Mantle transcriptome sequencing of *Mytilus* spp. and identification of putative biomineralization genes

Magdalena Malachowicz and Roman Wenne

Institute of Oceanology Polish Academy of Sciences, Sopot, Poland

ABSTRACT

In molluscs, the shell secreted by mantle tissue during the biomineralization process is the first barrier against predators and mechanical damage. Changing environmental conditions, such as ocean acidification, influence shell strength and thus protection of the soft body within. Mussels are marine bivalves with important commercial and ecological value worldwide. Despite this importance, the proteins involved in the biomineralization and pigmentation processes in *Mytilus* spp. remain unclear, as does taxonomy of *Mytilus* taxa, though there have been many molecular studies. To further understanding in these areas, this study aimed to characterize and compare mantle transcriptomes of four mussel taxa using next generation sequencing. Mussels representing four taxa, were collected from several localities and RNA from mantle tissue was extracted. RNA sequences obtained were assembled, annotated and potential molecular markers, including simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) were identified. Candidate contigs putatively related to biomineralization and pigmentation processes were then selected and several transcripts were chosen for phylogenetic analyses from the Bivalvia class. Transcriptome comparisons between *Mytilus* taxa, including gene ontology (GO) enrichment analysis and orthologues identification were performed. Of assembled contigs, 46.57%, 37.28% and 17.53% were annotated using NCBI NR, GO and Kyoto Encyclopedia of Genes and Genomes databases, respectively. Potential SSRs (483) and SNPs (1,497) were identified. Results presented a total of 1,292 contigs putatively involved in biomineralization and melanogenesis. Phylogenetic analyses of α-carbonic anhydrase, chitinase and tyrosinase revealed complex evolutionary history and diversity of these genes, which may be a result of duplication events or adaptation to different environments in mussels and other bivalves. Enrichment analyses revealed GO terms associated with pH and thermal response in *Mytilus edulis* from the North Sea and *M. galloprovincialis* from the Mediterranean Sea. The phylogenetic analysis within the genus *Mytilus* revealed *M. californianus* and *M. coruscus* to be genetically more distant from the other taxa: *M. trossulus, M. edulis, M. chilensis* and *M. galloprovincialis*. This work represents the first mantle transcriptome comparison between *Mytilus* taxa and provides contigs putatively involved in biomineralization.

Corresponding author
Magdalena Malachowicz,
mwarzecha@iopan.gda.pl

INTRODUCTION

Mussels of the genus *Mytilus* are widespread in both the Northern and Southern Hemispheres and are important components of coastal ecosystems (*Zbawicka et al., 2012*; *Smietanka et al., 2013*; *Lockwood, Connor & Gracey, 2015*). Mussels are a bivalve species of significant ecological and commercial value worldwide with annual production of 1.2M tonnes (*Gosling, 2003*; *Kijewski et al., 2009*; *Zbawicka, Trucco & Wenne, 2018*). Due to their filter-feeding abilities and sessility, they also play an important role as biosensors for pollution and other environmental changes (including ocean acidification) in coastal areas (*De Zwaan & Eertman, 1996*; *Fitzer et al., 2014*). Rising levels of atmospheric CO_2 affects the ability of molluscs to incorporate calcium carbonate, which is essential for building and maintaining shells (*Vendrami et al., 2016*). In molluscs biocalcitic structures fulfil a wide range of functions such as protecting the inner soft body from predators and environmental damage, detoxification, and attachment of muscles (*Lowenstam, Weiner & Veis, 1990*). The shell is an exoskeleton made up of a mineral phase (calcium carbonate, $CaCO_3$ ~95%) and organic macromolecules: proteins, glycoproteins, polysaccharides and lipids (*Lowenstam, Weiner & Veis, 1990*). Despite being a minor component (<5%), proteins in the shell, referred to as shell matrix proteins (SMPs), may play a key role in biomineralization including size regulation, nucleation, morphology and organization of the growing of $CaCO_3$ crystals (*Marin et al., 2008*).

Shells of bivalves differ in color, shape and marking, according to environmental conditions, for example the shell of the *Mytilus edulis* Linnaeus, 1758 in the intertidal zone is blue–black and heavy, while in the sublittoral region, where mussels are continuously immersed, it is thin and brown (*Gosling, 2003*). Shell morphology and thickness of various mussels from the genus *Mytilus* have also been associated with variability of environmental conditions, such as sediment types, trophic conditions, wave impact (*Steffani & Branch, 2003*; *Moschino et al., 2015*; *Telesca et al., 2018*).

The mussel shell is formed by the heavy fold mantle tissue, which encloses the internal organs of the animal (*Gosling, 2003*). The mantle is a complex structure underneath the shell, which consists of connective tissue with hemolymph, neural tissue, muscles and gonads (*Gosling, 2003*). Moreover, it has a sensory function and plays a role in directing particles onto the gills and in deflecting heavier material, storage of nutrient reserves, in bioaccumulation of metals and other compounds and in shell repair and growth (*Gosling, 2003*). To identify candidate biomineralization genes, over a dozen mantle transcriptomes have been assembled for mussels (*Freer et al., 2014*; *Bjärnmark et al., 2016*; *Yarra et al., 2016*), clams (*Clark et al., 2010*; *Arivalagan et al., 2016*; *Sleight et al., 2016*), scallops (*Shi et al., 2013*; *Sun et al., 2015*) and oysters (*Fang et al., 2011*; *Li et al., 2017a*). However, there remains a lack of mantle tissue transcriptome from *M. trossulus* Gould, 1850 and *M. chilensis* Hupé, 1854 as well as candidate biomineralization transcripts, and their comparative analysis within the genus *Mytilus*. In this study, the mantle transcriptomes of the four mussel taxa: *M. edulis*, *M. galloprovincialis* Lamarck, 1819—the Mediterranean Sea and Tasmania, *M. trossulus*, *M. chilensis*, were assembled and characterized using a Roche GS-FLX 454 pyrosequencing

system. The aim of the present study was twofold: (1) to identify transcripts putatively involved in biomineralization and melanogenesis, and (2) to perform a comparative transcriptome analysis among different taxa of the genus *Mytilus* based on mantle tissue.

A large number of proteins putatively involved in biomineralization and pigmentation processes were identified in different *Mytilus* taxa. This study is the first to present the mantle tissue transcriptome in *M. trossulus* and *M. chilensis*. The generated data (accession number PRJNA419475) add to the growing sequence database for mussels and provide the first comparison of the mantle tissue using mussel transcriptomes from different taxa.

