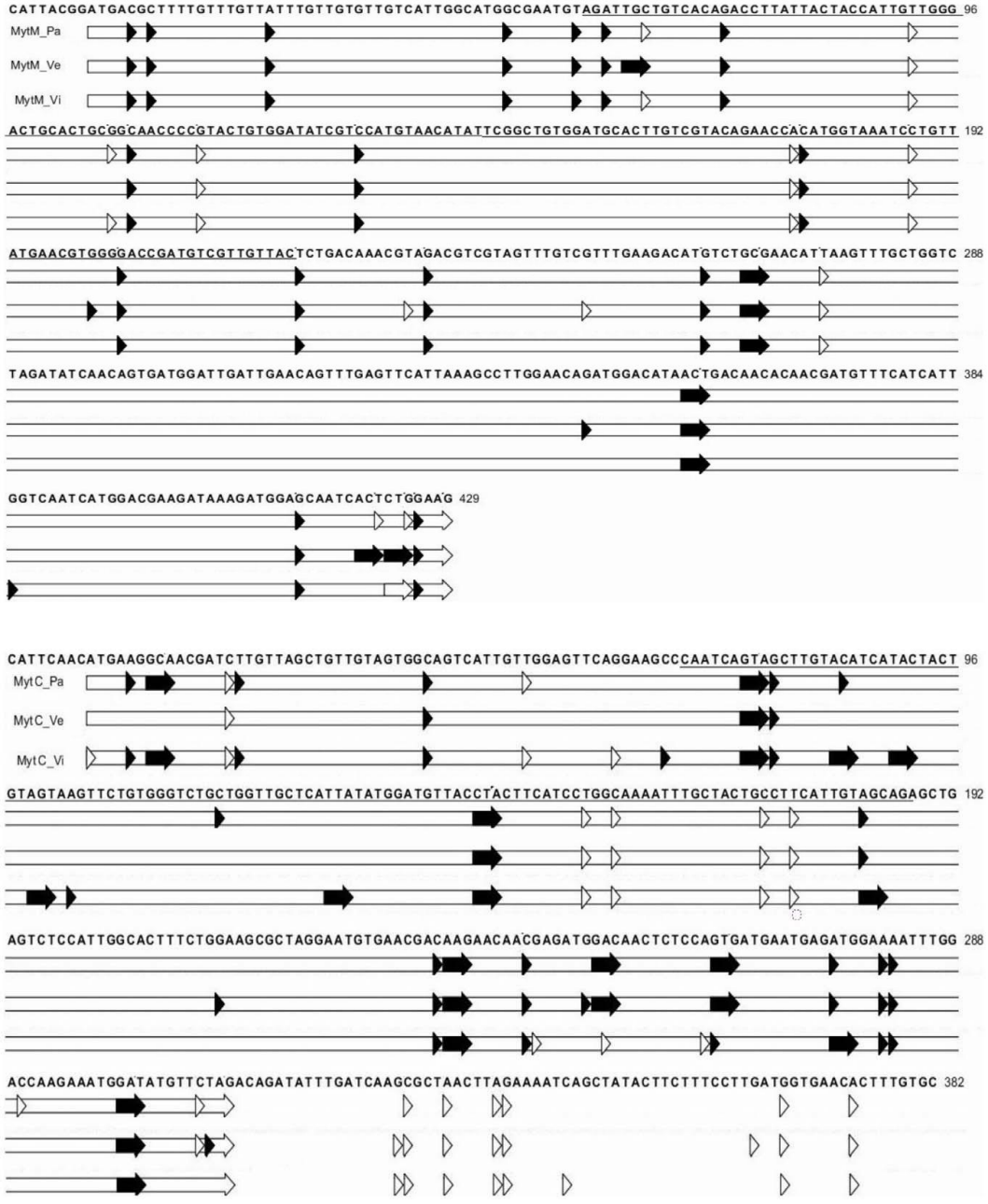
Partial gene sequences including the cds have been reported for MytC and denote 3 exons and 2 introns, with the mature peptide entirely located in exon 2 [29, 36].

The MytC pyrosequencing yielded 22,119 aligned reads fully covering the cds (71%, 88% and 79% of them differing from the original sequence in the Pa, Vi and Ve samples, respectively). The Pa, Vi and Ve MytC reads could be grouped in 844, 777 and 468 clusters or singletons (823, 752 and 466 of high quality), respectively. In total, 18113 high quality reads (82%) were translated in amino acids and produced 41-93 expected peptide variants (Table 4) without evidence of prevailing variation patterns.

Despite the remarkable number of sequence variants, 98.8% of the MytC peptide clusters

retained the typical cysteine array. Figure 5 locates all genuine SNCs along the original sequence of MytC.

Figure 4 and 5 - SNCs mapped along the MytM nd MytC sequence in the Pa, Ve and Vi samples.



Synonymous (empty) and non-synonymous (filled) changes are located on the AMP precursor sequence. Empty horizontal bars indicates cds, mature peptide sequence is underlined.

Amplicon pyrosequencing confirmed 3/5, 17/27 and 19/33 changes previously detected in the signal mature peptide, and C-terminal regions, respectively [36] and revealed additional changeable positions. As observed in MytM, different SNC combinations increase the total number of possible peptide variants.

The maps locating all synonymous and non-synonymous SNCs along the transcript sequence (cds) of the remaining AMPs are reported in Supplementary Materials (File S4) for comparison.

PlosOne, 2011

Discussion

This study intended to assess, by high-throughput amplicon sequencing, the natural variability of nine AMP precursor sequences found expressed in the Mediterranean mussel. For this purpose, we sampled mussels from farming sites subjected to common European regulations and sanitary controls in three producer countries.

Thirteen sequences almost covering the AMP cds (File S3) have been successfully amplified from hemocyte RNA samples representing mussels farmed in France, Spain and Italy, as well as native mussels of the Venice lagoon area, before and after Vibrio injection, for coupled comparison.

Sample preparation is a decisive step of the sequencing workflow, due to the difficulty in preparing a well-balanced unique amplicon pool for the emulsion PCR and subsequent pyrosequencing. For this purpose, we measured and equalized the concentrations of each amplicon with great attention before pooling. Although the two sequencing half plates produced similar read numbers per sample, a general variability of coverage depth between AMP amplicons was finally evident (Table 3, Figure 1). Nevertheless, we obtained at least 500x amplicon coverage with one only exception (MytM in the Ve ft sample).

Stringent criteria were then used to identify genuine SNCs, removing false positives without losing substantial information, and we could retain most of the output reads (82% and 86% to exemplify Myticin C and Mytimycin, respectively). At first look, the output reads of each AMP appeared very diverse, 88% with at least one SNC on average (100% for MGDt, no matter from which sample). The analysis of SNC frequency per base enabled us to rank the selected AMPs on the basis of the transcript variability (Figure 4). Despite the invariance of the cysteine array, each AMP showed typical levels of diversity irrespective of the geographical origin, with a majority of common SNCs present in all samples (86%, i.e. 134 SNCs) and the Vi sample showing the greatest number of SNCs (292 in total). Moreover, we did not see evidence of increased AMP sequence diversity in farmed and native mussels injected with a high dose of live *Vibrio* cells. Compared to the samples prepared from 100 mussels (Pa, Vi, Ve, Ve nc), those prepared from 40 mussels (Ve ft, Ve nt) showed fewer SNCs (-11%); a fact indicating that the sample size can limit the amount of detectable sequence variants.

The ratio of non-synonymous vs. synonymous changes (ω) substantiates the evolutionary diversification of the mussel AMP isotypes and suggests the functional advantage of transcript variability for most of the analyzed AMPs (ω values higher than 1, indicative of positively selected residues, were frequently detected). However, the ω values did not reflect precisely the classification based on SNC frequencies and some

AMP (e.g. MytID) may have been subjected to higher evolutionary pressure than others.

The virtual translation of the transcript consensuses resulting from read clustering allowed the identification of AMP isotypes with a relatively low number of SNCs and high sequence diversity according to different SNC combinations (e.g. MytB and MGDt). In the case of mytimycin, we confirmed two major sequence types previously described [39] with no evidence of additional variants. For the remaining AMPs, it was not possible to identify specific patterns of variation (amino acid changes combined together without scheme).

In conclusion, the sequence data reported in this study further emphasize the sequence diversity of mussel AMP precursors. Redundant expression of diverse AMPs with a broad range of action could be regarded as a strategy to reinforce the host response against invaders (foes trying in their turn to escape detection and the host reactions) while the immune system also has to maintain the organism homeostasis with appropriate responses to commensal microbes (friends) and to danger signals released by damaged host cells [54]. On the other hand, environmental factors act as selective force only if they change the distribution of host genotypes (affecting only some genotypes, not all), thus influencing the immune system evolution of the host in the context of its life-history and population traits [55].

The isotype diversity levels found in this study might result from events occurring at DNA level as well as post-transcriptional changes such as deaminase-mediated cytidine to uridine transitions [14, 56-57]. Hence, targeted sequence enrichment and extension strategies applied to genomic DNA could identify active and remnant gene copies of each AMP isotype, and reveal the mechanisms underlying the observed sequence variation.

PlosOne, 2011

Materials and Methods

Sampling sites, treatment and RNA extraction

Adult mussels (*Mytilus galloprovincialis*) with a shell length of 6–8 cm and mixed sex were obtained from commercial shellfish stocks near Palavas (Pa, Mediterranean Sea, France; 43°31'49 N, 03°54'53 E), Ria de Vigo (Vi, Atlantic Ocean, Spain; 42º14'32 N, 08º48'26 E) and off-shore Venice (Ve, North Adriatic Sea, Italy; 45°18'29.8 N, 12°21'32.0 E). In addition, we collected wild mussels from the industrial canals of Porto Marghera (Lagoon of Venice, Italy; 45°27'33.5 N, 12°15'41 E). More than 100 animals per group were sampled.

According to the EU Directive 91/492, mussels cultivated in waters classified A (e.g. ¼ mile off-shore in the Adriatic Sea) can be marketed without depuration and are assumed not to contain potential pathogens nor biotoxins. Due to heavy mixed pollution, shellfishing was prohibited since 1996 in the area from the industrial district (P. Marghera) to the Venice town, though the overall shellfish quality can be improved by 2 month-depuration in type A waters. Mussels farmed offshore or living in the industrial canals (Venice lagoon area) were acclimatized for one week in sea water collected at flood tide (32‰, 22°C) and fed with Isochrisis galbana. Following shell notching, 0.1 ml of exponentially growing bacteria (10^7 V. splendidus *LGP32* cells) were injected into the posterior adductor muscle (samples Ve ft and Ve nt).

Hemolymph (1 ml per animal) was withdrawn from the posterior adductor muscle with a syringe containing 0.2 ml of Alsever solution (27 mM sodium citrate, 2.6 mM citric acid, 114 mM glucose and 72 mM NaCl in distilled water) adjusted at pH 7.4, and used to compose pools, each representative of 10 animals.

Haemocytes were pelleted by 15 min centrifugation at 800 xg (4 °C), carefully resuspended in 200 µl of TRIZOL reagent (Invitrogen, Carlsbad, USA) and stored at -80°C until use. Total RNA was isolated according to the manufacturer's instructions and resuspended in RNAse-free water. A further purification step with LiCl 2 M was applied to remove possible contaminants. RNA concentration was measured by UV-spectrometry (ND1000, NanoDrop Technologies, Wilmington, USA) and the RNA integrity was verified by microcapillary electrophoresis (RNA 6000 Nano LabChip, Agilent Technologies, Palo Alto, USA).

Finally, equal quantities of each RNA pool (N=10) were mixed together to compose a unique pool per sample (N= 100 mussels for samples Pa, Vi, Ve and Ve nc; N= 40 for samples Ve ft and Ve nt).

cDNA

cDNA was synthesized from 1 µg of total RNA using SuperScript II enzyme and oligo(dT)$_{18}$ primers (Invitrogen), following the manufacturer's instruction. To increase cDNA yield, the reaction was extended for a second hour, adding 0.5 µl of enzyme. cDNA was then purified with MinElute PCR Purification Kit (Qiagen, Hilden, Germany).

