Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome description and inventory of immune-related genes of the endangered pen shell Pinna nobilis: a giant bivalve experiencing a mass mortality event

Robert Bunet (robert.bunet@institut-paul-ricard.org)
Institut oceanographique Paul Ricard https://orcid.org/0000-0003-4471-8818

Jean-Marc Prévot
Université de Toulon

Nardo Vicente
Institut oceanographique Paul Ricard

José Rafa García-March
IMEDMAR-UCV

Rajko Martinović
Institute of marine biology, University of Montenegro

José Tena-Medialdea
IMEDMAR-UCV

Danijela Joksimovic
Institute of marine biology, University of Montenegro

Jean-Luc Bonnefont
Institut oceanographique Paul Ricard

Stéphane Coupé
Mediterranean Institute of Oceanography, University of Toulon

Research article

Keywords: Bivalvia; Fan mussel; Pen shell; Pinna nobilis; genome; chemical defensome; immunity; mitochondria; repetitive elements

Posted Date: September 30th, 2019

DOI: https://doi.org/10.21203/rs.2.15332/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Background: The noble pen shell Pinna nobilis is a Mediterranean endemic and emblematic giant bivalve. Already registered as an endangered species in the late 20th century, it is facing a dramatic and rapidly expanding epizooty that decimates its populations since mid-2016. The ecological value of P. nobilis has urged important investigations for conservation purposes. In light of this, we report here the first draft genome of this animal.

Results: The whole-genome sequencing has been performed on an Illumina HiSeq X platform using a single paired-end library of short fragments (2x150 bp). The de novo contig assembly accounted for a total size of 584 Mb (96,738 contigs, N50 = 7.6 kb, with 0.4% of “N” nucleotides), representing 77.5% of the predicted genome size of 754 Mb. The pen shell genome is very AT-rich, with a GC-content of 35.6 %. Heterozygosity was found to be in the range of other sequenced bivalves (1%). Over one third (36.2 %) of the genome consisted of repeated elements with a surprising larger number of SINEs elements compared to other molluscan genomes. We were also able to reconstruct the full mitochondrial genome (~19 kb, with 12 protein-coding genes, 2 rRNA and 22 tRNA genes). In relation with the outbreak that affects P. nobilis, we paid a special attention on the innate immune and stress-related genes found in the sequence. We revealed that P. nobilis disposes of a complete chemical defensome, and a relatively sophisticated innate immune system.

Conclusion: In addition to offering a valuable resource for further research in comparative biology and evolution, access to the draft genome sequence is central to deepen our understanding of the vulnerability of P. nobilis to new diseases, which are likely to occur more often in the current scenario of a rapidly changing environment.

Background

The noble pen shell *P. nobilis* (Linnaeus, 1758) (Figure 1) is among the largest marine bivalve in the world, behind the tropical giant clam *Tridacna sp.* [1, 2]. Old adult specimens can reach 1.2 m anterior-posterior height and live up to 45 years [3]. Also called fan mussel, this species is endemic to the Mediterranean Sea where it inhabits the coastal fringe (from 0.5 m to 60 m depth). It lives partially buried in soft-bottoms (seagrass meadows, sediments, rough sand and gravel) with the anterior part of the shell anchored via externally secreted silk-like byssus filaments [4]. The superior part of the shell erects from the seabed to allow efficient filtration of surrounding water for feeding purposes. In addition to participate in the reduction of water turbidity by retaining significant quantity of organic matters from the suspended detritus [5], the fan mussel plays several essential ecological roles. Inside its shell body, it hosts crustacean symbionts such as the shrimp *Pontonia pinnophylax* or the crab *Nepinnotheres pinnotheres* [6]. On the other side, the shell provides a hard surface in a soft-bottom environment where epibionts can develop [7, 8] (see Figure 1).

Like other large sessile organisms, *P. nobilis* is particularly exposed to anthropogenic pressures such as harvesting, incidental trawling and anchoring or degradation of its natural habitat. In addition to direct threats, *P. nobilis* has to cope nowadays with marine pollution, consequences of global warming and ocean globalisation (*e.g.* introduction of new pathogens and emergent diseases) during its life cycle [9]. All these combined stressors, alone or in combination, have contributed during the last century to a decline in pen shell populations. Consequently, several Mediterranean countries have taken measures to protect the pen shell, e.g. by adopting in 1995 the Barcelona Convention (Annex II). At the European level, the pen shell was listed to the endangered and protected species under Annex IV of the Habitats Directive (Council Directive 92/43/EEC). After these protective actions, populations of *P. nobilis* seemed to regain vitality in several locations, *e.g.* in France [10, 11], Croatia [12] or Italy [13].

Unfortunately, since the autumn of 2016, a new threat has emerged causing a serious concern of extinction risk for this already endangered animal. A new disease associated with a newly identified species of *Haplosporidium* parasite, *Haplosporidium pinnae* [14] has first affected the Spanish Mediterranean coastline and Balearic Islands with mortality levels reaching up to 100% in scrutinized populations [15]. As anticipated by scientists, the reported mass-mortality events (MME) spread over several locations of the western Mediterranean basin: France including Corsica, Monaco, Tunisia, Italy, Greece, Malta, Cyprus and Turkey [16–20]. Interestingly, the collapse of tracked *P. nobilis* populations along the Tyrrhenian coastline of Italy was rather associated with a mycobacterial outbreak [16] suggesting that synergistic effects of emerging pathogens could be affecting the viability of *P. nobilis* sub-populations.

As a benthic sedentary organism that feeds by filtering water, *P. nobilis* is continually exposed to a wide range of biotic stressors, which may interact with the physical and chemical factors of its surrounding environment, challenging its homeostasis. It is now well known that the maintenance of homeostasis depends on a set of gene families that allow organisms to, along with other physiological functions, respond to pathogens [21] and deal with potentially toxic chemicals [22]. In bivalves, these genes are actually involved in the prominent role of the innate immune system [23, 24] and in the chemical sensing and detoxification process, also termed the chemical defensome [25]. Knowledge of such a gene repertoire is important to appreciate the adaptive potential of a studied species to current and future threats. In this context, genome and transcriptome sequencing give the opportunity to exhaustively look at the gene content of a specimen [24, 26–28]. These genome sequences then serve, for instance, to compare species that display differences in disease resistance, a phenomenon that has been observed among bivalves [29]. Indeed, mussels compared to oysters and clams seem much less susceptible to diseases and no mass mortality was detected in both natural and farmed populations of mussels [30–32].

In order to gain a better insight into the physiology and adaptive potential of *P. nobilis* regarding current climatic and zoonoses threats, we have undertaken the sequencing of its genome. In this study, we present the general features of the *P. nobilis* genome and focus our description towards the gene repertoire related to innate immunity and chemical defensome. This genome sequence represents the first

reference genome in the Pinnidae family and the second in the order of Pteriida, after the pearl oyster *Pinctada fucata,* and should be considered in future studies dealing with differential pathogens resistance across bivalve molluscs.

Materials And Methods

P. nobilis tissue sampling and nucleic acid preparation

A single adult individual of *Pinna nobilis* was sampled by freediving during summer 2016 in Les Embiez island (Six-Fours-les-Plages, France, 43° 4’41.89”, 5°47’33.12”E). A non-lethal sampling method was specifically developed to avoid manipulation of *P. nobilis.* Indeed, a small piece (14 mm^2, ~20 mg) of the mantle tissue was quickly excised using biopsy forceps before the fan mussel closes its valves. The piece of tissue was transferred in a tube filled with 95% ethanol and stored at –80°C until use. Total DNA was isolated using the Nucleospin® tissue kit from Macherey-Nagel. DNA quantity was measured using the Qubit™ dsDNA HS Assay kit (Invitrogen) and quality was checked on a standard 0.8% agarose gel.

Sequencing, genome assembly and size estimation

The purified genomic DNA (4 µg of starting material) was used to prepare a HiSeq X Ten DNA library. Sequencing was performed on an Illumina HiSeq 2x150 bp pair-end (Genewiz, USA). Trimming and *de novo* assembly were completed using CLC Genomics Server 9.0 and CLC Genomics Workbench 10.0 (Qiagen, Genewiz, USA). Before assembly, low-quality reads and sequencing adaptor contaminated reads were trimmed. For quality trimming, error rate was set at 0.01 whereas the maximal ambiguous nucleotides accepted was two. After trimming, reads with length < 100 bases and those of likely bacterial and viral origins were discarded from assembly. Final assembly statistics were determined using QUAST [33].

The haploid genome size was estimated from the trimmed PE reads with the k-mers frequency counting method, using JELLYFISH [34] that produced a histogram file of k-mers frequency distribution. An optimal k-mer size of 23 was chosen according to the formula k = log(200*genome size)/log(4) [35] and considering the genome sizes of the Pteriidae *Atrina pectinata* and *Pinctada fucata (*roughly 1,1 Gb), obtained from the Animal Genome Size database, release 2.0 [36] and [37]. We then used R software to plot the k-mer frequency distribution and manually estimated the genome size, by calculating the total number of k-mers divided by the position of peak depth, *i.e.* the value of the homozygous peak, as two peaks were obtained (Figure S2) [27]. The proportion of single-copy genome length was obtained by dividing the total number of k-mers under the peaks by the total number of k-mers under the whole curve. Additionally, KmerGenie [38] was used to confirm JELLYFISH results. The produced histogram file of k-mer frequency distribution was interpreted using Genomescope [39] to deduce the genome size and the proportion of unique sequences.

Nuclear genome annotation and completeness

The nature and occurrence of interspersed repeats and low complexity DNA sequences within our assembly were assessed using RepeatMasker v1.33 [40] run with nhmmscan version 3.1b2 (February 2015). Then, a second run of RepeatMasker v1.33 was performed using RepeatModeler 1.0.11, that is a package allowing a *de novo* identification and modelling of repeat sequence families. Simple sequence repeats (SSR) were specifically detected using the kmer-SSR package [41].

To annotate our genome, only contigs over 100X coverage were considered and blasted (BLASTN) against the nr database of the NCBI. Any remaining contigs corresponding to bacteria, protozoan or fungi organisms were manually sorted out. Afterward, a functional annotation was performed using Blast2GO workflow [42], designed to search for *i)* protein homologies (BLASTX) against the refseq non-redundant protein database (e-value cut-off of 10^{-6}), *ii)* homologies with protein motifs and domain in the InterPro and Gene Ontology (GO) databases (InterProScan) and *iii)* to associate GO to the best hits of the BLASTX predicted proteins and protein motifs.

Then, we ran BUSCO v.3 with default parameters, to look for orthologue genes among the unique “universal” eukaryotes and metazoan sets of genes43, 44, data sets: Eukaryota odb9 and the subset Metazoa odb9]. The proportion of “universal” genes within the genome assembly was considered as a proxy of the completeness. Finally, the number of unique genes was approximated by removing redundancies within the set of annotated contigs and GO results.

Defensome and innate immunity gene analysis

As the genome assembly was fragmented, a high number of annotated contigs might actually refer to a unique gene sequence. Hence, in order to prevent from irrelevant redundancy in the gene description, we first collected contigs that displayed a description and InterPro identifiers associated with a BLASTX e-value lower than 10^{-10} [24]. These contigs were manually examined according to the encoded protein domains, the sizes of the contigs and encoded peptides, and relative to the complete sequence of an orthologous gene and to the structure of orthologous proteins. For instance, within a gene family, two contigs with an identical description encoding two different domains of a protein, were considered as one gene. Two contigs with an identical description encoding the same domain were considered as one gene or two genes, according to the structure of an orthologous protein that might contain repeated domains. Considering the e-value cut-off, two contigs encoding similar domains but differentially described were considered as proxies of two genes.

Whenever possible, the detailed analysis of the structure of encoded peptides were assessed, especially for contigs encoding pathogen recognition receptors. In this aim, the putative complete or partial coding sequences was obtained using Augustus [45] and protein motifs were predicted using InterProScan [46] and ScanProsite [47].

Identification of mitochondrial genome and annotation

Contig descriptions that displayed a putative mitochondrial signature, either after BLASTX or BLASTN, were manually examined to retain those with an e-value lower than 10^{-10} or 10^{-6}, for BLASTX and BLASTN respectively, and exhibiting more than 75% identity and 60% coverage. One contig (Contig_66) that satisfied these criteria was further examined.

In order to double check the sequence and investigate a potential homoplasy, the trimmed reads were re-mapped against contig_66 using BOWTIE2 v2.3.5.1 [48] with the following options: “very-sensitive”, “end-to-end alignments”, “allowing up to two distinct alignments per read” and selecting reads with a minimum Phred-score of 30. Mapped reads were collected, and single-nucleotide and indel variants were visualized using SAMtools [49].

Gene content and arrangement of the *P. nobilis* mitochondrial sequence were assessed using both MITOS [50] and the beta version of MITOS2 [51] with an invertebrate genetic code. Then, the full-length coding sequences were deduced with ORFfinder [52] with search parameters set up as following: minimal ORF length of 300 nt, invertebrate mitochondrial genetic code, ORF start codons “ATG and alternative initiation codons”, nested ORFs ignored. Protein coding gene boundaries were manually assessed by aligning predicted protein sequences with available orthologues sequences (e.g. *Atrina pectinata,* and other related genera), in order to mainly visualize sequence length differences at the C-terminal ends. When such differences were found, a TBLASTN search was carried out with the 5’ end of the CDS to assess the sequence divergence with *A. pectinata,* thus increasing the reliability of the putative initiator codon. Transfer and ribosomal gene positions were considered exact without other confirmation. Repeat regions were detected using Dottup [53] and tandem repeat finder [54].

The full-length mitochondrial genome has been deposited in GenBank and released under the accession number MN432488.