MATERIALS AND METHODS

RNA extraction and pyrosequencing

Five populations of mussels (*Mytilus* spp.) representing four taxa were collected from several localities: *M. edulis* (North Sea, EduN), *M. trossulus* (Vancouver, TroV), *M. galloprovincialis* (Mediterranean Sea, GalM), *M. galloprovincialis* (Tasmania, GalT), *M. chilensis* (Chile, Chil) (Table 1). All taxa had been identified by single nucleotide polymorphism (SNP) genotyping (Zbawicka et al., 2012; Wenne et al., 2016; Larraín et al., 2017). Specimens chosen for RNA extraction were adults of indeterminate sex, spent. Tissue sampling was standardized across individuals: extraction of total RNA from the central and outer mantle tissue from right and left mussel side was performed using a GenElute Mammalian Total RNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO, USA) according to the manufacturer's protocol, with minor modifications. Homogenization was carried out in a custom buffer composed of 100 mM Tris, 1.4M NaCl, 20 mM EDTA, 2% CTAB, proteinase K (20 mg/ml) and 0.03 mM 2-mercaptoethanol (Malachowicz, Wenne & Burzynski, 2017). Samples were stored at −20 °C. RNA samples from two to four individuals were pooled for each group in equal proportions to make a total of five libraries. The quantity and quality of total RNA was determined at 260 nm on microplate using an Epoch Microplate Spectrophotometer (BioTek Instruments, Incorporated, Winooski, VT, USA) and gel electrophoresis. RNA integrity was checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). cDNA libraries were prepared using a cDNA Rapid Library Preparation Kit (Roche Applied Science, Basel, Switzerland) and pyrosequencing was performed by Macrogen Incorporation (Seoul, Korea), using Roche GS-FLX sequencing system (Roche, Basel, Switzerland) according to manufacturer's instructions.

De novo assembly, annotation, SSR and SNP discovery

All reads from the mantle tissue of five *Mytilus* groups were processed separately, through the same pipeline. Raw reads were pre-processed by removing adapters and low quality sequences (quality score = 0.05, discard reads <50 bp) with CLC Genomics Workbench software (v.7.5.5, CLC Bio, Qiagen, Aarhus, Denmark). Quality reads were assembled into contigs using de-Bruijn graphs in CLC Genomics Workbench with default parameters (Malachowicz, Kijewska & Wenne, 2015; Malachowicz, Wenne & Burzynski, 2017). Contigs were first searched against proteins from the NCBI non-redundant (NR) database using a BLASTX tool implemented in BLAST+ (v.2.2.29) (Altschul et al., 1990), with an

Table 1 Information about mussel (*Mytilus* spp.) samples included in the present study.

Group/library name	Species	Location	Year of sampling	N
EduN	*M. edulis*	The North Sea: The Oosterschelde estuary	2006	2
TroV	*M. trossulus*	Canada: Vancouver	2012	2
GalM	*M. galloprovincialis*	The Mediterranean Sea: Trieste and Chioggia and the Black Sea	2012	4
GalT	*M. galloprovincialis*	Australia: Tasmania, Spring Bay	2012	2
Chil	*M. chilensis*	Chile: Punta Arenas and Concepción	2012	2

Note:
N, number of specimens.

E-value threshold of 10^{-5} and other default parameters. A list of the top BLAST hits was achieved by using bit score (12th column) and *E*-value values (11th column) with the code: sort -k1,1 -k12,12 nr -k11,11 g -k3,3 gr blastout.txt sort -u -k1,1–merge > besthit. For functional annotation, gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were assigned to the transcripts using Blast2GO software (*Conesa et al., 2005*) with *E*-value Hit-Filter = 1.0 E^{-10} (other default parameters) and the single directional best-hit method in the KEGG Automatic Annotation Server (*Moriya et al., 2007*), respectively.

The software, MSDB (*Du et al., 2013*) was used to discover potential microsatellite markers. Di- to hexa- nucleotide motifs, with minimum thresholds of 6-5-4-4-4 repeats, respectively and 50 bp of flanking sequence length were identified. SNP analyses were carried out using CLC Genomics Workbench. For each transcriptome, clean reads were mapped back to corresponding contigs with an overlap criterion of 70% and a similarity of 90%. Candidate variants were called with the following minimum value parameters: neighborhood quality = 15; quality of central bases = 20; coverage = 20; count = 8; variant frequency = 35.0.

Identification of potential biomineralization and melanogenesis-related contigs

Candidate contigs related to biomineralization and melanogenesis were identified by searching top BLASTX hits with key words from a reference list of previously reported candidate biomineralization genes (Table S1). Melanogenesis-related transcripts were identified based on KEGG pathways analysis.

Several biomineralization proteins were selected for additional analysis: carbonic anhydrase (CA), chitinase (CHIT), tyrosinase (TYR), dermatopontin (DPT), arginine kinase (AK) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Proteins of selected sequences were predicted using OrfPredictor (*Min et al., 2005*). Functional domains and motifs were identified using Simple Modular Architecture Research Tool (SMART) software, allowing identification and annotation of domain architectures (http://smart. embl-heidelberg.de) (*Letunic, Doerks & Bork, 2015*). To further scrutinize and compare selected contigs, representative reference bivalve databases were created. Protein sequences

described as CA, CHIT, DPT, AK and GAPDH, available at the GenBank database (www.ncbi.nlm.nih.gov) were downloaded (217, 155, 8, 43, 24 sequences, respectively). Only sequences with annotated domains, were selected (199, 123, 8, 43, 24 sequences, respectively). Following this, mussel contigs from the present study were searched against the created databases with an E-value threshold of 10^{-15}. For phylogenetic reconstruction of *Mytilus* spp. TYR proteins, contigs from this study, amino acid sequences from GenBank (Acc. No: AKS48166, ALA16023, OPL3388), and sequences from previous studies (*Aguilera, McDougall & Degnan, 2014*; *Moreira et al., 2015*; *Hüning et al., 2016*; *Sleight et al., 2016*) were used. Homologous sequences were aligned using multiple alignment program MUSCLE (*Edgar, 2004*) with default parameters, and trimmed using TrimAl (*Capella-Gutierrez, Silla-Martinez & Gabaldon, 2009*) with an 80% threshold. Best substitution models were inferred using ModelFinder (*Kalyaanamoorthy et al., 2017*), as follow: CA—the Le Gascuel (LG) model (*Le & Gascuel, 2008*), gamma distribution, four discrete rate categories (+G4); GH18—the Whelan and Goldman (WAG) model (*Whelan & Goldman, 2001*) + G4 with invariant sites (+I); AK and GAPDH—WAG+ G4; DPT—the Block substitution matrices (Blosum62) model (*Henikoff & Henikoff, 1992*) + G4, with frequencies (+F); TYR—Blosum62+F+I+G4 model. Three phylogenetic models were run for each alignment. Neighbor-joining (NJ) trees were performed using MEGA X (*Kumar et al., 2018*) and the Jones–Taylor–Thornton (JTT) substitution model (*Jones, Taylor & Thornton, 1992*). Maximum likelihood reconstructions were performed using IQ-TREE (*Nguyen et al., 2015*) with 1,000 bootstrap replicates. Bayesian inferences were constructed using MrBayes v.3.2 (*Ronquist et al., 2012*) with parameters: ngen = 1,500,000; samplefreq = 100; printfreq = 100; nchain = 4; starttree = random. Trees were visualized and edited using MEGA X and manually combined to produce a consensus tree for each protein alignment.