Primer design

Primer design was performed on the raw EST sequences denoting 9 mussel AMP isoforms in Mytibase (http://mussel.cribi.unipd.it).

All available ESTs for each selected AMP precursor were aligned using ClustalW [58] and primers were designed on conserved regions flanking the cds, whenever possible, with Primer3 [59]. Due to the pyrosequencing limit of about 250 bp, read length in forward and reverse direction, the maximum length of the PCR products was set at 440 bp (File S3). Degenerated primers have been designed to consider the whole ESTs variability (MytIC, MGD1 and MGDt). Two partially overlapping amplicons were designed for longer cds (MytC, MytIB, and MytIC) or in cases where high sequence variability made the definition of a single primer pair difficult (MytA). Thereby, 13 amplicons were designed in total. Amplicon's specificity was tested using the BlastX algorithm [60].

A tagged sequencing strategy with 5'nucleotide barcodes was implemented to facilitate the parallel processing of multiple samples [61-62]. Briefly, the forward and reverse PCR primers were modified by 5'-addition of 39 unique 5-mer barcodes (File S5). Barcodes enable the identification of the 454 reads corresponding to specific AMP, amplicon and sample so that PCR amplicons derived from multiple reactions could be combined for the sequencing run. To reduce the likelihood of misidentification, barcodes were designed not to contain homopolymers and to differ each one by at least two bases according to Roche Life Science protocols. Finally, 19-mer sequences corresponding to either the 454 Roche A Adaptor (for forward primers) or B Adaptor (for reverse primers) were fused to each PCR primer (Fusion primer).

Thermodynamic properties of Fusion primers were controlled to avoid the formation of hetero- or homo-dymer (OligoAnalyzer 3.1, http://eu.idtdna.com). Melting temperatures were fixed according to Primer3 software [63], adding 5 (±1) °C following HF Phusion polymerase instructions (Finnzymes, Espoo, Finland).

Fusion primers were designed in two sets of 39 (13 amplicons x 3 samples), with each primer pair having a unique barcode.

PlosOne, 2011

PCR amplification

The PCR amplifications of 78 amplicons were carried out individually in a PCR volume of 50 µl with 20 ng of cDNA template, 1× Phusion HF buffer, 0.2 mM dNTPs, 1 U HF Phusion DNA polymerase, 1.5 µl DMSO and 0.2 µM of both forward and reverse primers. Amplification was performed in a Mastercycler Gradient Thermal Cycler (Eppendorf, Hamburg, Germany) programmed as follows: 98 °C for 30 s followed by 35 cycles of 98 °C for 10 s, 60-65 °C for 20 s, 72 °C for 30 s and a final extension step at 72 °C for 5 min. The resulting amplification products were run on a 2% agarose gel and visualized by SYBR Gold staining (Invitrogen) using UV light transillumination (Gel Doc XR System, Bio-Rad, Hercules, USA). Unspecific small products and primer–dimers were removed using the Agencourt AMPure system (AmPure PCR Purification kit, Brea, USA) and amplicons integrity was confirmed with Agilent 2100 Bioanalyzer (DNA-1000 chip). Good quality amplicons were finally used to compose an equimolecular pool; the number of molecules of each amplicon was calculated with the following formula:

$$\text{Molecules}/\mu l = C \times N_A / (bp_w \times 10^9 \times bp)$$

C : sample concentration (ng/µl)
N_A: Avogadro constant
bp_w : average pair basis weight (g)
bp : pair basis number

Massively parallel 454 pyrosequencing (FLX-System, Roche Life Sciences, Branford, USA) was performed by BMR-Genomics (www.bmr-genomics.it) using two PicoTiter half plates. Reads have been recorded at the Sequence Read Archive accessible at http://www.ncbi.nlm.nih.gov/Traces/sra (submission ID: SRA038518.3). Six 454-output files corresponding to the 6 sequenced samples are available (SRR286638.1, SRR286639.1, SRR286640.1, SRR286641.1, SRR287657.1 and SRR287658.1).

Data analysis

Tag sequence were used as keys to part the unprocessed 454 reads into the 6 different samples by means of the GS Amplicon Variant Analyzer Software (AVAST, Roche Life Sciences). Reads in raw format were trimmed using quality score (limit 0.05) and minimum length equal to 100 bp. Subsequently, the output reads were aligned to a backbone consisting of the 9 AMP original transcripts, as obtained from the 13 reference sequences, with CLC Genomic Workbench version 4.6 (CLC Bio, Katrinrbjerg, Denmark).

The total number of sequenced bases divided by the length of the amplified transcript provided the average base coverage per AMP.
The reads of each mapping (AMP isotype) showing the same length and 100% similarity were clustered together. Single nucleotide changes (SNCs) were detected considering all the aligned reads of each mapping. Non-specific and low quality matches are ignored during the process and SNCs were considered genuine only when covered at least 30x, with a minimum frequency of 3%, and setting the quality level of the changed base and surrounding bases to at least 20 and 15, respectively. SNCs located in the same codon were merged. The expected amino acid changes in the precursor and cds peptide sequences were deduced by virtual translation, and the ratio (ω) between non-synonymous and synonymous changes was also computed. SNC frequency per base in the cds region was calculated for each AMP and sample with the following formula:

$$\text{SNC frequency} = (n\ SNC_{(AMP)}) / (nt_{seq})$$

nt_{seq}: cds sequenced nucleotides
$n\ SNC_{(AMP)}$: number of genuine SNC per amplicon in each sample

To assess possible differences in levels of sequence diversity between the AMP precursors amplified from each of the 6 samples, data were analysed with 1-way ANOVA (α=0.001). The null hypothesis predicted that all AMPs had the same variation rate in all samples. If not, Tukey's Honestly Significant Difference test (HSD, α=0.05) could then discriminate different AMP groups. Genuine SNCs were mapped on the 9 AMP sequences using CLC Genomic Workbench. The cds, signal peptide, mature peptide with the cysteine array and C-terminal regions were systematically localized.
Once the correct sequence reading frame was defined the sequences covered at least 3x were virtually translated to investigate the overall patterns of variation for each AMP transcript precursor in the Pa, Vi and Ve samples. Leftover low quality read ends were manually trimmed and redundancy was removed by using Jalview [64].

Acknowledgments

A special thank to Francesca Dal Pero for helpful advice.

Conclusions and perspectives

Interest in bivalve genomics has emerged during the last decade owing to the importance of these organisms in aquaculture and to their role in marine environmental science. Last findings indicate *Mytilus galloprovincialis* as a significant model for the study of the innate immune responses, as it lives constantly surrounded by many potential microscopic invaders in coastal waters.

During my PhD, I studied the transcriptome of the bivalve *Mytilus galloprovincialis*, with particular attention to gene transcripts participating in the innate immune responses. From one side, the investigation of complex cell responses only by high-throughput RNA analysis is quite restrictive, but, on the other side, large-scale approaches to the cell transcriptomes can outline the molecular processes modulated in response to the pathogen attack. After a first overview, independent validation experiments and functional studies are necessary to understand in more detail the biological meaning of the observed transcriptome changes.

In the last year, next generation sequencers (NGS) have produced a massive quantity of data which, in term of productivity and costs, are no more comparable with the amounts produced by more traditional sequencing strategies. However, the extraction of key biological data by huge sequence datasets is often difficult, particularly in organisms without a sequenced genome.

Both traditional and innovative sequencing have been used to prepare custom DNA microarray and investigate the transcriptional footprints of bivalve species for which the available sequence data are scarce, fragmentary and often biased in terms of transcript coverage (Venier *et al*., 2011; Milan *et al*., 2011). Such approach allows the selection of data subsets significant in specific conditions, helpful for more targeted studies, thus by-passing the lack of an annotation. Also new sequenced transcriptomes and genomes of phylogenetically related organisms are precious resources for comparative gene-centred studies.

In general, the innate immune responses are quite complex and cannot be understood by using one experimental approach only. DNA microarrays give a global overview of expression changes for thousands of genes, but only a specific and detailed analysis of the emerging immuno-modulated sequences in an holistic view of the cellular and organism processes can provide an advanced understanding of innate defensive strategies. So, I worked on nine antimicrobial peptides (AMPs) and, through PCR amplification and high coverage sequencing, I could show the great constitutive molecular diversity of the antimicrobial transcripts, irrespective to the mussel geographical origin and experimental treatment of immune stimulation.

Mussel AMPs are very intriguing molecules possibly having other functions in addition to the expected antimicrobial activity. The production of recombinant mussel AMPs could open the way to functional and structural studies. For example, a newly produced recombinant myticin C showed chemotactic and cytokine-like activities, as it induced the expression of effector molecules like myticin B, mytilin B and lysozyme (Balseiro *et al*., 2011). Yet, the AMPs immuno modulatory activity and their role as pathogen recognition receptors need verification. Work focussed on a single peptide would better clarify its activity against specific bacteria or fungi, as in the case of the antifungal mytimycin (Sonthi *et al*., 2012). In discussing the innate immunity of invertebrates, Sonthi *et al*., also mentioned the intriguing concept of a primitive immuno memory, which was not verified in the case of mytimycin. One possible way to resolve the debate might be to quantify the co-expression of many genes following a first and a second challenge, focusing then on particular genes like those immune effectors.

A very recent work (Gerdol *et al*., 2012), revealed the presence of eight novel big defensins and five novel macins (mytimacins, a new described mussel AMP) in the transcriptome of the Mediterranean mussel, indicating that the repertoire of antimicrobial peptides for this organism is intricate and also demonstrating that the advent of next generation sequencing technologies provided valuable tools for the *in silico* identification of previously uncharacterized protein families. Until the mussel transcriptome will not be completed, the number and families of AMPs remains unclear. Unclear is also the mechanism underlying the striking transcript diversity evident in many AMPs. The hypothesis of post-transcriptional modifications is intriguing but should be confirmed. Also the gene copy number of bivalve AMPs, can be responsible at least of a part of the known AMP isoforms (e.g.: oyster defensins, Schmitt *et al*., 2010), however, knowledge on the mussel AMP genes is currently limited. Sonthi *et al*., (2011) and Vera *et al*. (2011) have reported the gene structure of mytimycin and myticin C, respectively, and supposed the presence of two gene copies per genome in both cases.

The study of immunorelated transcript sequences firstly described in Mytibase has already produced remarkable results: C1q domain-containing proteins (Gerdol *et al*., 2011), caspase gene family (Romero *et al*., 2011) and fibrinogen-related proteins (Romero *et al*., 2011), as examples. Interesting is also the exploration of yet non-elucidated innate defence mechanisms, like a macrofage inhibitory factor, well known in other organisms as regulatory cytokine released in cells during non-classical complement activation pathways (Parisi *et al*., 2011) or mytimycin, a first antifungal peptide reported in mussel (Sonthi *et al*., 2011 and 2012).

Increasing the sequence data, with the complete coverage of the mussel transcriptome

being expected, new structural and functional studies will be possible, and both actors and mechanisms of the innate mussel immunity will be elucidated. With them, also the secrets of the mussel resilience to typical bivalve disease will be hopefully solved.

References

- Aceto S, Formisano G, Carella F, De Vico G, Gaudio L. The metallothionein genes of Mytilus galloprovincialis: genomic organization, tissue expression and evolution. Mar Genomics 4(1):61-8, 2011.

- Balseiro P, Falcó A, Romero A, Dios S, Martínez-López A, Figueras A, Estepa A, Novoa B. Mytilus galloprovincialis myticin C: a chemotactic molecule with antiviral activity and immunoregulatory properties. PLoS One,6(8):e23140, 2011.Bonn, 2008

- Bower SM. Diseases and parasites of mussels. In The mussel Mytilus: ecology, physiology, genetics and culture, (ed. E.M. Gosling), pp. 543-563, 1992.

- Braby CE, Somero GN. Following the heart: temperature and salinity effects on heart rate in native and invasive species of blue mussels (genus Mytilus). J Exp Biol. Jul;209:2554-66, 2006.

- Breton S, Burger G, Stewart DT, Blier PU. Comparative analysis of gender-associated complete mitochondrial genomes in marine mussels (Mytilus spp.). Genetics. Feb;172(2):1107-19, 2006.