Results And Discussion

Sequencing and nuclear genome size determination

The sequencing of *Pinna nobilis* DNA generated about 799 million reads in pair, yielding 240 Gbases of sequence with high quality (Fig S1). After several trimming and filtration steps, reads could be assembled into 97,633 contigs, with a 304X coverage in average, and contig N50 and average length of 7,576 and 6,007 nucleotides, respectively (Table S1). The total size of the assembly yielded 586 Mb. In the subsequent steps of the assembly analysis, only contigs over 100X coverage were considered and any remaining contig with suspicion of bacterial, protozoan or fungi origin were eliminated. In the end, our assembly yielded ~584 Mb over 96,738 contigs.

Analysis of the k-mer frequency plots using JELLYFISH [34] showed a bimodal distribution (Figure S2), revealing the huge proportion of homozygous regions in the *P. nobilis* genome, of approximately 99%.

From the data of frequency distribution, we estimated a genome size of 778 Mb, of which unique sequences represent 575 Mb (73.9%). Such estimates were confirmed using KmerGenie [38] that predicted a genome size of 735 Mb, including 632 Mb of unique sequences. Thus, our assembly represents approximately 75% of the estimated genome size of the pen shell.

Among other species of Pteriida whose genome size has been elsewhere estimated, *P. nobilis* shows the smallest genome. It is respectively 23%, 32% and 34% shorter than those of the recently described *Pinctada fucuta marcensii,* which size has been estimated at 990 Mb [26], *Atrina rigida* (~1100 Mb from the Animal Genome Size Database) and *Pinctada fucata* (~1150 Mb, [37]).

Genome composition and completeness

The genome is very AT-rich (64%) while a fraction of the bases could not be determined (0.4% N) (Fig. S1 and Table S1). The repetitive elements (RE) accounted for approximately 36% of the genome, a regularly observed amount within bivalve genomes [27, 28, 37]. More than 60% of RE could be classified, and the observed distribution of each was also globally consistent with those observed in *Mytilus* [27] and *Pinctada* [26] genera (Figure 2A).

However, a high number of short interspersed nuclear elements or SINEs, accounting for more than 10 % of the genome assembly, was detected (Figure 2A and 2B). Strikingly, the proportion of SINEs in *P. nobilis* exceeds that of other marine mollusc organisms such as *Aplysia californica, Lottia gigantea P. fucuta, Crassostrea gigas,* or *M. galloprovinciallis* (see Figure 2A). In addition, within our assembly, SINEs were about four times more represented than long interspersed nuclear elements (LINEs) which has never been reported in bivalve genomes. Actually, the number of detected SINEs hardly reflects the number of genuine SINEs. Indeed, *P. nobilis* genome might contain numerous fragments of short non-autonomous retrotransposons, possibly fragments of LINEs, which actually resemble SINEs. Yet, as SINEs have evolutionary potentials [55–57], but also deleterious effects [58–60], clarifying the amount of SINE sequences within *P. nobilis* genome would contribute to understand the susceptibility of this animal to diseases and environmental stress.

Using BUSCO (Benchmarking Universal Single-Copy Orthologs) software [44], we could evidence the presence of 271 complete (27%) and 130 fragmented (13%) genes among the set of 978 eukaryote core orthologues. A highly similar result was obtained when considering the set of metazoan core orthologues instead (data not shown), suggesting that our genome assembly, like in other preceding mollusc genome sequencing, would be incomplete. Comparatively, bivalve genomes recently described showed a wide range of completeness, varying from ~16% in *M. galloprovinciallis* [27] to 87% in *Bathymodiolus platifrons* [61] for complete single copy gene models, and up to 43.3% and 97.7% for these species, respectively, while considering fragmented gene models as well. Strikingly, contrary to similar studies on bivalve genomes, BUSCO identified twice more complete than fragmented orthologues in our assembly, which was rather unexpected considering the contig sizes and sequencing coverage. This could be explained by our fragmented assembly, in which numerous contigs possibly contain only small length,

partial, and potentially too divergent, coding sequences, that BUSCO did not considered. For instance, the genome assemblies of *P. nobilis* and *C. gigas* are of similar sizes and share comparable sequencing coverages. However, their assembly statistics strongly differ. Indeed, the genome of the pacific oyster *C. gigas* was assembled in 11,969 scaffolds (N50 = 401,319) and its genome completeness reached 79% and 95% for complete and fragmented metazoan core orthologs, respectively [28]. Other similar observations support that genome completeness is inversely related to the number of scaffolds within assembly [27]. Moreover, completeness is most of the time achieved when both genome and transcriptome are concomitantly sequenced and investigated, as illustrated by genome descriptions of the deep sea mussel *B. platifrons* [61, 62] and *P. fucuta marcensii* [26] that evidenced almost full completeness. Hence, in addition to our study, *P. nobilis* genome would benefit from a future transcriptome sequencing to improve the completeness of the present description.

So far, using a BLASTX comparison against the protein refseq database, 51,317 (52.5%) out of the 97,633 contigs of our assembly matched an orthologue sequence. This figure was reduced to 21,773 unique annotated contigs after removing any redundancies, which is interpreted as an approximation of unique gene content. Reported to an assumed complete genome size, we would expect around 29,000 genes in the genome of *P. nobilis.* In the assembly of the pearl oyster *P. fucuta marcensii,* the estimated gene content varies from 21610 [63] to 32,937 [26]. Thus, our evaluation of the gene content within our assembly appears to be relevant and is comparable to that of *P. fucata* and other reported protein coding gene contents evidenced in *C. gigas* (28,027 genes, [28]) and *M. galloprovincialis* (25,000 expected *vs* 10,891 detected, [27]).

In the ontology annotation of Blast2GO, approximately 40 % of orthologues were found in only three recently sequenced bivalve species. The gene ontology (GO) distributions at level 3, among the three main categories of biological process, *i.e.* molecular functions and cellular components, together with the results described above, indicates the production of a comprehensive genome draft (Figure 3).

Insight into the gene repertoires of the innate immune system and chemical defensome

Within the Biological Process category, GO terms referring to “cellular response to stimulus” and “response to stress” were encountered in 2765 and 783 contigs respectively (Figure 3). A glimpse into the number of non-redundant descriptions for more specific GO terms (Table 1) revealed that *P. nobilis* would dispose of a well-developed defensome, challenging the explanation of the massive and deadly zoonoses recently experienced by this organism. Here, we report the first inventory of genes encoding key proteins specifically related to the innate immune response and to the acclimation to chemical stress, that all together account for the homeostasis maintenance of this organism.

1. Inventory of immune-related genes

Pathogen recognition receptors (PRRs) are molecular sensors that recognize pathogen structures and constitute a first defence line against pathogens. In invertebrates, diverging evolution of the immune system has led to a diverse and complex set of pathogen binding molecules [23]. Among them, several have been inventoried in various bivalves, such as the toll-like receptors (TLRs), peptidoglycan recognition receptors (PGRPs), glucan binding proteins (GNBPs), lectins and laminins [24, 64, 65]. Besides, an intracellular line of defence relies upon sensors of foreign RNAs, DEAD-box helicases, such as the retinoic acid inducible gene I (RIG I-like = DDX58) and the melanoma differentiation associated gene–5 (MDA–5) [66, 67]. In our *P. nobilis* genome sequence, we could evidence a wide range of genes related to immunity and we could highlight several families, from PRRs to effectors of the immune response that potentially harbored a rather diverse set of genes (Table S2).

1.1 Pathogen recognition receptors

The *toll-like receptor* (TLR) pathway is well conserved over the Metazoan. It plays a pivotal role in the control of microbiome homeostasis and pathogen infections, as well as in the regulation of development process ([68, 69] for review). The typical feature of TLRs consists in a Leucine-rich-region (LRR), a transmembrane domain (TM) and a Toll-interleukin receptor domain (TIR-domain), but structural variations have been reported, for instance in *M. edulis* [24]. The LRR recognizes protein motifs, mainly of microbial origins, while the TIR-domain ensure signal transduction by interacting with adaptor proteins [21, 65].

Although a wide TLR repertoire has been elucidated in some invertebrates such as the sea urchin *Strongylocentrotus purpuratus* in which more than two-hundred TLR have been identified70, 71], only a few have been reported in bivalves: twenty-seven putative TLR encoding contigs in *M. edulis* [24], twenty-three in *M. galloprovincialis [*65] though four have been reported in *C. gigas* [28] and only one detected in *Mya arenaria* and *Chlamys farreri* [72, 73].

In our sequence, we identified twenty-seven contigs encoding TLRs of various structures, all exhibiting a TIR-domain. Among them, twenty displayed a typical TLR structure, one also encoded a C1q motif at the C-terminal end of a typical TLR and one had two TIR-domains. We also found six and nine contigs encoding either a single TIR-domain or LRRs, respectively (Figure 4). However, the latter nine contigs could not be considered as additional putative TLRs since the TIR-domain was missing. Thus, we evidenced that *P. nobilis* disposes of a rather diverse set of TLRs, of a similar magnitude to the one reported in *M. edulis* [24]. When looking at the downstream key players of the TLR pathway, we could also detect contigs with MyD88, IRAK1, IRAK3 IRAK4 and TRAF patterns.

Peptidoglycan recognition receptors (PGRPs) and glucan binding proteins (GNBPs) are additional immune effectors involved in the recognition of pathogens. Unlike GNBPs, PGRPs are conserved from insect to mammals. However, both were evidenced in bivalves [24, 74, 75]. On the one hand, PGRPs recognise the peptidoglycan (PGN) cell wall of bacteria [76, 77], activate Toll receptors or the immune deficiency (IMD) pathway and induce phagocytosis. Moreover, PGRPs containing an amidase2 motif

present bactericidal moieties [78, 79]. On the other hand, GNBPs in insect and crustacean recognize and form a complex with lipopolysaccharides (LPS), gram-positive bacteria and β-glucans of fungal cell walls, subsequently inducing a degranulation and the activation of prophenoloxidase [80]. At last, gene expression of PGRPs and GNBPs are activated following PGN or bacterial challenge, evidencing the key role of those proteins in the bacterial clearance [75].

The genome assembly of *P. nobilis* contains six contigs encoding PGRPs, of which four exhibited a PGRP/amidase2 motif and two encoded a lysine motif (LysM). Noteworthy, we could evidence one transmembrane PGRPs, confirming previous observation of Philipp and collaborators [24]that bivalves would produce both soluble and membrane bound PGRPs. Besides PGRPs, we found eight annotations referring to GNBPs. The encoded peptides all exhibited a glycosyl hydrolase motif, and no transmembrane domain could be detected.

Lectins are sugar-binding proteins, structurally made of an oligomeric assemblage of peptide sub-units that possess a carbohydrate recognition domain (CRD) and eventually a transmembrane region [81]. They are involved in cellular interactions, protein synthesis and transport, signal transduction, and some of them mediate pathogen recognition, using Ca^{2+}-dependant carbohydrate-recognition domain [23, 82]. Animal lectins have been classified into thirteen groups according to specific conserved amino acid sequences within the CRD, carbohydrate specificity and functions: C-, F-, I-, L-, M-, P- and R-type, F-box lectins, chitinase-like lectins, calnexins, galectins, interlectins and ficolins [83]. In bivalves, an additional type of lectin, represented by C1q domain containing proteins (C1qDC), also binds carbohydrates. Among those lectins, C1qDC, C-type, F-type, ficolin and interlectins have important implications in pathogen recognition and innate immune processes [23, 84–86].

We found ninety-two annotations for C1q motifs and twenty-six contigs encoding a peptide homologous to C-type lectins or a C-type motif (*i.e.* IPR001304). Using “fucolectin” as a keyword for F-type lectins, three contigs were evidenced. Finally, we found four contigs annotated for ficolin while no contig for interlectin could be retrieved. Those results are congruent with the reported lectins diversity among bivalves, and evidence the predominance of C1qDC proteins within the lectin repertoire in *P. nobilis,* as similarly reported in *C. gigas* [23].

Laminins are one of the extracellular matrix (ECM) components. Laminins are large proteins of approximately 800 kDa, composed of three polypeptide chains (α, β and γ) exhibiting five globular domains that interact with some of the ECM components such as integrins, proteoglycans or receptors, and a coil-coiled domain [87]. Interactions between laminins and their specific receptors activate the intracellular signalling pathways that regulate cell growth, differentiation, migration and adhesion to the extracellular matrix and other cells [88–90]. They also influence the migration and functions of immune cells [91]. Noteworthy, laminins are host cell targets for a wide range of pathogens. Some recent research indeed evidenced reductions of infection rates when the expression of laminin was lowered [92] or the protein partially deleted [93], underling the importance of laminins in the process of pathogen infection. Moreover, they even could help protozoan escape the host innate immunity, as evidenced in some

susceptible mosquitos to *Plasmodium* ookinetes [94] and in other host-protozoan pathogen interactions [95]. This observation has to be considered in the light of the infection of *P. nobilis* by the protozoan *Haplosporidum pinnae* [96]. In the genome assembly of *P. nobilis,* we identified seven contigs encoding laminins (4, 2, 1). Ultimately, while looking at sensors of foreign viral RNAs, we could highlight two contigs, encoding one RIG I-likei.e. DDX58)and oneMDA–5i.e. interferon-induced with helicase C: I-fish)gene.

1.2 Signalling

Cytokines are low molecular weight proteins secreted by immune cells in response to pathogen infections and cell injuries. They are essential mediators of the immune response acting through a series of conserved signalling pathways, which for instance can lead to contrasted outcomes such as apoptosis or cell survival. We found twenty eight cytokines within the genome assembly of *P. nobilis,* of which nineteen were homologous to bivalves interleukines–17 (twelve contigs) or tumour necrosis factor (TNF) superfamily members displaying conserved TNF_2 domains (seven contigs). We also identified one contig encoding both a well-conserved macrophage inhibitory factor (MIF), considered as a key mediator of the innate immune system across taxa, which has already been evidenced in some bivalves [97–99], and three interferon regulatory factors.

