

RESEARCH ARTICLE

A First Insight into the Genome of the Filter-Feeder Mussel *Mytilus galloprovincialis*

Maria Murgarella¹, Daniela Puiu², Beatriz Novoa³, Antonio Figueras³, David Posada¹, Carlos Canchaya^{1*}

1 Department of Biochemistry, Genetics and Immunology and Unidad Asociada CSIC, University of Vigo, Vigo, Spain, **2** Center for Computational Biology. McKusick-Nathans, Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, **3** Instituto de Investigaciones Marinas, Consejo Superior de Investigaciones Científicas, Vigo, Spain

* canchaya@uvigo.es



OPEN ACCESS

Citation: Murgarella M, Puiu D, Novoa B, Figueras A, Posada D, Canchaya C (2016) A First Insight into the Genome of the Filter-Feeder Mussel *Mytilus galloprovincialis*. PLoS ONE 11(3): e0151561. doi:10.1371/journal.pone.0151561

Editor: John A Craft, Glasgow Caledonian University, UNITED KINGDOM

Received: April 7, 2015

Accepted: March 1, 2016

Published: March 15, 2016

Copyright: © 2016 Murgarella et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Illumina read sequences used in this study can be downloaded from the NCBI Sequence Read Archive under the accessions SRR1598987, SRR1598945 and SRR1598943. Assembled sequences and annotations have been submitted to NCBI under the submission code SUB1006464 (PRJNA262617) following NCBI WGS requirements.

Funding: This work was supported by the Spanish Agencia Galega de Innovación para o desenvolvemento (<http://gain.xunta.es>) Project INCITE 10PXIB402096PR awarded to AF y CC. CC was funded by "Isidro Parga Pondal" Spanish

Abstract

Mussels belong to the phylum Mollusca, one of the largest and most diverse taxa in the animal kingdom. Despite their importance in aquaculture and in biology in general, genomic resources from mussels are still scarce. To broaden and increase the genomic knowledge in this family, we carried out a whole-genome sequencing study of the cosmopolitan Mediterranean mussel (*Mytilus galloprovincialis*). We sequenced its genome (32X depth of coverage) on the Illumina platform using three pair-end libraries with different insert sizes. The large number of contigs obtained pointed out a highly complex genome of 1.6 Gb where repeated elements seem to be widespread (~30% of the genome), a feature that is also shared with other marine molluscs. Notwithstanding the limitations of our genome sequencing, we were able to reconstruct two mitochondrial genomes and predict 10,891 putative genes. A comparative analysis with other molluscs revealed a gene enrichment of gene ontology categories related to multixenobiotic resistance, glutamate biosynthetic process, and the maintenance of ciliary structures.

Introduction

Mussels belong to the phylum Mollusca, one of the largest and most diverse taxa in the animal kingdom, only second to Insecta. The number of molluscan species has been estimated to be 93,000 [1], with 25% of them being marine. Among all classes belonging to this phylum, bivalves—where mussels belong—show the most highly modified body plan, flattened side-to-side, over evolutionary time. Morphologically, they are characterized by the presence of a bivalve shell, filtering gills, no differentiated head, and a lack of radula. Other anatomical features such as adult byssal attachment and mantle fusion may have also played an important role in their adaptation as filter feeders and burrowers, respectively [2]. Some bivalves show an atypical double uniparental inheritance (DUI) of mitochondria. In these species, all progeny inherit one mitochondrial genome from the mother (F-type), while males also receive a mitochondrial genome from their father (M-type). This DUI, initially described in *M. edulis* [3], has

research fellowships from Xunta de Galicia (<http://www.xunta.es>).

Competing Interests: The authors have declared that no competing interests exist.

been extensively studied in the genus *Mytilus* [4,5]. Another remarkable characteristic of mussels is their natural resistance to diseases. Unlike vertebrates, that have also developed an adaptive immune system, the immune system of bivalves is solely based on innate defences, which play a prominent role in protecting these animals against invading microorganisms. However, differences in disease resistance have been observed among bivalves. Compared to other edible bivalves, like oysters and clams, mussels seem far less susceptible to mass mortalities and diseases [6,7]. Interestingly, previous mass mortalities in mussels have not been linked to any micro-organisms [8], suggesting that the innate immune system in mussels is highly efficient.

Marine mussels have also a significant commercial value. Their production corresponds to 50% of global EU aquaculture in weight and about 30% in value [9]. In this regard, the most important mussel species are the Mediterranean (*Mytilus galloprovincialis*, Lamarck 1819) and the blue (*Mytilus edulis*) mussels. In Spain, the 2012 farmed production of *M. galloprovincialis* was 231,754 TM [10], of which 227,229 TM were produced in Galicia, NW Spain [11]. Indeed, the relevance of its farming in Galicia is not only economical but also social: mussel culture employs about 15,000 people in 2,400 familiar enterprises [9].

Despite the commercial and scientific interest in mussels in biology and aquaculture, the number of genomic resources available in public databases for these organisms is quite limited, and usually restricted to their transcriptomes. For instance, in *M. galloprovincialis*, some transcriptome studies using ESTs [12–14] and high throughput cDNA [15,16] are available. Clearly, molecular tools still need to be developed for the management, molecular breeding and genetic manipulation of *Mytilus* spp. in aquaculture [17].

We present here the first genome sequencing study in this genus, namely a low-coverage whole-genome study of the Mediterranean mussel *M. galloprovincialis*. Indeed, low-coverage sequencing of non-model organisms can provide valuable information about their genomes [18,19] regarding important features such as gene content, functional elements and repetitive sequences [20–22]. Here, we offer a first insight into the general features and complexity of the genome of *M. galloprovincialis*, providing a starting point for future genomic research on this important bivalve. Moreover, the availability of this genome sequence together with those of four other molluscs already sequenced (the California sea hare *Aplysia californica*, the owl limpet *Lottia gigantea* [23], the pacific oyster *Crassostrea gigas* [24] and the pearl oyster *Pinctada fucata* [25]) should improve the knowledge of this important phylum through genomic comparisons at multiple levels.

Materials and Methods

Sequencing, k-mer analyses, and assembly

For sequencing, we extracted 4µg of DNA from muscle tissue from a single mussel extracted from the Ria of Vigo, Spain. Using this DNA, three sequencing libraries with insert sizes of 180, 500 and 800 bp were constructed and sequenced at BGI (Beijing Genomics Institute—China). These libraries were sequenced with the Illumina HiSeq2000 high-throughput platform using paired-end sequencing (100-bp reads). To clean the initial set of reads, we filtered out raw reads if they fulfilled any of these conditions: a) >5% ambiguous bases (represented by the letter N); b) poly-A structures; c) >= 20 bases with low quality scores; d) adapter contamination: reads with more than 10 bp aligned to the adapter sequence (no more than 3-bp mismatch allowed); or e) small insert-size reads in which paired reads overlapped more than or equal to 10 bp (10% mismatch allowed).

We used Jellyfish [26] for counting k-mers and obtaining their frequency distributions. With these data, we drew frequency plots using k-mer lengths of 15, 17, 19 and 21. To assign the “true” coverage peak, we compared these plots to identify the peak that changed in height

(“heterozygous peak”) and the one that did not (“coverage peak”). The latter was then used to calculate the genome size as the total k-mer number divided by the coverage-peak depth [27]. Finally, we assembled *de novo* the reads resulting from the quality filtering step using SOAPdenovo v1.05 [27] with parameters `-K 31 -d 1 -M 1 -F-R`. Then, we ran the Assemblathon 2 script [28] to obtain assembly statistics. Using this script, we compared the genome assemblies of *M. galloprovincialis* with those of *A. californica*, *L. gigantea*, *P. fucata*, and *C. gigas* (S1 File). Genome surveys of other molluscs with scarce sequencing depth [22] were not included in these comparisons. We confirmed the identification of the studied mussel as *M. galloprovincialis* by scanning the assembled sequences with two *Mytilus* genetic markers, *Glu-5'* [29] and *EFbis* [30], using BLASTN [31] and Geneious version 6.1.8 [32].

Isolation of mitochondrial sequences and variant calling

For the mitochondrial genome analysis, first we detected contigs in our assembly that matched *M. galloprovincialis* mitochondrial DNA sequences. For this, we aligned our contigs against F (GenBank NC_006886, MgF) and M (GenBank AY363687, MgM) *M. galloprovincialis* mitochondrial sequences using BLASTN and *nucmer* (from MUMMER version 3.23 [33]). We filtered out BLASTN alignments with e-value above 1×10^{-6} , identity < 90%, or contig alignment coverage < 90% to remove non-specific alignments. The program *nucmer* was used with arguments `-maxmatch` and `-c 100`. We mapped the contigs that aligned (using *nucmer*) with MgF and constructed its corresponding tiling path (S1 Fig). For both BLAST and *nucmer*, we also calculated the proportion of MgF and MgM nucleotides that aligned with the assembly. Second, for variant calling, we mapped the cleaned reads against several mitochondrial genomes of *Mytilus* spp. (S1 File) using Bowtie2 v2.0.6 [34] with the option “very-sensitive”, identifying single-nucleotide variants (SNVs) with SAMtools version 0.1.18 [35].

Repeat Sequence Analysis

To estimate the amount and composition of repetitive elements in our assembled genome, we carried out three different analyses using RepeatMasker (<http://www.repeatmasker.org>). First, we ran RepeatMasker against our working assembly using the default repeat sequence entries of the genus *Mytilus*. For this, we used the Repbase database [36] version 20120418 (we called this search “MYTILUS”). To overcome the limitation of MYTILUS due to the low number of *M. galloprovincialis* entries present in Repbase, a second analysis was performed using the Repbase entries from the phylum Mollusca (search “MOLLUSCA”). Finally, a third RepeatMasker analysis (search “INHOUSE”) was carried out using our own in-house library of repetitive elements. This library was built from putative repetitive sequences obtained using RepeatModeler (<http://www.repeatmasker.org/RepeatModeler.html>). We repeated similar analyses with the other four molluscan genomes studied (S1 File).

Gene Prediction and Annotation

Before starting the gene prediction and annotation steps, we searched for potential contaminant sequences in our assembly. We looked for sequences with bacterial signatures using Kraken 0.10.5-beta [37] and BLASTX. First, we selected contigs with bacterial k-mer matches using Kraken. Second, we calculated the accumulated length of the matched regions in each of these contigs. Third, we putatively annotated as “bacterial” those contigs whose aligned regions added up more than 10% of the total contig length. Fourth, we confirmed the putative bacterial origin by matching the resulting contigs using BLASTX against the non-redundant (nr) database with a cut-off e-value of 1.0×10^{-6} . Finally, we annotated as bacterial those resulting contigs

that best matched proteins of bacterial origin. We did not include these putative bacterial sequences in the subsequent analyses.

Next, to estimate the quality of the predicted mussel gene repertoire, we used the program CEGMA [38] with default options. This tool quantifies the completeness of a gene repertoire within a genome assembly by aligning its sequences to a "universal" set of 248 eukaryotic core proteins. For CEGMA, significant sequence alignments spanning more than 90% of the length of any CEGMA protein are classified as being "complete" in the analysed set of sequences. To estimate the gene repertoire size in *M. galloprovincialis*, we assigned the percentage of CEGMA sequences found in our assembly as the percentage of the *M. galloprovincialis* gene repertoire in our assembly [38].

We used MAKER2 [39] to predict *in silico* the gene models in our assembly. Additional *M. galloprovincialis* experimental sequence data coming from two other different sources were used to improve the gene prediction with MAKER2. These data consisted of our in-house RNA-Seq *de novo* assemblies from 4 different tissues, and protein sequences from the NCBI database. We used these 2 different datasets to build a first *ab initio* gene prediction model using the program SNAP [40]. Finally, MAKER2 was used with the resulting previous SNAP results to accomplish a refined, more reliable gene prediction. To characterize and validate the resulting protein sequences, we aligned the contigs with mpiBLAST [41] against the *nr* database (e-value cut-off of 1×10^{-6}). Finally, we used Blast2GO [42] with the BLAST *nr* database and InterProScan 5RC7 [43] to obtain a more complete functional protein annotation and description of our genome survey. Using this bioinformatic tool, we also compared the functional protein annotation of *M. galloprovincialis* with those of the other four molluscs (S1 File). The comparison was done using the Fisher's exact test, with a False Discovery Rate (FDR) of 0.05. We included only those ontologies that belonged to the "biological function" category and presented significant differences with other molluscan datasets.