Comparative transcriptomic analysis

Homologous transcripts present in all five groups were identified by using bi-directional tBLASTx analyses conducted between each transcriptome with an E-value cut-off of 10^{-15}. A list of top BLAST hits was achieved by using bit score (12th column) and E-value values (11th column) with the code: sort -k1, 1 -k12, 12 nr -k11, 11 g -k3,3 gr blastout.txt | sort -u -k1,1–merge>besthit. Further, results from this study were compared with available *Mytilus* spp. mantle sequences. The accession numbers are: for *M. edulis*: PRJEB4516 (*Freer et al., 2014*), PRJNA30561 (*Yarra et al., 2016*); for *M. galloprovincialis*: PRJNA230138 (*Moreira et al., 2015*), PRJNA295512 (*Bjärnmark et al., 2016*). For this purpose, a de novo transcriptome assembly was performed using the raw reads deposited in the NCBI Sequence Read Archive (SRA). The Illumina HiSeq raw reads were filtered using Trimmomatic v.0.32 (*Bolger, Lohse & Usadel, 2014*) and retained reads were assembled using Trinity program with default parameters (*Grabherr et al., 2011*).

To find enriched GO terms between groups, Fisher's exact test was performed with Blast2GO software, applying a one-tailed test, removing double IDs, with a false discovery rate (FDR) cut-off of 0.05. As a test-set, list of contigs from each group was used. As a background (reference-set), a fusion of the five transcriptomes was used.

To identify putative orthologs among the genus of *Mytilus*, transcripts from this study and from *M. coruscus* (*Xu et al., 2016*), *M. californianus* (*Romiguier et al., 2014*) were analyzed using OrthoMCL software with default settings (*Li, Stoeckert & Roos, 2003*). The putative orthologous clusters were aligned with a MUSCLE tool and then concatenated in CLC Genomics Workbench. A splits network was created with SplitsTree 4.13.1 (*Huson & Bryant, 2006*), using the neighbor-net algorithm under pdistance model. A phylogenetic tree was inferred using a NJ model in MEGA X, with the JTT + G4 substitution model and 1,000 bootstrap replicates.

RESULTS

Assembly, annotation, SSR and SNP detection

A total of 753,135 raw reads with an average read length of 342 bp were generated from sequencing of the *Mytilus* spp. mantle transcriptomes. After filtering, 743,773 (98.75%) high-quality reads were used for de novo assembly. In total, 20,982 contigs with an average length of 529 bp were obtained (Table S2). Mussel raw sequences from this study were deposited in NCBI SRA under accession number PRJNA419475.

BLAST analysis against NCBI NR database showed that between 43.19% (TroV) and 52.18% (GalM) of transcripts had a significant hit (Tables S2 and S3). As expected, most of the annotated transcripts were homologous to proteins from the Mollusca phylum: Pacific oyster (*C. gigas*, av. 50%), *Mytilus* spp. (av. 10%) and owl limpet (*L. gigantea*, av. 5%) (Fig. S1). KEGG analysis showed that a total of 3,679 (17.53%) transcripts were assigned to 271 biochemical pathways. The number of sequences assigned to each pathway was different across groups (Table S2). However, the most represented pathways in all groups included focal adhesion, phagosome, ribosome and apoptosis (Table S4). The signal transduction, endocrine system, immune system and cellular community were the most represented KO (KEGG orthology) subcategories in all groups (Fig. S2). In total, 35,383 GO terms were assigned to 7,824 (37.28%) contigs (Table S2). The main GO category distribution was similar in all *Mytilus* spp. groups, the dominant was the biological process category (BP), followed by cellular component and molecular function (MF) (Fig. S3). In this study, 483 potential simple sequence repeats (SSRs) were identified in 645 contigs (3.07% of the total number of contigs) (Table S5A). Among the identified putative markers, the most highly represented in the EduN, TroV, GalM and GalT groups were di-nucleotide repeats, and tri-nucleotide in Chil. AT was the dominating motif in all groups (Table S5B). In addition, a total of 1,497 putative SNPs were detected in all groups. The A/G variation type was dominant in EduN and TroV groups, whereas in other groups the C/T was dominant (Fig. S4).

Contigs potentially involved in biomineralization and melanogenesis

BLASTx searching revealed 1,292 contigs with significant sequence similarity to candidate biomineralization proteins (Table S6). SMPs such as, nacrein, calmodulin, collagens, perlucin, perlustrin, PIF were found among identified contigs (see Table S6). Transcripts involved in ion and acid-base regulation, protein homeostasis and anti-oxidation were also found including V-type proton ATPases, sodium/potassium-transporting ATPases,
heat shock 70 kDa protein (HSP70) and glutathione peroxidase. Other annotations included C1q-domain-containing protein, cAMP-response element-binding protein and cold-shock protein (CSP) (Table S6). Of annotated biomineralization contigs, 68.65% had domains, including chitin binding, EF hand, sushi, Von Willebrand factor A (Table S6).