Regarding IL–17 and TNF superfamily in detail, our observations are consistent with previously reported evidences of a central ancient IL–17 signalling pathway in the immune response of bivalves [23, 100]. IL–17 is a pro-inflammatory cytokine mainly involved in the defence against extracellular pathogen. Moreover, the IL–17 gene might be an initiator of a more complex, integrated, immune response, as evidenced by the rapid upregulation of IL–17 gene expression observed in *C. gigas* haemocytes submitted to a bacterial challenge [101]. TNF-α is also mainly involved in resistance against pathogens, though in certain cases, circulating TNF might be permissive towards parasitic, bacterial, and viral infections [102]. The transcriptome of *M. edulis* showed twelve TNF superfamily members, similar to the number found in *P. nobilis,* and a high number of transcripts encoding TNF protein motifs, which emphasises the central role of TNF superfamily in the innate immune response in bivalves [24].

Besides cytokines, the immune response may also likely be fine-tuned by a set of molecules of the neuroendocrine system that can rapidly trigger an adaptive response to various biotic and abiotic stressors [103–106]. The genome assembly of *P. nobilis* contains a number of contigs consistent with a likely well-developed neuroendocrine system. We could indeed confirm the presence of several enzymes of the catecholaminergic, cholinergic, serotoninergic and GABAergic systems, of the nitric oxide signalling and neuropeptides (data not shown).

1.3 Destruction of pathogens

In the fight against pathogens, innate immunity relies upon the complement cascade, activated by either the classical, lectin or alternative pathway, and ultimately results in the formation of a pore complex within the pathogen membrane [107] and the production of pro-inflammatory molecules [108, 109]. Vertebrates have evolved a complement system composed of the complement factors C1–9 and factor B (Bf), C3 being central since it integrates all three activation pathways and activates subsequent steps of the cascade, that further lead to the lysis of pathogen cells. In ancient taxonomic lineages, only few factors could be evidenced, suggesting an ancient origin of the complement system [110]. In bivalves, no Bf, C2, C4, C5 or C7–9 factors could be reported, challenging the comprehensive view of the complement cascade occurring in these organisms [24, 110]. Consistently, we were also unable to detect these factors. However, our genome assembly contains contigs matching the complement factors C3 and C6 as observed in *M. edulis* [24]. Noteworthy one contig matched an orthologous C8b protein of *Rattus norvegicus* (BlastX, e-value 10^{-4}), and four contigs were positive for the membrane attack complex domain IPR020864. Hence, our results corroborate the idea of a rather divergent complement system compared to that of vertebrates [110].

Antimicrobial peptides (AMPs) are small secreted peptides involved in the defence against the invasion of a wide range of pathogens [111]. AMPs typically contain one hydrophilic, positively charged, domain allowing fixation on the cell plasma membrane of a pathogen, and a hydrophobic domain that embed into the plasma membrane, disrupting the membrane or intracellular functions, ultimately inducing cell death [112, 113]. AMPs are a very diverse family and have been found from prokaryotes to vertebrates. Several groups of AMPs have been reported in bivalves, i.e. defensins, mytilins, myticins and mytimycins [114–116]. Within the *P. nobilis* genome, no orthologue sequences of AMPs were discerned using keywords or InterPro identifiers. Nonetheless, six contigs were spotted to contain sequences similar to soluble scavenger receptor cysteine-rich peptides and c*ysteine-rich secretory proteins, which are likely involved i*n the innate immune response [117–120].

1.4 Apoptosis

Apoptosis is also a prominent and efficient mechanism that limits the progression of pathogen infections, while preventing inflammation of the invaded tissue. Most of the time, apoptosis is triggered following interactions between the host immune cell and the pathogen, but obligate intracellular microorganisms have evolved strategies to circumvent apoptosis and promote survival [121–123]. The genome of *P. nobilis* contains the main actors of apoptosis, i.e. members of the BCl–2 family, caspases and inhibitors and activators of apoptosis. Noteworthy, we also found 38, 34 and 21 contigs encoding a DEATH, BIR and CARD domains, respectively. These results suggest a well-developed apoptotic machinery in molluscs, as previously discussed [24].

2. Inventory of stress-related genes

A wide range of structurally diverse chemical stressors is encountered in the marine environment. Heavy metals, natural exogenous compounds (*e.g.* toxic aldehydes), microbial and algal products or manmade molecules, such as pesticides and hydrocarbons, may considerably affect the physiology of marine organisms and challenge their persistence [124]. This is especially true when considering sessile organisms that are unable to escape their contaminated environment. Therefore, a variety of proteins have evolved towards the protection against potentially toxic chemicals and have been grouped under the term of chemical defensome [25]. More specifically, it mainly consists in sets of enzymes that, collectively, ensure the cell detoxification through four main processes : *i)* the modification of chemicals to less toxic and more easily excretable substances, *ii)* the elimination of toxicants and transformed products out of cells, *iii)* the molecular protection against radicals including the reactive oxygen species and *iv)* the sensing of environmental contamination or cellular damages. The *P. nobilis* genome assembly possesses most of the genes linked to the chemical defensome. Altogether, they account for roughly 1.5% (337/21,773) of the predicted genes in *P. nobilis* genome, similar to the proportion observed in other organisms [125]. Thereafter, we present an insight into the *P. nobilis* chemical defensome (Table S3) and detail the family diversity of some key genes, according to the four processes presented above.

2.1 Reductive, conjugative and oxidative biotransformation

Proteins involved in chemical biotransformation are required to mitigate the toxicity of chemicals and to modify these toxicants to more hydrophilic substances prior to their excretion. Among those, the cytochromes P450 (CYPs), the sulfotransferases (SULTs) and the UDP-glucuronosyl and glycosyl transferases (UGTs)represented expanded families.

Cytochromes P450 are enzymes responsible for oxidative biotransformation. The diversification of CYP genes is very important in almost all species and genes are organised in numerous families, of which the CYP1, CYP2, CYP3, CYP4, CYP6 and CYP9 families account for the most important ones [126]. The genome assembly of *P. nobilis* contains ninety-eight contigs annotated for CYP genes, which overall presented a relatively high mean similarity with orthologous sequences (approximately 70%), allowing annotation at the family level, revealing thirty-seven CYP2, thirteen CYP3 and twenty-four CYP4. According to the classification of CYP Clan [127], *P. nobilis* CYPs were found distributed among Clan 2, Clan 3 and Clan 4 (Figure 5) resembling the distribution in other organisms [25, 128]. Noteworthy, no CYP19, i.e. the aromatase, was identified which is also consistent with other reported observations in invertebrates [129, 130].

Sulfotransferases (SULTs) are a variety of either cytosolic or membrane-anchored enzymes that catalyse the conjugation of sulfuryl groups to many diverse molecules, with a target specificity according to their localization [131, 132]. For instance, the cytosolic SULTs mediate the transfer of a sulfuryl group to endogenous metabolites and xenobiotics [133], although membrane SULTs catalyse the sulfation of carbohydrates and tyrosyl residues of larger proteins or peptides [134]. In bivalves, SULTs represent a large family. For instance, twenty-eight and thirty-one genes have been detected in the oyster and pearl

oyster genomes, respectively and up to eighty-three genes in the *Chlamys farreri* genome [28, 37, 135]. We found twenty-eight genes encoding SULTs in the genome of *P. nobilis,* distributed in nine subfamilies. However, most SULT genes belonged to subfamily 1 (eight genes), 10 (seven genes) and 11 (seven genes). Moreover, we could evidence a total of seventeen membrane anchored SULTs in our assembly.

Glutathion S-transferase (GSTs) are typically small enzymes of less than 250 amino acids [136], present in all three branches of life (Archaea, Eubacteria and Eukaryota). Once activated upon exposure to a range of toxins or to oxidative damages, they transfer reduced glutathione (GSH) to hydrophobic xenobiotics [137–139]. The classification of the GSTs family separates the cytosolic, mitochondrial and microsomal GSTs, and distinguishes fourteen subclasses. In eukaryotes, the subfamilies alpha, delta, epsilon, omega, pi, sigma, theta, mu and zeta are cytosolic GSTs, while kappa GSTs are the mitochondrial and peroxisomal ones. Structurally, all GSTs are homo- or heterodimers of the same GST class. They display conserved C- and N-terminal domains that encompass two types of binding sites, conserved G-sites which bind GSH and H-sites that are highly variable, allowing interactions with a wide range of hydrophobic substances [140, 141]. The mitochondrial kappa GSTs are homodimers exhibiting structural similarities to cytosolic GSTs, thus likely have similar molecular specificities [142]. Microsomal GSTs, on the contrary, result from the association of three identical monomers that each contain an alpha helix and a GSH binding domain [143, 144].

Overall, we found fifteen contigs encoding GSTs in our *P. nobilis* assembled sequence. Eleven GSTs clustered with the cytosolic GSTs: one alpha, one mu, two omega, one pi and three theta, while three of them could not be specifically assigned. Alpha GSTs protect against therapeutic drugs, environmental toxins and product of oxidative stress, while mu GSTs protect against molecules such as polycyclic aromatic hydrocarbon metabolites [145, 146]. Compared to other subclass of GSTs, omega GSTs have additional conjugative and biotransformation abilities and perform a wide range of thiol transferase and reduction reactions, such as the biotransformation of reactive α-haloketones to nontoxic acetophenones [142]. The theta GSTs are mainly involved in counteracting general oxidative stress [141, 147–150]. Eventually, one mitochondrial kappa GST and three microsomal GSTs (isoforms 1, 2 and 3) could be distinguished in our sequence. Therefore, *P. nobilis* disposes of a rather complete set of GSTs, with a relatively low level of gene expansion within the GSTs classes, compared to other marine organisms, such as the sea urchin, sea anemone and crustaceans [25, 125, 151].

2.2 Efflux transporters

Efflux transporters are complexes of proteins mainly represented by the large family of the *ATP-binding cassette (ABC)* transporters. Using ATP as a source of energy, they catalyse the transport of various substrates such as sugars, lipids, amino acids, antibiotics or siderophores but also xenobiotics or pollutants, across membranes [152–156]. To date, eight sub-families (from A to H) have been described although some family could be absent from a genome [152, 157]. Considering xenobiotic stress, the first line of cellular defence is represented by the ABC*B* (p-glycoproteins), ABC*G* (mitoxantrone resistance

protein) and the ABC*C* (multidrug resistance proteins) that export both unmodified and modified substrates out of the cell [158–161]. A search within the genome assembly revealed fifty-two contigs encoding typical features of ABC transporters, of which thirty-seven corresponded to the three multidrug transporter subfamilies B, C, and G. Similar total number and proportion have been described in human, sea urchin, anemone and the copepod *Tigriopus japonicus* [25, 125, 162, 163] (Figure 6).

2.3 Antioxidant proteins and metal detoxification

Antioxidant proteins counteract the deleterious effects of excessive oxidative stress within cells, endogenously generated by the presence of metals, the detoxification processes or even the global increase in oxygen consumption [164–166]. We could identify at least one contig containing a coding sequence associated to each of the following enzymes: catalase, superoxide dismutase, glutathione peroxidase, glutathione synthase and glutathione reductase. Concerning the metal-binding proteins, we found a single contig for metallothionein, which, to the best of our knowledge, seems to be a general feature among bivalves. However, we detected seven heavy metal binding proteins (HIPs) in the genome of *P. nobilis,* similar to the histidine-rich glycoproteins (HRG), which have been evidenced in the blood plasma of several bivalves and where they represent more than 40 % of the total protein weight [167, 168]. Moreover, *P. nobilis* was also determined to contain ferritin, ceruloplasmin and biliverdin reductase sequences, but it seems to lack a heme oxygenase and phytochelatins. Based on these results, we could expect a good ability of *P. nobilis* to cope with metallic elements.

2.4 Heat shock proteins

Heat shock proteins (HSP) are expressed as a general response to many stressful challenges [169–171]. We could evidence seven hsp70, one hsp90, one heat shock cognate protein and one heat shock factor in the genome assembly of *P. nobilis.* Noteworthy, seventeen co-chaperones dnaJ/hsp40 were also detected, consistent with previous observations across a wide range of organisms, from unicellular to human [172–174]. These proteins are conserved throughout the evolution and regulate the ATP-dependant activity of hsp70, by stimulating the hydrolysis of ATP. Hence, they appear pivotal actors of translation, (un)folding, translocation and degradation processes [175, 176]. Additionally, we could recognise five mitochondrial HSPs: one HSP10, two HSP60 and two HSP70.

2.5 Nuclear receptors and transcription factors

Finally, we searched for nuclear receptors (NR) and transcription factors (TF), as some of them are known to be involved in the sensing of xenobiotics and other environmental pollutants, in stress signalling and in the regulation of the inflammatory response [25]. NRs are effector proteins, characterized by a ligand-binding domain and a DNA-binding domain and constitute a large family, organized in six subfamilies [177, 178]. Once activated, NRs regulate the expression of genes that control development, metabolism or homeostasis [179]. The number of NRs found in *P. nobilis,* namely thirty-two, is in agreement with the

number of NRs found in other species, i.e. twenty-one in *Drosophila melanogaster [180], thirty-one in the copepod Tigriopus japonicus [181], thirty-three in the* owl limpet *Lottia gigantea [182], thirty-nine in the* gastropod snail *Biomphalaria glabrata [182], forty-three in the bivalve C. gigas [183] and forty-eight in humans [184]. Noteworthy, many of the NRs involved in the defensome, such as the e*cdysone receptor*, PXR, the estrogen receptor (ERR), PPAR and retinoic acid receptors [185, 186] were discerned in the P. nobilis genome.*

We also detected seventeen transcription factors involved in the chemical defensome in our assembly. Activated transcription factors bind specific responsive elements (REL) within regulatory regions and, in turn, modulate the expression of the responsive genes. For instance, hif–1 and ARNT bind hypoxia REL [187], mtf1 binds metal and hypoxia REL [188] and AhR/ARNT bind xenobiotic REL [187]. Some of these transcription factors actually interact with other proteins to form active dimers, such as Maf proteins that heterodimerize with nfe–2 or Bach proteins [189].