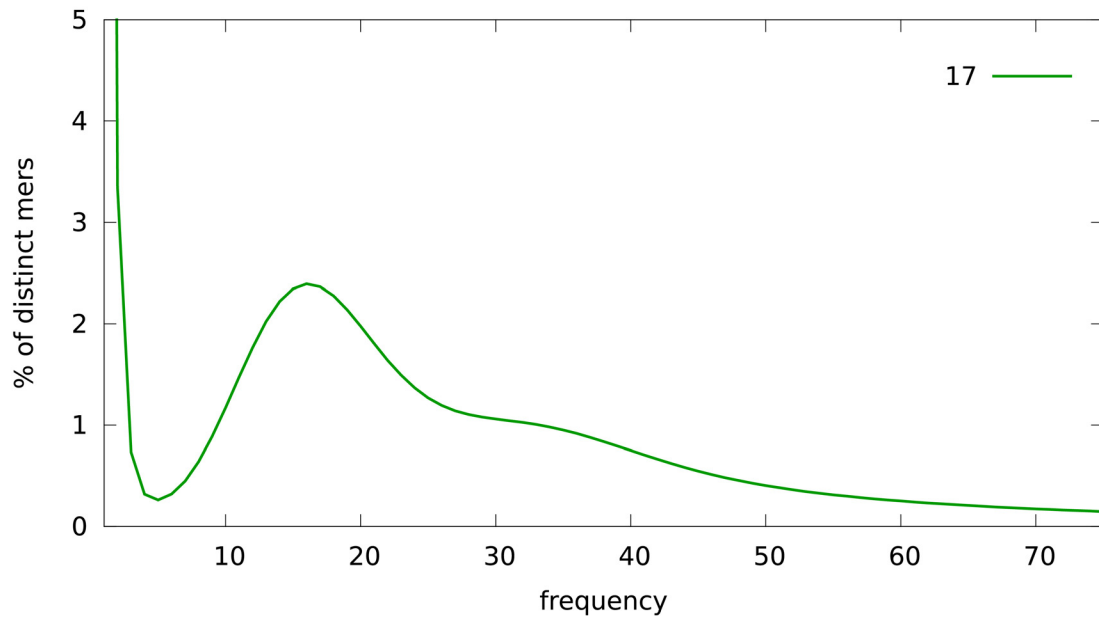
Results and Discussion

Genome sequence composition and size

The obtained k-mer frequency plots for *M. galloprovincialis* were bimodal, with two clear peaks at 16X and 32X (Fig 1A; 17-mers). A pattern like this has been previously reported for other genomes [44,45], and in simulations [46], and it is thought to be a direct consequence of heterozygosity in diploid genomes. To identify which of these two peaks was the coverage peak, we generated several k-mer frequency distributions with different k-mer sizes (Fig 1B). We observed that the peak height at 16X changed considerably for different k-mer sizes, while the peak height at 32X remained more or less stable. Therefore, we concluded that the latter peak was the coverage peak, containing k-mers from homozygous regions, while the former was the heterozygous peak, containing k-mers from heterozygous regions. The observed sharp difference in height between the homozygous and heterozygous peaks (Fig 1B; e.g. 21-mer plot) may be produced by a seemingly high heterozygosity within the *M. galloprovincialis* genome. This is not surprising, as species within the genus *Mytilus* have been previously reported to have high genetic diversity [47–51]. In a genomic survey [52] of 76 non-model organisms using RNA-Seq data, *M. trossulus* occupied the second highest value of genetic diversity while *M. galloprovincialis* occupied the 15th.

Conveniently, genome size can also be estimated from the k-mer count data [46]. Using this method, we estimated the genome size of *M. galloprovincialis* to be 1.6 Gb. Interestingly, discrepancies between genome sizes estimated from sequencing and experimental data have been previously reported [53]. Using flow cytometry, *M. galloprovincialis* ($2n = 28$ [54]) was proposed to have a genome size of either 1.4 Gb [55] or 1.9 Gb [56], while our 1.6 Gb estimate fits

A)



B)

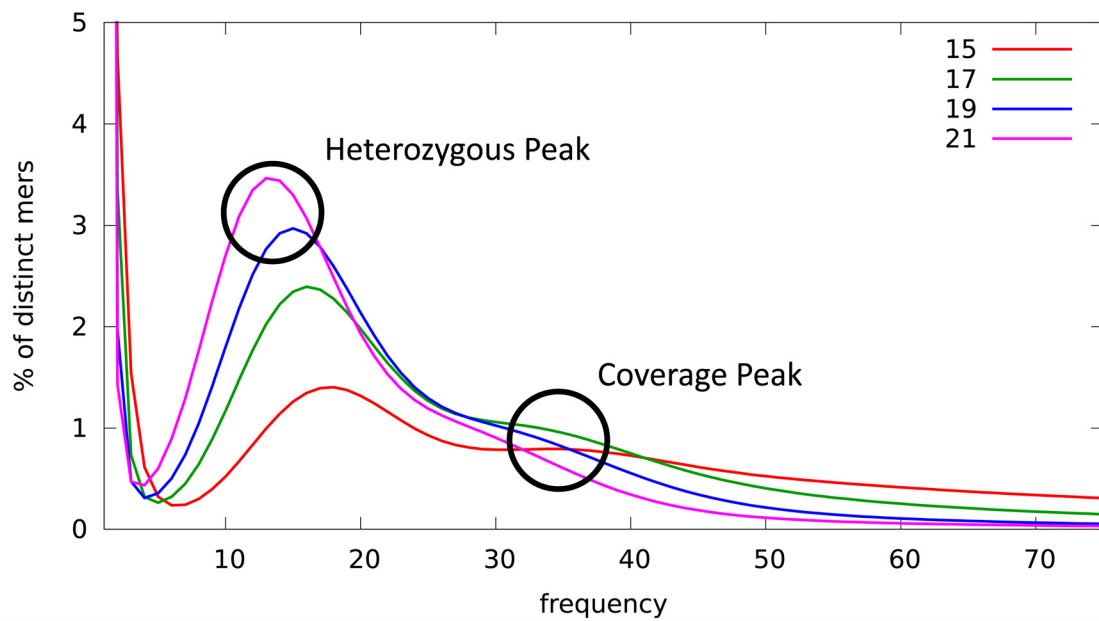


Fig 1. A) Histogram of 17 k-mers. **B)** K-mer frequency distributions ad different k-mer lengths.

doi:10.1371/journal.pone.0151561.g001

nicely in the middle. When compared with other bivalves, a 1.6 Gb genome size for *M. galloprovincialis* is located approximately in the middle tier not only among bivalves (Fig 2A) but also among other Mytiloidea (Fig 2B). Moreover, the genome size range of Mytiloidea is one of the highest within the superclass Pteriomorpha when sorted by their median. These relative positions of the *M. galloprovincialis*' genome size highlights the representativeness of this genome size within the taxa included in our comparisons and the possibility to use this genome as a "model" for Mytiloidea and Bivalvia taxa for some genomic features such as repeat elements and gene content.

Assembly fragmentation in the *M. galloprovincialis* genome

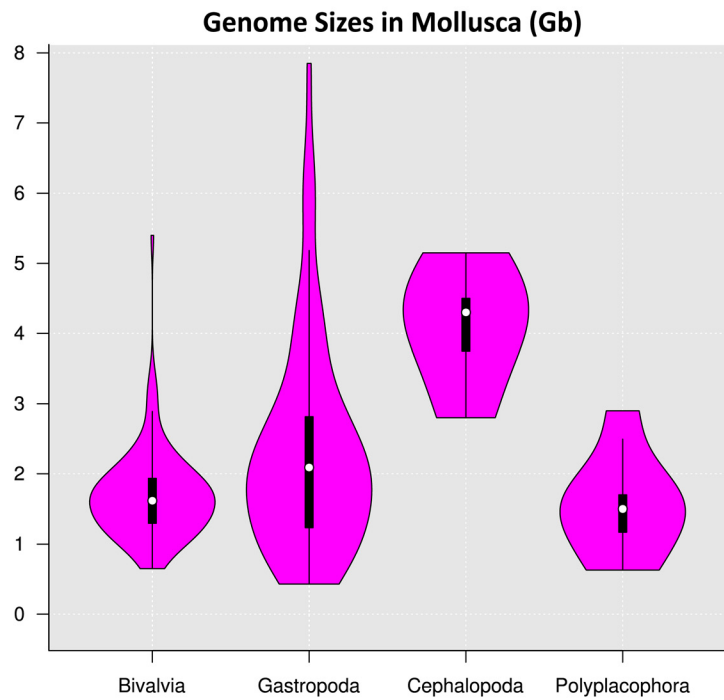
In a genome sequencing survey, most of the biological information is obtained by analysing the final assembly. Base calls and quality values from reads of sequencing libraries with different insert sizes provide the information needed to assemble small reads into larger structures such as contigs and scaffolds. For *M. galloprovincialis*, we used the *de novo* assembler SOAPdenovo for genome assembly. This assembler has been previously used for *de novo* assembling dozens of plant and animal genomes, including panda [27], duck [57], potato [45] and cucumber [58], among others. In Table 1, we show assembly statistics of *M. galloprovincialis* and the other four marine molluscs. These statistics revealed a highly fragmented assembly. For instance, the assembly contained a large number of sequences (1,746,447) and low N50 (2,651 bp) when compared to *A. californica*, *P. fucata*, *C. gigas* and *L. gigantea* assemblies.

Two reasons could explain the differences among assembly statistics for these genomes. The first one is genome size. The difficulty of assembling a genome increases with its size. Genomes of large sizes contain correspondingly a high number of internal repetitions such as paralogues, duplications, structural rearrangements, and mobile elements [59]. Reads from these DNA regions can match more than one genome position, decreasing the contiguity of assemblies. The genome size of *M. galloprovincialis* is only comparable with that of *A. californica*, while those of *P. fucata*, *C. gigas* and *L. gigantea* are 33, 66 and 75% smaller, respectively. The second reason is the sequencing technology used. For instance, the assemblies of *L. gigantea* and *A. californica*, despite their low coverage (8X and 11X, respectively), showed much better assembly statistics than that of *M. galloprovincialis*. Larger reads obtained by Sanger sequencing technology for these two genomes surely contributed to the lower number of scaffolds and larger N50 obtained. For *P. fucata*, with a final 35X coverage, both Illumina and 454 sequencing were used. However, despite using the same short-read sequencing technology as in *M. galloprovincialis*, the assembly statistics for *C. gigas* were superior. The assembly of this organism, using reads from Illumina mate-pair and pair-end libraries, was improved with the addition of fosmid libraries in the scaffolding step. In our case, using only pair-end sequencing libraries did prevent contigs from assembling into larger scaffold sequences.