Several biomineralization proteins were selected for more comprehensive analyses. In *M. galloprovincialis*, *M. edulis* and *M. trossulus* groups, five contigs (GalT_1066, GalT_1436, EduN_2761, GalM_2384, TroV_1436) were identified as DPT with DERM domain. A phylogenetic analysis of DPT proteins in Bivalvia generated a separate, well-supported clade of Mytiloida members, suggesting that these contigs are *Mytilus* spp. orthologues of DPT (Fig. 1A). Contigs with 81–95% sequence identity to AK were identified in all groups. The phylogenetic analysis showed two separate clades of Pterimorphia and Heteroconchia subclasses (Fig. 1B). Contigs Chil_958, EduN_492, GalM_470 and TroV_123 encoded proteins with similarity to GAPDH and possessed the Gp_dh_C domain. Alignment with other known GAPDHs from Ostreoida, Pectinoida and Pterioida members revealed many conserved regions among these species (Fig. S5). However, a phylogenetic analysis showed two well-supported separate clades of Mytiloida and Ostreoida with Pectinoida orders (Fig. 1C). TYR amino acid sequences derived were aligned with available molluscan TYR sequences. A phylogenetic analysis revealed that mussel TYR, in this study, created two separate clades: clade A and clade B (Fig. 2), described earlier as ancestral and bivalve-specific (*Aguilera, McDougall & Degnan, 2014*). It is possible that there were two small expansions within *Mytilus*, which is a common feature in bivalves. In *M. galloprovincialis* there were TyrA2, TyrA3 orthologues shared with the genera *Crassostrea* and *Pinctada*, indicating gene duplication before the divergence of lineages A and B. Searches in mussel transcriptomes revealed five contigs (in EduN, TroV and GalT groups) that pertain to the CA gene family and 10 contigs (in all transcriptomes) that showed similarity to the glycoside hydrolase family 18 (GH18) CHIT members (Table 2). A SMART analysis revealed CA domains in annotated CA sequences, and GH18_CHIT-like superfamily domain in GH18 sequences. The chitin binding peritrophin-A domain (CBM_14 superfamily) was found in transcripts recognized as acidic mammalian CHIT (GalT_650, GalT_649 and TroV_158). The phylogenetic analysis of CA in the Bivalvia class revealed three major clusters: α-carbonic anhydrases related proteins (α-CARPs), cytosolic/extracellular α-carbonic anhydrases (α-CAs) and membrane-bound α-CAs. There were two visible sub-clusters: CA2 and CA14 (Fig. 3A). Two major clades, A and B were resolved in the GH18 CHIT family phylogenetic tree. The A cluster consisted of CHIT domain-containing proteins (CHIDs, with GH18_SI-CLP domain), and Di-N-acetylchitobiases. In contrast to A, cluster B contained genes involved in chitin degradation (CHIT and CHIT-like proteins), creating three sub-clades (B1, B2, B3) (Fig. 3B). Presented phylogenetic reconstructions were supported by the high values of the nodes.

A total of 42, 99, 19 and 62 contigs from the mussel mantle transcriptomes were assigned to 13, 40, 8 and 15 KO terms in melanogenesis, endocytosis, TYR and calcium signaling KEGG pathways, respectively (Table S7).
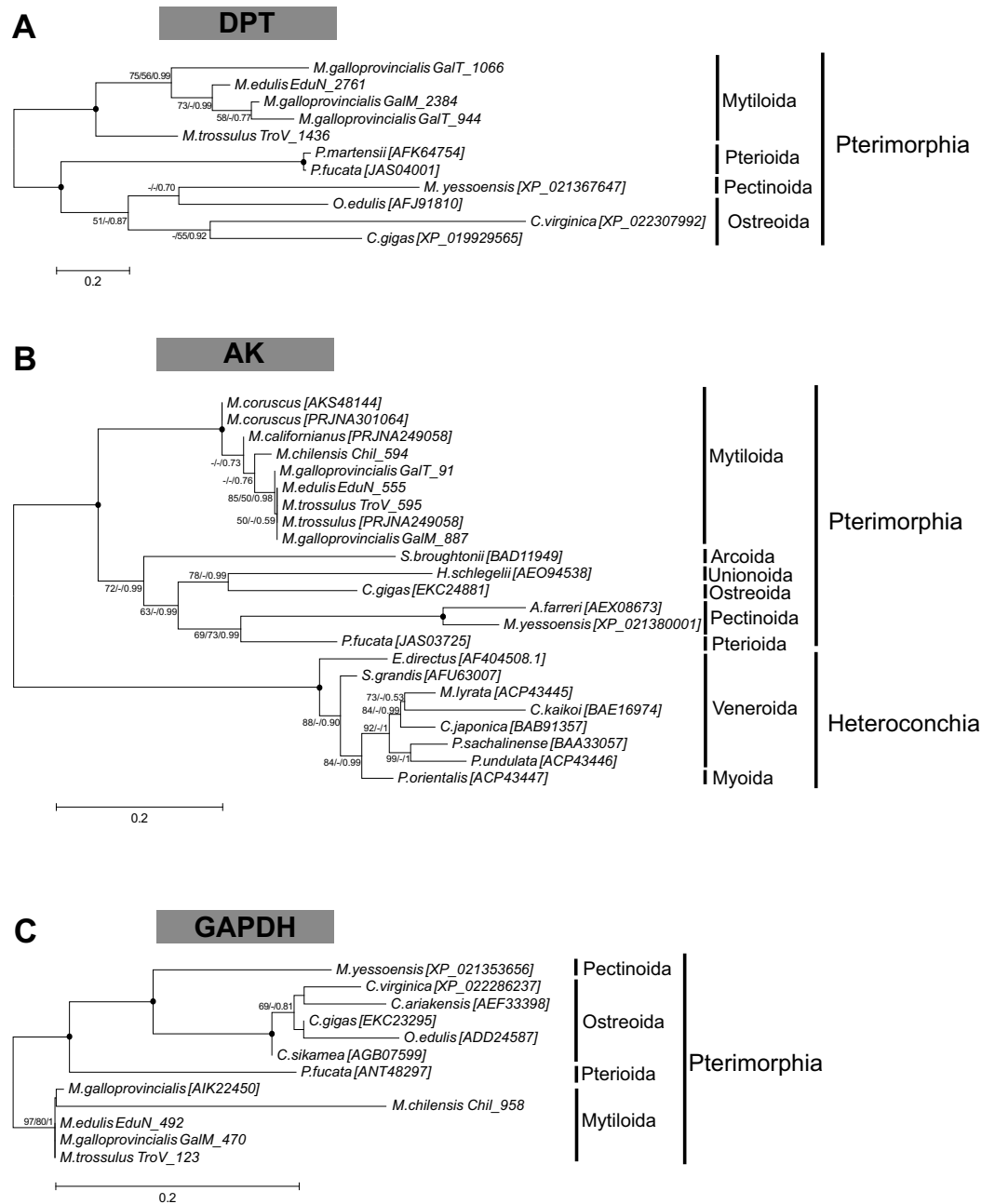
Figure 1 Molecular phylogenetic analysis of dermatopontin (A), arginine kinase (B) and glyceraldehyde-3-phosphate dehydrogenase (C) protein sequences in Bivalvia class. A consensus tree based on maximum likelihood (ML) topology. The numbers of the nodes show the bootstrap values (BV) >50% and Bayesian posterior probabilities (BPP) >0.50, as follows: ML BV/NJ BV/BIs BPP. A black dot at the node represents BV >90% and BPP >0.90.