Overall, we report the existence of forty-nine NRs and TFs in the draft genome of *P. nobilis,* of which many are involved in the chemical defensome, unmasking the capacity of this marine bivalve to early detect and respond to harmful chemicals.

Mitochondrial genome

Mitochondria are essential components of the cell where they provide energy. As a reminiscence of symbiotic organisms, they possess their own genome. Conveniently, genome sequence data in eukaryotes also contain a large number of mitochondrial sequences due to the process of total DNA isolation that capture organelle nucleic acids.

Within our assembly, we identified one contig (Contig_66, 18799 nt) that matched the complete genome of the closely related penshell *Atrina pectinata* (NC_020028.1, [190]), over its entire length (68% covering, 80% identity, null e-value). The subsequent mapping of the trimmed raw reads against the mitochondrial contig, could confirm the unambiguously reconstructed mitochondrial genome with a mean coverage of approximately 6000X, and the absence of an alternative consensus sequence, that could have resulted from a putative homoplasy, due to accumulating mutations and/or the simultaneous presence of male and female mitochondrial genomes. Actually, our variant calling analysis highlighted only seven potential variations of low statistical significance (i.e. one indel and six SNPs) (Table S4 data).

We could map twelve genes encoding proteins of the respiratory chain [cytochrome c oxidase subunit I-III (*cox1, cox2* and *cox3),*, apocytochrome b (*cytb),*, ATPase 6 (*atp6),*, NADH dehydrogenase subunits 1–6, and 4L (*nad1–6* and *nad4L)]*], two for rRNA (*rrnS* and *rrnL),*, twenty-two genes for tRNA (trn) and a long non-coding region between tRNA-Cys and tRNA-Met, including a direct repeat region of 740 nt (Figure 7). Only ATP synthase subunit 8 (*atp8)* was missing from our annotation. However, this gene also seemed to be absent in other bivalve species. Whether *P. nobilis* and other bivalves are truly lacking the highly

divergent *atp8* gene remains debated [191]. The complete mitogenome DNA sequence should be useful for future research regarding the structuration of *P. nobilis* populations [192, 193].

Conclusion

Pinna nobilis is an emblematic and endemic bivalve of the Mediterranean Sea where it is distributed all around the basin, but for how long? At the end of the past century, recurrent local population collapses due to human activities have risen concerns about the persistence of this animal, leading to efficient conservation measures that allowed populations to recover. However, the actual ongoing MME caused by emerging pathogens resurrects major worries for the survival of this species at a global level. In Spain, where the outbreak started in early autumn 2016, and where high mortality rates reaching up to 100% of the surveyed populations were observed [15], authorities changed the status of *P. nobilis* from “endangered” to “endangered with extinction” (Orden TEC/596/2019, Ministerio para la transición Ecológica, 8 of April, 2019). The MME has since spread throughout the Mediterranean Sea [19, 20, Vicente, personal communication] and IUCN is now considering adding *P. nobilis* on the Red List of threatened species. Nevertheless, a few populations or individuals subsist in some locations, even in places surrounded by the epizooty, e.g., Alfacs Bay and Mar Menor in Spain [17], the Thau Basin in south of France, the Diana basin in Corsica (Vicente pers. obs. Sep. 2019) or in the inner part of the bay of Kotor in Montenegro (Martinović pers. data August 2019). So far, no clear evidence has been provided to explain the reasons behind the witnessed resistance of these populations (or individuals) to the disease. Nonetheless, this phenomenon could obviously be associated to either the challenging physical and chemical environmental conditions (e.g. significant and frequent variation of temperature and/or salinity) in these specific sites that would be not suitable for the dissemination of the pathogens or a local genetic adaptation of these *P. nobilis* populations (individuals). Yet, all these sites are under strong anthropogenic pressures. Therefore, this situation is all the more worrying that the organism physiology is stressed by chemical and physical factors (pollutants, habitat degradation, climate change) usually weakening the immune response [195–198].

In this study we reveal that *P. nobilis* has a complete chemical defensome, that globally resembles that of other sessile marine organisms. Similarly, *P. nobilis* has a sophisticated innate immune system, based on an extended number of recognition receptors (i.e. 38 TLR, 27 lectins, 16 laminins) and cytokines (IL–17, MIF), a well conserved TLR pathway, the presence of the main effectors of complement cascade and a complex apoptosis pathway. However, we were unable to clearly detect AMPs, orthologous to those evidenced in the *Mytilus* genera, although several contigs were found to encode cysteine-rich peptides. A transcriptomic study performed on fan mussels challenged with pathogenic bacteria would likely help clarifying with point, and potentially reveals species-specific AMPs. Overall, the gene diversity of the innate immune system appears quite developed, similar to that observed in the common mussel *M. edulis* [24].

The extensive number of recognition receptors seems obvious with respect to the highly diverse, commensal and pathogenic microbial fauna surrounding and hosted by *P. nobilis* individuals.

Nevertheless, it questions about how, at the molecular level, *H. pinnae* invade tissues and cells, and circumvent the innate immunity. The genome comparison of resistant *versus* sensitive individuals, or species across the *Pinna* genera (e.g. *P. nobilis vs Pinna rudis* since *P. rudis* seems so far to be unaffected by the outbreak, Garcia-March and Vicente, pers. obs.), will help to find putative regions under selection for these traits. Actually, the resistance to pathogen or abiotic stress do not only rely upon gene content but also on how gene expression is finely regulated. Regarding gene regulation, notwithstanding trans-regulating elements, *P. nobilis* would also likely methylate DNA, as we could detect a DNA methyl transferase and predict several hundred CpG islands (data not shown). Hence, the next steps forward the comprehension of pathogen resistance will be to investigate other genomes and transcriptomes.

To conclude, this genome assembly represents the first reference genome in the Pinnidae family and the second in the order of Pteriida, after the pearl oyster *P. fucata.* This reference genome (including the mitogenome) should be considered in future research regarding comparative evolution, comparative genomics or studies aiming to understand the genetic variability of *P. nobilis* populations, in order to evolve effective conservation plans.

Declarations

Availability of supporting data

Illumina read sequences used in this study can be retrieved from the NCBI Sequence Read Archive under the accessions XXXXXXXXX and XXXXXXXXX. Contigs have been submitted to NCBI under the submission code XXXXXXXX. The full annotated mitochondrial genome has been deposited under GenBanK accession number MN432488.

Acknowledgement

This work was supported by The Foundation Prince Albert II of Monaco under the grant n°1796 related to the PinnaSpot project: “BF/HEM 15–1662, The study, protection and possible breeding of pen shell (*Pinna nobilis)* in the Boka Kotorska Bay”.

Ethic Statement

No experimentation involving removal, dislocation or killing of *Pinna nobilis* individuals was performed. The authorization to sample *P. nobilis* was obtained from the authority responsible for the concerned maritime area, *i.e.* the DDTM (Direction Départementale des Territoires et de la Mer du Var).

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

RB, SC, JLB, NV, JRGM, RM, JTM and DJ designed and supervised the study. RB, SC and NV participated in the sampling and experiments. RB and SC lead the analysis of the data. JMP, SC and RB contributed to computer analysis tools and conducted bioinformatic analyses. RB and SC wrote the paper. All authors read and approved the final manuscript. RB and SC contributed equally to this work.

Abbreviations

AMP: Antimicrobial petide

CRD: Carbohydrate recognition domain

CYP: Cytochrome P450

ECM: Extracellular matrix

GNBP: Glucan binding protein

GO: Gene ontology

GSH: Glutathione

GST: Glutathion S-transferase

HSP: Heat shock protein

IL: Interleukin

LPS: Lipopolysaccharide

LRR: Leucin-rich region

MIF: Macrophage inhibitory factor

MME: Mass mortality event

NR: Nuclear receptor

PGN: Peptidoglycan

PGRP: Peptidoglycan recognition receptor

PRR: Pathogen recognition receptor

RE: Repetitive elements

REL: Responsive element

SSR: Simple sequence repeats

SULT : Sulfotransferase

TIR: Toll-interleukin receptor

TLR: Toll-like receptor

TM: Transmembrane domain

TNF: Tumor necrosis factor

UGT : UDP-glucuronosyl and glucosyl transferase

References

1.Vicente N, Moreteau J: *Statut de Pinna nobilis L. en Mediterranee (mollusque eulamellibranche). Les espèces marines à protéger en Méditerranée* 1991:159–168.

2.Zavodnik D, Hrs-Brenko M, Legac M: *Synopsis on the fan shell Pinna nobilis L. in the eastern Adriatic Sea. Les espèces marines à protéger en Méditerranée* 1991:169–178.

3.Rouanet E, Trigos S, Vicente N: *From youth to death of old age: the 50-year story of a Pinna nobilis fan mussel population at Port-Cros Island (Port-Cros National Park, Provence, Mediterranean Sea). Scientific Reports of Port–Cros National Park* 2015, *29:*209–222.

4.Butler A, Vicente N, De Gaulejac B: *Ecology of the pterioid bivalves Pinna bicolor Gmelin and Pinna nobilis L. Marine Life* 1993, *3(*1–2):37–45.

5.Trigos S, García-March J, Vicente N, Tena J, Torres J: *Utilization of muddy detritus as organic matter source by the fan mussel Pinna nobilis. Mediterranean Marine Science* 2014, *15(*3):667–674.

6.Rabaoui L, Zouari ST, Hassine OKB: *Two species of Crustacea (Decapoda) associated with the fan mussel, Pinna nobilis Linnaeus, 1758 (Mollusca, Bivalvia). Crustaceana* 2008:433–446.

7.Addis P, Secci M, Brundu G, Manunza A, Corrias S, Cau A: *Density, size structure, shell orientation and epibiontic colonization of the fan mussel Pinna nobilis L. 1758 (Mollusca: Bivalvia) in three contrasting habitats in an estuarine area of Sardinia (W Mediterranean). Scientia Marina* 2009, *73(*1):143–152.

8.Rabaoui L, Tlig-Zouari S, Cosentino A, Hassine OKB: *Associated fauna of the fan shell Pinna nobilis (Mollusca: Bivalvia) in the northern and eastern Tunisian coasts. Scientia Marina* 2009, *73(*1):129–141.

9.Basso L, Vazquez-Luis M, Garcia-March JR, Deudero S, Alvarez E, Vicente N, Duarte CM, Hendriks IE: *The pen shell, Pinna nobilis: A review of population status and recommended research priorities in the Mediterranean Sea. Advances in marine biology* 2015, *71:*109–160.

10.Médioni E. V, N.: *Kinetic study of the populations of Pinna nobilis L.1758, on the French Mediterranean coasts. Premier séminaire international sur la grande nacre de Méditerranée 10–12 octobre 2002, Ile des Embiez, France* 2002, *Mémoires de l'Institut océanographique Paul Ricard:*43–48.

11.Vicente N: *La grande nacre de Méditerranée Pinna nobilis. Présentation générale. Premier séminaire international sur la grande nacre de Méditerranée 10–12 octobre 2002, Ile des Embiez, France* 2002, *Mémoires de l'Institut océanographique Paul Ricard:*7–16.

12.Siletic T, Peharda M: *Population study of the fan shell Pinna nobilis L. in Malo and Veliko Jezero of the Mljet National Park (Adriatic Sea). Scientia Marina* 2003, *67(*1):91–98.

13.Centoducati G, Tarsitano E, Bottalico A, Marvulli M, Lai OR, Crescenzo G: *Monitoring of the endangered Pinna nobilis Linne, 1758 in the Mar Grande of Taranto (Ionian sea, Italy). Environmental Monitoring and Assessment* 2007, *131(*1–3):339–347.

14.Darriba S: *First haplosporidan parasite reported infecting a member of the Superfamily Pinnoidea (Pinna nobilis) during a mortality event in Alicante (Spain, Western Mediterranean). Journal of invertebrate pathology* 2017, *148:*14–19.

15.Vázquez-Luis M, Álvarez E, Barrajón A, García-March JR, Grau A, Hendriks IE, Jiménez S, Kersting D, Moreno D, Pérez M: *SOS Pinna nobilis: a mass mortality event in western Mediterranean Sea. Frontiers in Marine Science* 2017, *4:*220.

16.Carella F, Aceto S, Pollaro F, Miccio A, Iaria C, Carrasco N, Prado P, De Vico G: *A mycobacterial disease is associated with the silent mass mortality of the pen shell Pinna nobilis along the Tyrrhenian coastline of Italy. Scientific reports* 2019, *9(*1):2725.

17.García-March JR T-MJ, Henandis-Caballero S, Vázquez-Luis M, López D, Téllez C, Prado P, Navas JI, Bernal J, Catanese G, Grau A, López-Sanmartín M, Nebot-Colomer E, Ortega A, Planes S, Kersting D, Jimenez S, Hendriks I, Moreno D, Giménez-Casalduero F, Pérez M, Izquierdo A, Sánchez J, Vicente N, Sanmarti N, Guimerans Mt, Crespo JL, Valencia JM, Torres J, Barrajon A, Álvarez E, Peyran C, Morage T, Deudero S,: *Can we save a marine species affected by a highly-infective-highly-lethal-waterborne disease from extinction? Biological Conservation (*In review).