Ab initio prediction of repetitive sequences identifies a large diversity of repetitive elements

Repetitive elements (REs) are an important part of most eukaryotic genomes [60]. From humans to plants [61], a high proportion of these genomes consists of REs (i.e. interspersed repeats and low complexity DNA sequences). Although originally considered as "junk" DNA, they can play an important role in the adaptation [62] and evolution [63,64] of eukaryotes. To measure the extent of REs in the genome of *M. galloprovincialis*, we used RepeatMasker (Table 2) with Repbase and in-house repeat libraries. Using only Repbase, results of both "MYTILUS" and "MOLLUSCA" analyses found low repetitive content (1.4% and 1.57% respectively). However, identification of REs in a genome is not a trivial task. Difficulties arise

A)



B)

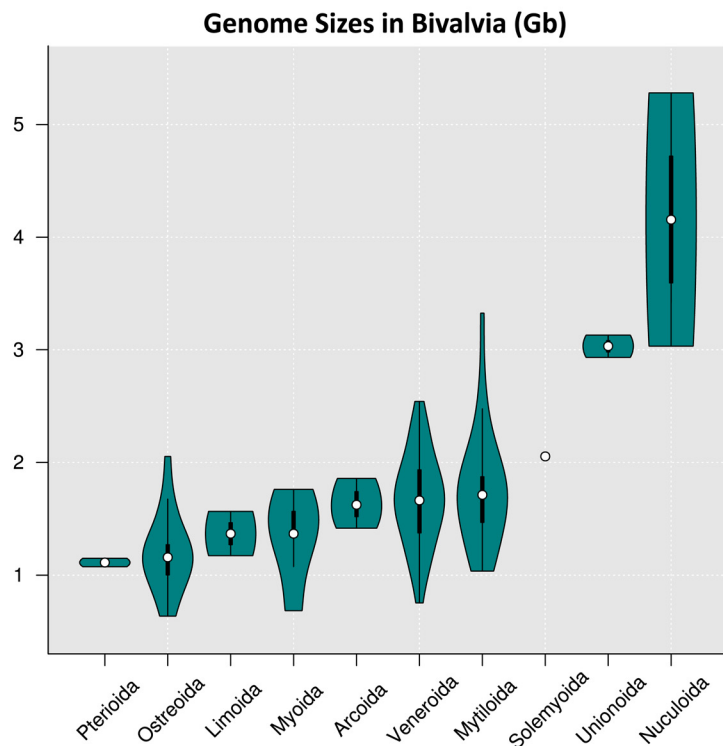


Fig 2. Genome size distribution of Molluscs (A) and Bivalves (B). Violin plots were built using DNA mass content data from Genome Size Animal Database (<http://www.genomesize.com/>) converted into number of

base pairs. Kernel density for Solemyoidea is absent since there is only one record in this group. Data for Pterioidea comes from the genome size estimation of *P. fucata* and *Atrina rigida* (belonging to Mytiloidea on the original dataset)

doi:10.1371/journal.pone.0151561.g002

when they do not share any similarity at sequence level with any other repeat sequences in curated databases as Repbase. To overcome this issue, *ab initio* methods (e.g. RepeatModeler) build libraries of new repeats [65] from scratch. The resulting libraries are then used to identify more precisely the repeat content of a genome. For example, using an *ab initio* prediction method, 36% of the *C. gigas* genome contained REs [24]. To test whether this high percentage of REs is a unique feature of *C. gigas*, we used RepeatModeler to build libraries of repetitive sequences in *M. galloprovincialis*. The INHOUSE search found 36.13% of REs in our assembly, corresponding to 1276 different families. Noteworthy, 30.16% of the genome corresponded to "Unclassified" *de novo* REs belonging to 1059 clusters. On the other hand, from the "classified" part, the most representative fraction (2.27% of the genome) was made of repetitive DNA elements.

Similar repeat content screenings were done in *C. gigas*, *L. gigantea*, *P. fucata* and *A. californica*. In these analyses, the organisms with more REs corresponded to *A. californica* with 43.70%, and the lowest to *L. gigantea* with 22.47%. Coincidentally, these organisms had the largest and smallest genome sizes, respectively. These results went in accordance with the long known positive correlation between genome size and REs content in eukaryotic genomes [66].

Table 1. Assembly statistics of *M. galloprovincialis* and four other molluscan genome sequencing projects. Data shown was obtained with the Assemblathon 2 metrics script.

	<i>M. galloprovincialis</i>	<i>C. gigas</i>	<i>P. fucata</i>	<i>L. gigantea</i>	<i>A. californica</i>
Estimated Genome Size	1,600 Mb	545 Mb	1150 Mb	359,5 Mb	1,800 Mb
Number of scaffolds	1,746,447	11,969	800,982	4,475	8,766
Total size of scaffolds	1,599,211,957	558,601,156	1,413,178,538	359,512,207	715,791,924
Total scaffold length as percentage of known genome size	100.0%	102.5%	122.9%	100.0%	39.8%
Longest scaffold	67,529	1,964,558	698,791	9,386,848	1,784,514
Shortest scaffold	100	100	100	1000	5,001
Number of scaffolds > 500 nt	676,492 (38.7%)	6,484 (54.2%)	323,197(40.4%)	4,475 (100.0%)	8,766 (100.0%)
Number of scaffolds > 1K nt	393,685 (22.5%)	5,788 (48.4%)	142,882 (17.8%)	4,471 (99.9%)	8,766 (100.0%)
Number of scaffolds > 10K nt	12,859 (0.7%)	3,172 (26.5%)	27,367 (3.4%)	1,318 (29.5%)	5,269 (60.1%)
Number of scaffolds > 100K nt	0 (0.0%)	1,353 (11.3%)	629 (0.1%)	291 (6.5%)	2,079 (23.7%)
Number of scaffolds > 1M nt	0 (0.0%)	60 (0.5%)	0 (0.0%)	98 (2.2%)	27 (0.3%)
Mean scaffold size	916	46,671	1,764	80,338	81,655
Median scaffold size	258	824	402	3,622	13,763
N50 scaffold length	2,651	401,319	14,455	1,870,055	264,327
Percentage of assembly in scaffolded contigs	18.5%	95.7%	75.4%	99.0%	93.9%
Percentage of assembly in unscaffolded contigs	81.5%	4.3%	24.6%	1.0%	6.1%
Average number of contigs per scaffold	1.1	2.8	1.3	4.1	7.4
Sequencing Coverage	32X	155X	40X	8,87X	11X
Sequencing Technology	Illumina	Illumina	454 + Illumina	Sanger	Sanger

doi:10.1371/journal.pone.0151561.t001

Table 2. Percentage of bases masked in the assembly using different RepeatMasker libraries.

	% Bases masked			Families found (using RepeatModeler Library)
	RepeatMasker Own Species Library	RepeatMasker Mollusca Library	RepeatModeler Library	
<i>M. galloprovincialis</i>	1.40%	1.57%	36.13%	1276
<i>C. gigas</i>	2.71%	2.81%	31.90%	870
<i>P. fucata</i>	1.19%	1.37%	37.46%	1524
<i>L. gigantea</i>	1.37%	1.42%	22.47%	621
<i>A. californica</i>	11.45%	11.53%	43.70%	938

doi:10.1371/journal.pone.0151561.t002

In *P. fucata*, this *ab initio* prediction of REs found almost four times (37.46%) as much as the percentage predicted using only the Repbase database [25]. Comparatively, the *M. galloprovincialis* genome contained proportionally the largest content (>80%) of unknown REs in Molluscs (Fig 3). Moreover, the genomes of *P. fucata* and *M. galloprovincialis* harboured the largest number of unknown families of REs among the molluscs studied: 1325 and 1059 respectively. One possible explanation for these high numbers would be that these unknown families are artefacts resulting from low coverage sequencing. However, a second possibility is that the unknown families come from multiple novel species-specific REs. Long-read sequencing over repeat-containing genomic regions will help to distinguish between these two alternative explanations.

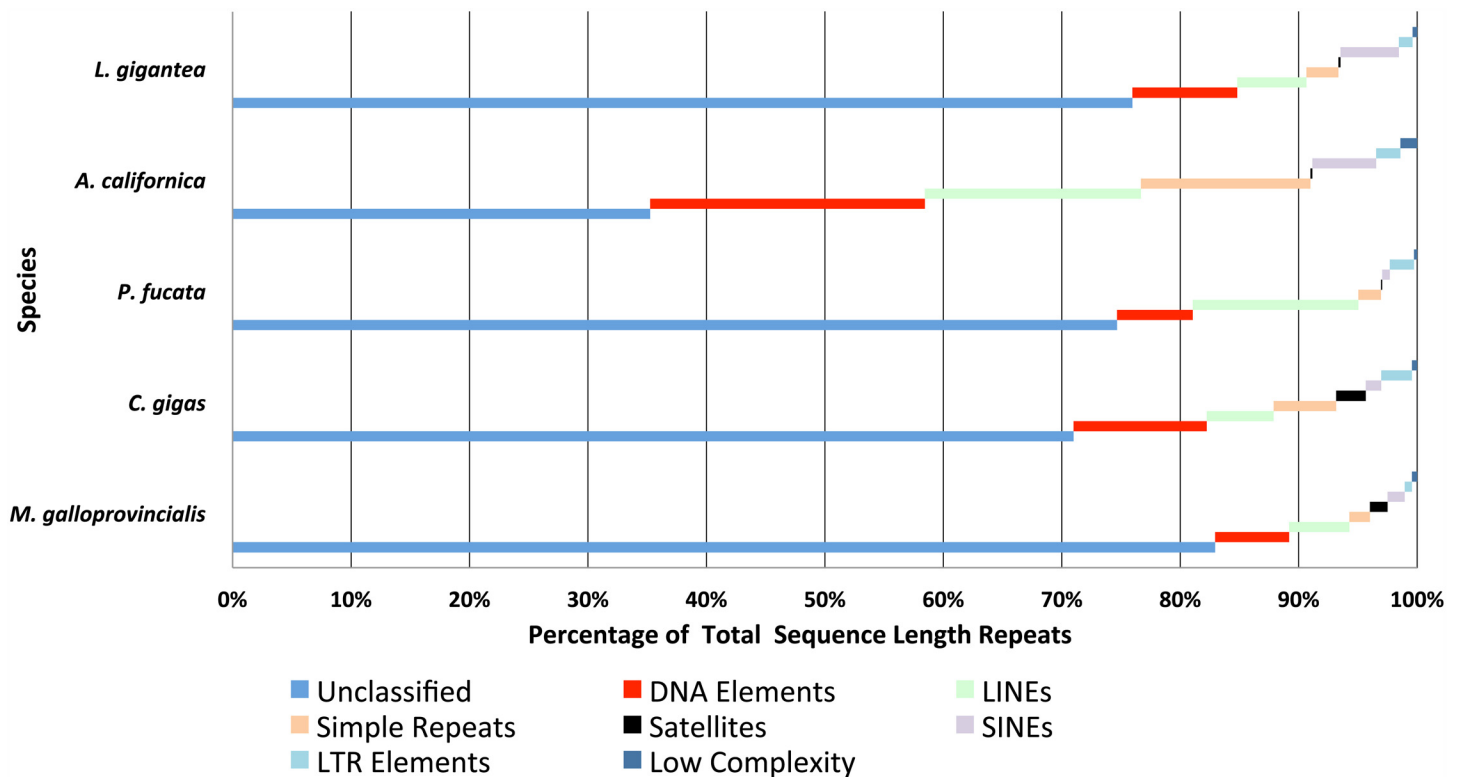


Fig 3. Percentage of each repetitive element relative to the total sequence length occupied by them.

doi:10.1371/journal.pone.0151561.g003

A large group of REs in a genome are transposable elements (TEs). In our analyses, the molluscan genome with the largest content of DNA TEs and retrotransposons corresponded to *A. californica*, encompassing 10.15% and 11.23% of its genome, respectively. In *M. galloprovincialis*, the large proportion of unknown repetitive sequences made it difficult to extract conclusions about the diversity of TEs. On the other hand, the most abundant TEs belonged to the *Penelope* family [67] with 86,269 copies. The *Penelope* retrotransposon induces a hybrid dysgenesis syndrome in *Drosophila virilis*, maybe as a first step for reproductive isolation among populations [68]. In addition, we found only one transposon group significantly more abundant in *M. galloprovincialis* than in the other molluscs studied, the DNA transposon Tip100 family [69]. This transposon family belongs to the hobo-Ac-Tam3 (hAT) superfamily. Interestingly, hobo transposons are also involved in hybrid dysgenesis in *D. melanogaster* [70]. Elliot and Gregory [71] proposed that larger genomes evolve primarily through the expansion of only a small subset of existing TEs. The Tip100 family can be part of this subset of TEs in our genome. Further studies are needed to test whether members of *Penelope* or Tip100 families work similarly in *M. galloprovincialis*.