- Craft JA, Gilbert JA, Temperton B, Dempsey KE, Ashelford K, Tiwari B, Hutchinson TH, Chipman JK. Pyrosequencing of Mytilus galloprovincialis cDNAs: tissue-specific expression patterns. PLoS One, 5(1):e8875, 2010.

- Daguin C, Borsa P. Genetic relationships of Mytilus galloprovincialis Lmk. Populations worldwide: evidence from nuclear DNA markers. Bivalve Systematics and Evolution. Geological Society of London Special Publications 177,389-397, 2000.

- FAO data (accessible at www.fao.org/fi)

- Gérard K, Bierne N, Borsa P, Chenuil A, Féral JP. Pleistocene separation of mitochondrial lineages of Mytilus spp. mussels from Northern and Southern Hemispheres and strong genetic differentiation among southern populations. Mol Phylogenet Evol. 49(1):84-91, 2008.

- Gerdol M, De Moro G, Manfrin C, Venier P, Pallavicini A. Big defensins and mytimacins, new AMP families of the Mediterranean mussel Mytilus galloprovincialis. ev Comp Immunol,36(2):390-9, 2012.

- Goldberg ED, Bertine KK. Beyond the Mussel Watch--new directions for monitoring marine pollution. Sci Total Environ, 20;247(2-3):165-74, 2000.

- Gosling EM. Genetics of *Mytilus*. In *The mussel* Mytilus*:* ecology, physiology, genetics and culture*,* ed. E.M. Gosling, pp. 1-20, 1992.

- Gosling EM. Bivalve molluscs: biology, ecology, culture. Fishing News Books, Oxford, 2003.

- Gregory TR. Animal Genome Size Database, 2005

- Hilbish TJ, Mullinax A, Dolven SI, Meyer A, Koehn RK, Rawson PD. Origin of the antitropical distribution pattern in the marine mussels (Mytilus spp.): routes and timing of transequatorial migration. Mar. Biol, 136, 2000.

- Hoeh WR, Stewart DT, Saavedra C, Sutherland BW, Zouros E. Phylogenetic evidence for role-reversals of gender-associated mitochondrial DNA in Mytilus (Bivalvia: Mytilidae). Mol Biol Evol, 14(9):959-67, 1997.

- Hunt CD, Slone E. Long-term monitoring using resident and caged mussels in Boston Harbor yield similar spatial and temporal trends in chemical contamination. Mar Environ Res. Dec;70(5):343-57, 2010.

- Kimbrough KL, Commey S, Apeti DA, Lauenstein GG. Chemical contamination assessment of the Hudson-Raritan Estuary as a result of the attacks on the World Trade Center: analysis of trace elements. Mar Pollut Bull. Dec;60(12):2289-96, 2010.

- Lockwood BL, Sanders JG, Somero GN. Transcriptomic responses to heat stress in invasive and native blue mussels (genus Mytilus): molecular correlates of invasive success. J. Exp. Biol. 213(Pt 20): 3548-3558, 2010.

- Luckas B, Dahlmann J, Erler K, Gerdts G, Wasmund N, Hummert C, Hansen PD. Overview of key phytoplankton toxins and their recent occurrence in the North and Baltic Seas. Environ Toxicol 20: 1–17, 2005.

- Lutz A and Kennish MJ. Ecology and morphology of larval and early postlarval mussels. In The mussel Mytilus: ecology, physiology, genetics and culture, (ed. E.M. Gosling), pp. 87-169, 1992.

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

- Mitchelmore CL, Chipman JK. DNA strand breakage in aquatic organisms and the potential value of the comet assay in environmental monitoring. Mutat Res 399: 135–147, 1998.

- Moore RC. Treatise of invertebrate paleontology. Geological society of America and university of Kansas press, pp 1953-1983, 1983.

- Pantzartzi C, Drosopoulou E, Yiangou M, Drozdov I, Tsoka S, Ouzounis CA, et al. Promoter complexity and tissue-specific expression of stress response components in Mytilus galloprovincialis, a sessile marine invertebrate species. PLoS Comput. Biol. 6: e1000847, 2010.

- Parisi MG, Toubiana M, Mangano V, Parrinello N, Cammarata M, Roch P. MIF from mussel: Coding sequence, phylogeny, polymorphism, 3D model and regulation of expression. Dev Comp Immunol, Nov 4, 2011.

- Romero A, Estévez-Calvar N, Dios S, Figueras A, Novoa B. New insights into the apoptotic process in mollusks: characterization of caspase genes in Mytilus galloprovincialis. PLoS One, 11;6(2):e17003, 2011.

- Romero A, Dios S, Poisa-Beiro L, Costa MM, Posada D, Figueras A, Novoa B. 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(3):334-44, 2011.

- Saavedra C, Bachère E. Bivalve genomics. Aquaculture, 256: 1-14, 2006.

- Scarpato A, Romanelli G, Galgani F, Andral B, Amici M, et al. Western Mediterranean coastal waters--monitoring PCBs and pesticides accumulation in *Mytilus galloprovincialis* by active mussel watching: the Mytilos project. J Environ Monit. Apr;12(4):924-35, 2010.

- Schmitt P, Wilmes M, Pugnière M, Aumelas A, Bachère E, Sahl HG, Schneider T, Destoumieux-Garzón D. Insight into invertebrate defensin mechanism of action: oyster defensins inhibit peptidoglycan biosynthesis by binding to lipid II. J Biol Chem, 17;285(38):29208-16, 2010.

- Shumway SE. Mussels and public health. In The mussel Mytilus: ecology, physiology, genetics and culture, (ed. E.M. Gosling), pp. 511-542, 1992.

References

- Seed R, Suchanek TH. Population and community ecology of Mytilus. In The mussel Mytilus: ecology, physiology, genetics and culture, (ed. E.M. Gosling), pp. 87-169, 1992.
- Seed R. The ecology of Mytilus edulis L. (Lamellibranchiata) on exposed rocky shores. Growth and mortality. Oecologia, 3, 317-350, 1969.
- Sprung M. Physiological energetics of mussel larvae (*Mytilus edulis*). I. Shell growth and biomass. *Marine Ecology Progress Series*, 17, 295-305, 1984.
- Sonthi M, Toubiana M, Pallavicini A, Venier P, Roch P. Diversity of Coding Sequences and Gene Structures of the Antifungal Peptide Mytimycin (MytM) from the Mediterranean Mussel, *Mytilus galloprovincialis*. Mar Biotechnol (NY) 13:857–867, 2011.
- Sonthi M, Cantet F, Toubiana M, Trapani MR, Parisi MG, Cammarata M, Roch P. Gene expression specificity of the mussel antifungal mytimycin (MytM). Fish Shellfish Immunol,32(1):45-50, 2012.
- Thiesen BF. The growth of *Mytilus edulis* L. (Bivalvia) from Disko and Thule district, Greenland. *Ophelia*, 12, 59-77, 1973.
- Thompson IS, Richardson CA, Seed R, Walker G. Quantification of mussel (*Mytilus edulis*) growth from power station cooling waters in response to chlorination procedures. Biofouling,16, 1-15, 2000.
- Venier P, De Pittà C, Pallavicini A, Marsano F, Varotto L, Romualdi C, *et al*. Development of mussel mRNA profiling: Can gene expression trends reveal coastal water pollution? Mutat. Res. 602: 121-134, 2006.
- Vera M, Martínez P, Poisa-Beiro L, Figueras A, Novoa B. Genomic organization, molecular diversification, and evolution of antimicrobial peptide myticin-C genes in the mussel (*Mytilus galloprovincialis*). PLoS One,6(8):e24041, 2011.
- Widdows J, Donkin P, Staff FJ, Matthiessen P, Law RJ. Measurement of stress effects (scope for growth) and contaminant levels in mussels (*Mytilus edulis*) collected from the Irish Sea. Mar Environ Res 53: 327–356, 2002.

1. Lefranc MP, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, Wu Y, Gemrot E, Brochet X, Lane J, Regnier L, Ehrenmann F, Lefranc G, Duroux P, IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res* 2009, 37(Database issue):D1006-12.
2. Ortutay C, Vihinen M, Immunome knowledge base (IKB): an integrated service for immunome research. *BMC Immunol* 2009, 9:10:3.
3. *Cossins A., Somero G, Guest editors'introduction.* J Exp Biol *2007,* 210 (Pt 9): 1491.
4. Hedgecock D, Lin JZ, DeCola S, Haudenschild CD, Meyer E, Manahan DT, Bowen B, Transcriptomic analysis of growth heterosis in larval Pacific oysters (*Crassostrea gigas*). *Proc Natl Acad Sci U S A.* 2007, 104(7):2313-8.
5. Waterhouse RM, Kriventseva EV, Meister S, Xi Z, Alvarez KS, Bartholomay LC, Barillas-Mury C, Bian G, Blandin S, Christensen BM, Dong Y, Jiang H, Kanost MR, Koutsos AC, Levashina EA, Li J, Ligoxygakis P, Maccallum RM, Mayhew GF, Mendes A, Michel K, Osta MA, Paskewitz S, Shin SW, Vlachou D, Wang L, Wei W, Zheng L, Zou Z, Severson DW, Raikhel AS, Kafatos FC, Dimopoulos G, Zdobnov EM, Christophides GK, Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* 2007, 316(5832):1738-43.
6. Goldstone JV, Environmental sensing and response genes in cnidaria: the chemical defensome in the sea anemone *Nematostella vectensis*. *Cell Biol Toxicol* 2008, 24(6):483-502.
7. Gracey AY, Chaney ML, Boomhower JP, Tyburczy WR, Connor K, Somero GN, Rhythms of gene expression in a fluctuating intertidal environment. *Curr Biol* 2008, 18(19):1501-7.
8. Fleury E, Huvet A, Lelong C, de Lorgeril J, Boulo V, Gueguen Y, Bachère E, Tanguy A, Moraga D, Fabioux C, Lindeque P, Shaw J, Reinhardt R, Prunet P, Davey G, Lapègue S, Sauvage C, Corporeau C, Moal J, Gavory F, Wincker P, Moreews F, Klopp C, Mathieu M, Boudry P, Favrel P, Generation and analysis of a 29,745 unique Expressed Sequence Tags from the Pacific oyster (*Crassostrea gigas*) assembled into a publicly accessible database: the GigasDatabase. *BMC Genomics* 2009, 10:341.
9. Messier-Solek C, Buckley KM, Rast JP, Highly diversified innate receptor systems and new forms of animal immunity. *Semin Immunol* 2010, 22(1):39-47.
10. Thompson FL, Gevers D, Thompson CC, Dawyndt P, Naser S, Hoste B, Munn CB, Swings J, Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. *Appl Environ Microbiol* 2005, 71(9):5107-15.
11. Gómez-León J, Villamil L, Lemos ML, Novoa B, Figueras A*, Isolation of Vibrio alginolyticus* and *Vibrio splendidus* from aquacultured carpet shell clam (*Ruditapes decussatus*) larvae associated with mass mortalities. *Appl Environ Microbiol* 2005, 71(1):98-104.
12. Igbinosa EO, Okoh AI, Emerging *Vibrio* species: an unending threat to public health in developing countries. *Res Microbiol* 2008, 159(7-8):495-506.
13. Smolarz K, Wołowicz M, Stachnik M, First record of the occurrence of "gill disease" in *Mytilus edulis trossulus* from the Gulf of Gdańsk (Baltic Sea, Poland) *J Invertebr Pathol* 2006, 93(3):207-9.
14. Samain JF, McCombie H: *Summer mortality of Pacific oyster Crassostrea virginica. The Morest project*. Versailles Cedex: Editions Quae; 2007 (http://www.ifremer.fr/morest-gigas).
15. Francisco CJ, Hermida MA, Santos MJ, Parasites and symbionts from *Mytilus galloprovincialis* (Lamark, 1819) (Bivalves: Mytilidae) of the Aveiro Estuary Portugal. *J Parasitol* 2010, 96(1):200-5.
16. Watermann BT, Herlyn M, Daehne B, Bergmann S, Meemken M, Kolodzey H, Pathology and mass mortality of Pacific oysters, *Crassostrea gigas* (Thunberg), in 2005 at the East Frisian coast, Germany. *J Fish Dis* 2008, 31(8):621-30.