18.IUCN: *https://www.iucn.org/news/mediterranean/201907/mediterranean-noble-pen-shell-crisis-pinna-nobilis-june–2019-update.* 2019.

19.Katsanevakis S, Tsirintanis K, Tsaparis D, Doukas D, Sini M, Athanassopoulou F, Kolygas M, Tontis D, Koutsoubas D, Bakopoulos V: *The cryptogenic parasite Haplosporidium pinnae invades the Aegean Sea*

and causes the collapse of Pinna nobilis populations. Aquatic Invasions 2019, *14(*2):150–164.

20.Panarese R, Tedesco P, Chimienti G, Latrofa MS, Quaglio F, Passantino G, Buonavoglia C, Gustinelli A, Tursi A, Otranto D: *Haplosporidium pinnae associated with mass mortality in endangered Pinna nobilis (Linnaeus 1758) fan mussels. Journal of invertebrate pathology* 2019, *164:*32–37.

21.Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, Funayama N, Agata K, Bosch TC: *The innate immune repertoire in Cnidaria - ancestral complexity and stochastic gene loss. Genome Biology* 2007, *8(*4):R59.

22.Ponting CP: *The functional repertoires of metazoan genomes. Nature Reviews Genetics* 2008, *9(*9):689.

23.Gerdol M: *Immune-related genes in gastropods and bivalves: a comparative overview. Invertebrate Survival Journal* 2017(1):103–118%V 114.

24.Philipp EE, Kraemer L, Melzner F, Poustka AJ, Thieme S, Findeisen U, Schreiber S, Rosenstiel P: *Massively parallel RNA sequencing identifies a complex immune gene repertoire in the lophotrochozoan Mytilus edulis. PLoS One* 2012, *7(*3):e33091.

25.Goldstone JV, Hamdoun A, Cole BJ, Howard-Ashby M, Nebert DW, Scally M, Dean M, Epel D, Hahn ME, Stegeman JJ: *The chemical defensome: environmental sensing and response genes in the Strongylocentrotus purpuratus genome. Dev Biol* 2006, *300(*1):366–384.

26.Du X, Fan G, Jiao Y, Zhang H, Guo X, Huang R, Zheng Z, Bian C, Deng Y, Wang Q: *The pearl oyster Pinctada fucata martensii genome and multi-omic analyses provide insights into biomineralization. GigaScience* 2017, *6(*8):gix059.

27.Murgarella M, Puiu D, Novoa B, Figueras A, Posada D, Canchaya C: *A First Insight into the Genome of the Filter-Feeder Mussel Mytilus galloprovincialis. PLoS One* 2016, *11(*3):e0151561-e0151561.

28.Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H: *The oyster genome reveals stress adaptation and complexity of shell formation. Nature* 2012, *490(*7418):49.

29.Gestal C, Roch P, 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. Reviews in Fisheries Science* 2008, *16(*sup1):133–156.

30.Gerdol M, Venier P: *An updated molecular basis for mussel immunity. Fish & shellfish immunology* 2015, *46(*1):17–38.

31.Venier P, Varotto L, Rosani U, Millino C, Celegato B, Bernante F, Lanfranchi G, Novoa B, Roch P, Figueras A: *Insights into the innate immunity of the Mediterranean mussel Mytilus galloprovincialis. BMC genomics* 2011, *12(*1):69.

32.Watermann B, 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. Journal of Fish Diseases* 2008, *31(*8):621–630.

33.Gurevich A, Saveliev V, Vyahhi N, Tesler G: *QUAST: quality assessment tool for genome assemblies. Bioinformatics* 2013, *29(*8):1072–1075.

34.Marçais G, Kingsford C: *A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics* 2011, *27(*6):764–770.

35.Kelley DR, Schatz MC, Salzberg SL: *Quake: quality-aware detection and correction of sequencing errors. Genome Biology* 2010, *11(*11):R116.

36.Gregory TR: *Animal Genome Size Database. http://www.genomesize.com.* 2019.

37.Takeuchi T, Kawashima T, Koyanagi R, Gyoja F, Tanaka M, Ikuta T, Shoguchi E, Fujiwara M, Shinzato C, Hisata K: *Draft genome of the pearl oyster Pinctada fucata: a platform for understanding bivalve biology. DNA research* 2012, *19(*2):117–130.

38.Chikhi R, Medvedev P: *Informed and automated k-mer size selection for genome assembly. Bioinformatics* 2013, *30(*1):31–37.

39.Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC: *GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics* 2017, *33(*14):2202–2204.

40.*http://www.repeatmasker.org.*

41.Pickett BD, Miller JB, Ridge PG: *Kmer-SSR: a fast and exhaustive SSR search algorithm. Bioinformatics* 2017, *33(*24):3922–3928.

42.Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M: *Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics* 2005, *21(*18):3674–3676.

43.Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: *BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics* 2015, *31(*19):3210–3212.

44.Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM: *BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular biology and evolution* 2017, *35(*3):543–548.

45.Stanke M, Steinkamp R, Waack S, Morgenstern B: *AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic acids research* 2004, *32(*suppl_2):W309-W312.

46.Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G *et al: InterProScan 5: genome-scale protein function classification. Bioinformatics* 2014, *30(*9):1236–1240.

47.De Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N: *ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic acids research* 2006, *34(*suppl_2):W362-W365.

48.Langmead B, Salzberg SL: *Fast gapped-read alignment with Bowtie 2. Nature Methods* 2012, *9:*357.

49.Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP: *The Sequence Alignment/Map format and SAMtools. Bioinformatics* 2009, *25(*16):2078–2079.

50.Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF: *MITOS: improved de novo metazoan mitochondrial genome annotation. Molecular phylogenetics and evolution* 2013, *69(*2):313–319.

51.*http://mitos2.bioinf.uni-leipzig.de.*

52.*https://www.ncbi.nlm.nih.gov/orffinder/.*

53.Rice P, Longden I, Bleasby A: *EMBOSS: the European molecular biology open software suite.* In.*:* Elsevier current trends; 2000.

54.Benson G: *Tandem repeats finder: a program to analyze DNA sequences. Nucleic acids research* 1999, *27(*2):573–580.

55.Gong C, Maquat LE: *lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. Nature* 2011, *470(*7333):284–288.

56.Makałowski W: *Genomic scrap yard: how genomes utilize all that junk. Gene* 2000, *259(*1):61–67.

57.Ponicsan SL, Kugel JF, Goodrich JA: *Genomic gems: SINE RNAs regulate mRNA production. Current opinion in genetics & development* 2010, *20(*2):149–155.

58.Bartolomé C, Maside X, Charlesworth B: *On the abundance and distribution of transposable elements in the genome of Drosophila melanogaster. Molecular biology and evolution* 2002, *19(*6):926–937.

59.Finnegan DJ: *Transposable elements. Current Opinion in Genetics & Development* 1992, *2(*6):861–867.

60.Hollister JD, Gaut BS: *Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Research* 2009,

*19(*8):1419–1428.

61.Sun J, Zhang Y, Xu T, Zhang Y, Mu H, Zhang Y, Lan Y, Fields CJ, Hui JHL, Zhang W: *Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. Nature ecology & evolution* 2017, *1(*5):0121.

62.Wong YH, Sun J, He LS, Chen LG, Qiu J-W, Qian P-Y: *High-throughput transcriptome sequencing of the cold seep mussel Bathymodiolus platifrons. Scientific reports* 2015, *5:*16597.

63.Xu M, Huang J, Shi Y, Zhang H, He M: *Comparative transcriptomic and proteomic analysis of yellow shell and black shell pearl oysters, Pinctada fucata martensii. BMC genomics* 2019, *20(*1):469.

64.Gerdol M, Manfrin C, De Moro G, Figueras A, Novoa B, Venier P, Pallavicini A: *The C1q domain containing proteins of the Mediterranean mussel Mytilus galloprovincialis: A widespread and diverse family of immune-related molecules. Developmental & Comparative Immunology* 2011, *35(*6):635–643.

65.Toubiana M, Gerdol M, Rosani U, Pallavicini A, Venier P, Roch P: *Toll-like receptors and MyD88 adaptors in Mytilus: complete cds and gene expression levels. Developmental & Comparative Immunology* 2013, *40(*2):158–166.

66.Barral PM, Sarkar D, Su Z-z, Barber GN, DeSalle R, Racaniello VR, Fisher PB: *Functions of the cytoplasmic RNA sensors RIG-I and MDA–5: key regulators of innate immunity. Pharmacology & therapeutics* 2009, *124(*2):219–234.

67.Cao X: *New DNA-sensing pathway feeds RIG-I with RNA. Nature immunology* 2009, *10(*10):1049.

68.Aderem A, Ulevitch RJ: *Toll-like receptors in the induction of the innate immune response. Nature* 2000, *406(*6797):782.

69.Vijay K: *Toll-like receptors in immunity and inflammatory diseases: past, present, and future. International immunopharmacology* 2018, *59:*391–412.

70.Smith LC, Rast JP, Brockton V, Terwilliger DP, Nair SV, Buckley KM, Majeske AJ: *The sea urchin immune system. Invertebrate Survival Journal* 2006, *3(*1):25–39.

71.Tu Q, Cameron RA, Worley KC, Gibbs RA, Davidson EH: *Gene structure in the sea urchin Strongylocentrotus purpuratus based on transcriptome analysis. Genome research* 2012, *22(*10):2079–2087.

72.Mateo DR, Greenwood SJ, Araya MT, Berthe FCJ, Johnson GR, Siah A: *Differential gene expression of γ-actin, Toll-like receptor 2 (TLR–2) and interleukin–1 receptor-associated kinase 4 (IRAK–4) in Mya arenaria haemocytes induced by in vivo infections with two Vibrio splendidus strains. Developmental and Comparative Immunology* 2010, *34(7*):710–714.

73.Qiu L, Song L, Xu W, Ni D, Yu Y: *Molecular cloning and expression of a Toll receptor gene homologue from Zhikong Scallop, Chlamys farreri. Fish & Shellfish Immunology* 2007, *22(*5):451–466.

74.Itoh N, Takahashi KG: *A novel peptidoglycan recognition protein containing a goose-type lysozyme domain from the Pacific oyster, Crassostrea gigas. Molecular Immunology* 2009, *46(*8):1768–1774.

75.Ni D, Song L, Wu L, Chang Y, Yu Y, Qiu L, Wang L: *Molecular cloning and mRNA expression of peptidoglycan recognition protein (PGRP) gene in bay scallop (Argopecten irradians, Lamarck 1819). Developmental & Comparative Immunology* 2007, *31(*6):548–558.

76.Dziarski R: *Peptidoglycan recognition proteins (PGRPs). Molecular immunology* 2004, *40(*12):877–886.

77.Thörnqvist P-O, Johansson MW, Söderhäll K: *Opsonic activity of cell adhesion proteins and β–1, 3-glucan binding proteins from two crustaceans. Developmental & Comparative Immunology* 1994, *18(*1):3–12.

78.Li X, Wang S, Qi J, Echtenkamp SF, Chatterjee R, Wang M, Boons G-J, Dziarski R, Gupta D: *Zebrafish peptidoglycan recognition proteins are bactericidal amidases essential for defense against bacterial infections. Immunity* 2007, *27(*3):518–529.

79.Royet J, Dziarski R: *Peptidoglycan recognition proteins: pleiotropic sensors and effectors of antimicrobial defences. Nature Reviews Microbiology* 2007, *5(*4):264–277.

80.Vargas-Albores F, Yepiz-Plascencia G: *Beta glucan binding protein and its role in shrimp immune response. Aquaculture* 2000, *191(*1–3):13–21.

81.Taylor ME, Drickamer K: *Structure–function analysis of C-type animal lectins.* In: *Methods in enzymology.* vol. 363: Elsevier; 2003: 3–16.

82.McGreal EP, Martinez-Pomares L, Gordon S: *Divergent roles for C-type lectins expressed by cells of the innate immune system. Molecular Immunology* 2004, *41(*11):1109–1121.

83.*www.imperial.ac.uk/research/animallectins/ctld/lectins.html.*

84.Iordache F, Ionita M, Mitrea LI, Fafaneata C, Pop A: *Antimicrobial and antiparasitic activity of lectins. Current pharmaceutical biotechnology* 2015, *16(*2):152–161.

85.Fujita T, Matsushita M, Endo Y: *The lectin-complement pathway–its role in innate immunity and evolution. Immunological reviews* 2004, *198(*1):185–202.

86.Elumalai P, Rubeena AS, Arockiaraj J, Wongpanya R, Cammarata M, Ringø E, Vaseeharan B: *The Role of Lectins in Finfish: A Review. Reviews in Fisheries Science & Aquaculture* 2019, *27(*2):152–169.

87.Aumailley M: *The laminin family. Cell adhesion & migration* 2013, *7(*1):48–55.

88.Ekblom P, Lonai P, Talts JF: *Expression and biological role of laminin–1. Matrix Biology* 2003, *22(*1):35–47.

89.Kleinman HK, Cannon FB, Laurie GW, Hassell JR, Aumailley M, Terranova VP, Martin GR, DuBois-Dalcq M: *Biological activities of laminin. Journal of Cellular Biochemistry $V 27* 1985(4):317–325.

90.Mercurio AM: *Laminin receptors: achieving specificity through cooperation. Trends in Cell Biology* 1995, *5(*11):419–423.

91.Simon T, Bromberg JS: *Regulation of the immune system by laminins. Trends in immunology* 2017, *38(*11):858–871.