Mitochondrial genomes

Mitochondria are essential components of the cell where they produce energy through oxidative phosphorylation. In addition, they can also mediate phenotypes such as lifespan, fertility, starvation resistance, altitude adaptation, and temperature regulation [72]. Conveniently, genome sequencing data in eukaryotes contain also mitochondrial genome sequences. This is because methods for DNA/RNA isolation capture also organelle nucleic acids. For instance, a bioinformatic pipeline has been recently proposed to extract mitochondrial reads from genome sequencing data and assemble them in organisms without a reference mitochondrial genome [73]. Two types of mitochondrial genomes, female (MgF) and male (MgM), have been reported for *M. galloprovincialis* [74]. BLASTN and *nucmer* alignments of the assembly against MgF produced significant matches with 56 (67.6% of coverage) and 51 (68.5% of coverage, S1 Fig) contigs, respectively. In addition BLASTN and *nucmer* alignments of the assembly against MgM produced a low number of significant matches with only 9 (22.2% of coverage) and 9 (20.5% of coverage) contigs, respectively. The incomplete coverage of both MgF and MgM could be explained by the high stringent conditions used in the assembly (due to the heterozygosity of the *M. galloprovincialis* genome). These conditions may have prevented the assembler to behave normally in the presence of reads from two rather similar mitochondrial genomes and in different proportions within the studied tissue. Indeed, this incomplete coverage may be also due to the low similarity between our mitochondrial sequences in the assembly with MgF and MgM.

To test the latter possibility, we searched further for mitochondrial sequences using the repertoire of reads from *M. galloprovincialis*. We directly mapped them onto different mitochondrial genomes from the genus *Mytilus*. Read mapping against mitochondrial genomes of *Mytilus* spp. produced a full coverage of several mitochondrial genomes. The largest number of reads (113,824) mapped onto an F mitochondrial genome from *M. edulis* (MeF, GenBank KM192128, Fig 4A), slightly more (111 more) than the number of reads that mapped to MgF. Moreover, fewer variants were found when mapping against MeF (228 variants) than MgF (302 variants). These two mitochondrial genomes were 99.26% identical to each other at nucleotide level. Surprisingly, a well-covered mapping against an M mitochondrial genome of *M. edulis* (MeM, Genbank KM192129) was also observed (Fig 4B). The average mapping coverage against the latter was 195X, about 10 times lower than that against MeF (1156X). A likely explanation of a better read mapping over *M. edulis* mitochondrial genomes is introgression.

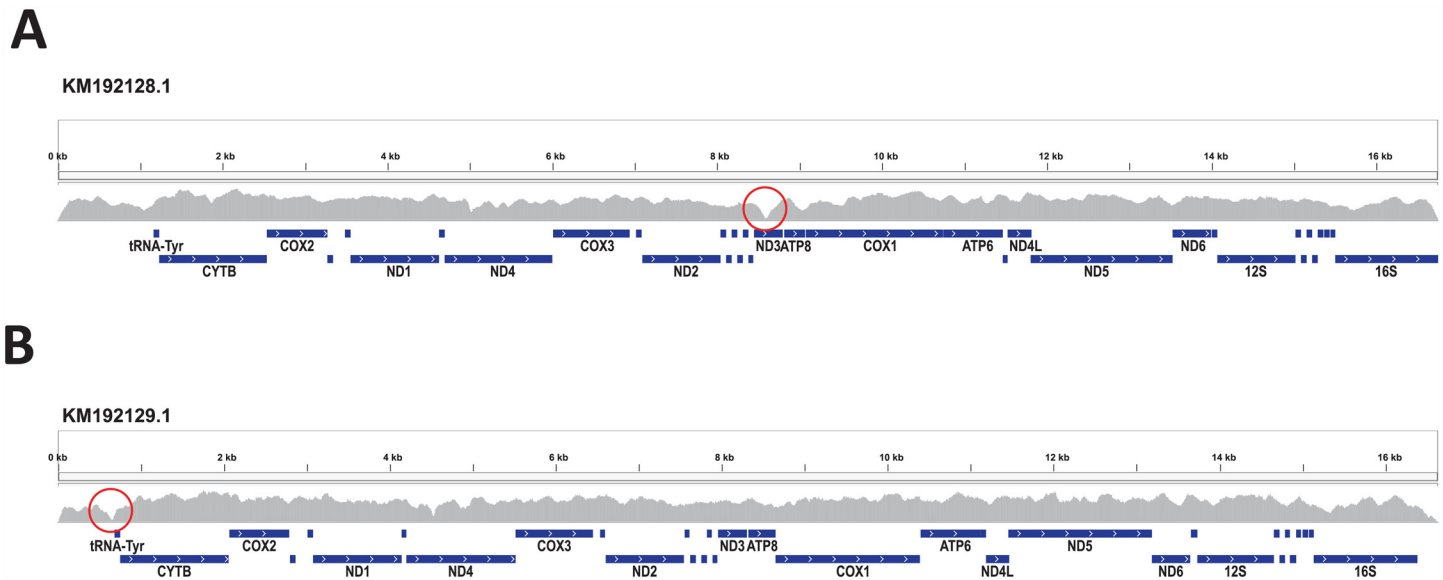


Fig 4. Read mapping against the mitochondrial genomes A) MeF (GenBank KM192128) and B) MeM (GenBank KM192129). Red circles highlight the regions of least coverage.

doi:10.1371/journal.pone.0151561.g004

Introgressed *M. edulis* mitochondrial haplotypes in *M. galloprovincialis* and *M. trossulus* populations have already been described [75,76]. Šmietanka et al. [76] reported a predominance of *M. edulis* mitochondrial haplotypes in the Atlantic *M. galloprovincialis* population. On the other hand, despite the presence of both M and F haplotypes in our data, it was not possible to determine the sex of our sequenced individual based only on its mitochondrial sequences. Though individuals having both M and F haplotypes are most likely males [3,77], presence of the M haplotype is not causally linked to masculinity [78]. Moreover, presence of M haplotypes in female individuals outside of hybrid zones due to disruption of DUI has also been reported [79].

Read mapping was not evenly distributed along mitochondrial genomes. Sequence motifs such as inverted repeats and GGC motifs could produce some coverage bias [80]. When mapping our reads against MeM, the region least covered (18X) corresponded to a short sequence in the control region (CR), upstream of the tRNA^{Tyr} gene, within its corresponding variable domain 2 (VD2) [81]. This could be due to the presence of homopolymers of guanines in VD2 that could have biased the sequencing step [80]. On the other hand, the region least covered of MeF was located within a coding gene, the NADH dehydrogenase 3 (ND3) (Fig 4A). However, this region did not contain homopolymers. One explanation for this low coverage could be the presence of DNA secondary structures in this region. Illumina technology is vulnerable to bias due to secondary structures formed during the amplification step [82]. Rodakis et al. [83] postulated the presence of a hairpin at positions 8545 to 8573 of MgF, only 9 bp from the start of the corresponding region of low coverage. This hairpin structure might be the origin of replication of the light strand in F haplotypes under the asymmetrical model of mtDNA replication. Though this hairpin structure is not located exactly over the low covered region, its closeness may have biased the amplification and sequencing of the DNA fragment covering this mitochondrial region (Fig 4A).

Although mitochondrial evolution is usually assumed to be neutral, evidence of selective pressures on mitochondria has been reported in different organisms [84–87]. Relevance of mitochondrial nucleotide variation on the fitness of individuals has been a matter of dispute.

To characterize this variation, we searched for single-nucleotide variants (SNVs) within our mitochondrial DNA sequences. Variant calling using MeF showed that 192 out of 228 (84,95%) variants were located in coding regions. Both M and F mitochondria in *M. galloprovincialis* contain two rRNA, 13 protein-coding, and 23 tRNA genes. In our mapping over MeF, the gene containing the highest ratio of variants was ND4 (S1 Table). Interestingly, ND4 was previously reported as the most abundant expressed transcript in mussels [88]. Thus, the high ratio of variants in ND4 agreed with previous evidence of elevated number of mutation rates in highly expressed genes [89,90]. However, adaptive processes could produce some of these ND4 variants. Signatures of adaptive variation on this proton pump were reported in other marine organisms [84,91]. On the other hand, when mapping our reads against MeM, the coding gene with the highest percentage of variants per nucleotide was cytochrome b (CYTB) gene. This gene has been previously reported to be downregulated on male mussels when exposed to 17beta-estradiol E2 [92]. In mussels, the adaptive variants in ND4 and CYTB could regulate the pH gradient affecting respiratory control. Further studies need to be done to test the role of these genes in the hypoxia tolerance of mussels in the intertidal zone.

Gene models

One of the major objectives of a genome-sequencing project is gene prediction. Due to the low sequencing depth of *M. galloprovincialis* (32X), we expected that the corresponding assembly would be either incomplete or extensively fragmented. Accordingly, gene prediction in our assembled genome sequencing data yielded an incomplete gene repertoire. To assess the amount of missing genes in our working assembly, the completeness of the gene content was measured with CEGMA [38]. Results of the CEGMA analyses showed that only 39 (15.73%) of the 248 core eukaryotic genes (CEGs) were considered “complete” in our assembly. When the CEGMA analysis was extended to include also partial but significant matches, 107 (43.27%) proteins aligned. This percentage is approximately three times as much as the number of complete CEGs found. Table 3 shows the result of the completeness analyses in other mollusc genomes using the same approach. *L. gigantea* showed the highest percentage of completeness (85.89%). This may be explained by its reduced genome size and the use of Sanger technology for sequencing. Second was *C. gigas*, where about three quarters of its genes (78.63%) gave “complete” matches. Noteworthy, in *M. galloprovincialis* and *P. fucata* the percentage of “partial” matches were approximately three times as much as the number of “complete” CEGs. This observation could be explained also by the low coverage sequencing of these two genomes.

Table 3. Results for the CEGMA completeness analyses for 5 molluscan species.

Species	Completeness		Genome Size	Genome Coverage	
	Proteins	Percentage			
<i>M. galloprovincialis</i>	Complete	39	15.73%	1,600 Mb	40X
	Partial	107	43.27%		
<i>C. gigas</i>	Complete	195	78.63%	545 Mb	155X
	Partial	236	95.16%		
<i>P. fucata</i>	Complete	63	25.00%	1,150 Mb	40X
	Partial	153	61.69%		
<i>L. gigantea</i>	Complete	213	85.89%	359,5 Mb	8.87X
	Partial	242	97.58%		
<i>A. californica</i>	Complete	90	36.29%	1,800 Mb	11X
	Partial	207	83.47%		

doi:10.1371/journal.pone.0151561.t003

Fortunately, a preliminary view of the gene repertoire in *M. galloprovincialis* may be informative enough even at low sequencing depth. This repertoire consisted of 10,891 protein-coding genes. In addition, we can extrapolate the percentage of “partial” mapped CEGs as the completeness of the whole gene repertoire in our genome assembly [38]. Therefore, the expected number of genes in the *M. galloprovincialis* genome would be about 25,000 genes.

Finally, although the Blast2GO annotation resulted in 2,397 predicted protein sequences with no BLAST hits against the non-redundant protein database, it succeeded in effectively annotating about a quarter (2,800 sequences) of the total number of proteins in our assembly. The species that contributed most annotations was *C. gigas* (Fig 5).