17. Gestal C, Roch F, Renault T, Pallavicini A, Paillard C, Novoa B, Oubella R, Venier P, Figueras A, Study of diseases and the immune system of bivalves using molecular biology and genomics. *Rev Fisheries Sci* 2008, 16(S1): 131-154.
18. Lhafi SK, Kühne M, Occurrence of *Vibrio* spp. in blue mussels (*Mytilus edulis*) from the German Wadden Sea. *Int J Food Microbiol* 2007, 116(2):297-300.
19. Magnadottir B, Immunological Control of Fish Diseases. *Mar Biotechnol (NY)* 2010, 12(4):361-79.
20. Schneeweiss H, Renwrantz L, Analysis of the attraction of haemocytes from *Mytilus edulis* by molecules of bacterial origin. *Dev Comp Immunol* 1993, 17(5):377-87.
21. Hubert F, Noel T, Roch P, A member of the arthropod defensin family from edible Mediterranean mussels (*Mytilus galloprovincialis*). *Eur J Biochem* 1996, 240(1):302-6. Erratum in: *Eur J Biochem* 1996, 240(3):815.
22. Carballal MJ., Villalba A., Lopez C, Seasonal variation and effects of age, food availability, size gonadal development, and parasitism on the hemogram of *Mytilus galloprovincialis*. *J Invertebr Pathol* 1998, 72: 304-312.
23. Guillou F, Mitta G, Galinier R, Coustau C, Identification and expression of gene transcripts generated during an anti-parasitic response in *Biomphalaria glabrata*. *Dev Comp Immunol* 2007, 31(7):657-71.
24. Kuchel RP, Aladaileh S, Birch D, Vella N, Raftos DA., Phagocytosis of the protozoan parasite, *Marteilia sydneyi*, by Sydney rock oyster (*Saccostrea glomerata*) hemocytes. *J Invertebr Pathol* 2010, 104(2):97-104.
25. Endo Y, Nakazawa N, Iwaki D, Takahashi M, Matsushita M, Fujita T, Interactions of ficolin and mannose-binding lectin with fibrinogen/fibrin augment the lectin complement pathway. *J Innate Immun* 2010, 2(1):33-42.
26. García-García E, Prado-Alvarez M, Novoa B, Figueras A, Rosales C, Immune responses of mussel hemocyte subpopulations are differentially regulated by enzymes of the PI 3-K, PKC, and ERK kinase families. *Dev Comp Immunol* 2008, 32(6):637-53.
27. Donaghy L, Kim BK, Hong HK, Park HS, Choi KS, Flow cytometry studies on the populations and immune parameters of the hemocytes of the Suminoe oyster, *Crassostrea ariakensis*. *Fish Shellfish Immunol.* 2009, 27(2):296-301.
28. Le Foll F, Rioult D, Boussa S, Pasquier J, Dagher Z, Leboulenger F, Characterisation of *Mytilus edulis* hemocyte subpopulations by single cell time-lapse motility imaging. *Fish Shellfish Immunol* 2010, 28(2):372-86.
29. Hughes TK Jr, Smith EM, Leung MK, Stefano GB, Evidence for the conservation of an immunoreactive monokine network in invertebrates. *Ann N Y Acad Sci* 1992, 650:74-80.
30. Liu DW, Opioid peptides and innate immune response in mollusc. *Protein Pept Lett* 2008, 15(7):683-6.
31. Hagger JA, Jones MB, Lowe D, Leonard DR, Owen R, Galloway TS, Application of biomarkers for improving risk assessments of chemicals under the Water Framework Directive: a case study. *Mar Pollut Bull* 2008, 56(6):1111-8.
32. Auffret M, Rousseau S, Boutet I, Tanguy A, Baron J, Moraga D, Duchemin M: A multiparametric approach for monitoring immunotoxic responses in mussels from contaminated sites in Western Mediterranea. *Ecotoxicol Environ Saf* 2006, 63(3):393-405.
33. Canesi L, Barmo C, Fabbri R, Ciacci C, Vergani L, Roch P, Gallo G, Effects of vibrio challenge on digestive gland biomarkers and antioxidant gene expression in *Mytilus galloprovincialis*. Comp *Biochem Physiol C Toxicol Pharmacol* 2010, 152(3):399-406.
34. Tanguy A, Bierne N, Saavedra C, Pina B, Bachère E, Kube M, Bazin E, Bonhomme F, Boudry P, Boulo V, Boutet I, Cancela L, Dossat C, Favrel P, Huvet A, Jarque S, Jollivet D, Klages S, Lapègue S, Leite R,

Moal J, Moraga D, Reinhardt R, Samain JF, Zouros E, Canario A, Increasing genomic information in bivalves through new EST collections in four species: development of new genetic markers for environmental studies and genome evolution. *Gene* 2008, 408(1-2):27-36.

35. Venier P, De Pittà C, Pallavicini A, Marsano F, Varotto L, Romualdi C, Dondero F, Viarengo A, Lanfranchi G, Can gene expression trends reveal coastal water pollution? *Mutat Res* 2006, 602: 121-134.
36. Craft JA, Gilbert JA, Temperton B, Dempsey KE, Ashelford K, Tiwari B, Hutchinson TH, Chipman JK, Pyrosequencing of *Mytilus galloprovincialis* cDNAs: tissue-specific expression patterns. *PLoS One* 2010, 5(1):e8875.
37. Venier P, De Pittà C, Bernante F, Varotto L, De Nardi B, Bovo G, Roch P, Novo B, Figueras A, Pallavicini A, Lanfranchi G, MytiBase: a knowledgebase of mussel (*M. galloprovincialis)* transcribed sequences. *BMC Genomics* 2009, 10:72.
38. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C, InterPro: the integrative protein signature database *Nucleic Acids Res* 2009, 37:D224-228.
39. Steinstraesser L, Kraneburg U, Jacobsen F, Al-Benna S, Host defense peptides and their antimicrobial-immunomodulatory duality. *Immunobiology*, in press.
40. Charlet M, Chernysh S, Philippe H, Hetru C, Hoffmann JA, Bulet P, Innate immunity. Isolation of several cysteine-rich antimicrobial peptides from the blood of a mollusc, *Mytilus edulis*. *J Biol Chem* 1996, 271(36):21808-13.
41. Mitta G, Vandenbulcke F, Roch P, Original involvement of antimicrobial peptides in mussel innate immunity. *FEBS Lett* 2000, 486(3):185-90.
42. Yeaman MR Yount NY, Unifying themes in host defence effector polypeptides. *Nat Rev Microbiol* 2007, 5(9):727-40.
43. Pallavicini A, Costa M. M, Gestal C, Dreos R, Figueras A, Venier P, Novoa B, High sequence variability of myticin transcripts in hemocytes of immune-stimulated mussels suggests ancient host-pathogen interactions. *Dev Comp Immunol* 2008, 32(3):213-26.
44. Shapiro L, Scherer PE, The crystal structure of a complement-1q family protein suggests an evolutionary link to tumor necrosis factor. *Curr Biol* 1998, 8(6):335-8.
45. Carland TM, Gerwick L, The C1q domain containing proteins: Where do they come from and what do they do? *Dev Comp Immunol* 2010, 34(8):785-90.
46. Nayak A, Ferluga J, Tsolaki AG, Kishore U, The non-classical functions of the classical complement pathway recognition subcomponent C1q. *Immunol Lett* 2010, 131(2):139-50.
47. Kojouharova M, Reid K, Gadjeva M, New insights into the molecular mechanisms of classical complement activation. *Mol Immunol* 2010, 47(13):2154-60.
48. Liu HL, Liu SF, Ge YJ, Liu J, Wang XY, Xie LP, Zhang RQ, Wang Z, Identification and characterization of a biomineralization related gene PFMG1 highly expressed in the mantle of *Pinctada fucata*. *Biochemistry* 2007, 46(3):844-51.
49. Gestal C, Pallavicini A, Venier P, Novoa B, Figueras A, MgC1q, a novel C1q-domain-containing protein involved in the immune response of *Mytilus galloprovincialis*. *Dev Comp Immunol* 2010, 34(9):926-34.
50. Dodds AW, Matsushita M, The phylogeny of the complement system and the origins of the classical pathway. *Immunobiology* 2007, 212(4-5):233-43.
51. Huang S, Yuan S, Guo L, Yu Y, Li J, Wu T, Liu T, Yang M, Wu K, Liu H, Ge J, Yu Y, Huang H, Dong M, Yu C, Chen S, Xu A, Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary

innate complexity and diversity. *Genome Res* 2008, 18(7):1112-26.

52. Kilpatrick DC, Animal lectins: a historical introduction and overview. *Biochim Biophys Acta* 2002, 1572(2-3):187-97.
53. McGreal EP, Martinez-Pomares L, Gordon S, Divergent roles for C-type lectins expressed by cells of the innate immune system. *Mol Immunol* 2004, 41(11):1109-21.
54. Vasta G.R, Roles of galectins in infection. *Nat Rev Microbiol* 2009, 7(6):424-38.
55. Dodd RB, Drickramer K, Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. *Glycobiology* 2001, 11(5):71R-79R.
56. Gabius HJ, Cell surface glycans: the why and how of their functionality as biochemical signals in lectin-mediated information transfer. *Crit Rev Immunol* 2006, 26(1):43-79.
57. Weiss IM, Kaufmann S, Mann K, Fritz M, Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the mollusc *Haliotis laevigata*. *Biochem Biophys Res Commun* 2000, 267(1):17-21.
58. Schulenburg H, Hoeppner MP, Weiner J 3rd, Bornberg-Bauer E, Specificity of the innate immune system and diversity of C-type lectin domain (CTLD) proteins in the nematode *Caenorhabditis elegans*. *Immunobiology* 2008, 213(3-4):237-50.
59. Matsushita M, Ficolins: complement-activating lectins involved in innate immunity. *J Innate Immun* 2010, 2(1):24-32.
60. Zhang SM, Zeng Y, Loker ES, Expression profiling and binding properties of fibrinogen-related proteins (FREPs), plasma proteins from the schistosome snail host *Biomphalaria glabrata*. *Innate Immun* 2008, 14(3):175-89.
61. Söderhäll I, Wu C, Novotny M, Lee BL, Söderhäll K, A novel protein acts as a negative regulator of prophenoloxidase activation and melanization in the freshwater crayfish *Pacifastacus leniusculus*. *J Biol Chem* 2009, 284(10):6301-10.
62. Mali B, Möhrlen F, Frohme M, Frank U, A putative double role of a chitinase in a cnidarian: pattern formation and immunity. *Dev Comp Immunol* 2004, 28(10):973-81.
63. Lee CG, Chitin, chitinases and chitinase-like proteins in allergic inflammation and tissue remodeling. *Yonsei Med J* 2009, 50(1):22-30.
64. Martinon F, Mayor A, Tschopp J, The inflammasomes: guardians of the body. *Annu Rev Immunol* 2009, 27:229-65.
65. Fitzgerald KA, NLR-containing inflammasomes: central mediators of host defense and inflammation. *Eur J Immunol* 2010, 40(3):595-8.
66. Kawai T, Akira S, The roles of TLRs, RLRs and NLRs in pathogen recognition. *Int Immunol* 2009, 21(4):317-37.
67. Sabin LR, Hanna SL, Cherry S, Innate antiviral immunity in *Drosophila*. *Curr Opin Immunol* 2010, 22(1):4-9.
68. Hibino T, Loza-Coll M, Messier C, Majeske AJ, Cohen AH, Terwilliger DP, Buckley KM, Brockton V, Nair SV, Berney K, Fugmann SD, Anderson MK, Pancer Z, Cameron RA, Smith LC, Rast JP, The immune gene repertoire encoded in the purple sea urchin genome. *Dev Biol* 2006, 300(1):349-65.
69. Satake H, Sasaki N, Comparative overview of toll-like receptors in lower animals. *Zoolog Sci* 2010, 27(2):154-61.
70. Lodowski DT, Palczewski K, Chemokine receptors and other G protein-coupled receptors. *Curr Opin HIV AIDS* 2009, 4(2):88-95.
71. Walker AJ, Lacchini AH, Sealey KL, Mackintosh D, Davies AJ, Spreading by snail (*Lymnaea stagnalis*) defence cells is regulated through integrated PKC, FAK and Src signalling. *Cell Tissue Res* 2010,

341(1):131-45.