92.Nde PN, Simmons KJ, Kleshchenko YY, Pratap S, Lima MF, Villalta F: *Silencing of the Laminin γ–1 Gene Blocks Trypanosoma cruzi Infection. Infection and Immunity* 2006, *74(*3):1643–1648.

93.Van Wijk XM, Döhrmann S, Hallström BM, Li S, Voldborg BG, Meng BX, McKee KK, Van Kuppevelt TH, Yurchenco PD, Palsson BO: *Whole-genome sequencing of invasion-resistant cells identifies laminin α2 as a host factor for bacterial invasion. MBio* 2017, *8(*1):e02128–02116.

94.Warburg A, Shtern A, Cohen N, Dahan N: *Laminin and a Plasmodium ookinete surface protein inhibit melanotic encapsulation of Sephadex beads in the hemocoel of mosquitoes. Microbes and infection* 2007, *9(*2):192–199.

95.Vasta GR: *Roles of galectins in infection. Nature Reviews Microbiology* 2009, *7(*6):424.

96.Catanese G, Grau A, Valencia JM, Garcia-March JR, Vázquez-Luis M, Alvarez E, Deudero S, Darriba S, Carballal MJ, Villalba A: *Haplosporidium pinnae sp. nov., a haplosporidan parasite associated with mass mortalities of the fan mussel, Pinna nobilis, in the Western Mediterranean Sea. Journal of invertebrate pathology* 2018, *157:*9–24.

97.Li F, Huang S, Wang L, Yang J, Zhang H, Qiu L, Li L, Song L: *A macrophage migration inhibitory factor like gene from scallop Chlamys farreri: Involvement in immune response and wound healing. Developmental & Comparative Immunology* 2011, *35(*1):62–71.

98.Parisi M-G, Toubiana M, Mangano V, Parrinello N, Cammarata M, Roch P: *MIF from mussel: coding sequence, phylogeny, polymorphism, 3D model and regulation of expression. Developmental & Comparative Immunology* 2012, *36(*4):688–696.

99.Wang B, Zhang Z, Wang Y, Zou Z, Wang G, Wang S, Jia X, Lin P: *Molecular cloning and characterization of macrophage migration inhibitory factor from small abalone Haliotis diversicolor supertexta. Fish & Shellfish Immunology* 2009, *27(*1):57–64.

100.Rosani U, Varotto L, Gerdol M, Pallavicini A, Venier P: *IL–17 signaling components in bivalves: comparative sequence analysis and involvement in the immune responses. Developmental &*

Comparative Immunology 2015, *52(*2):255–268.

101.Roberts S, Gueguen Y, de Lorgeril J, Goetz F: *Rapid accumulation of an interleukin 17 homolog transcript in Crassostrea gigas hemocytes following bacterial exposure. Developmental & Comparative Immunology* 2008, *32(*9):1099–1104.

102.Betti M, Ciacci C, Lorusso LC, Canonico B, Falcioni T, Gallo G, Canesi L: *Effects of tumour necrosis factor α (TNFα) on Mytilus haemocytes: role of stress-activated mitogen-activated protein kinases (MAPKs). Biology of the Cell* 2006, *98(*4):233–244.

103.Lacoste A, Jalabert F, Malham SK, Cueff A, Poulet SA: *Stress and stress-induced neuroendocrine changes increase the susceptibility of juvenile oysters (Crassostrea gigas) to Vibrio splendidus. Appl Environ Microbiol* 2001, *67(*5):2304–2309.

104.Liu Z, Li M, Yi Q, Wang L, Song L: *The Neuroendocrine-Immune Regulation in Response to Environmental Stress in Marine Bivalves. Frontiers in Physiology* 2018, *9(*1456).

105.Shaliapina V: *The functional swings in the neuroendocrine regulation of stress. Fiziologicheskii zhurnal imeni IM Sechenova* 1996, *82(*4):9–14.

106.Wang L, Song X, Song L: *The oyster immunity. Developmental & Comparative Immunology* 2018, *80:*99–118.

107.Janeway Jr CA, Travers P, Walport M, Shlomchik MJ: *The complement system and innate immunity.* In: *Immunobiology: The Immune System in Health and Disease 5th edition.* Garland Science; 2001.

108.Gasque P: *Complement: a unique innate immune sensor for danger signals. Molecular Immunology* 2004, *41(*11):1089–1098.

109.Ward PA: *Role of Complement, Chemokines, & Regulatory Cytokines in Acute Lung Injury.* 1996.

110.Nonaka M: *The complement C3 protein family in invertebrates. Invertebrate Survival Journal* 2011, *8(*1):21–32.

111.Vizioli J, Salzet M: *Antimicrobial peptides from animals: focus on invertebrates. Trends in pharmacological sciences* 2002, *23(*11):494–496.

112.Brogden KA: *Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nature reviews microbiology* 2005, *3(*3):238.

113.Sitaram N, Nagaraj R: *Host-defense antimicrobial peptides: importance of structure for activity. Current pharmaceutical design* 2002, *8(*9):727–742.

114.Gerdol M, De Moro G, Manfrin C, Venier P, Pallavicini A: *Big defensins and mytimacins, new AMP families of the Mediterranean mussel Mytilus galloprovincialis. Developmental & Comparative*

Immunology 2012, *36(*2):390–399.

115.Padhi A, Verghese B: *Molecular diversity and evolution of myticin-C antimicrobial peptide variants in the Mediterranean mussel, Mytilus galloprovincialis. Peptides* 2008, *29(*7):1094–1101.

116.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. Marine biotechnology* 2011, *13(*5):857–867.

117.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. Journal of Biological Chemistry* 1996, *271(*36):21808–21813.

118.Furukawa R, Matsumoto M, Kaneko H: *Characterization of a scavenger receptor cysteine-rich-domain-containing protein of the starfish, Asterina pectinifera: ApSRCR1 acts as an opsonin in the larval and adult innate immune systems. Developmental & Comparative Immunology* 2012, *36(*1):51–61.

119.Liu Z, Dong S, Xu J, Zeng M, Song H, Zhao Y: *Production of cysteine-rich antimicrobial peptide by digestion of oyster (Crassostrea gigas) with alcalase and bromelin. Food Control* 2008, *19(*3):231–235.

120.Sarrias MR, Gronlund J, Padilla O, Madsen J, Holmskov U, Lozano F: *The Scavenger Receptor Cysteine-Rich (SRCR) domain: an ancient and highly conserved protein module of the innate immune system. Critical Reviews™ in Immunology* 2004, *24(*1).

121.Everett H, McFadden G: *Apoptosis: an innate immune response to virus infection. Trends in Microbiology* 1999, *7(*4):160–165.

122.Miao EA, Leaf IA, Treuting PM, Mao DP, Dors M, Sarkar A, Warren SE, Wewers MD, Aderem A: *Caspase–1-induced pyroptosis is an innate immune effector mechanism against intracellular bacteria. Nature Immunology* 2010, *11:*1136.

123.Sokolova I: *Apoptosis in molluscan immune defense. Invertebrate Survival Journal* 2009, *6(*1):49–58.

124.Kültz D: *MOLECULAR AND EVOLUTIONARY BASIS OF THE CELLULAR STRESS RESPONSE. Annual Review of Physiology* 2005, *67(*1):225–257.

125.Goldstone JV: *Environmental sensing and response genes in cnidaria: the chemical defensome in the sea anemone Nematostella vectensis. Cell biology and toxicology* 2008, *24(*6):483–502.

126.Werck-Reichhart D, Feyereisen R: *Cytochromes P450: a success story. Genome biology* 2000, *1(*6):reviews3003. 3001.

127.Nelson DR: *Cytochrome P450 Nomenclature, 2004.* In: *Cytochrome P450 protocols.* Springer; 2006: 1–10.

128.Zanette J, Goldstone JV, Bainy AC, Stegeman JJ: *Identification of CYP genes in Mytilus (mussel) and Crassostrea (oyster) species: first approach to the full complement of cytochrome P450 genes in bivalves. Marine environmental research* 2010, *69:*S1-S3.

129.Goldstone JV, Sundaramoorthy M, Zhao B, Waterman MR, Stegeman JJ, Lamb DC: *Genetic and structural analyses of cytochrome P450 hydroxylases in sex hormone biosynthesis: Sequential origin and subsequent coevolution. Molecular Phylogenetics and Evolution* 2016, *94:*676–687.

130.Hallmann A, Konieczna L, Swiezak J, Milczarek R, Smolarz K: *Aromatisation of steroids in the bivalve Mytilus trossulus. PeerJ* 2019, *7:*e6953.

131.Blanchard RL, Freimuth RR, Buck J, Weinshilboum RM, Coughtrie MW: *A proposed nomenclature system for the cytosolic sulfotransferase (SULT) superfamily. Pharmacogenetics and Genomics* 2004, *14(*3):199–211.

132.Coughtrie MW: *Function and organization of the human cytosolic sulfotransferase (SULT) family. Chemico-biological interactions* 2016, *259:*2–7.

133.Bojarová P, Williams SJ: *Sulfotransferases, sulfatases and formylglycine-generating enzymes: a sulfation fascination. Current opinion in chemical biology* 2008, *12(*5):573–581.

134.Bundgaard JR, Rehfeld JF: *Chapter 250 - Tyrosylprotein sulfotransferases.* In: *Handbook of Biologically Active Peptides (Second Edition).* Edited by Kastin AJ. Boston: Academic Press; 2013: 1829–1834.

135.Li Y, Sun X, Hu X, Xun X, Zhang J, Guo X, Jiao W, Zhang L, Liu W, Wang J *et al: Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. Nature Communications* 2017, *8(*1):1721.

136.Frova C: *Glutathione transferases in the genomics era: New insights and perspectives. Biomolecular Engineering* 2006, *23(*4):149–169.

137.Hayes JD, Flanagan JU, Jowsey IR: *GLUTATHIONE TRANSFERASES. Annual Review of Pharmacology and Toxicology* 2005, *45(*1):51–88.

138.Ranson H, Hemingway J: *Mosquito glutathione transferases. Methods in enzymology* 2005, *401:*226–241.

139.Sheehan D, Meade G, Foley VM: *Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. Biochemical Journal* 2001, *360(*1):1–16.

140.Che-Mendoza A, Penilla RP, Rodríguez DA: *Insecticide resistance and glutathione S-transferases in mosquitoes: a review. African Journal of Biotechnology* 2009, *8(*8).

141.Sandamalika WG, Priyathilaka TT, Lee S, Yang H, Lee J: *Immune and xenobiotic responses of glutathione S-Transferase theta (GST-θ) from marine invertebrate disk abalone (Haliotis discus discus): With molecular characterization and functional analysis. Fish & shellfish immunology* 2019, *91:*159–171.

142.Board PG, Menon D: *Glutathione transferases, regulators of cellular metabolism and physiology. Biochimica et biophysica acta (bba)-general subjects* 2013, *1830(*5):3267–3288.

143.Holm PJ, Bhakat P, Jegerschöld C, Gyobu N, Mitsuoka K, Fujiyoshi Y, Morgenstern R, Hebert H: *Structural basis for detoxification and oxidative stress protection in membranes. Journal of molecular biology* 2006, *360(*5):934–945.

144.Molina DM, Wetterholm A, Kohl A, McCarthy AA, Niegowski D, Ohlson E, Hammarberg T, Eshaghi S, Haeggström JZ, Nordlund P: *Structural basis for synthesis of inflammatory mediators by human leukotriene C 4 synthase. Nature* 2007, *448(*7153):613.

145.Best JH, Pflugmacher S, Wiegand C, Eddy FB, Metcalf JS, Codd GA: *Effects of enteric bacterial and cyanobacterial lipopolysaccharides, and of microcystin-LR, on glutathione S-transferase activities in zebra fish (Danio rerio). Aquatic toxicology* 2002, *60(*3–4):223–231.

146.Sundberg K, Widersten M, Seidel A, Mannervik B, Jernström B: *Glutathione conjugation of bay-and fjord-region diol epoxides of polycyclic aromatic hydrocarbons by glutathione transferases M1–1 and P1–1. Chemical research in toxicology* 1997, *10(*11):1221–1227.

147.Blackburn AC, Matthaei KI, Lim C, Taylor MC, Cappello JY, Hayes JD, Anders M, Board PG: *Deficiency of glutathione transferase zeta causes oxidative stress and activation of antioxidant response pathways. Molecular pharmacology* 2006, *69(*2):650–657.

148.Enayati AA, Ranson H, Hemingway J: *Insect glutathione transferases and insecticide resistance. Insect Molecular Biology* 2005, *14(*1):3–8.

149.McLellan LI, Wolf CR: *Glutathione and glutathione-dependent enzymes in cancer drug resistance. Drug Resistance Updates* 1999, *2(*3):153–164.

150.Revathy KS, Umasuthan N, Lee Y, Choi CY, Whang I, Lee J: *First molluscan theta-class Glutathione S-Transferase: Identification, cloning, characterization and transcriptional analysis post immune challenges. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 2012, *162(*1–3):10–23.

151.Roncalli V, Cieslak MC, Passamaneck Y, Christie AE, Lenz PH: *Glutathione S-transferase (GST) gene diversity in the crustacean Calanus finmarchicus–contributors to cellular detoxification. PLoS One* 2015, *10(*5):e0123322.

152.Dean M, Hamon Y, Chimini G: *The human ATP-binding cassette (ABC) transporter superfamily. Journal of lipid research* 2001, *42(*7):1007–1017.

153.Della Torre C, Bocci E, Focardi SE, Corsi I: *Differential ABCB and ABCC gene expression and efflux activities in gills and hemocytes of Mytilus galloprovincialis and their involvement in cadmium response. Marine environmental research* 2014, *93:*56–63.