Gene Functional Annotation

In the ontology assignments of Blast2GO there was not any biological process category at level three overrepresented in *M. galloprovincialis* relative to the other molluscs (Fig 6A). The most abundant level-three gene ontology (GO) terms found in the other molluscs corresponded also to the most abundant GO terms found in *M. galloprovincialis*: organic substance metabolic process (GO:0071704, 953 genes), cellular metabolic process (GO:0044237, 931 genes) and single-organism cellular process (GO:0044763, 924 genes). However, half of the common level-

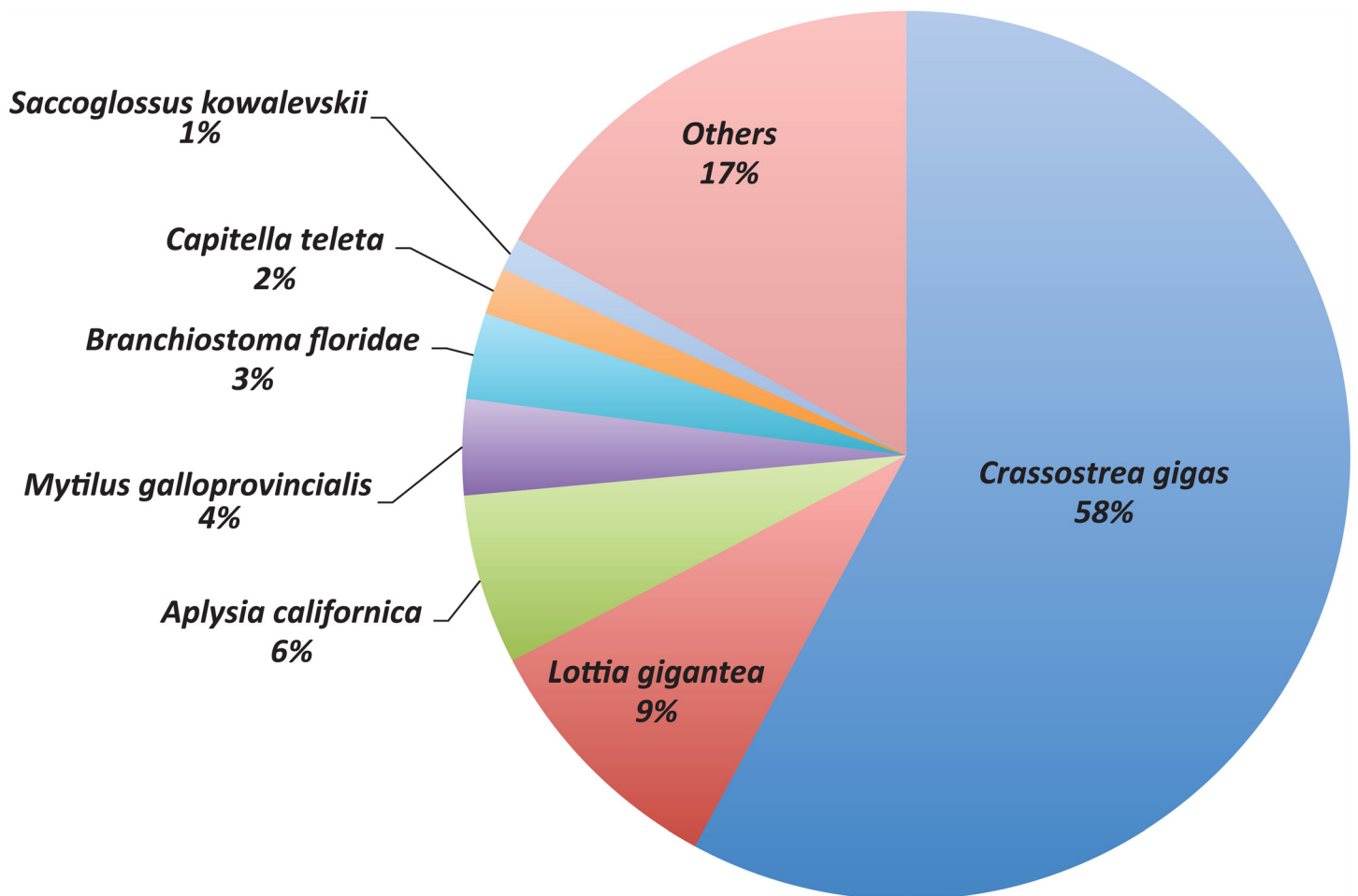
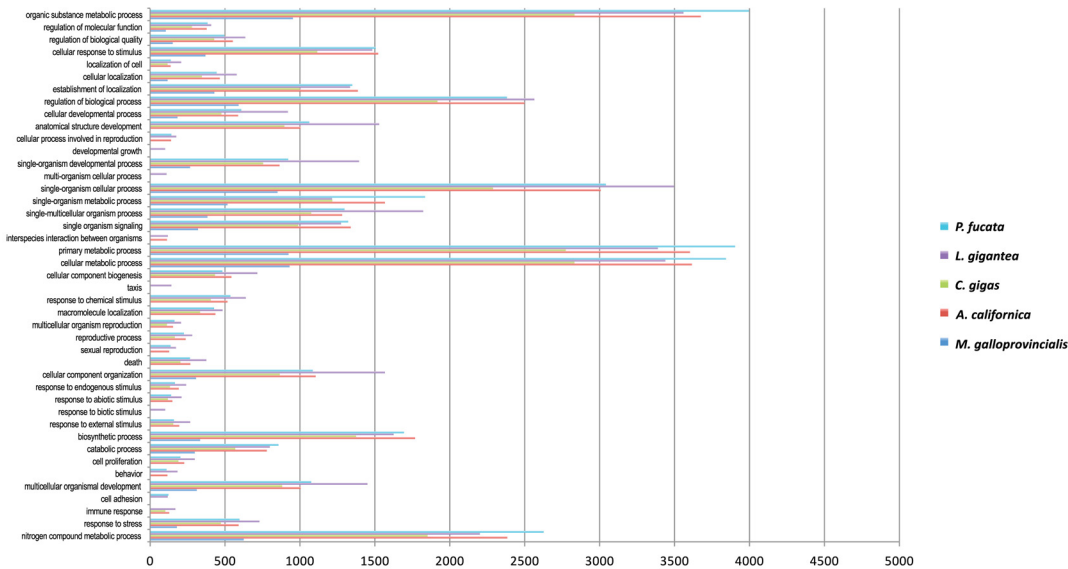


Fig 5. Distribution of best protein BLAST hits by species. Only species appearing in more than 1% of BLAST hits are represented.

doi:10.1371/journal.pone.0151561.g005

A)

Biological Process (L3)



B)

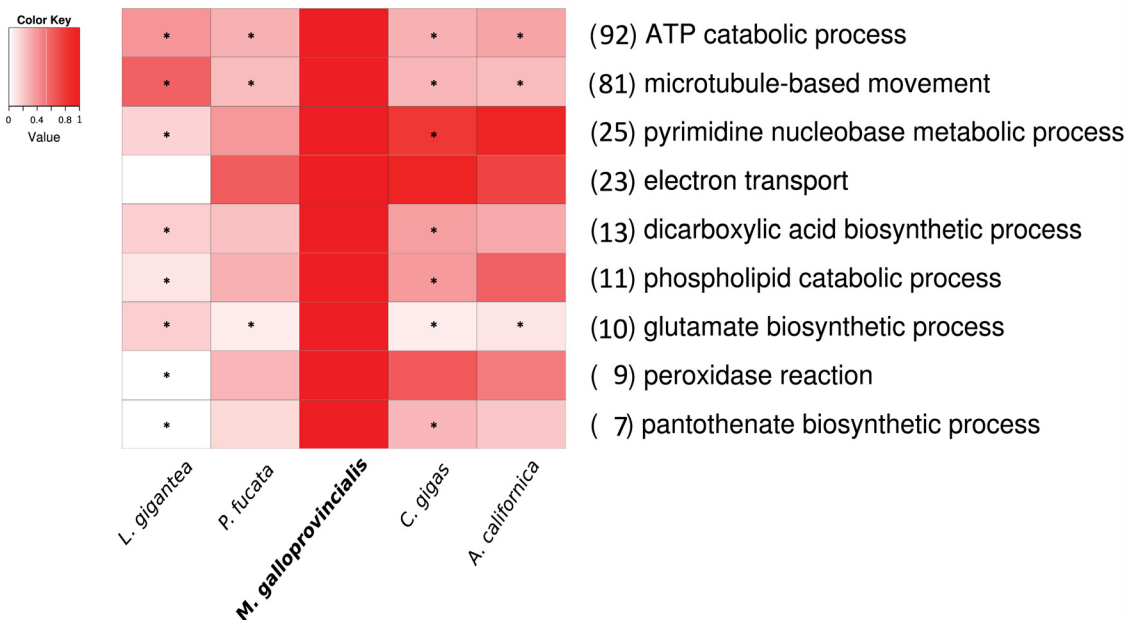


Fig 6. A) Gene Ontology predictions involved in biological processes at Level Three for the mollusc genomes studied. **B)** Biological Process Gene Ontologies with significant differences between *M. galloprovincialis* and other molluscs. Heatmap shows those ontologies where *M. galloprovincialis*

contained the largest percentage of genes per genome. Heatmap values were normalized assigning value one to *M. galloprovincialis*. Numbers next to each GO term indicate the number of genes in mussels with that annotation. Asterisks indicate the genome comparisons where mussel genes are statistically overrepresented as obtained with Blast2GO.

doi:10.1371/journal.pone.0151561.g006

three GO terms (21 out of 43 genes) in other molluscs were not found in *M. galloprovincialis*. The most likely explanation was the limited completeness of our genome.

Despite the completeness limitation, we were able to find genes with functional annotations that would explain how mussels cope with the specific characteristics of their environment, such as those related to immunity, resistance to hypoxia, shell formation, and adhesion to surfaces (S2 File). First, we looked for genes with immune-related functions in mussels that provide defence against bacteria and other environmental challenges [93–95]. GO terms related to these functions were “immune response” (GO:0006955, twenty genes), “immune system process” (GO:0002376, fifty five genes), and “defence response” (GO:0006952, thirty five genes). These genes had functions related to toll-like receptor signalling pathway, inflammatory response, regulation of innate immune response, and defence response to fungus, virus and bacteria. Second, other genes important for adaptation in mussels are genes that confer resistance to hypoxia and oxidative stress during tidal emersion and resubmersion cycles [96,97]. We found seven genes annotated with the GO term “resistance to hypoxia” (GO:0001666), such as isocitrate dehydrogenase [98] and ubiquitin carboxyl-terminal hydrolase [99] genes. In addition, another 25 genes were annotated with the GO term “response to oxidative stress” (GO:0006979). Finally, we manually searched for genes that provide physical protection against the environment. We found genes involved in byssus attachment to surfaces [100] and shell formation [101] (S2 File). In summary, our *M. galloprovincialis* genome contained genes whose functions may help this organism to adapt to its environment. They provide a starting point to test experimentally their role in these functions and their relevance in mussel biology.

Contrasting the GO annotations of *M. galloprovincialis* with those of other molluscs can reveal unique characteristics of the former. One hundred and forty ontologies presented significant differences in number (either positive or negative) in the *M. galloprovincialis*-versus-all comparison (S2 Table). In three of these ontologies mussel genes were significantly overrepresented (Fig 6B). The first of these three corresponded to GO:0006200, which is related to energy production based on ATP consumption. There were two main groups of gene functions within this GO term in *M. galloprovincialis*: ATP-binding cassette (ABC) (eight genes) and multidrug resistance-associated (16 genes). Interestingly, ABC transporters in *Mytilus* spp. [102] and other organisms [103–105] were previously reported as protection against multixenobiotics. These genes would represent the first line of defence against natural and anthropogenic toxicants in the marine environment. The second overrepresented GO term corresponded to “glutamate biosynthetic process” (GO:0006537, ten genes). Five of these were glutamate synthase genes. Similar annotations were also found in transcripts from hepatopancreas in crustaceans [106]. Glutamate synthase is mainly involved in the synthesis of glutamic acid from its precursor α -ketoglutarate [107]. Finally, the third overrepresented GO term corresponded to “microtubule-based movement” (GO:0007018). Almost half of the 81 proteins in this GO category were annotated as axonemal dynein heavy chain. The second most abundant gene function within this GO term corresponded to kinesin, a motor protein [108]. Both axonemal dyneins and kinesins participate in the active transport of molecules along ciliary structures [109,110].

At least two of the three aforementioned overrepresented ontologies in *M. galloprovincialis* had annotations linked to specific biological functions relevant for *M. galloprovincialis*. Indeed, genes contributing to either maintenance of ciliary structures (GO:0007018) or multixenobiotic

resistance (GO:0006200) [111,112] might be subject to high selective pressures from the environment. Therefore, the multiplicity of these genes in its gene repertoire might represent a genomic adaptation of *M. galloprovincialis* to sedentary filter-feeding life style that forces them to deal with a variety of changing environments and ecological characteristics. Further studies of the gene content, annotation and expression of these genes in *M. galloprovincialis* should be carried out to validate this hypothesis.