Comparative transcriptomics

Bi-directional Best Hit (BBH) method was used to identify 6,130 homologous pairs between transcriptomes of which 552 were shared between all five groups (Fig. S6). The highest number of homologs were found between pairs: EduN_TroV, EduN_GalT and EduN_GalM (1,029, 946, 818, respectively). Putative homologous pairs were ordered
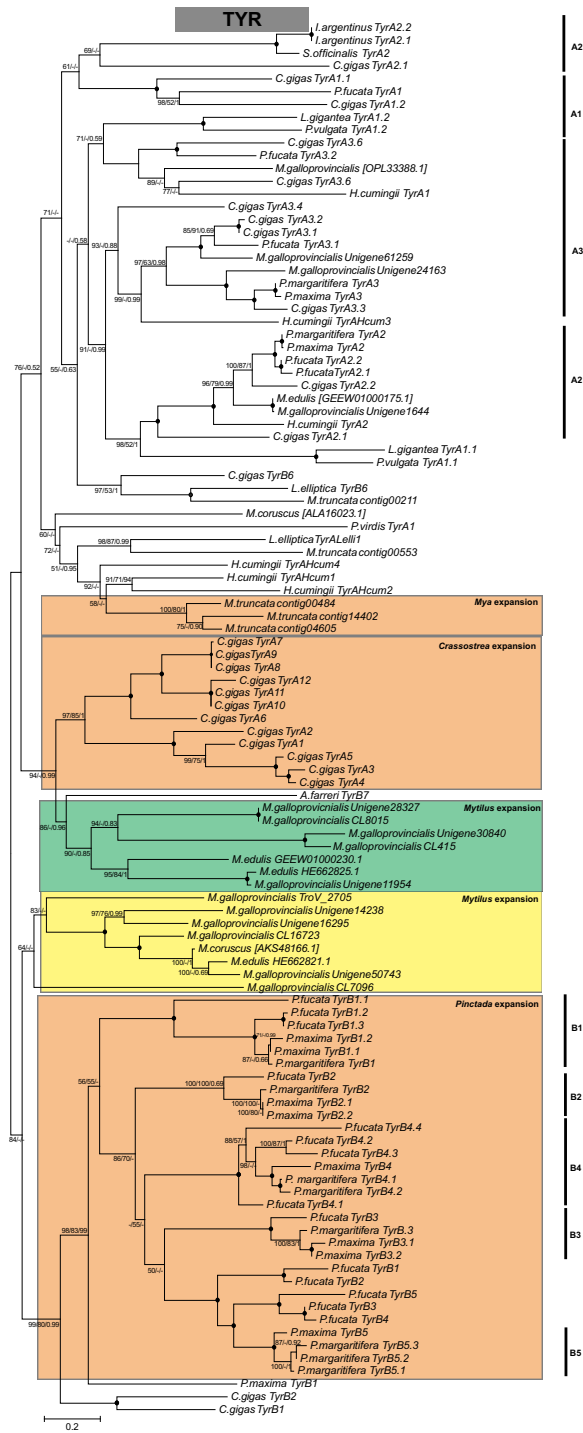
Figure 2 Molecular phylogenetic analysis of tyrosinase (TYR) protein sequences in Bivalvia class. A consensus tree based on maximum likelihood (ML) topology. The numbers of the nodes show the bootstrap values (BV) >50% and Bayesian posterior probabilities (BPP) >0.50, as follows: ML BV/NJ BV/ BIs BPP. A black dot at the node represents BV >90% and BPP >0.90. Bivalve TyrA ortholog groups are annotated as A1–A3, *Pinctada* spp. TyrB ortholog groups are annotated as B1–B5. Color code: green and yellow—two possible *Mytilus* expansions; orange—*Crassostrea*, *Mya*, *Pinctada* expansions, described in previous studies (*Aguilera, McDougall & Degnan, 2014*; *Sleight et al., 2016*).

Table 2 α-CA and GH18 proteins identified in mantle transcriptomes, according to BLASTX searches.