154.Huang L, Liu S-L, Zheng J-W, Li H-Y, Liu J-S, Yang W-D: *P-glycoprotein and its inducible expression in three bivalve species after exposure to Prorocentrum lima. Aquatic Toxicology* 2015, *169:*123–132.

155.Kingtong S, Chitramvong Y, Janvilisri T: *ATP-binding cassette multidrug transporters in Indian-rock oyster Saccostrea forskali and their role in the export of an environmental organic pollutant tributyltin. Aquatic Toxicology* 2007, *85(*2):124–132.

156.Leslie EM, Deeley RG, Cole SPC: *Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense. Toxicology and Applied Pharmacology $V 204* 2005(3):216–237.

157.Popovic M, Zaja R, Loncar J, Smital T: *A novel ABC transporter: the first insight into zebrafish (Danio rerio) ABCH1. Marine environmental research* 2010, *69:*S11-S13.

158.Achard M, Baudrimont M, Boudou A, Bourdineaud J: *Induction of a multixenobiotic resistance protein (MXR) in the Asiatic clam Corbicula fluminea after heavy metals exposure. Aquatic toxicology* 2004, *67(*4):347–357.

159.Acs A, Imre K, Kiss G, Csaba J, Gyori J, Vehovszky A, Farkas A: *Evaluation of multixenobiotic resistance in dreissenid mussels as a screening tool for toxicity in freshwater sediments. Arch Environ Contam Toxicol* 2015, *68(*4):707–717.

160.Bonnafé E, Sroda S, Budzinski H, Valière A, Pedelluc J, Marty P, Geret F: *Responses of cytochrome P450, GST, and MXR in the mollusk Corbicula fluminea to the exposure to hospital wastewater effluents. Environmental Science and Pollution Research* 2015, *22(*14):11033–11046.

161.Deeley RG, Westlake C, Cole SP: *Transmembrane transport of endo-and xenobiotics by mammalian ATP-binding cassette multidrug resistance proteins. Physiological reviews* 2006, *86(*3):849–899.

162.Dean M, Annilo T: *Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu Rev Genomics Hum Genet* 2005, *6:*123–142.

163.Jeong C-B, Kim B-M, Lee J-S, Rhee J-S: *Genome-wide identification of whole ATP-binding cassette (ABC) transporters in the intertidal copepod Tigriopus japonicus. BMC Genomics* 2014, *15(*1):651.

164.Manduzio H, Rocher B, Durand F, Galap C, Leboulenger F: *The point about oxidative stress in molluscs. Invertebrate Survival Journal* 2005(2):91–104%V 102.

165.Sies H: *Oxidative stress: oxidants and antioxidants. Experimental Physiology: Translation and Integration* 1997, *82(*2):291–295.

166.Valavanidis A, Vlahogianni T, Dassenakis M, Scoullos M: *Molecular biomarkers of oxidative stress in aquatic organisms in relation to toxic environmental pollutants. Ecotoxicology and environmental safety* 2006, *64(*2):178–189.

167.Abebe AT, Devoid SJ, Sugumaran M, Etter R, Robinson WE: *Identification and quantification of histidine-rich glycoprotein (HRG) in the blood plasma of six marine bivalves. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 2007, *147(*1):74–81.

168.Devoid SJ, Etter R, Sugumaran M, Wallace GT, Robinson WE: *Histidine-rich glycoprotein from the hemolymph of the marine mussel Mytilus edulis L. binds Class A, Class B, and borderline metals. Environmental Toxicology and Chemistry: An International Journal* 2007, *26(*5):872–877.

169.Feder ME, Hofmann GE: *HEAT-SHOCK PROTEINS, MOLECULAR CHAPERONES, AND THE STRESS RESPONSE: Evolutionary and Ecological Physiology. Annual Review of Physiology* 1999, *61(*1):243–282.

170.Kim HP, Morse D, Choi AM: *Heat-shock proteins: new keys to the development of cytoprotective therapies. Expert opinion on therapeutic targets* 2006, *10(*5):759–769.

171.Place SP, Zippay ML, Hofmann GE: *Constitutive roles for inducible genes: evidence for the alteration in expression of the inducible hsp70 gene in Antarctic notothenioid fishes. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 2004, *287(*2):R429-R436.

172.Acharya P, Kumar R, Tatu U: *Chaperoning a cellular upheaval in malaria: heat shock proteins in Plasmodium falciparum. Molecular and biochemical parasitology* 2007, *153(*2):85–94.

173.Sahi C, Kominek J, Ziegelhoffer T, Yu HY, Baranowski M, Marszalek J, Craig EA: *Sequential duplications of an ancient member of the DnaJ-family expanded the functional chaperone network in the eukaryotic cytosol. Molecular biology and evolution* 2013, *30(*5):985–998.

174.Vos MJ, Hageman J, Carra S, Kampinga HH: *Structural and functional diversities between members of the human HSPB, HSPH, HSPA, and DNAJ chaperone families. Biochemistry* 2008, *47(*27):7001–7011.

175.Cheetham ME, Caplan AJ: *Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. Cell stress & chaperones* 1998, *3(*1):28.

176.Qiu X-B, Shao Y-M, Miao S, Wang L: *The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. Cellular and Molecular Life Sciences CMLS* 2006, *63(*22):2560–2570.

177.Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schütz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P *et al: The nuclear receptor superfamily: the second decade. Cell* 1995, *83(*6):835–839.

178.Sluder AE, Maina CV: *Nuclear receptors in nematodes: themes and variations. TRENDS in Genetics* 2001, *17(*4):206–213.

179.Laudet V, Gronemeyer H: *The nuclear receptor factsbook:* Gulf Professional Publishing; 2002.

180.Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF: *The genome sequence of Drosophila melanogaster. Science* 2000, *287(*5461):2185–2195.

181.Hwang D-S, Lee B-Y, Kim H-S, Lee MC, Kyung D-H, Om A-S, Rhee J-S, Lee J-S: *Genome-wide identification of nuclear receptor (NR) superfamily genes in the copepod Tigriopus japonicus. BMC genomics* 2014, *15(*1):993.

182.Kaur S, Jobling S, Jones CS, Noble LR, Routledge EJ, Lockyer AE: *The nuclear receptors of Biomphalaria glabrata and Lottia gigantea: implications for developing new model organisms. PLoS One* 2015, *10(*4):e0121259-e0121259.

183.Vogeler S, Galloway TS, Lyons BP, Bean TP: *The nuclear receptor gene family in the Pacific oyster, Crassostrea gigas, contains a novel subfamily group. BMC genomics* 2014, *15(*1):369.

184.Robinson-Rechavi M, Carpentier A-S, Duffraisse M, Laudet V: *How many nuclear hormone receptors are there in the human genome? TRENDS in Genetics* 2001, *17(*10):554–556.

185.Cruzeiro C, Lopes-Marques M, Ruivo R, Rodrigues-Oliveira N, Santos MM, Rocha MJ, Rocha E, Castro LFC: *A mollusk VDR/PXR/CAR-like (NR1J) nuclear receptor provides insight into ancient detoxification mechanisms. Aquatic Toxicology* 2016, *174:*61–69.

186.Janošek J, Hilscherová K, Bláha L, Holoubek I: *Environmental xenobiotics and nuclear receptors—Interactions, effects and in vitro assessment. Toxicology in Vitro* 2006, *20(*1):18–37.

187.Kewley RJ, Whitelaw ML, Chapman-Smith A: *The mammalian basic helix–loop–helix/PAS family of transcriptional regulators. The International Journal of Biochemistry & Cell Biology* 2004, *36(*2):189–204.

188.Murphy BJ, Sato BG, Dalton TP, Laderoute KR: *The metal-responsive transcription factor–1 contributes to HIF–1 activation during hypoxic stress. Biochemical and Biophysical Research Communications $V 337* 2005(3):860–867.

189.Igarashi K, Sun J: *The heme-Bach1 pathway in the regulation of oxidative stress response and erythroid differentiation. Antioxidants & redox signaling* 2006, *8(*1–2):107–118.

190.Yan J-k, Wu B, Yang A-g, Zhou L-q, Liu Z-h: *The complete mitochondrial genome of the pen shell Atrina pectinata (Mollusca: Bivalvia: Pinnidae): The first representative from the family Pinnidae. Mitochondrial DNA* 2013, *24(*4):368–369.

191.Lubośny M, Przyłucka A, Śmietanka B, Breton S, Burzyński A: *Actively transcribed and expressed atp8 gene in Mytilus edulis mussels. PeerJ* 2018, *6:*e4897-e4897.

192.Sanna D, Cossu P, Dedola GL, Scarpa F, Maltagliati F, Castelli A, Franzoi P, Lai T, Cristo B, Curini-Galletti M: *Mitochondrial DNA reveals genetic structuring of Pinna nobilis across the Mediterranean Sea. PLoS One* 2013, *8(*6):e67372.

193.Wesselmann M, González-Wangüemert M, Serrão EA, Engelen AH, Renault L, García-March JR, Duarte CM, Hendriks IE: *Genetic and oceanographic tools reveal high population connectivity and diversity in the endangered pen shell Pinna nobilis. Scientific reports* 2018, *8(*1):4770.

194.*http://wolfe.ucd.ie/GenomeVx/.*

195.Adamo SA: *The effects of the stress response on immune function in invertebrates: an evolutionary perspective on an ancient connection. Hormones and behavior* 2012, *62(*3):324–330.

196.GIRON PEREZ MI: *Relationships between innate immunity in bivalve molluscs and environmental pollution. CONACYT* 2010.

197.Malagoli D, Casarini L, Sacchi S, Ottaviani E: *Stress and immune response in the mussel Mytilus galloprovincialis. Fish & Shellfish Immunology* 2007, *23(*1):171–177.

198.Mydlarz LD, Jones LE, Harvell CD: *Innate Immunity, Environmental Drivers, and Disease Ecology of Marine and Freshwater Invertebrates. Annual Review of Ecology, Evolution, and Systematics* 2006, *37(*1):251–288.

Table 1

Due to technical limitations, Table 1 is only available as a download in the supplemental files section.

Figures



Figure 1

A specimen of P. nobilis covered with epibionts in a posidonia oceanica seagrass meadow.

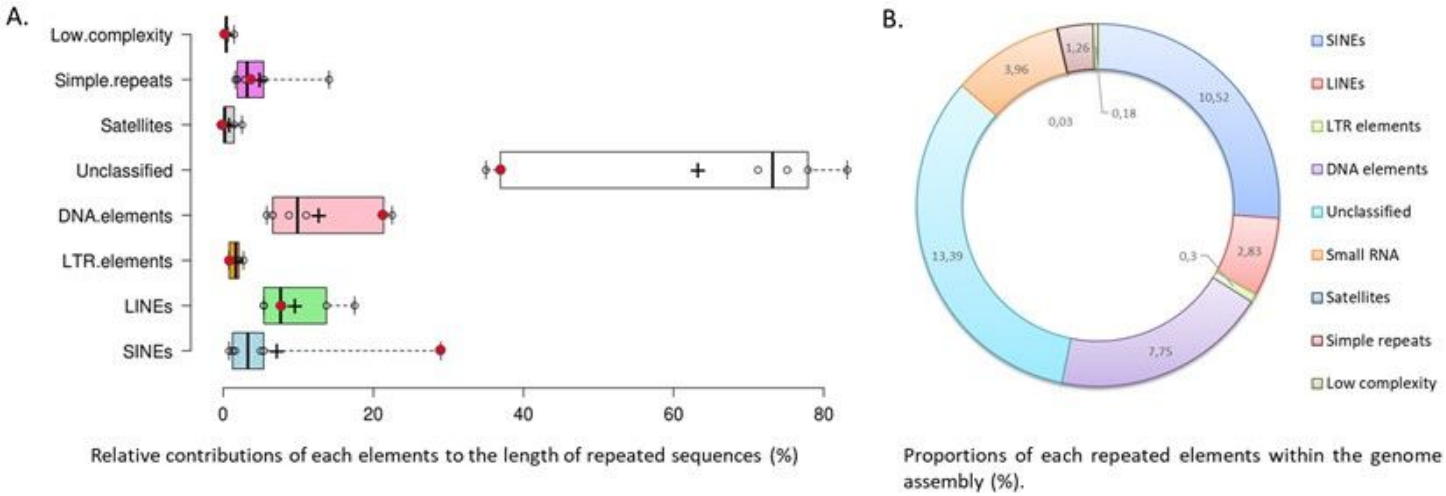


Figure 2

Overview of repetitive elements abundances within bivalve genomes. A. Distribution of the relative contribution of repetitive elements across multiple mollusc species : L. gigantea, A. californica, P. fucuta, C. gigas, M. galloprovinciallis and P. nobilis (highlighted by red dots). Limits of the boxplots represent the first, median and third quartiles, with Spear whisker extents (dashed lines). Means are represented by a plus sign. B. Proportions of each repetitive element length within the genome assembly of P. nobilis.