Conclusion

Next-Generation Sequencing (NGS) technologies have already significantly increased our understanding of many genomes across the Tree of Life. Next-Generation Sequencing has the potential to increase our basic knowledge on the genomes of non-model organisms such as marine molluscs, where genomic resources are scarce. As shown here with the Mediterranean mussel *Mytilus galloprovincialis*, *de novo* genome surveys at low-level sequencing depth can be used to provide first insights into the composition and structure of genomes in non-model organisms [18,19]. This study has shed some light onto the genome complexity, abundance of REs and (partial) gene repertoire of *M. galloprovincialis*. The comparative analyses of the genomic features observed in *M. galloprovincialis* with other marine molluscs have shown that an important part of the genome in these organisms contains a large number of repetitive sequences. Most of the REs found in *M. galloprovincialis* are unknown and need to be quantified and classified in more detail. Moreover, our analysis of the gene content in *M. galloprovincialis* has put into evidence the limits of low sequencing depth projects for gene annotation in complex genomes. Despite these limitations, through comparison with other molluscan genomes, we managed to identify two biological functions, detoxification and ciliary structure maintenance, where *M. galloprovincialis* has a large number of genes, most likely as a consequence of its condition of filter-feeder. This low-coverage genome survey will help in the design of additional sequencing and novel assembly strategies to obtain a more complete view of the mussel genome and the evolutionary forces that may have shaped its architecture and composition.

Availability of Supporting Data

Illumina read sequences used in this study can be downloaded from the NCBI Sequence Read Archive under the accessions SRR1598987, SRR1598945 and SRR1598943. Assembled sequences and annotations have been submitted to NCBI under the submission code SUB1006464 (PRJNA262617) following NCBI WGS requirements.

Supporting Information

S1 Fig. Nucmer alignments over the Mitochondrial F haplotype of *M. galloprovincialis*.

Rectangles depict direct (red) and reverse (blue) matches.

(PDF)

S1 File. Assemblies and Mitogenomes—List of web sites containing genome assemblies of the studied molluscs and GenBank accession numbers of all the mitochondrial genome sequences analysed.

(PDF)

S2 File. Summary of gene functional annotations. List of genes with functions related to immunity, resistance to hypoxia and stress, shell formation and adhesion to surfaces.

(PDF)

S1 Table. Variants in mitochondrial genes. The table shows the number of variants present in the mitochondrial genes. The following information is given: gene name, size of the gene in base-pairs, absolute number of variants found and percentage of variants related to the gene length.

(PDF)

S2 Table. GO terms under the category “biological process” with significant differences between *M. galloprovincialis* and at least one mollusc protein set. The following information is given: GO-ID, GO-term, Number of proteins in the organism for that GO annotation, False Discovery Rate value, P-value and whether it is up or downregulated. For those non-significant terms, the symbol \approx is used. When significant, “++” is used for proteins overrepresented in *M. galloprovincialis*, and “--” for those underrepresented.

(PDF)

Acknowledgments

Thanks to Steven Salzberg at Johns Hopkins University for providing technical advice and support on genome assembly. This work was supported by the Axencia Galega de Innovacion para o Desenvolvemento (Project INCITE 10PXIB402096PR) (Spain). CC was funded by an “Isidro Parga Pondal” research fellowship from Xunta de Galicia (Spain).

Author Contributions

Conceived and designed the experiments: BN AF DP (fifth author) CC. Performed the experiments: CC DP (second author). Analyzed the data: MM CC. Contributed reagents/materials/analysis tools: BN AF DP (fifth author) CC. Wrote the paper: MM BN AF DP (fifth author) CC.

References

1. Brusca RC, Brusca GJ. Invertebrates. Second. Sinauer Associates Incorporated; 2003.
2. Stanley SM. Adaptive themes in the evolution of the Bivalvia (Mollusca). *Annu Rev Earth Planet Sci.* 1975; 3: 361.
3. Skibinski DOF, Gallagher C, Beynon CM. Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. *Genetics.* 1994; 138: 801–809. PMID: [7851776](#)
4. Zouros E. The exceptional mitochondrial DNA system of the mussel family Mytilidae. *Genes Genet Syst.* 2000; 75: 313–318. PMID: [11280005](#)
5. Breton S, Beaupré HD, Stewart DT, Hoeh WR, Blier PU. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? *Trends in Genetics.* 2006; 23: 465–474. doi: [10.1016/j.tig.2007.05.011](#)
6. Watermann BT, Herlyn M, Daehne B, Bergmann S, Meemken M, Kolodzey H. Pathology and mass mortality of Pacific oysters, *Crassostrea gigas* (Thunberg), in 2005 at the East Frisian coast, Germany. *Journal of Fish Diseases.* 2008; 31: 621–630. doi: [10.1111/j.1365-2761.2008.00953.x](#) PMID: [18700939](#)
7. Gestal C, Roch P, Renault T, Pallavicini A, Paillard C, Novoa B, et al. Study of Diseases and the Immune System of Bivalves Using Molecular Biology and Genomics. *Reviews in Fisheries Science.* 2008; 16: 133–156. doi: [10.1080/10641260802325518](#)
8. Myrand B, Guderley H, Himelman J. Reproduction and summer mortality of blue mussels *Mytilus edulis* in the Magdalen Islands, southern Gulf of St. Lawrence. *Mar Ecol Prog Ser.* 2000; 197: 193–207.
9. Robert R, Sanchez JL, Perez-Paralle L. A glimpse on the mollusc industry in Europe. *Aquaculture.* 2013.
10. National Statistics Institute. Spanish Statistical Office. INEBASE [Internet]. 2014. Available: <http://www.ine.es/inebmenu/indice.htm>
11. Xunta de Galicia. Anuario de Acuicultura [Internet]. 2013. Available: <http://www.pescadegalicia.com/Publicaciones/AnuarioAcuicultura2013/>

12. Venier P, Pallavicini A, De Nardi B, Lanfranchi G. Towards a catalogue of genes transcribed in multiple tissues of *Mytilus galloprovincialis*. *Gene*. 2003; 314: 29–40. doi: [10.1016/S0378-1119\(03\)00708-X](https://doi.org/10.1016/S0378-1119(03)00708-X) PMID: [14527715](https://pubmed.ncbi.nlm.nih.gov/14527715/)
13. Venier P, De Pittà C, Bernante F, Varotto L, De Nardi B, Bovo G, et al. MytiBase: a knowledgebase of mussel (*M. galloprovincialis*) transcribed sequences. *BMC Genomics*. 2009; 10: 72. doi: [10.1186/1471-2164-10-72](https://doi.org/10.1186/1471-2164-10-72) PMID: [19203376](https://pubmed.ncbi.nlm.nih.gov/19203376/)
14. Romero A, Estévez-Calvar N, Dios S, Figueras A, Novoa B. New Insights into the Apoptotic Process in Mollusks: Characterization of Caspase Genes in *Mytilus galloprovincialis*. Bergmann A, editor. *PLoS ONE*. 2011; 6: e17003. doi: [10.1371/journal.pone.0017003.t003](https://doi.org/10.1371/journal.pone.0017003.t003) PMID: [21347300](https://pubmed.ncbi.nlm.nih.gov/21347300/)
15. Rosani U, Varotto L, Rossi A, Rossi A, Roch P, Novoa B, et al. Massively parallel amplicon sequencing reveals isotype-specific variability of antimicrobial peptide transcripts in *Mytilus galloprovincialis*. Ouzounis CA, editor. *PLoS ONE*. 2010; 6: e26680–e26680. doi: [10.1371/journal.pone.0026680](https://doi.org/10.1371/journal.pone.0026680)
16. Suárez-Ulloa V, Fernández-Tajes J, Aguiar-Pulido V, Rivera-Casas C, González-Romero R, Ausió J, et al. The CHROMEVALOA database: a resource for the evaluation of Okadaic Acid contamination in the marine environment based on the chromatin-associated transcriptome of the mussel *Mytilus galloprovincialis*. *Marine Drugs*. Multidisciplinary Digital Publishing Institute; 2013; 11: 830–841. doi: [10.3390/md11030830](https://doi.org/10.3390/md11030830)
17. Figueras AJ. *Biología y cultivo del mejillón (Mytilus galloprovincialis) en Galicia*. Editorial CSIC—CSIC Press; 2007.
18. Rasmussen DA, Noor MAF. What can you do with 0.1x genome coverage? A case study based on a genome survey of the scuttle fly *Megaselia scalaris* (Phoridae). *BMC Genomics*. 2008; 10: 382–382. doi: [10.1186/1471-2164-10-382](https://doi.org/10.1186/1471-2164-10-382)
19. Leese F, Brand P, Rozenberg A, Mayer C, Agrawal S, Dambach J, et al. Exploring Pandora's Box: Potential and Pitfalls of Low Coverage Genome Surveys for Evolutionary Biology. Mans BJ, editor. *PLoS ONE*. 2012; 7: e49202. PMID: [23185309](https://pubmed.ncbi.nlm.nih.gov/23185309/)
20. Wicker T, Narechania A, Sabot F, Stein J, Vu GT, Graner A, et al. Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. *BMC Genomics*. 2008; 9: 518–15. doi: [10.1186/1471-2164-9-518](https://doi.org/10.1186/1471-2164-9-518) PMID: [18976483](https://pubmed.ncbi.nlm.nih.gov/18976483/)
21. Henkel CV, Dirks RP, Minegishi Y. First draft genome sequence of the Japanese eel, *Anguilla japonica*. *Gene*. Elsevier B.V.; 2012; 511: 195–201. doi: [10.1016/j.gene.2012.09.064](https://doi.org/10.1016/j.gene.2012.09.064)
22. Yoshida M-A, Ishikura Y, Moritaki T, Shoguchi E, Shimizu KK, Sese J, et al. Genome structure analysis of molluscs revealed whole genome duplication and lineage specific repeat variation. *Gene*. Elsevier B.V.; 2011; 483: 63–71. doi: [10.1016/j.gene.2011.05.027](https://doi.org/10.1016/j.gene.2011.05.027)
23. Simakov O, Marletaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U, et al. Insights into bilaterian evolution from three spiralian genomes. *Nature*. Nature Publishing Group; 2012; 493: 1–6. doi: [10.1038/nature11696](https://doi.org/10.1038/nature11696)
24. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*. Nature Publishing Group; 2013; 490: 49–54. doi: [10.1038/nature11413](https://doi.org/10.1038/nature11413)
25. Takeuchi T, Kawashima T, Koyanagi R, Gyoja F, Tanaka M, Ikuta T, et al. Draft genome of the pearl oyster *Pinctada fucata*: a platform for understanding bivalve biology. *DNA Research*. Oxford University Press; 2012; 19: 117–130. doi: [10.1093/dnares/dss005](https://doi.org/10.1093/dnares/dss005)
26. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics Methods and Protocols*. Oxford University Press; 2011; 27: 764–770. doi: [10.1093/bioinformatics/btr011](https://doi.org/10.1093/bioinformatics/btr011)
27. Li R, Fan W, Tian G, Zhu H, He L, Cai J, et al. The sequence and *de novo* assembly of the giant panda genome. *Nature*. Nature Publishing Group; 2010; 463: 311–317. doi: [10.1038/nature08696](https://doi.org/10.1038/nature08696)
28. Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, et al. Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species. *Gigascience*. 2013; 2: 10. doi: [10.1186/2047-217X-2-10](https://doi.org/10.1186/2047-217X-2-10) PMID: [23870653](https://pubmed.ncbi.nlm.nih.gov/23870653/)
29. Rawson PD, Joyner KL, Meetze K, Hilbish TJ. Evidence for intragenic recombination within a novel genetic marker that distinguishes mussels in the *Mytilus edulis* species complex. *Heredity*. 1996; 77: 599–607. doi: [10.1038/hdy.1996.187](https://doi.org/10.1038/hdy.1996.187) PMID: [8972080](https://pubmed.ncbi.nlm.nih.gov/8972080/)
30. Bierre N, David P, Boudry P, Bonhomme F. Assortative Fertilization and Selection at Larval Stage in the Mussels *Mytilus edulis* and *M. galloprovincialis*. *Evolution*. 2002; 56: 292–298. doi: [10.1111/j.0014-3820.2002.tb01339.x](https://doi.org/10.1111/j.0014-3820.2002.tb01339.x) PMID: [11926497](https://pubmed.ncbi.nlm.nih.gov/11926497/)
31. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215: 403–410. doi: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/)