72. Yang S, Wu X, Identification and functional characterization of a human sTRAIL homolog, CasTRAIL, in an invertebrate oyster *Crassostrea ariakensis*. *Dev Comp Immunol* 2010, 34(5):538-45.
73. Ciacci C, Betti M, Canonico B, Citterio B, Roch P, Canesi L, Specificity of anti-Vibrio immune response through p38 MAPK and PKC activation in the hemocytes of the mussel *Mytilus galloprovincialis*. *J Invertebr Pathol* 2010, 105(1):49-55.
74. Roberts S, Gueguen Y, de Lorgeril J, Goetz F, Rapid accumulation of an interleukin 17 homolog transcript in Crassostrea gigas hemocytes following bacterial exposure. *Dev Comp Immunol* 2008, 32(9):1099-104.
75. Costa MM, Prado-Alvarez M, Gestal C, Li H, Roch P, Novoa B, Figueras A, Functional and molecular immune response of Mediterranean mussel (*Mytilus galloprovincialis*) haemocytes against pathogen-associated molecular patterns and bacteria. *Fish Shellfish Immunol* 2009, 26(3):515-23.
76. Kelly EK, Wang L, Ivashkiv LB, Calcium-Activated Pathways and Oxidative Burst Mediate Zymosan-Induced Signalling and IL-10 Production in Human Macrophages. *J Immunol* 2010, 184(10):5545-52.
77. Hoarau P, Damiens G, Roméo M, Gnassia-Barelli M, Bebianno MJ Cloning and expression of a GST-pi gene in *Mytilus galloprovincialis*. Attempt to use the GST-pi transcript as a biomarker of pollution. *Comp Biochem Physiol C Toxicol Pharmacol* 2006, 143(2):196-203.
78. Dondero F, Piacentini L, Marsano F, Rebelo M, Vergani L, Venier P, Viarengo A, Gene transcription profiling in pollutant exposed mussels (*Mytilus* spp.) using a new low-density oligonucleotide microarray. *Gene* 2006, 376 (1), 24-36.
79. Kolaiti RM, Lucas JM, Kouyanou-Koutsoukou S, Molecular cloning of the ribosomal P-proteins MgP1, MgP2, MgP0, and superoxide dismutase (SOD) in the mussel *Mytilus galloprovincialis* and analysis of MgP0 at stress conditions. *Gene* 2009, 430(1-2):77-85.
80. Bilbao E, Cajaraville MP, Cancio I, Cloning and expression pattern of peroxisomal beta-oxidation genes palmitoyl-CoA oxidase, multifunctional protein and 3-ketoacyl-CoA thiolase in mussel *Mytilus galloprovincialis* and thicklip grey mullet *Chelon labrosus*. *Gene* 2009, 443(1-2):132-42.
81. Song L, Wu L, Ni D, Chang Y, Xu W, Xing K, The cDNA cloning and mRNA expression of heat shock protein 70 gene in the haemocytes of bay scallop (*Argopecten irradians*, Lamarck 1819) responding to bacteria challenge and naphthalin stress. *Fish Shellfish Immunol* 2006, 21(4):335-45.
82. Cellura C, Toubiana M, Parrinello N, Roch P, Specific expression of antimicrobial peptide and HSP70 genes in response to heat-shock and several bacterial challenges in mussels. *Fish Shellfish Immunol* 2007, 22(4):340-50.
83. Gao Q, Zhao J, Song L, Qiu L, Yu Y, Zhang H, Ni D, Molecular cloning, characterization and expression of heat shock protein 90 gene in the haemocytes of bay scallop *Argopecten irradians*. *Fish Shellfish Immunol* 2008, 24(4):379-85.
84. Paquette N, Broemer M, Aggarwal K, Chen L, Husson M, Ertürk-Hasdemir D, Reichhart JM, Meier P, Silverman N, Caspase-mediated cleavage, IAP binding, and ubiquitination: linking three mechanisms crucial for *Drosophila* NF-kappaB signalling. *Mol Cell* 2010, 37(2):172-82.
85. Kim JY, Park SC, Hwang I, Cheong H, Nah JW, Hahm KS, Park Y, Protease inhibitors from plants with antimicrobial activity. *Int J Mol Sci* 2009, 10(6):2860-72.
86. Augustin R, Siebert S, Bosch TC, Identification of a kazal-type serine protease inhibitor with potent anti-staphylococcal activity as part of *Hydra*'s innate immune system. *Dev Comp Immunol* 2009, 33(7):830-7.
87. Groettrup M, Kirk CJ, Basler M, Proteasomes in immune cells: more than peptide producers? *Nat Rev Immunol* 2010, 10(1):73-8.
88. Li H, Parisi MG, Toubiana M, Cammarata M, Roch P, Lysozyme gene expression and hemocyte behaviour in the Mediterranean mussel, *Mytilus galloprovincialis*, after injection of various bacteria or temperature

stresses. *Fish Shellfish Immunol* 2008, 25(1-2):143-52.

89. Li H, Venier P, Prado-Alvárez M, Gestal C, Toubiana M, Quartesan R, Borghesan F, Novoa B, Figueras A, Roch P, Expression of *Mytilus* immune genes in response to experimental challenges varied according to the site of collection. *Fish Shellfish Immunol* 2010, 28(4):640-8.
90. Ottaviani E, Franceschi C, The neuroimmunology of stress from invertebrates to man. *Prog Neurobiol* 1996, 48(4-5):421-40.
91. Kasahara S, Cooper EL, Nervous, endocrine, immune systems as a target for complementary and alternative medicine. *Adv Exp Med Biol* 2004, 546:405-23.
92. Hauser F, Susanne N, Williamson M, Predel R, Tanaka Y, Grimmelikhuijzen CJ, Genomics and peptidomics of neuropeptides and protein hormones present in the parasitic wasp *Nasonia vitripennis*. *J Proteome Res* 2010, 9(10):5296-310.
93. Rast JP, Messier-Solek C, Marine invertebrate genome sequences and our evolving understanding of animal immunity. *Biol Bull* 2008, 214(3):274-83.
94. Li W, Gracey AY, Mello LV, Brass A, Cossins AR, ExprAlign--the identification of ESTs in non-model species by alignment of cDNA microarray expression profiles. *BMC Genomics* 2009, 10:560.
95. Nemeth E, Baird AW, O'Farrelly C, Microanatomy of the liver immune system. *Semin Immunopathol* 2009, 31(3):333-43.
96. Schäffler A, Schölmerich J, Innate immunity and adipose tissue biology. *Trends Immunol* 2010, 31(6):228-35.
97. Thomas PG, Doherty PC, New approaches in immunotherapy. *Science* 2010, 327(5963):249.
98. Venier P, Maron S, Canova S, Detection of micronuclei in gill cells and haemocytes of mussels exposed to benzo[a]pyrene. *Mutat Res* 1997, 390(1-2):33-44.
99. Li H, Toubiana M, Monfort P, Roch P, Influence of temperature, salinity and *E. coli* tissue content on immune gene expression in mussel: results from a 2005-2008 survey. *Dev Comp Immunol* 2009, 33(9):974-9.
100. Pruzzo C, Gallo G, Canesi L, Persistence of vibrios in marine bivalves: the role of interactions with haemolymph components. *Environ Microbiol* 2005, 7(6):761-72.
101. Labreuche Y, Soudant P, Gonçalves M, Lambert C, Nicolas JL, Effects of extracellular products from the pathogenic *Vibrio aestuarianus* strain 01/32 on lethality and cellular immune responses of the oyster *Crassostrea gigas*. *Dev Comp Immunol* 2006, 30(4):367-79.
102. Saulnier D, De Decker S, Haffner P, Cobret L, Robert M, Garcia C, A Large-Scale Epidemiological Study to Identify Bacteria Pathogenic to Pacific Oyster *Crassostrea gigas* and Correlation Between Virulence and Metalloprotease-like Activity. *Microb Ecol* 2010, 59(4):787-98.
103. Travers MA, Le Bouffant R, Friedman CS, Buzin F, Cougard B, Huchette S, Koken M, Paillard C, Pathogenic *Vibrio harveyi*, in contrast to non-pathogenic strains, intervenes with the p38 MAPK pathway to avoid an abalone haemocyte immune response. *J Cell Biochem* 2009, 106(1):152-60.
104. Gourgou E, Aggeli IK, Beis I, Gaitanaki C, Hyperthermia-induced Hsp70 and MT20 transcriptional upregulation are mediated by p38-MAPK and JNKs in *Mytilus galloprovincialis* (Lamarck); a pro-survival response. *J Exp Biol* 2010, 213(2):347-57.
105. Parisi MG, Li H, Jouvet LB, Dyrynda EA, Parrinello N, Cammarata M, Roch P, Differential involvement of mussel hemocyte sub-populations in the clearance of bacteria. *Fish Shellfish Immunol* 2008, 25(6):834-40.
106. Bettencourt R, Pinheiro M, Egas C, Gomes P, Afonso M, Shank T, Santos RS.High-throughput sequencing and analysis of the gill tissue transcriptome from the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus*. *BMC Genomics* 2010, 11:559.

107. Schmucker D Chen B, Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. *Genes Dev.* 2009, 23(2):147-56.
108. Dheilly NM, Nair SV, Smith LC, Raftos DA, Highly variable immune-response proteins (185/333) from the sea urchin, *Strongylocentrotus purpuratus*: proteomic analysis identifies diversity within and between individuals. *J Immunol* 2009, 182(4):2203-12.
109. Costa MM, Dios S, Alonso-Gutierrez J, Romero A, Novoa B, Figueras A, Evidence of high individual diversity on myticin C in mussel (*Mytilus galloprovincialis*). *Dev Comp Immunol* 2009, 33(2):162-70.
110. Pålsson-McDermott EM, O'Neill LA, Building an immune system from nine domains. *Biochem Soc Trans* 2007, 35(Pt 6):1437-44.
111. Gorbushin AM, Iakovleva NV, A new gene family of single fibrinogen domain lectins in *Mytilus*. *Fish Shellfish Immunol* 2011, 30(1):434-8.
112. Romero A, Dios S, Poisa-Beiro L, Costa MM, Posada D, Figueras A, Novoa B, 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,* in press.
113. Harris MA, Developing an ontology. *Methods Mol Biol* 2008, 452:111-24.
114. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R, InterProScan: protein domains identifier. *Nucleic Acids Res* 2005, 33:W116-20.
115. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, 25(17):3389-3402.
116. Edgar RC, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004, 32(5):1792-7.
117. Hall TA, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 1999, 41:95-98.
118. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ, Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009, 25(9):1189-1191.
119. Eddy SR, A new generation of homology search tools based on probabilistic inference. *Genome Inform* 2009, 23(1):205-11.
120. Bendtsen JD, Nielsen H, von Heijne G, Brunak S, Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004, 340:783-795.
121. Rouillard JM, Zuker M, Gulari E, OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res* 2003, 31(12):3057-62.
122. Philips T, Eberwine JA, Antisense RNA amplification: a linear amplification method for analyzing the mRNA population from single living cells. *Methods* 1996, 10:283–288.
123. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 2003, 34:374–378.
124. Tusher VG, Tibshirani R, Chu G, Significance Analysis of Microarrays applied to the ionizing radiation exposure. *Proc Natl Acad Sci U S A* 2001, 98(9): 5116-5121.
125. Storey JD, Tibshirani R, Statistical methods for identifying differentially expressed genes in DNA microarrays. *Methods Mol Biol* 2003, 224:149-57.
126. Dysvik B, Jonassen I, J-Express: exploring gene expression data using Java. *Bioinformatics* 2001, 17: 369–370.

Domeneghetti *et al.*, 2011

Balseiro P, Falcó A, Romero A, Dios S, Martínez-López A, Figueras A, *et al., Mytilus galloprovincialis* Myticin C: a chemotactic molecule with antiviral activity and immunoregulatory properties. PLoS One. 2011;6(8):e23140.