Contig	Size (nt)	Acc. number (NCBI)	Homology in NCBI (NR) database		
			Description	NCBI Acc. code	E-value
EduN_1982	697	MG827120	*M. coruscus* carbonic anhydrase-like protein	AKS48148.1	1E-46
EduN_2638	1,264	MG827121	*C. virginica* carbonic anhydrase 14-like isoform X3	XP_022339697.1	8E-81
EduN_4714	404	MG827122	*M. galloprovincialis* carbonic anhydrase II	ALF62133.1	3E-65
TroV_766	1,475	MG827123	*C. gigas* carbonic anhydrase 14-like isoform X2	XP_011435377.1	2E-59
GalT_2236	347	MG827124	*C. gigas* carbonic anhydrase 2-like	XP_011429671.1	5E-40
EduN_7774	362	MG827131	*M. yessoensis* chitinase 3	OWF51745.1	7E-37
EduN_2068	1,194	MG827125	*M. yessoensis* Acidic mammalian chitinase	OWF47140.1	1E-52
TroV_279	1,678	MG827132	*M. yessoensis* Acidic mammalian chitinase	OWF47140.1	4E-62
TroV_158	1,647	MG827133	*M. coruscus* mytchitin 5	AIF74559.1	0.00
GalM_153	195	MG827126	*M. coruscus* mytchitin 4	AIF74558.1	2E-25
Chil_296	1,026	MG827134	*M. coruscus* chitinase-like protein-1	AKS48165.1	8E-157
Chil_188	500	MG827127	*M. galloprovincialis* chitinase-like protein-3	AKS48199.1	9E-88
GalT_650	2,725	MG827128	*M. yessoensis* Acidic mammalian chitinase	OWF47140.1	3E-122
GalT_649	705	MG827129	*M. yessoensis* Acidic mammalian chitinase	OWF47140.1	1E-36
GalT_1838	847	MG827130	*M. yessoensis* probable chitinase 10	XP_021358847.1	2E-33
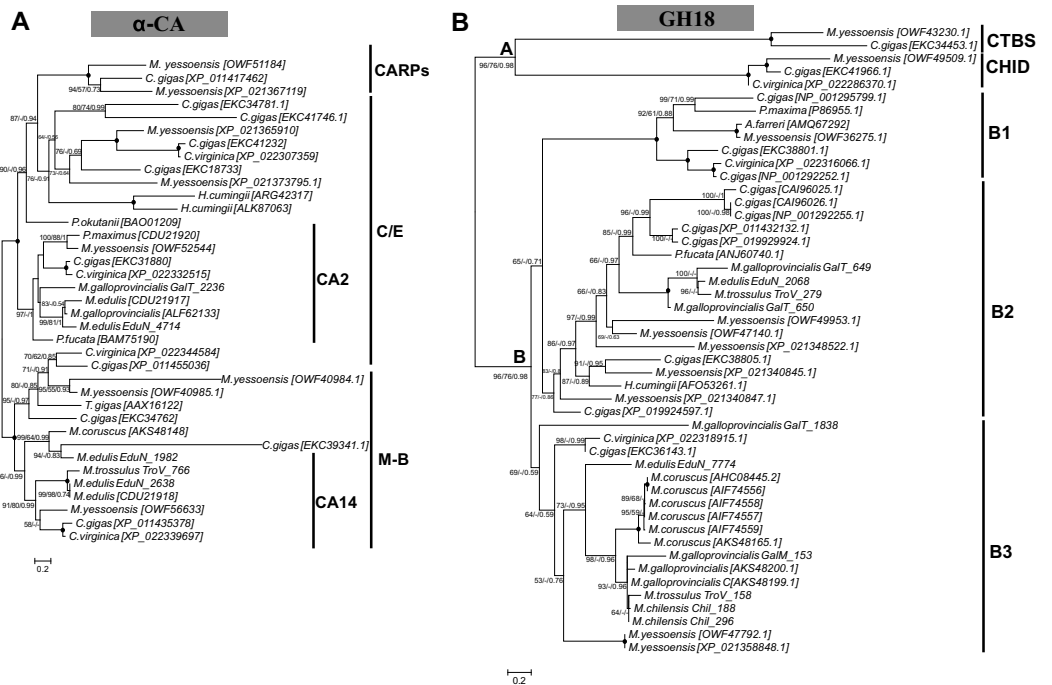
Figure 3 Molecular phylogenetic analysis of α-carbonic anhydrase (A) and glycoside hydrolase family 18 (B) protein sequences in Bivalvia class. A consensus tree based on maximum likelihood (ML) topology. The numbers of the nodes show the bootstrap values (BV) >50% and Bayesian posterior probabilities (BPP) >0.50, as follows: ML BV/NJ BV/BIs BPP. A black dot at the node represents BV >90% and BPP >0.90. Abbreviations: CARPs, carbonic anhydrase related proteins; C/E, cytosolic/extracellular; M-B, membrane-bound.

according to sequence similarity of BLAST alignments (Fig. S6). In general most homologous pairs exhibited a sequence similarity of 90–100%. The highest similarity per no. of homologs was found between EduN_GalM (85.45%), followed by EduN_Chil (84.89%), GalM_Chil (84.55%), GalM_GalT (84.12%). The lowest similarity to other groups was shown by TroV. Transcriptomes obtained from this study were compared to available mantle sequences for *M. edulis* and *M. galloprovincialis* using tBLASTx. In total, 18,456 (87.96%) contigs from this study were homologous with one of the four created mantle transcriptomes, but 12.04% might represent new sequence information. Additionally, between 47.8% and 81.5% of homologous contigs had high similarity (>90%; Fig. S7).

For more accurate analysis, Fisher's exact test was applied to discover significantly over-represented GO terms in each group. The most interesting GO enrichments associated with BP and MF were identified only in EduN and GalM groups. The results were reduced to the most specific terms (FDR = 0.01) (Figs. S8 and S9). A total of 25 GO terms were enriched in the *M. edulis* from the North Sea including cellular response to chemical stimulus (GO:0070887), positive regulation of cellular process (GO:0048522) in the BP category, ATP binding (GO:0005524), calcium ion binding (GO:0005509) in the MF category (Fig. S8). A total of 46 terms were enriched in the *M. galloprovincialis* from the Mediterranean Sea such as organonitrogen compound biosynthetic process (GO:1901566), regulation of biological quality (GO:0065008) in the BP category, metal ion binding (GO:0046872), purine ribonucleotide binding (GO:0032555) in the MF category (Fig. S9).

Both the neighbor-net splits network (Fig. 4A) and NJ phylogenetic tree (Fig. 4B) showed that *M. chilensis* was the most similar to the *M. edulis* taxon, followed by *M. galloprovincialis* and *M. trossulus*. A separate clade was identified containing *M. californianus* Conrad, 1837 and *M. coruscus* Gould, 1861 (presently accepted as *M. unguiculatus* Valenciennes, 1858) mussels (Fig. 4).

DISCUSSION

The mollusc shell, formed by calcium carbonate deposition (secreted by the mantle tissue), forms the first barrier against marine predators, mechanical or toxic damage and protects the soft body. This calcified structure is sensitive to environmental conditions: CO_2 level, temperature, humidity. Moreover, mussels have adapted to different environments that affect the shell formation process and the stability of the secreted composite biomineral (*Mann & Jackson, 2014*). The sequencing of mantle transcriptome from multiple taxa of *Mytilus* spp. has added to the growing transcriptome resources for mussels and has provided contigs putatively involved in biomineralization. High-quality reads were assembled totaling 20,982 contigs, 9,773 (46.57%) of them were annotated by BLAST searches against NCBI NR protein database (Tables S2 and S3). The top hits showed homology with other bivalves and most annotations were extracted from *C. gigas* Thunberg, 1793 (Fig. S1), as expected due to the availability of an oyster fully sequenced genome. Transcriptome-based markers are a valuable resource for determining functional genetic variation (*Malachowicz, Wenne & Burzynski, 2017*). In this study,