32. Kearsse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics Methods and Protocols*. 2012; 28: 1647–1649. doi: [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199)
33. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004; 5: R12. doi: [10.1186/gb-2004-5-2-r12](https://doi.org/10.1186/gb-2004-5-2-r12) PMID: [14759262](https://pubmed.ncbi.nlm.nih.gov/14759262/)
34. Langmead Ben, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9: 357–359. doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923) PMID: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/)
35. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics Methods and Protocols*. Oxford University Press; 2009; 25: 2078–2079. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352)
36. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 2015; 6: 11–11. doi: [10.1186/s13100-015-0041-9](https://doi.org/10.1186/s13100-015-0041-9) PMID: [26045719](https://pubmed.ncbi.nlm.nih.gov/26045719/)
37. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. BioMed Central Ltd; 2014; 15: R46–R46. doi: [10.1186/gb-2014-15-3-r46](https://doi.org/10.1186/gb-2014-15-3-r46)
38. Parra G, Bradnam K, Ning Z, Keane T, Korf I. Assessing the gene space in draft genomes. *Nucleic Acids Research*. 2008; 37: 289–297. doi: [10.1093/nar/gkn916](https://doi.org/10.1093/nar/gkn916) PMID: [19042974](https://pubmed.ncbi.nlm.nih.gov/19042974/)
39. Holt C, Yandell M. The Generic Genome Browser: A Building Block for a Model Organism System Database. *BMC Bioinformatics*. 2010; 12: 491–491. doi: [10.1186/1471-2105-12-491](https://doi.org/10.1186/1471-2105-12-491)
40. Korf I. Gene finding in novel genomes. *BMC Bioinformatics*. 2004; 5: 59. doi: [10.1186/1471-2105-5-59](https://doi.org/10.1186/1471-2105-5-59) PMID: [15144565](https://pubmed.ncbi.nlm.nih.gov/15144565/)
41. Darling A, Carey L, Feng W. The design, implementation, and evaluation of mpiBLAST. *Proceedings of ClusterWorld*. 2003.
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: 3674–3676. doi: [10.1093/bioinformatics/bti610](https://doi.org/10.1093/bioinformatics/bti610) PMID: [16081474](https://pubmed.ncbi.nlm.nih.gov/16081474/)
43. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, et al. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Research*. 2011; 40: D306–D312. doi: [10.1093/nar/gkr948](https://doi.org/10.1093/nar/gkr948) PMID: [22096229](https://pubmed.ncbi.nlm.nih.gov/22096229/)
44. Zheng W, Huang L, Huang J, Wang X, Chen X, Zhao J, et al. High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nature Communications*. Nature Publishing Group; 2013; 4: 1–10. doi: [10.1038/ncomms3673](https://doi.org/10.1038/ncomms3673)
45. Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, et al. Genome sequence and analysis of the tuber crop potato. *Nature*. 2011; 475: 189–195. doi: [10.1038/nature10158](https://doi.org/10.1038/nature10158) PMID: [21743474](https://pubmed.ncbi.nlm.nih.gov/21743474/)
46. Liu B, Shi Y, Yuan J, Hu X, Zhang H, Li N, et al. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv.org*. 2013. p. 2012.
47. Skibinski DOF, Cross TF, Ahmad M. Electrophoretic investigation of systematic relationships in the marine mussels *Modiolus modiolus* L., *Mytilus edulis* L., and *Mytilus galloprovincialis* Lmk. (Mytilidae; Mollusca). *Biol J Linn Soc Lond*. Blackwell Publishing Ltd; 1980; 13: 65–73. doi: [10.1111/j.1095-8312.1980.tb00070.x](https://doi.org/10.1111/j.1095-8312.1980.tb00070.x)
48. Koehn RK, Gaffney PM. Genetic heterozygosity and growth rate in *Mytilus edulis*. *Mar Biol*. 1984; 82: 1–7. doi: [10.1007/BF00392757](https://doi.org/10.1007/BF00392757)
49. Grant WS, Cherry MI. *Mytilus galloprovincialis* Lmk. in Southern Africa. *Journal of Experimental Marine Biology and Ecology*. 1985; 90: 179–191. doi: [10.1016/0022-0981\(85\)90119-4](https://doi.org/10.1016/0022-0981(85)90119-4)
50. Diz AP, Presa P. Regional patterns of microsatellite variation in *Mytilus galloprovincialis* from the Iberian Peninsula. *Mar Biol*. 2008; 154: 277–286. doi: [10.1007/s00227-008-0921-3](https://doi.org/10.1007/s00227-008-0921-3)
51. Lallias D, Stockdale R, Boudry P, Lapègue S, Beaumont AR. Characterization of Ten Microsatellite Loci in the Blue Mussel *Mytilus edulis*. *Journal of Shellfish Research*. 2009; 28: 547–551. doi: [10.2983/035.028.0317](https://doi.org/10.2983/035.028.0317)
52. Romiguier J, Bernard A, Chenuil A, Dernat R, Duret L, Faivre N, et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*. Nature Publishing Group; 2014; 1–12. doi: [10.1038/nature13685](https://doi.org/10.1038/nature13685)
53. Elliott TA, Gregory TR. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philos Trans R Soc Lond, B, Biol Sci*. 2015; 370: 20140331. doi: [10.1098/rstb.2014.0331](https://doi.org/10.1098/rstb.2014.0331) PMID: [26323762](https://pubmed.ncbi.nlm.nih.gov/26323762/)
54. Thiriot-Quiévreux C, Ayraud N. Les caryotypes de quelques espèces de bivalves et de gastéropodes marins. *Mar Biol*. 1982; 70: 165–172.

55. Ieyama H, Kameoka O, Tan T, Yamasaki J. Chromosomes and nuclear DNA contents of some species in Mytilidae. *Venus, Jap Journ Malac. Japanese Journal of . . .*; 1994; 53: 327–331.
56. Rodríguez-Juíz AM, Torrado M, Méndez J. Genome-size variation in bivalve molluscs determined by flow cytometry. *Mar Biol.* 1996; 126: 489–497.
57. Huang Y, Li Y, Burt DW, Chen H, Zhang Y, Qian W, et al. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nature Genetics.* 2013; 45: 776–783. doi: [10.1038/ng.2657](https://doi.org/10.1038/ng.2657) PMID: [23749191](https://pubmed.ncbi.nlm.nih.gov/23749191/)
58. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, et al. The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics.* 2009; 41: 1275–1281. doi: [10.1038/ng.475](https://doi.org/10.1038/ng.475) PMID: [19881527](https://pubmed.ncbi.nlm.nih.gov/19881527/)
59. Gregory TR. Synergy between sequence and size in Large-scale genomics. *Nat Rev Genet.* 2005; 6: 699–708. doi: [10.1038/nrg1674](https://doi.org/10.1038/nrg1674) PMID: [16151375](https://pubmed.ncbi.nlm.nih.gov/16151375/)
60. Jurka J, Kapitonov VV, Kohany O, Jurka MV. Repetitive Sequences in Complex Genomes: Structure and Evolution. *Annu Rev Genom Human Genet.* 2007; 8: 241–259. doi: [10.1146/annurev.genom.8.080706.092416](https://doi.org/10.1146/annurev.genom.8.080706.092416)
61. Fedoroff NV. Transposable Elements, Epigenetics, and Genome Evolution. *Science.* 2012; 338: 758–767. PMID: [23145453](https://pubmed.ncbi.nlm.nih.gov/23145453/)
62. Chénais B, Caruso A, Hiard S, Casse N. The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Gene. Elsevier B.V.*; 2012; 509: 7–15. doi: [10.1016/j.gene.2012.07.042](https://doi.org/10.1016/j.gene.2012.07.042)
63. Feschotte C, Pritham EJ. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu Rev Genet.* 2007; 41: 331–68. doi: [10.1146/annurev.genet.40.110405.090448](https://doi.org/10.1146/annurev.genet.40.110405.090448) PMID: [18076328](https://pubmed.ncbi.nlm.nih.gov/18076328/)
64. Lisch D. How important are transposons for plant evolution? *Nat Rev Genet. Nature Publishing Group*; 2013; 14: 49–61. doi: [10.1038/nrg3374](https://doi.org/10.1038/nrg3374)
65. Lerat E. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity. Nature Publishing Group*; 2009; 104: 520–533. doi: [10.1038/hdy.2009.165](https://doi.org/10.1038/hdy.2009.165)
66. Metcalfe CJ, Casane D. Accommodating the load. *Mob Genet Elements.* 2014; 3: e24775. doi: [10.4161/mge.24775](https://doi.org/10.4161/mge.24775)
67. Evgen'ev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, Lankenau DH, et al. Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. *Proc Natl Acad Sci USA.* 1997; 94: 196–201. doi: [10.1073/pnas.94.1.196](https://doi.org/10.1073/pnas.94.1.196) PMID: [8990185](https://pubmed.ncbi.nlm.nih.gov/8990185/)
68. Lozovskaya ER, Scheinker VS, Evgen'ev MB. A hybrid dysgenesis syndrome in *Drosophila virilis*. *Genetics. Genetics Society of America*; 1990; 126: 619–623.
69. Habu Y, Hisatomi Y, Lida S. Molecular characterization of the mutable flaked allele for flower variegation in the common morning glory. *Plant Journal.* 1998; 16: 371–376. doi: [10.1046/j.1365-313x.1998.00308.x](https://doi.org/10.1046/j.1365-313x.1998.00308.x) PMID: [9881157](https://pubmed.ncbi.nlm.nih.gov/9881157/)
70. Yannopoulos G, Stamatis N, Monastiriotti M, Hatzopoulos P, Louis C. hobo is responsible for the induction of hybrid dysgenesis by strains of *Drosophila melanogaster* bearing the male recombination factor 23.5MRF. *Cell.* 1987; 49: 487–495. doi: [10.1016/0092-8674\(87\)90451-X](https://doi.org/10.1016/0092-8674(87)90451-X) PMID: [3032457](https://pubmed.ncbi.nlm.nih.gov/3032457/)
71. Elliott TA, Gregory TR. Do larger genomes contain more diverse transposable elements? *BMC Evolutionary Biology.* 2014; 15: 69–69. doi: [10.1186/s12862-015-0339-8](https://doi.org/10.1186/s12862-015-0339-8)
72. Ballard JWO, Melvin RG. Linking the mitochondrial genotype to the organismal phenotype. *Mol Ecol.* 2010; 19: 1523–1539. doi: [10.1111/j.1365-294X.2010.04594.x](https://doi.org/10.1111/j.1365-294X.2010.04594.x) PMID: [20345689](https://pubmed.ncbi.nlm.nih.gov/20345689/)
73. Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research.* 2013; 41: e129. doi: [10.1093/nar/gkt371](https://doi.org/10.1093/nar/gkt371) PMID: [23661685](https://pubmed.ncbi.nlm.nih.gov/23661685/)
74. Mizi A, Zouros E, Moschonas N, Rodakis GC. The complete maternal and paternal mitochondrial genomes of the Mediterranean mussel *Mytilus galloprovincialis*: implications for the doubly uniparental inheritance mode of mtDNA. *Molecular Biology and Evolution.* 2005; 22: 952–967. doi: [10.1093/molbev/msi079](https://doi.org/10.1093/molbev/msi079) PMID: [15647523](https://pubmed.ncbi.nlm.nih.gov/15647523/)
75. Rawson PD, Hilbish TJ. Asymmetric Introgression of Mitochondrial DNA Among European Populations of Blue Mussels (*Mytilus* spp.). *Evolution.* 1998; 52: 100. doi: [10.2307/2410924](https://doi.org/10.2307/2410924)
76. Śmietanka B, Burzyński A, Hummel H, Wenne R. Glacial history of the European marine mussels *Mytilus*, inferred from distribution of mitochondrial DNA lineages. *Heredity.* 2014; 113: 250–258. doi: [10.1038/hdy.2014.23](https://doi.org/10.1038/hdy.2014.23) PMID: [24619178](https://pubmed.ncbi.nlm.nih.gov/24619178/)
77. Zouros E, Ball AO, Saavedra C, Freeman KR. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proc Natl Acad Sci USA.* 1994; 91: 7463–7467. PMID: [8052604](https://pubmed.ncbi.nlm.nih.gov/8052604/)