Banni M, Negri A, Mignone F, Boussetta H, Viarengo A, Dondero F. Gene Expression Rhythms in the Mussel *Mytilus galloprovincialis* (Lam.) across an Annual Cycle. PLoS One 6(5): e18904, 2011.

Ben Said O, Goñi-Urriza M, El Bour M, Aissa P, Duran R. Bacterial community structure of sediments of the Bizerte lagoon (Tunisia), a southern Mediterranean coastal anthropized lagoon. Microb Ecol. 59: 445-456, 2010.

Carland TM, Gerwick L. The C1q domain containing proteins: Where do they come from and what do they do? Dev. Comp. Immunol. 34: 785-790, 2010.

Chen JJ, Hsueh HM, Delongchamp RR, Lin CJ, Tsai CA. Reproducibility of microarray data: a further analysis of microarray quality control (MAQC) data. BMC Bioinformatics 8: 412, 2007.

Costa F, Alba R, Schouten H, Soglio V, Gianfranceschi L, Serra S, *et al*. Use of homologous and heterologous gene expression profiling tools to characterize transcription dynamics during apple fruit maturation and ripening. BMC Plant Biol. 10: 229, 2010.

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

Craft JA, Gilbert JA, Temperton B, Dempsey KE, Ashelford K, Tiwari B, *et al.* Pyrosequencing of *Mytilus galloprovincialis* cDNAs: Tissue-Specific Expression Patterns. Plos One 5(1): e8875, 2010.

Dodd RB, Drickamer K. Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. Glycobiology. 11(5):71R-79R, 2001.Dodds AW, Matsushita M: The phylogeny of the complement system and the origins of the classical pathway. Immunobiology 212: 233-243, 2007.

Evans TG, Somero GN. Phosphorylation events catalyzed by major cell signaling proteins differ in response to thermal and osmotic stress among native (*Mytilus californianus* and *Mytilus trossulus*) and invasive (*Mytilus galloprovincialis*) species of mussels. Physiol. Biochem. Zool. 83: 984-996, 2010.

GEO. The Gene Expression Omnibus can be accessed at http://www.ncbi.nlm.nih.gov/geo.

Gerdol M, Manfrin C, De Moro G, Figueras A, Novoa B, Venier P, *et al*. The C1q domain containing proteins of the Mediterranean mussel *Mytilus galloprovincialis*: a widespread and diverse family of immune-related molecules. Dev. Comp. Immunol. 35: 635-643, 2011.

Gestal C, Pallavicini A, Venier P, Novoa B, Figueras A. MgC1q, a novel C1q-domain-containing protein involved in the immune response of *Mytilus galloprovincialis*. Dev. Comp. Immunol. 34: 926-34, 2010.

Gibson G, Muse SV. A Primer of Genome Science. Sinauer Associates Inc., Sunderland, MA USA, 2004.

Gibson G. The environmental contribution to gene expression profiles. Nat. Rev. Genet. 9: 575-581 2008.

Gorbushin AM, Iakovleva NV. A new gene family of single fibrinogen domain lectins in *Mytilus*. Fish Shellfish Immunol. 30: 434-438, 2011.

Gracey AY, Chaney ML, Boomhower JP, Tyburczy WR, Connor K, Somero GN. Rhythms of gene expression in a fluctuating intertidal environment. Curr. Biol. 18: 1501-1507, 2008.

Hook SE. Promise and progress in environmental genomics: a status report on the applications of gene expression-based microarray studies in ecologically relevant fish species. J. Fish Biol. 77: 1999-2022, 2010.

Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. Brief. Funct. Genomic. Proteomic. 8: 174-183, 2009.

IMAQUANIM. Facts on this FP6 Integrated Project are at http://imaquanim.dfvf.dk/info.

Kerr MK, Churchill GA. Experimental design for gene expression microarrays. Biostatistics 2: 183-201, 2001.

Larsen PF, Schulte PM, Nielsen EE. Gene expression analysis for the identification of selection and local adaptation in fishes. J. Fish Biol. 78: 1-22, 2011.

Lenoir T, Giannella E. The emergence and diffusion of DNA microarray technology. J. Biomed. Discov. Collab. 1: 11, 2006.

Li H, Parisi MG, Parrinello N, Cammarata M, Roch P. Molluscan antimicrobial peptides, a review from activity-based evidences to computer-assisted sequences. Inv. Surv. J. 8: 85-97, 2011.

Li H, Venier P, Prado-Alvárez M, Gestal C, Toubiana M, Quartesan R, *et al*. Expression of *Mytilus* immune genes in response to experimental challenges varied according to the site of collection. Fish Shellfish Immunol. 28: 640-648, 2010.

Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, *et al*. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. 14: 1675-1680, 1996.

Lockwood BL, Sanders JG, Somero GN. Transcriptomic responses to heat stress in invasive and native blue mussels (genus *Mytilus*): molecular correlates of invasive success. J. Exp. Biol. 213(Pt 20): 3548-3558, 2010.

Lockwood BL, Somero GN. Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*). Mol. Ecol. 20: 517-29, 2011.

Malagoli D. Cytokine network in invertebrates: the very next phase of comparative immunology. Inv. Sur. J. 7: 146-148, 2010.

Manfrin C, Dreos R, Battistella S, Beran A, Gerdol M, Varotto L, *et al*. Mediterranean mussel gene expression profile induced by okadaic acid exposure. Environ. Sci. Technol. 44: 8276-8283, 2010.

Martín-Subero JI, Esteller M. Profiling epigenetic alterations in disease. Adv. Exp. Med. Biol. 711: 162-177, 2011.

MESH. Current MEdical Subject Headings can be accessed at http://www.ncbi.nlm.nih.gov/mesh.

Morozova O, Hirst M, Marra MA. Applications of new sequencing technologies for transcriptome analysis. Annu. Rev. Genomics Hum. Genet. 10: 135-151, 2009.

Nazar RN, Chen P, Dean D, Robb J. DNA chip analysis in diverse organisms with unsequenced genomes. Mol. Biotechnol. 44: 8-13, 2010.

NCT. The National Center for Toxicogenomics is at the National Institute of Environmental Health Sciences (http://www.niehs.nih.gov/research/atniehs/nct.cfm).

Padhi A, Verghese B. Molecular diversity and evolution of myticin-C antimicrobial peptide variants in the Mediterranean mussel, *Mytilus galloprovincialis*. Peptides 229: 1094-10101, 2008.

Pallavicini A, Costa MM, Gestal C, Dreos R, Figueras A, *et al*. High sequence variability of myticin transcripts in hemocytes of immune-stimulated mussels suggests ancient host-pathogen interactions. Dev. Comp. Immunol. 32: 213-226, 2008.

Pantzartzi C, Drosopoulou E, Yiangou M, Drozdov I, Tsoka S, Ouzounis CA, *et al.* Promoter complexity and tissue-specific expression of stress response components in *Mytilus galloprovincialis*, a sessile marine invertebrate species. PLoS Comput. Biol. 6: e1000847, 2010.

Ptitsyn A, Schlater A, Kanatous S. Transformation of metabolism with age and lifestyle in Antarctic seals: a case study of systems biology approach to cross-species microarray experiment. BMC Syst. Biol. 4: 133, 2010.

Rogers S, Cambrosio A. Making a new technology work: the standardization and regulation of microarrays. Yale J. Biol. Med. 80: 165-178, 2007.

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

Immunol. 35: 334-344, 2011.

Romero A, Estévez-Calvar N, Dios S, Figueras A, Novoa B. New insights into the apoptotic process in mollusks: characterization of caspase genes in *Mytilus galloprovincialis*. PLoS One 6(2):e17003, 2011.

Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467-470, 1995.

Shaw JP, Dondero F, Moore MN, Negri A, Dagnino A, Readman JW, *et al.* Integration of biochemical, histochemical and toxicogenomic indices for the assessment of health status of mussels from the Tamar Estuary, U.K. Mar. Environ. Res. [in press] 2011.

Shi L, Perkins RG, Fang H, Tong W. Reproducible and reliable microarray results through quality control: good laboratory proficiency and appropriate data analysis practices are essential. Curr. Opin. Biotechnol. 19: 10-18, 2008.

Somero GN. Comparative physiology: a "crystal ball" for predicting consequences of global change. Am. J. Physiol. Regul. Integr. Comp. Physiol. 301: R1-R14, 2011

Sonthi M, Toubiana M, Pallavicini A, Venier P, Roch P. Diversity of coding sequences and gene structures of the antifungal peptide mytimycin (MytM) from the Mediterranean mussel, *Mytilus galloprovincialis*. Mar. Biotechnol. (NY) [in press] 2011.

Steinberg CE, Stürzenbaum SR, Menzel R. Genes and environment - striking the fine balance between sophisticated biomonitoring and true functional environmental genomics. Sci. Total Environ. 400: 142-161, 2008.

Thomas M, Lieberman J, Lal A. Desperately seeking microRNA targets. Nat. Struct. Mol. Biol. 17: 1169-1174, 2010.

Venier P, De Pittà C, Bernante F, Varotto L, De Nardi B, Bovo G, *et al.* MytiBase: a knowledge base of mussel (*M. galloprovincialis*) transcribed sequences. BMC Genomics 10: 72, 2009.

Venier P, De Pittà C, Pallavicini A, Marsano F, Varotto L, Romualdi C, *et al*. Development of mussel mRNA profiling: Can gene expression trends reveal coastal water pollution? Mutat. Res. 602: 121-134, 2006.

Venier P, Pallavicini A, De Nardi B, Lanfranchi G. Towards a catalogue of genes transcribed in multiple tissues of *Mytilus galloprovincialis*. Gene 314: 29-40, 2003.

Venier P, Varotto L, Rosani U, Millino C, Celegato B, Bernante F, *et al.* Insights into the innate immunity of the Mediterranean mussel *Mytilus galloprovincialis*. BMC Genomics 12: 69, 2011.

Waterhouse RM, Kriventseva EV, Meister S, Xi Z, Alvarez KS, Bartholomay LC, *et al.* Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. Science 316: 1738-1743, 2007.

Yeaman MR, Yount NY. Unifying themes in host defence effector polypeptides. Nat. Rev. Microbiol. 5: 727-740, 2007.

Zhang SM, Zeng Y, Loker ES. Expression profiling and binding properties of fibrinogen-related proteins (FREPs), plasma proteins from the schistosome snail host *Biomphalaria glabrata*. Innate Immun. 14: 175-189, 2008.