A

B
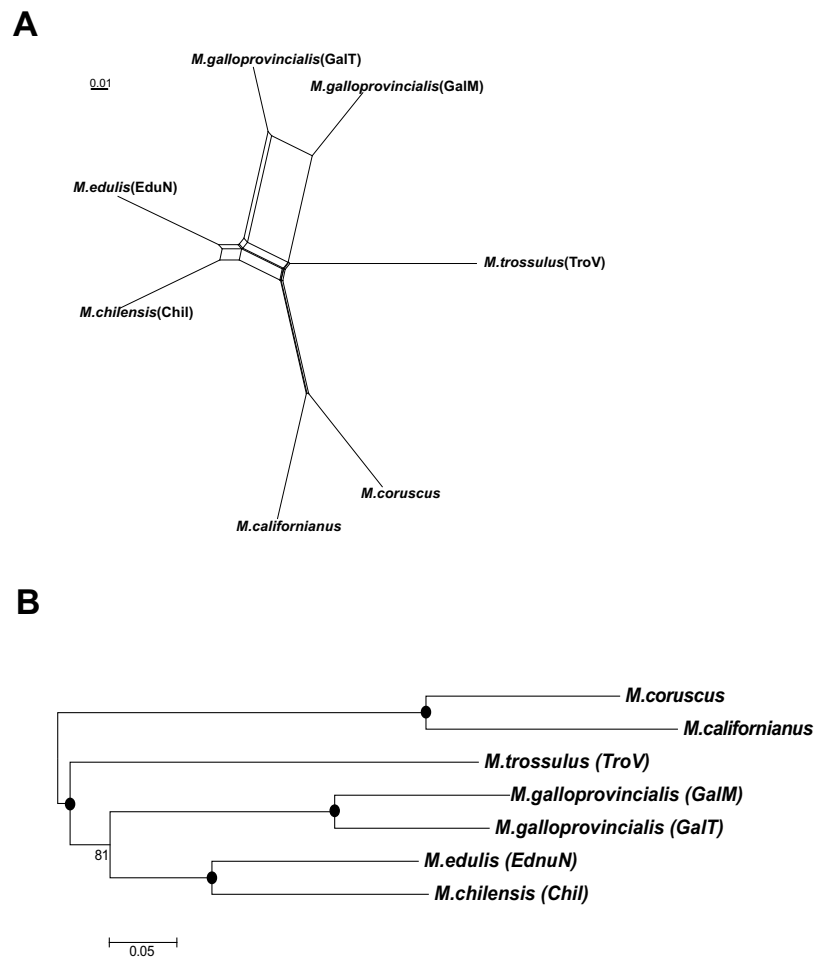
Figure 4 The splits network (A) and phylogeny (B) of six mussels taxa. A black dot at the node represents BV >90%.

483 potential SSRs and 1,497 putative mantle-related SNPs were identified. SSRs (*Vidal et al., 2009*) and SNPs (*Vendrami et al., 2016*) have previously been developed in mussels using the transcriptome approach, however, this study aimed to identify molecular markers in mantle tissue from several *Mytilus* taxa. These results might be a resource for molecular marker studies associating candidate biomineralization genes with different environmental conditions and/or taxa.

Both GO and KEGG analyses showed a wide array of categories and pathways potentially involved in biomineralization and pigmentation mechanisms (Tables S4 and S7). In all groups, cellular processes: focal adhesion, tight junction, adherens junction and signaling pathways: insulin, calcium, chemokine, ErbR, MAPK and VEGF were identified (Table S4). The importance of these pathways in shell formation has been proposed by *Wang et al. (2013)* and *Liu et al. (2015)*. Tyrosine metabolism and melanogenesis pathways detected in the mantle transcriptomes presented here might play a fundamental role in the shell pigmentation process. Numerous genes in those pathways were identified such as calmodulin, protein kinase, TYR, β-catenin (Table S7).

Moreover, it seems that the Wnt signaling pathway regulation of the transcription factor essential for melanocyte development (MITF) might play a crucial role in melanogenesis (melanocyte development) in bivalves. Another important pathway may be endocytosis, a mechanism for cells to remove nutrients, ligands and to influence pigment granule formation (*Feng et al., 2015*). *Feng et al. (2015)* further suggested, that, endocytosis and genes VPS or Rab might have an effect on shell coloration in *Crassostrea gigas* (Thunberg, 1793). These results are in line with previous studies in scallop (*Sun et al., 2015*) and oysters (*Li et al., 2017a*).

A total of 1,292 contigs putatively involved in the biomineralization mechanism were also identified. Although most of these transcripts have been reported in *M. edulis* (*Hüning et al., 2016*) and *M. galloprovincialis* (*Gao et al., 2015*; *Moreira et al., 2015*) mantle transcriptomes, new transcripts were identified for *M. trossulus* and *M. chilensis*. This data may contribute to future comparison between different localities of the same taxon, to assess the influence of environment on the shell formation. All presented contigs should be considered as candidate biomineralization proteins since their functions were hypothesized mainly based on expression data, but were not validated (*Herlitze et al., 2018*). In mussels, the shell is composed of two polymorph forms of calcium carbonate (calcite and aragonite) (*Liao et al., 2015*). Some matrix proteins, potentially related to the formation of calcite (e.g., TYR) as well as aragonite crystals (e.g., Pif, nacrein) and a number of enzymes metabolizing chitin protein, including CHIT (may be responsible for chitin degradation) and chitin synthase (may be active in construction of the chitin framework) (*Feng et al., 2017*) were identified in the presented transcriptomes. Perlucin (nucleates growth of $CaCO_3$ crystals) (*Blank et al., 2003*) and calponin (calcium binding protein, may ensure shell elasticity) (*Sleight et al., 2016*) were also confirmed in the transcriptomes of *Mytilus* spp. by this study (Table S6). In EduN, TroV, GalM and GalT groups, 15 contigs were identified rich in immunoglobulin domains with homology from 34% to 69% sequence identity to hemicentin from *M. yessonis* Jay, 1857, *Crassostrea virginica* Gmelin, 1791 and *C. gigas*. Hemicentin is an extracellular ion-binding protein with an unclear function in shell formation, however, it was previously found in the coral skeletal organic matrix (*Ramos-Silva et al., 2013*). Another significant protein is chorion peroxidase (TroV_3244) with an unknown role in biomineralization, though it has been suggested that it might be involved in melanogenesis in *S. officinalis* Linnaeus, 1758 (*Gesualdo et al., 1997*).

In this study, several candidate biomineralization proteins were selected for more comprehensive analysis: DPT, AK and GAPDH, since they were represented in at least four of the five *Mytilus* groups. DPT is an extracellular matrix protein and may participate in nacre formation (*Jiao et al., 2012*). GAPDH is an enzyme involved in carbohydrate metabolism, which catalyze the formation of 1,3-bisphospho glycerate from glyceraldehyde 3-phosphate and might be involved in the shell growth process (*Arivalagan et al., 2016*). AK is a protein with ATP-guanido phosphotransferases domain, which may play a role in buffering intracellular pH, important in the extrapallial space with the supersaturation of shell matrix components, including calcium ions (*Clark et al., 2010*). A phylogenetic reconstruction of these proteins showed a

well-supported, separate clade of Mytiloida order in a sister relationship to other Bivalves (Fig. 1). However, the trees presented did not support relationships between the four orders: Mytiloida, Ostreoida, Pterioida and Pectinoida. Only AK phylogeny clearly pictures the relationships among major bivalve lineages Pterimorphia and Heteroconchia, although it should be noted that there is a lack of sequences for DPT and GAPDH proteins. In the case of AK and GAPDH high conservation was observed between *M. edulis, M. galloprovincialis* and *M. trossulus* taxa, suggesting the importance of these candidate biomineralization genes in mussels. Only DPT phylogeny confirmed the expected evolutionary relationships of *M. trossulus, M. edulis* and *M. galloprovincialis*.