78. Kenchington EL, Hamilton L, Cogswell A, Zouros E. Paternal mtDNA and maleness are co-inherited but not causally linked in mytilid mussels. PLoS ONE. Public Library of Science; 2008; 4: e6976–e6976. doi: [10.1371/journal.pone.0006976](https://doi.org/10.1371/journal.pone.0006976)
79. Brannock PM, Roberts MA, Hilbish TJ. Ubiquitous heteroplasmy in *Mytilus* spp. resulting from disruption in doubly uniparental inheritance regulation. Mar Ecol Prog Ser. 2013; 480: 131–143. doi: [10.3354/meps10228](https://doi.org/10.3354/meps10228)
80. Ekblom R, Smeds L, Ellegren H. Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. BMC Genomics. BioMed Central Ltd; 2014; 15. doi: [10.1186/1471-2164-15-467](https://doi.org/10.1186/1471-2164-15-467)
81. Cao L, Kenchington E, Zouros E, Rodakis GC. Evidence that the large noncoding sequence is the main control region of maternally and paternally transmitted mitochondrial genomes of the marine mussel (*Mytilus* spp.). Genetics. 2004; 167: 835–850. doi: [10.1534/genetics.103.026187](https://doi.org/10.1534/genetics.103.026187) PMID: [15238532](https://pubmed.ncbi.nlm.nih.gov/15238532/)
82. Stein A, Takasuka TE, Collings CK. Are nucleosome positions in vivo primarily determined by histone-DNA sequence preferences? Nucleic Acids Research. 2009; 38: 709–719. doi: [10.1093/nar/gkp1043](https://doi.org/10.1093/nar/gkp1043) PMID: [19934265](https://pubmed.ncbi.nlm.nih.gov/19934265/)
83. Rodakis GC, Cao L, Mizi A, Kenchington ELR, Zouros E. Nucleotide content gradients in maternally and paternally inherited mitochondrial genomes of the mussel *Mytilus*. Journal of Molecular Evolution. 2007; 65: 124–136. doi: [10.1007/s00239-005-0298-6](https://doi.org/10.1007/s00239-005-0298-6) PMID: [17632681](https://pubmed.ncbi.nlm.nih.gov/17632681/)
84. Consuegra S, John E, Verspoor E, de Leaniz CG. Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. Genetics Selection Evolution 2015 47:1. BioMed Central; 2015; 47: 1. doi: [10.1186/s12711-015-0138-0](https://doi.org/10.1186/s12711-015-0138-0)
85. Yang Y, Xu S, Xu J, Guo Y, Yang G. Adaptive Evolution of Mitochondrial Energy Metabolism Genes Associated with Increased Energy Demand in Flying Insects. Oliveira PL, editor. PLoS ONE. Public Library of Science; 2014; 9: e99120. doi: [10.1371/journal.pone.0099120](https://doi.org/10.1371/journal.pone.0099120)
86. Grossman LI, Wildman DE, Schmidt TR, Goodman M. Accelerated evolution of the electron transport chain in anthropoid primates. Trends in Genetics. 2003; 20: 578–585. doi: [10.1016/j.tig.2004.09.002](https://doi.org/10.1016/j.tig.2004.09.002)
87. Foote AD, Morin PA, Durban JW, Pitman RL, Wade P, Willerslev E, et al. Positive selection on the killer whale mitogenome. Biol Lett. 2011; 7: 116–118. doi: [10.1098/rsbl.2010.0638](https://doi.org/10.1098/rsbl.2010.0638) PMID: [20810427](https://pubmed.ncbi.nlm.nih.gov/20810427/)
88. Craft JA, Gilbert JA, Temperton B, Dempsey KE, Ashelford K, Tiwari B, et al. Pyrosequencing of *Mytilus galloprovincialis* cDNAs: Tissue-Specific Expression Patterns. Aziz RK, editor. PLoS ONE. 2010; 5: e8875. doi: [10.1371/journal.pone.0008875](https://doi.org/10.1371/journal.pone.0008875) PMID: [20111607](https://pubmed.ncbi.nlm.nih.gov/20111607/)
89. Park C, Qian W, Zhang J. Genomic evidence for elevated mutation rates in highly expressed genes. EMBO reports. EMBO Press; 2012; 13: 1123–1129. doi: [10.1038/embor.2012.165](https://doi.org/10.1038/embor.2012.165)
90. Chen X, Zhang J. Yeast mutation accumulation experiment supports elevated mutation rates at highly transcribed sites. Proc Natl Acad Sci USA. 2014; 111: E4062–E4062. doi: [10.1073/pnas.1412284111](https://doi.org/10.1073/pnas.1412284111) PMID: [25217566](https://pubmed.ncbi.nlm.nih.gov/25217566/)
91. Teacher AG, André C, Merilä J, Wheat CW. Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. BMC Evolutionary Biology. 2011; 12: 248–248. doi: [10.1186/1471-2148-12-248](https://doi.org/10.1186/1471-2148-12-248)
92. Ciocan CM, Cubero-Leon E, Minier C, Rotchell JM. Identification of reproduction-specific genes associated with maturation and estrogen exposure in a marine bivalve *Mytilus edulis*. Laudet V, editor. PLoS ONE. 2010; 6: e22326–e22326. doi: [10.1371/journal.pone.0022326](https://doi.org/10.1371/journal.pone.0022326)
93. Bassim S, Genard B, Gauthier-Clerc S, Moraga D, Tremblay R. Ontogeny of bivalve immunity: assessing the potential of next-generation sequencing techniques. Rev Aquacult. 2014; 1–21. doi: [10.1111/raq.12064](https://doi.org/10.1111/raq.12064)
94. Venier P, Varotto L, Millino C, Celegato B, Bernante F, Lanfranchi G, et al. Insights into the innate immunity of the Mediterranean mussel *Mytilus galloprovincialis*. BMC Genomics. BioMed Central Ltd; 2011; 12: 69. doi: [10.1186/1471-2164-12-69](https://doi.org/10.1186/1471-2164-12-69)
95. Bibby R, Widdicombe S, Parry H, Spicer J, Pipe R. Effects of ocean acidification on the immune response of the blue mussel *Mytilus edulis*. Aquat Biol. 2008; 2: 67–74. doi: [10.3354/ab00037](https://doi.org/10.3354/ab00037)
96. Letendre J, Chouquet B, Manduzio H, Marin M, Bultelle F, Leboulenger F, et al. Marine Environmental Research. Marine Environmental Research. Elsevier Ltd; 2009; 67: 69–74. doi: [10.1016/j.marenvres.2008.11.003](https://doi.org/10.1016/j.marenvres.2008.11.003)
97. Woo S, Denis V, Won H, Shin K, Lee G, Lee T-K, et al. Expressions of oxidative stress-related genes and antioxidant enzyme activities in *Mytilus galloprovincialis* (Bivalvia, Mollusca) exposed to hypoxia. Zool Stud. 2013; 52: 15. doi: [10.1186/1810-522X-52-15](https://doi.org/10.1186/1810-522X-52-15)
98. Wise DR, Ward PS, Shay JES, Cross JR, Gruber JJ, Sachdeva UM, et al. Hypoxia promotes isocitrate dehydrogenase-dependent carboxylation of α -ketoglutarate to citrate to support cell growth and

- viability. *Proc Natl Acad Sci USA*. 2011; 108: 19611–19616. doi: [10.1073/pnas.1117773108](https://doi.org/10.1073/pnas.1117773108) PMID: [22106302](https://pubmed.ncbi.nlm.nih.gov/22106302/)
99. Brahim-Horn C, Pouysségur J. When hypoxia signalling meets the ubiquitin-proteasomal pathway, new targets for cancer therapy. *Critical Reviews in Oncology/Hematology*. 2005; 53: 115–123. doi: [10.1016/j.critrevonc.2004.09.003](https://doi.org/10.1016/j.critrevonc.2004.09.003) PMID: [15661562](https://pubmed.ncbi.nlm.nih.gov/15661562/)
 100. Silverman HG, Roberto FF. Understanding marine mussel adhesion. *Mar Biotechnol*. 2007; 9: 661–681. doi: [10.1007/s10126-007-9053-x](https://doi.org/10.1007/s10126-007-9053-x) PMID: [17990038](https://pubmed.ncbi.nlm.nih.gov/17990038/)
 101. Suzuki M, Nagasawa H. Mollusk shell structures and their formation mechanism 1. *Can J Zool*. 2013; 91: 349–366. doi: [10.1139/cjz-2012-0333](https://doi.org/10.1139/cjz-2012-0333)
 102. Luckenbach T, Epel D. ABCB- and ABCC-type transporters confer multixenobiotic resistance and form an environment-tissue barrier in bivalve gills. *AJP: Regulatory, Integrative and Comparative Physiology*. 2008; 294: R1919–R1929. doi: [10.1152/ajpregu.00563.2007](https://doi.org/10.1152/ajpregu.00563.2007)
 103. Kurelec B. The multixenobiotic resistance mechanism in aquatic organisms. *Crit Rev Toxicol*. 1992; 22: 23–43. doi: [10.3109/10408449209145320](https://doi.org/10.3109/10408449209145320) PMID: [1352103](https://pubmed.ncbi.nlm.nih.gov/1352103/)
 104. 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*. 2005; 204: 216–237. doi: [10.1016/j.taap.2004.10.012](https://doi.org/10.1016/j.taap.2004.10.012) PMID: [15845415](https://pubmed.ncbi.nlm.nih.gov/15845415/)
 105. Epel D, Luckenbach T, Stevenson CN, Macmanus-Spencer LA, Hamdoun A, Smital T. Efflux transporters: Newly appreciated roles in protection against pollutants. *Environ Sci Technol*. 2008; 42: 3914–3920. PMID: [18589945](https://pubmed.ncbi.nlm.nih.gov/18589945/)
 106. Wang W, Wu X, Liu Z, Zheng H, Cheng Y. Insights into hepatopancreatic functions for nutrition metabolism and ovarian development in the crab *Portunus trituberculatus*: gene discovery in the comparative transcriptome of different hepatopancreas stages. Buratti E, editor. *PLoS ONE*. 2014; 9: e84921. doi: [10.1371/journal.pone.0084921](https://doi.org/10.1371/journal.pone.0084921) PMID: [24454766](https://pubmed.ncbi.nlm.nih.gov/24454766/)
 107. Lea PJ, Mifflin BJ. Glutamate synthase and the synthesis of glutamate in plants. *Plant Physiol Bioch*. 2003; 41: 555–564. doi: [10.1016/S0981-9428\(03\)00060-3](https://doi.org/10.1016/S0981-9428(03)00060-3)
 108. Verhey KJ, Dishinger J, Kee HL. Kinesin motors and primary cilia. *Biochem Soc Trans*. Portland Press Ltd; 2011; 39: 1120–1125. doi: [10.1042/BST0391120](https://doi.org/10.1042/BST0391120)
 109. Bray D. Cell Movements. Ben-Jacob E, editor. *PLoS ONE*. Garland Science; 2001. doi: [10.1371/journal.pone.0002648](https://doi.org/10.1371/journal.pone.0002648)
 110. Hirokawa N, Noda Y, Tanaka Y, Niwa S. Kinesin superfamily motor proteins and intracellular transport. *Nat Rev Mol Cell Biol*. 2009; 10: 682–696. doi: [10.1038/nrm2774](https://doi.org/10.1038/nrm2774) PMID: [19773780](https://pubmed.ncbi.nlm.nih.gov/19773780/)
 111. Gómez-Mendikute A, Elizondo M, Cajaraville MP. Characterization of mussel gill cells in vivo and in vitro. *Cell Tissue Res*. 2005; 321: 131–140. doi: [10.1007/s00441-005-1093-9](https://doi.org/10.1007/s00441-005-1093-9) PMID: [15902503](https://pubmed.ncbi.nlm.nih.gov/15902503/)
 112. Bošnjak I, Hrabar J, Petrić M, Lušić J, Šver L, Mladineo I. Multixenobiotic Resistance Mechanism in Gills of Reared vs. Wild Mediterranean Mussel *Mytilus galloprovincialis* (Lamarck, 1819). *Water Air Soil Pollut*. Springer International Publishing; 2014; 225: 2073–13. doi: [10.1007/s11270-014-2073-z](https://doi.org/10.1007/s11270-014-2073-z)