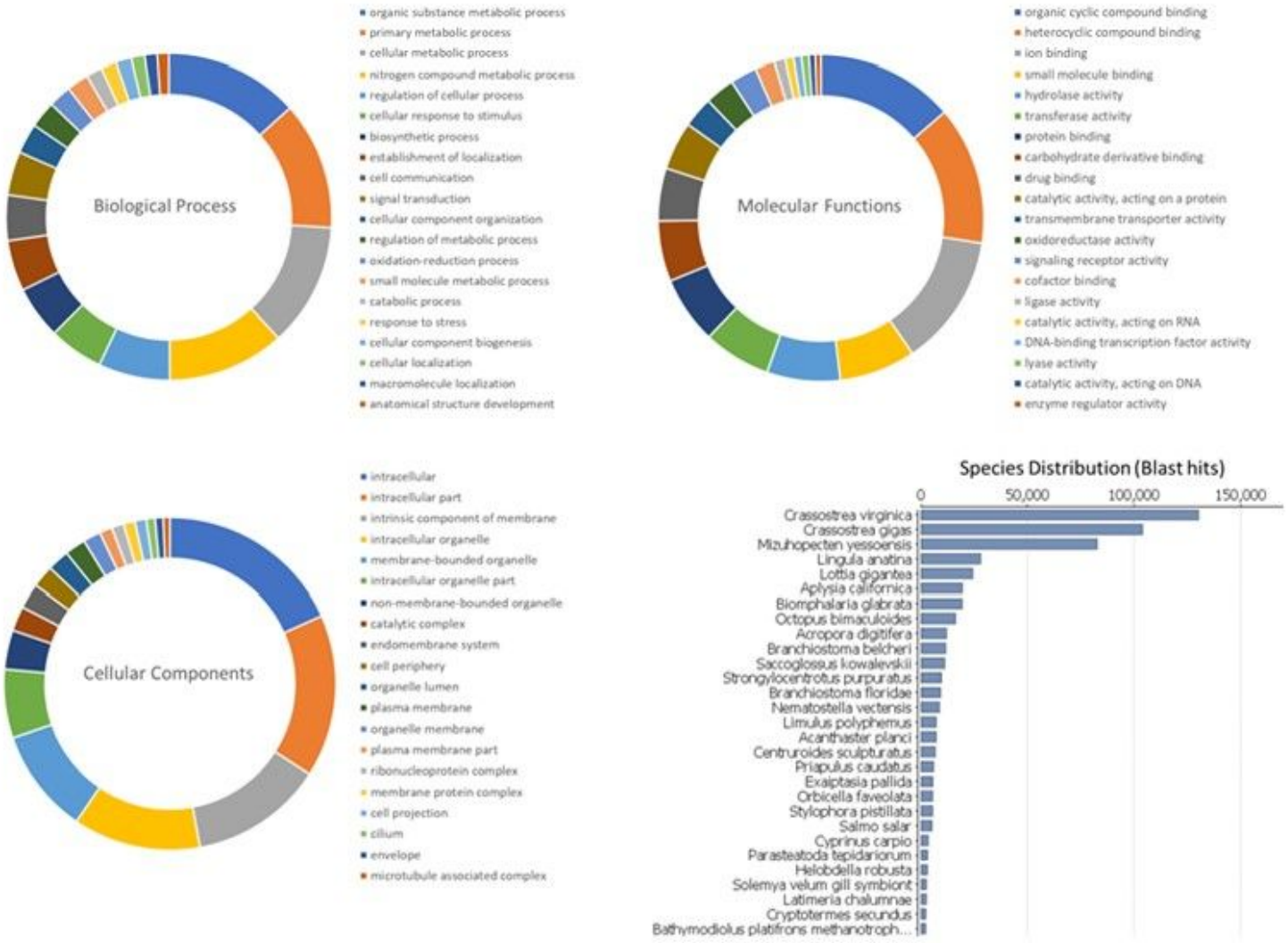


Figure 3

Overview of the proportions of physiological functions according to gene ontology and distribution of Blast Hits (BlastX) according to species, at level 3.

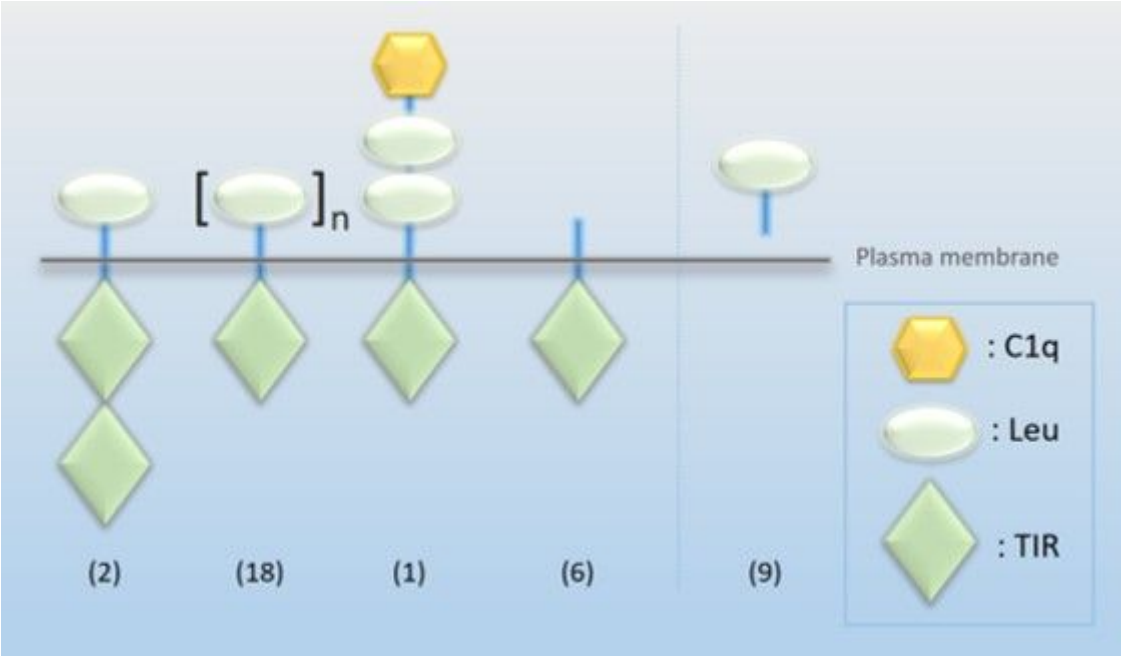


Figure 4

Schematic representation of encoded TLRs found within the P. nobilis genome assembly. The numbers in brackets represent the number of contigs encoding a TLR with the above represented structural feature. The “n” stands for the number of Leucine motifs within the extracellular part of the receptor and varies from 1 to 10 repeats. Nine contigs were annotated as TLR (Blast2GO) but only contained Leu domain without transmembrane region and no TIR motif, hence, they were not included in the TLR repertoire.

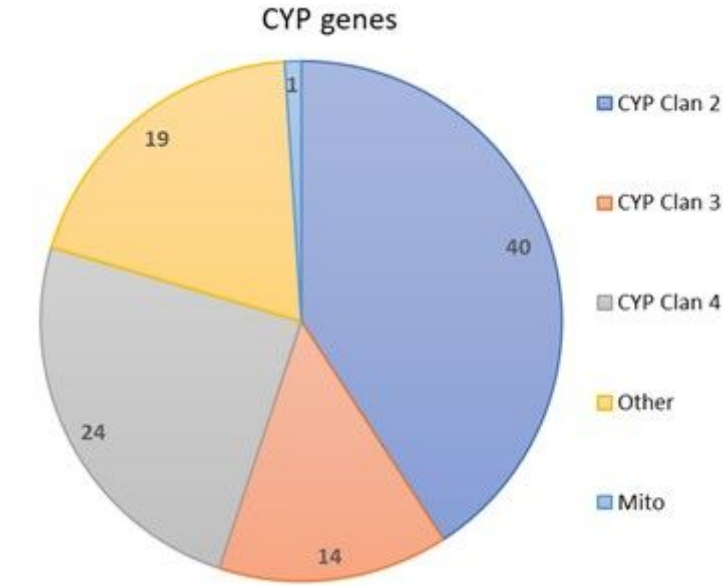


Figure 5

Relative proportions of CYP clan families.

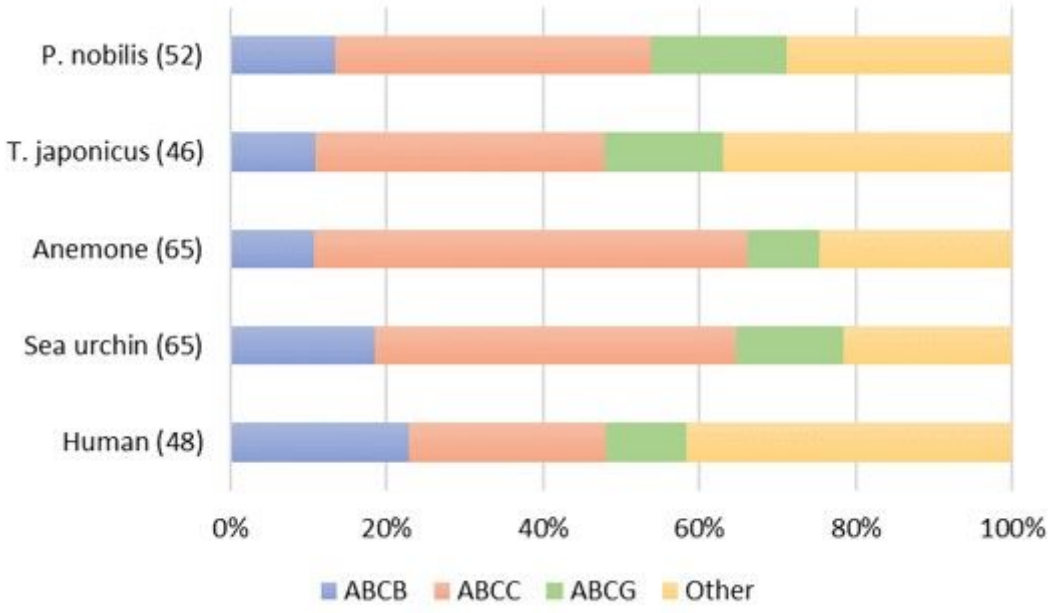


Figure 6

Relative proportion of xenobiotic transporter genes in various organisms. The number in brackets represents the total number of ABC genes detected.

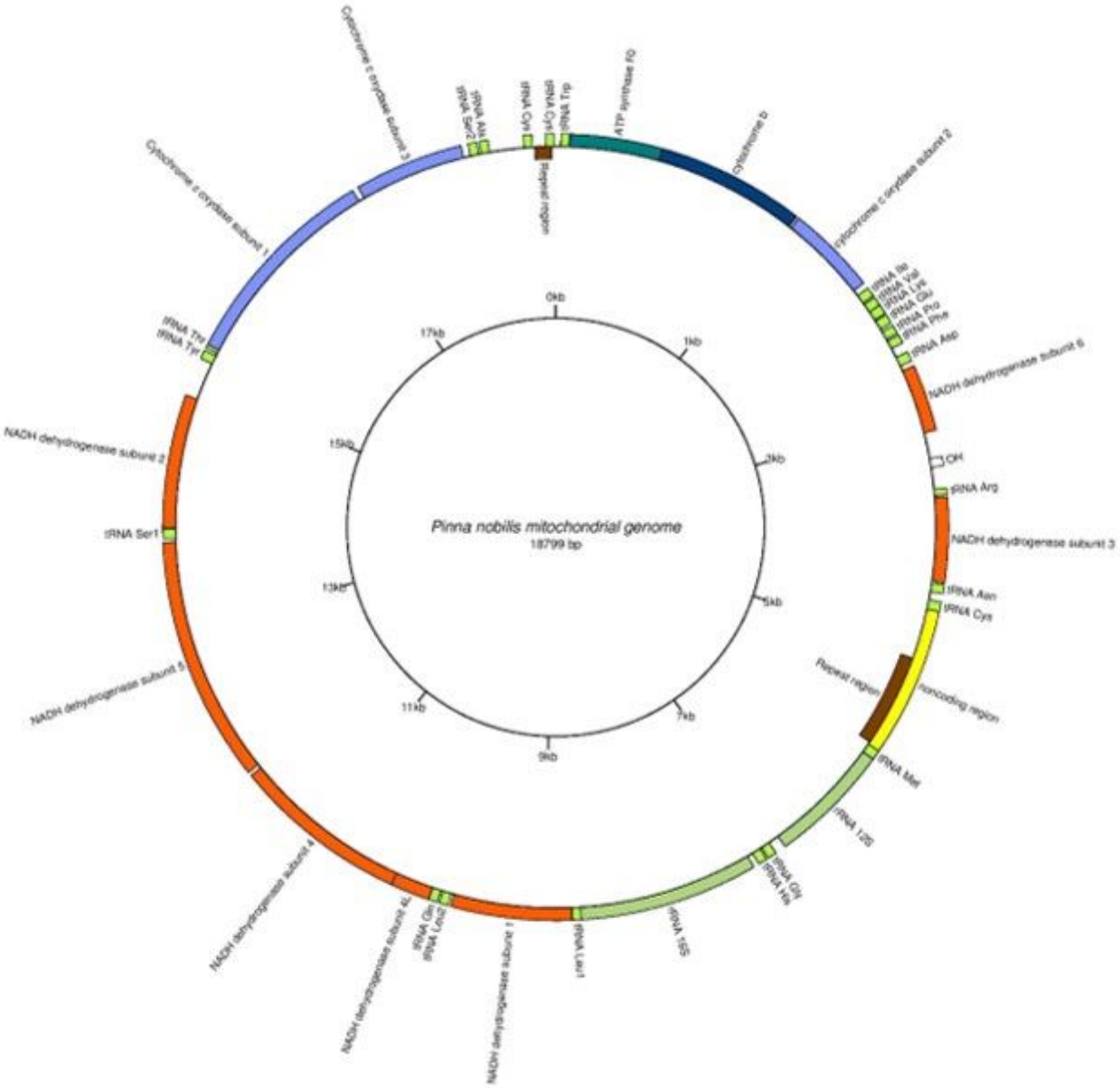


Figure 7

Structure of the P. nobilis mitochondrial genome. The structure has been drawn using GenomeVx [194] and arbitrarily circularized. Genes are all in the same orientation (clockwise).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supportinginformation.docx
- Table1.jpg