1. Gérard K, Bierne N, Borsa P, Chenuil A, Féral JP (2008) Pleistocene separation of mitochondrial lineages of *Mytilus* spp. mussels from Northern and Southern Hemispheres and strong genetic differentiation among southern populations. Mol Phylogenet Evol 49(1):84-91.
2. Gilg MR, Hilbish TJ (2003) Patterns of larval dispersal and their effect on the maintenance of a blue mussel hybrid zone in southwestern England. Evolution 57(5):1061-77.
3. Galler S (2008) Molecular basis of the catch state in molluscan smooth muscles: a catchy challenge. J Muscle Res Cell Motil 29(2-5):73-99.
4. Babarro JM, De Zwaan A (2008) Anaerobic survival potential of four bivalves from different habitats. A comparative survey. Comp Biochem Physiol A Mol Integr Physiol 151(1):108-13.
5. Bers AV, D'Souza F, Klijnstra JW, Willemsen PR, Wahl M (2006) Chemical defence in mussels: antifouling effect of crude extracts of the periostracum of the blue mussel *Mytilus edulis*. Biofouling 22(3-4):251-9.
6. Akaishi FM, St-Jean SD, Bishay F, Clarke J, da S Rabitto I, et al. (2007) Immunological responses, histopathological finding and disease resistance of blue mussel (*Mytilus edulis*) exposed to treated and untreated municipal wastewater. Aquat Toxicol 82(1):1-14.
7. Moore MN (2008) Autophagy as a second level protective process in conferring resistance to environmentally-induced oxidative stress. Autophagy 4(2):254-6.
8. Rayyan A, Damianidis P, Antoniadou C, Chintiroglou CC (2006) Protozoan parasites in cultured mussels *Mytilus galloprovincialis* in the Thermaikos Gulf (north Aegean Sea, Greece). Dis Aquat Organ 70(3):251-4.
9. Francisco CJ, Hermida MA, Santos MJ (2010) Parasites and symbionts from *Mytilus galloprovincialis* (Lamark, 1819) (Bivalves: Mytilidae) of the Aveiro Estuary Portugal. J Parasitol 96(1):200-5.
10. Gomez DK, Baeck GW, Kim JH, Choresca CH Jr, Park SC (2008) Molecular detection of betanodaviruses from apparently healthy wild marine invertebrates. J Invertebr Pathol 97(3):197-202.
11. Takahashi D, Shukla SK, Prakash O, Zhang G (2010) Structural determinants of host defense peptides for antimicrobial activity and target cell selectivity. Biochimie 92(9):1236-41.
12. Bucki R, Leszczyńska K, Namiot A, Sokołowski W (2010) Cathelicidin LL-37: a multitask antimicrobial peptide. Arch Immunol Ther Exp (Warsz) 58(1):15-25.
13. Tian C, Gao B, Fang Q, Ye G, Zhu S (2010) Antimicrobial peptide-like genes in *Nasonia vitripennis*: a genomic perspective. BMC Genomics 11:187.
14. Yeaman MR, Yount NY (2007) Unifying themes in host defence effector polypeptides. Nat Rev Microbiol 5(9):727-40.
15. Pujol N, Zugasti O, Wong D, Couillault C, Kurz CL, et al. (2008) Anti-fungal innate immunity in *C. elegans* is enhanced by evolutionary diversification of antimicrobial peptides. PLoS Pathog 4(7):e1000105.
16. Smith LC (2010) Diversification of innate immune genes: lessons from the purple sea urchin. Disease Models & Mechanisms 3:1-6.
17. Buckley KM, Terwilliger DP, Smith LC (2008) Sequence variations in 185/333 messages from the purple sea urchin suggest posttranscriptional modifications to increase immune diversity. J Immunol 181(12):8585-94.
18. Hamilton CE, Papavasiliou FN, Rosenberg BR (2010) Diverse functions for DNA and RNA editing in the immune system. RNA Biol 7(2):220-8.
19. Linzmeier RM, Ganz T (2005) Human defensin gene copy number polymorphisms: comprehensive

analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23. Genomics 86(4):423-30.

20. Chen Q, Hakimi M, Wu S, Jin Y, Cheng B, et al. (2010) Increased genomic copy number of DEFA1/DEFA3 is associated with susceptibility to severe sepsis in Chinese Han population. Anesthesiology 112(6):1428-34.
21. Sperstad SV, Haug T, Blencke HM, Styrvold OB, Li C, Stensvåg K (2011) Antimicrobial peptides from marine invertebrates: Challenges and perspectives in marine antimicrobial peptide discovery. Biotechnol Adv in press.
22. Charlet M, Chernysh S, Philippe H, Hétru C, Hoffmann J, et al. (1996) Innate immunity: Isolation of several cysteine-rich antimicrobial peptides from the blood of a mollusc, *Mytilus edulis*. J. Biol. Chem 271:21808–21813.
23. Hubert F, Noel T, Roch P (1996) A member of the arthropod defensin family from edible Mediterranean mussels (*Mytilus galloprovincialis*). Eur J Biochem 15;240(1):302-6.
24. Mitta G, Vandenbulcke F, Roch P (2000) Original involvement of antimicrobial peptides in mussel innate immunity. FEBS Lett 486(3):185-90.
25. Venier P, De Pittà C, Bernante F, Varotto L, De Nardi B, et al. (2009) MytiBase: a knowledge base of mussel (*M. galloprovincialis*) transcribed sequences. BMC Genomics 10:72.
26. Mitta G, Hubert F, Noël T, Roch P (1999) Myticin, a novel cysteine-rich antimicrobial peptide isolated from haemocytes and plasma of the mussel *Mytilus galloprovincialis*. Eur J Biochem 265(1):71-8.
27. Mitta G, Vandenbulcke F, Hubert F, Salzet M, Roch P (2000) Involvement of mytilins in mussel antimicrobial defense. J Biol Chem 275(17):12954-62.
28. Mitta G, Vandenbulcke F, Noël T, Romestand B, Beauvillain JC, et al. (2000) Differential distribution and defence involvement of antimicrobial peptides in mussel. J Cell Sci 113(Pt 15):2759-69.
29. Costa MM., Dios S, Alonso-Gutierrez J, Romero A, Novoa B, et al. (2009) Evidence of high individual diversity on myticin C in mussel (*Mytilus galloprovincialis*). Dev Comp Immunol 33(2):162-70.
30. Berkowitz BA, Bevins CL, Zasloff MA (1990) Magainins: a new family of membrane-active host defense peptides. Biochem Pharmacol 39(4):625-9.
31. Li P, Chan HC, He B, So SC, Chung YW, et al. (2001) An antimicrobial peptide gene found in the male reproductive system of rats. Science 291(5509):1783-5.
32. Cellura C, Toubiana M, Parrinello N, Roch P (2007) Specific expression of antimicrobial peptide and HSP70 genes in response to heat-shock and several bacterial challenges in mussels. Fish Shellfish Immunol 22(4):340-50.
33. Li H, Toubiana M, Monfort P, Roch P (2009) Influence of temperature, salinity and *E. coli* tissue content on immune gene expression in mussel: results from a 2005-2008 survey. Dev Comp Immunol 33(9):974-9.
34. Li H, Venier P, Prado-Alvárez M, Gestal C, Toubiana M, et al. (2010) Expression of *Mytilus* immune genes in response to experimental challenges varied according to the site of collection. Fish Shellfish Immunol 28(4):640-8.
35. Venier P, Varotto L, Rosani U, Millino C, Celegato B, et al. (2011) Insights into the innate immunity of the Mediterranean mussel *Mytilus galloprovincialis*. BMC Genomics 12:69.
36. Pallavicini A, Costa MM, Gestal C, Dreos R, Figueras A, et al. (2008) High sequence variability of myticin transcripts in hemocytes of immune-stimulated mussels suggests ancient host-pathogen interactions. Dev Comp Immunol 32(3): 213-26.
37. Parisi MG, Li H, Toubiana M, Parrinello N, Cammarata M, et al. (2009) Polymorphism of mytilin B mRNA is not translated into mature peptide. Mol Immunol 46(3):384-92.

38. Mitta G, Hubert F, Dyrynda EA, Boudry P, Roch P (2000) Mytilin B and MGD2, two antimicrobial peptides of marine mussels: gene structure and expression analysis. Dev Comp Immunol 24(4):381-93.
39. Sonthi M, Toubiana M, Pallavicini A, Venier P, Roch P (2011) Diversity of Coding Sequences and Gene Structures of the Antifungal Peptide Mytimycin (MytM) from the Mediterranean Mussel, *Mytilus galloprovincialis*. Mar Biotechnol (NY) 13:857–867.
40. Yang YS, Mitta G, Chavanieu A, Calas B, Sanchez JF, et al. (2000) Solution structure and activity of the synthetic four-disulfide bond Mediterranean mussel defensin (MGD-1). Biochemistry 39(47):14436-47.
41. Roch P, Yang Y, Toubiana M, Aumelas A. (2008) NMR structure of mussel mytilin, and antiviral-antibacterial activities of derived synthetic peptides. Dev Comp Immunol 32(3):227-38.
42. Pantzartzi C, Drosopoulou E, Yiangou M, Drozdov I, Tsoka S, et al. (2010) Promoter complexity and tissue-specific expression of stress response components in *Mytilus galloprovincialis*, a sessile marine invertebrate species. PLoS Comput Biol 6(7):e1000847.
43. Aceto S, Formisano G, Carella F, De Vico G, Gaudio L. (2011) The metallothionein genes of *Mytilus galloprovincialis*: genomic organization, tissue expression and evolution. Mar Genomics 4(1):61-8.
44. Romero A, Estévez-Calvar N, Dios S, Figueras A, et al.. (2011) New insights into the apoptotic process in mollusks: characterization of caspase genes in *Mytilus galloprovincialis*. PLoS One 6(2):e17003.
45. Gracey AY, Chaney ML, Boomhower JP, Tyburczy WR, Connor K, et al. (2008) Rhythms of gene expression in a fluctuating intertidal environment. Curr Biol 18(19):1501-7.
46. Banni M, Negri A, Mignone F, Boussetta H, Viarengo A et al. (2011) Gene expression rhythms in the mussel *Mytilus galloprovincialis* (Lam.) across an annual cycle. PLoS One 6(5):e18904.
47. Lockwood BL, Somero GN (2011) Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*). Mol Ecol 20(3):517-29.
48. Manfrin C, Dreos R, Battistella S, Beran A, Gerdol M, et al. (2011) Mediterranean mussel gene expression profile induced by okadaic acid exposure. Environ Sci Technol 44(21):8276-83.
49. Craft JA, Gilbert JA, Temperton B, Dempsey KE, Ashelford K, et al (2010) Pyrosequencing of *Mytilus galloprovincialis* cDNAs: Tissue-Specific Expression Patterns. PLoS One 5(1):e8875.
50. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74:5463–67.
51. Thomas RK, Nickerson E, Simons JF, Jänne PA, Tengs T, et al. (2006) Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. Nat Med 12(7):852-5.
52. Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW (2007) Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res 17(8):1195-201.
53. Benaglio P, Rivolta C (2010) Ultra high throughput sequencing in human DNA variation detection: a comparative study on the NDUFA3-PRPF31 region. PLoS One 5(9):e13071.
54. Lazzaro BP, Rolff J (2011) Immunology. Danger, microbes, and homeostasis. Science 332(6025):43-4.
55. Schulenburg H, Kurtz J, Moret Y, Siva-Jothy M (2009) Introduction. Ecological immunology. Philos Trans R Soc Lond B Biol Sci 364:3–14.
56. Gao B, Zhu S (2010) Identification and characterization of the parasitic wasp *Nasonia* defensins: positive selection targeting the functional region? Dev Comp Immunol 34(6):659-68.

57. Zhang Y, He X, Li X, Fu D, Chen J, et al. (2011) The second bactericidal permeability increasing protein (BPI) and its revelation of the gene duplication in the Pacific oyster, *Crassostrea gigas*. Fish Shellfish Immunol 30(3):954-63.
58. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 11; 22(22):4673-4680.
59. Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. *Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press, Totowa, *NJ*, 365-386.
60. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389-402.
61. Binladen J, Gilbert MT, Bollback JP, Panitz F, Bendixen C, et al. (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS One 2(2):e197.
62. Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B et al. (2007) A pyrosequencing tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res* 35:130.
63. Breslauer KJ, Frank R, Blöcker H, Marky LA (1986) Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci., 83(11): 3746-50.
64. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9):1189-1191.

Supporting information

Additional files Venier et al., 2011.

Files are deposited at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3039611/

Additional file 1 - The 1820 putative immune-related sequences selected from Mytibase.
From left to right: ID and sequence data; first-hit similarities resulting from BLAST searches vs. UniProt/SW database; IPR domains from InterproScan analysis and GO terms; KEGG biochemical pathways and EC enzyme nomenclature based on BLAST similarity searches vs. annotated subsets of EMBL UniProtKB.

Additional file 2 - Differentially expressed genes in mussel hemocytes at 3 h post-injection of live V. splendidus.
Probe ID, sequence information and ordered expression values (log2 of normalized test/control values) are reported. Similarities resulting from InterproScan Analysis are reported in brackets (* annotation based on manual inspection of other relevant similarities)

Additional file 3 - Differentially expressed genes in mussel hemocytes at 48 h post-injection of live V. splendidus.
Probe ID, sequence information and ordered expression values (log2 of normalized test/control values) are reported. Similarities resulting from InterproScan Analysis are reported in brackets (* annotation based on manual inspection of other relevant similarities)

Additional file 4 - Common differentially expressed genes in mussel hemocytes at 3 and 48 h post-injection of live V. splendidus.
Probe ID, sequence information and ordered expression values (log2 of normalized test/control values). Similarities resulting from InterproScan Analysis are reported in brackets (* annotation based on manual inspection of other relevant similarities)

List of Abbreviations
AIF, Allograft Inflammatory Factor; **AKT**, RAC serine/threonine-protein kinase; **AMP**, Anti Microbial Peptides; **APAF1**, Apoptotic Peptidase Activating Factor 1; **ASC**, Apoptosis-associated Speck-like protein containing a CARD; **BCL2**, Baculoviral apoptosis regulator 2; **C1-5** Complement component; **CALR**, Calreticulin; **CASP**, Caspase; **CD63/LIMP**, Tetraspanin-7 (lysosome membrane protein); **CLR**, C-type Lectin Receptor; **CLR**, C-type Lectin Receptor; **CRD**, Carbohydrate Recognition Domain; **DAMPs**, damage-associated molecular patterns; **EST**, Expressed Sequence Tag; **FADD**, FAS (TNFRSF)-associated via death domain; **FNBP1**, Formin-Binding Protein 1; **FREP**, Fibrinogen-Related Protein; **GO**, Gene Ontology; **GRP**, Glucose-Regulated Protein; **HSC70**, Heat Shock Cognate 70; **HSP**, Heat Shock Protein; **IAP**, Inhibitor of Apoptosis Protein; **IAP**, Inhibitor of Apoptosis Proteins; **Ig**, Immunoglobulin; **IKBα**, Inhibitor of nuclear factor Kappa-B kinase alpha; **IKK**, Inhibitor of nuclear factor Kappa-B Kinase complex; **IL**, InterLeukin; **INX**, Innexin; **IPR**, Identifiable PRotein feature based on the InterPro database; **IRAK4**, Interleukin Receptor-Associated Kinase 4; **JAK**, Janus kinase 1; **KLHL**, Kelch-like protein; **LAMP**, Lysosomal-Associated Membrane Protein; **LBP**, Lipopolysaccharide Binding Protein; **LDLR**, Low-Density Lipoprotein Receptor; **LITAF**, LPS-Induced TNF-Alpha Factor; **LPS**, LipoPolySaccharide; **LRR**, Leucine Rich Repeat; **M6PR**, Mannose 6 Phosphate Receptor; **MACPF**, Membrane Attack Complex/Perforin; **MAPKs**, Mitogen-Activated Protein Kinases; **MASP**, Mannan-binding lectin Serine Protease; **MBL**, Mannose Binding Lectin; **MGCs**, Mytibase consensuses or singletons (*M. galloprovincialis* transcripts); **MIF**, Migration Inhibitory Factor; **MNK**, MAP kinase-interacting serine/threonine-protein kinase;

MR1, Mannose Receptor 1; **MyD88**, Myeloid Differentiation primary response gene 88; **NALP**, NATCH, LRR, and PYR containing proteins; **NCBI**, National Centre for Biotechnological Information; **NFkB**, Nuclear Factor of kappa light polypeptide gene enhancer in B-cells; **NLR**, NOD-Like Receptor; **NOD**, Nucleotide Binding Oligomerization Domain; **P13K**, Phosphatidylinositol-4,5-bisphosphate 3-Kinase; **P2X7**, ATP-gated ionotropic P2X purinoceptor subunit 7; **PA28**, Proteasome Activator subunit 28; **PAC2**, Proteasome Assembly Chaperone 2; **PAMP**, Pathogen Associated Molecular Pattern; **PGRP**, Peptidoglycan Recognition Protein; **PI31**, Proteasome Inhibitor PI31 subunit; **Pim**, proto-oncogene serine/threonine-protein kinase Pim; **ProPO**, ProPhenolOxidase; **PRR**, Pathogen Recognition Receptors; **RAB**, Ras-related gtp-Binding protein; **RIG**, Retinoic acid-Inducible Gene-I; **RIP**, Receptor-Interacting serine-threonine kinase; **RLR**, RIG-Like Receptor; **ROS**, Reactive Oxygen Species; **SEC22**, vesicle transport protein SEC22; **SOD**, SuperOxide Dismutase; **SP**, serine proteases; **SRCR**, Scavenger Receptor Cysteine-Rich protein precursor; **TAB**, TAK-binding protein; **TAK**, mitogen-activated protein kinase kinase; **TIMP3**, Tissue Inhibitors of MetalloProteinase 3; **TIR**, Toll/Interleukin-1 Receptor; **TIRAP**, Toll-Interleukin Receptor (TIR) domain containing Adaptor Protein; **TLR**, Toll-Like Receptor; **TNF**, Tumour Necrosis Factor; **TNFR**, Tumour Necrosis Factor Receptor; **TOLLIP**, TOLL-Interacting Protein; **TRADD**, TnfRsf-Associated via Death Domain; **TRAF6**, TNF receptor-associated factor 6; **Ub**, Ubiquitin; **UBR5**, Ubiquitin protein Ligase E3 (component n-recognin 5); **UniProtKB**, UniProt Knowledgebase; **α2**, proteasome subunit alpha type-2; **β4, β5**, Proteasome subunit beta type-4,-5.

Additional files Domeneghetti *et al.*, 2011

Abbreviation list (Figure 4):

AIF: Allograft Inflammatory Factor;
APAF1: Apoptotic Peptidase Activating Factor 1;
BCL2: Baculoviral apoptosis regulator 2;
C1-C5: Complement component 1-5;
CALR: Calreticulin;
CASP: Caspase;
CD63/LIMP: Tetraspanin-7 (lysosome membrane protein);
CLR: C-type Lectin Receptor;
CLR: C-type Lectin Receptor;
DAMPs: damage-associated molecular patterns;
FADD: FAS (TNFRSF)-Associated via Death Domain;
FNBP1: Formin-Binding Protein 1;
GRP 78/94: Glucose-Regulated Protein 78/94;
HSC70: Heat Shock Cognate 70;
HSP70/90: Heast Shock Protein 70/90;
IAP: Inhibitor of Apoptosis Proteins;
IKBα: Inhibitor of nuclear factor Kappa-B kinase alpha;
IKK: Inhibitor of nuclear factor Kappa-B Kinase complex;
IL: InterLeukin;
IRAK4: Interleukin Receptor-Associated Kinase 4;
JAK: Janus kinase;
KLHL: Kelch-like protein;
LDLR: Low-Density Lipoprotein Receptor;
LITAF: LPS-Induced TNFAlpha Factor;
LPS: LipoPolySaccharide;
MAPKs: Mitogen-Activated Protein Kinases;
MBL: Mannose Binding Lectin;
MGD1/2*: Mytilus galloprovincialis* Defensin 1/2
MIF: Migration Inhibitory Factor;
MNK: MAP kinase-interacting serine/threonine-protein kinase;
MR1: Mannose Receptor 1;
MyD88: Myeloid Differentiation primary response gene 88;
NALPs: NATCH, LRR, and PYR containing proteins;
NFkB: Nuclear Factor of kappa light polypeptide gene enhancer in B-cells;
NLR: NOD-Like Receptor;
NOD: Nucleotide Binding Oligomerization Domain;

P13K: Phosphatidylinositol-4,5-bisphosphate 3-Kinase;
PAC2: Proteasome Assembly Chaperone 2;
PAMPs: Pathogen Associated Molecular Patterns;
PGRP: Peptidoglycan Recognition Protein;
PI31: Proteasome Inhibitor PI31 subunit;
Pim: proto-oncogene serine/threonine-protein kinase Pim;
PRR: Pathogen Recognition Receptors;
RAB: Ras-related gtp-Binding protein;
RIP: Receptor-Interacting serine-threonine kinase;
ROS: Reactive Oxygen Species;
SEC22: vesicle transport protein SEC22;
SOD: SuperOxide Dismutase;
STAT: Signal Transducer and Activator of Transcription protein;
SRCR: Scavenger Receptor Cysteine-Rich protein precursor;
TAK: mitogen activated protein kinase kinase;
TIMP3: Tissue Inhibitors of MetalloProteinase 3;
TNF: Tumour Necrosis Factor;
TNFR: Tumour Necrosis Factor Receptor;
TRAF6: TNF receptor-associated factor 6;
Ub: Ubiquitin;
UBR5: Ubiquitin protein Ligase E3 (component n-recognin 5);
α2: proteasome subunit alpha type 2;
β4, β5: Proteasome subunit beta type 4/5

Additional files Rosani *et al*., 2011.

Files are deposited at:

http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0026680

File S1. Number of reads per AMP and sample.

File S2. Detailed list of all SNCs detected per AMP precursor in the 6 samples, with type of change and related frequency.

File S3. The 13 reference sequences designed to cover 9 AMP precursor transcripts of M. galloprovincialis. Sequence number and expected amplicon length with related transcript ID, precursor length, and number of clustered ESTs. All reference sequences are also reported in detail.

File S4. The maps locating all synonymous and non-synonymous SNCs along the transcript sequence (cds) of the AMPs.

File S5. Fusion primer pairs used for the AMP amplification, primer length, melting temperature and GC percentage.

PhD activities

Workshop e Meeting

- "Functional Genomics and System Biology", Hinxton, UK, 29 novembre- 1 dicembre 2011

  Poster: Rosani U, Varotto L, Roch P, Novoa B, Figueras A, Pallavicini A, Venier P. The intriguing sequence variability of antimicrobial peptides expressed in *M. galloprovincialis*.
- "XII congresso SIICS", Monteortone, 20-22 Febbraio 2011
- "71° congresso Unione Zoologica Italiana", Palermo, 20-22 settembre 2010

  Poster: U Rosani, L Varotto, A Rossi, B Novoa, P Roch, A Figueras, P Venier. Studio della variabilità di peptidi antimicrobici in *Mytilus galloprovincialis*.

  Vincitore premio UZI per la produzione scientifica orientata a tematiche di Biodiversità ed Evoluzione
- "XI congresso SIICS", Modena, 24-26 febbraio 2010

  Presentazione: U Rosani, L Varotto, A Rossi, B Novoa, P Roch, A Figueras, P Venier. Variability of antimicrobial peptides in *Mytilus galloprovincialis*.
- "microRNAs in physiology and disease", MACS, Bologna, 1 dicembre 2009
- "Genome Informatics", Cold Spring Harbor Laboratorty, NY, USA, 27-30 ottobre 2009.

  Poster: U Rosani, L Varotto, P Venier. Studying the variability of antimicrobial peptides in *Mytilus galloprovincialis*.
- "The Impact of the Environment on Innate Immunity", European Science Foundation, Innsbruck (Austria), 4-9 maggio 2009

  Poster: U Rosani, L Varotto, S Domeneghetti, G Lanfranchi, F Bernante, A Pallavicini and P Venier. Tools for studying the innate responses of *Mytilus galloprovincialis*.
- "X congresso SIICS", Urbino, 18-20 febbraio 2009

  Presentazione: U Rosani, L Varotto, S Domeneghetti, A Pallavicini, G Lanfranchi, P Venier. New knowledge of antimicrobial peptides in Mediterranean mussel.

Pubblicazioni

- U. Rosani, L. Varotto, A. Rossi, P. Roch, B. Novoa, A. Figueras, A. Pallavicini, P. Venier. Massively parallel amplicon sequencing reveals isotype-specific variability of antimicrobial peptide transcripts in *Mytilus galloprovincialis*. PlosOne 6: e26680, 2011.
- S Domeneghetti, C Manfrin, L Varotto, U Rosani, M Gerdol, G De Moro, A Pallavicini, P Venier. How gene expression profiles disclose vital processes and immune responses in *Mytilus* spp. ISJ, 8: 179-189, 2011.
- Venier P, Varotto L, Rosani U, Millino C, Celegato B, Bernante F, Lanfranchi G, Novoa B, Roch P, Figueras A, Pallavicini A. Insights into the innate immunity of the Mediterranean mussel *Mytilus galloprovincialis*. BMC Genomics. Jan 26;12:69, 2011.
- Varotto L, Domeneghetti S, Rosani U, Pallavicini A, Bisol P, Lanfranchi G, Venier P. DNA microarray analysis in *M. galloprovincialis* from the Venice lagoon. Scientific Research and Safeguarding of Venice. Research Programme 2006-2007. P.P. Campostrini Ed., CORILA (Venezia).