Rising atmospheric CO_2 emissions and the direct consequences of it, ocean acidification and ocean warming, have the potential to limit the ability of calcifying marine organisms to produce their protective exoskeleton and shell, and lead to decalcification (*Fitzer et al., 2014*; *Goffredo et al., 2014*). Changes in the structural order potentially impact on mussel shell strength, limiting protection from predators and the environment. In order to project the future implications of ocean acidification on marine calcifying organisms the biomineralization mechanism need to be fully understood. Previous studies have explored the impact of acidification on shell formation and on the activity of candidate biomineralization molecules: α-CA, CHIT and TYR in mussels (*Hüning et al., 2013*; *Fitzer et al., 2014*). The evolutionary relationships of these genes in metazoan have been previously studied (*Aguilera, McDougall & Degnan, 2014*; *Le Roy et al., 2014*; *Sleight et al., 2016*; *Feng et al., 2017*), however, there has still been a lack of sequence resources especially from mussels. Given the importance of these genes in biomineralization process, phylogenetic reconstructions using contigs obtained from this study were performed. α-CAs are metalloenzymes which catalyze the reversible hydration of CO_2 to form bicarbonate ion and proton (*Le Roy et al., 2014*). In mammals the α-CA family is characterized by 13 different enzymatically active isoforms and three CARPs which appear to lack CA activity (*Aspatwar, Tolvanen & Parkkila, 2010*). CHIT is hydrolyzing β-1,4-linkages in chitin, which is a major component of the organic matrix in the shell. TYR is a type-3 copper protein superfamily, a multifunctional, shell associated protein, melanocyte-specific marker, potentially involved in pigmentation and essential for the varied coloration of shell in bivalves (*Feng et al., 2015*; *Sun et al., 2015*). It has been suggested that TYR represents a protective mechanism against external shell corrosion (*Fitzer et al., 2014*). This study identified several copies (isoforms) of α-CA, CHIT and TYR proteins in the genus *Mytilus*, which might have been a result of several duplication events and diversification resulting in multiple functionality: immune response, potentially biomineralization and pigmentation (Figs. 2 and 3). This diversity might be also correlated with wide geographical distribution, adaptation to different environments and different selection pressures driving evolution of these specific genes. The contigs identified in this study have provided a new resource for future studies on α-CAs, CHIT and TYR, however, the phylogeny reveals a complex evolutionary history of these genes that requires more data.

To identify homologs and similarity of mantle transcriptomes between different taxa and localizations, BBH method was used. The highest similarity (>80%), was observed

between *M. edulis* (EduN), *M. galloprovincialis* (GalM, GalT) and *M. chilensis* (Chil) taxa (Fig. S6). Respectively, 848, 488, 549, 343 and 298 contigs in EduN, TroV, GalM, Chil, GalT groups, did not show homology (*E*-value cut-off 10^{-15}) to mantle transcriptomes from previous studies in mussels (Fig. S7). These contigs may represent novel or specific genes for each group or might be related to environment conditions. To further investigate diversity between different *Mytilus* spp. groups, a GO Fisher exact test was carried out. In *M. edulis* from the North Sea (EduN), GO terms related to heat stress response were over-represented, including: apoptotic process, regulation of programmed cell death, cellular protein complex assembly, microtubule-based process and calcium ion binding (Fig. S8). These terms were associated with high pH exposure in the Antarctic pteropod (*Johnson & Hofmann, 2016*) and thermal response in the genus *Crassostrea* (*Zhang et al., 2012*, *Li et al., 2017b*). In *M. galloprovincialis* from the Mediterranean Sea (GalM) metal ion binding, purine ribonucleoside triphosphate binding, lipid transport and oxidative phosphorylation were enriched (Fig. S9). These terms might be associated with pH and temperature response, as found in *P. fucata* Gould, 1850 (*Li et al., 2016*) and *C. gigas* (*Zhang et al., 2012*). Thermal and pH response of mantle tissue in both EduN and GalM groups, might be a direct consequence of increasing carbon dioxide (connected with ocean acidification and global warming), which significantly affects marine calcifiers (*Li et al., 2016*). However, these results should be interpreted with care, since different number of sequences were used to retrieve GO terms.

Over the last decade, understanding of *Mytilus* spp. evolutionary characteristics significantly increased. Several approaches using molecular methods were used, such as molecular markers (*Kijewski et al., 2011*), SNP genotyping (*Gardner et al., 2016*) and mitochondrial DNA (*Smietanka, Burzynski & Wenne, 2009*). In this study phylogenetic relationships were investigated between the six taxa using transcriptome data (Fig. 4). Our analyses showed that *M. edulis*, *M. chilensis*, *M. galloprovincialis* and *M. trossulus* created a separate clade from *M. californianus* and *M. coruscus*, which is consistent with earlier studies of their mitochondrial DNA (*Gaitán-Espitia et al., 2016*).

CONCLUSIONS

This study investigated mantle transcriptomes of five groups from four *Mytilus* spp. using 454 pyrosequencing technology. Of generated transcripts 46.57% were annotated by BLAST search. Of these 6.16% were contigs potentially involved in biomineralization and pigmentation processes, such as CA, CHIT and TYR. GO enrichment analysis revealed stress response in *M. edulis* (from the North Sea) and *M. galloprovincialis* (from the Mediterranean Sea), which might be related to ocean acidification and global warming. To our knowledge, this study is the first to describe mantle transcriptomes from multiple *Mytilus* spp and provide a basis for further work (using RNA-seq and gene expression patterns) on biomineralization proteins, which may help to further account for mussels diversity and responses to environment conditions.

ACKNOWLEDGEMENTS

The authors thank Dr. Tomasz Sanko for preparation of RNA samples for 454 sequencing. This research was supported in part by PL-Grid Infrastructure.

Funding

This research was funded in part by the 2011/01/B/NZ9/04352 NCN project to Professor Roman Wenne and statutory topic IV in the Institute of Oceanology, Polish Academy of Sciences. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
NCN project to Professor Roman Wenne: 2011/01/B/NZ9/04352.
Institute of Oceanology Polish Academy of Sciences.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Magdalena Malachowicz performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Roman Wenne conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

Data Availability

The raw data is available on Sequence Read Archive under accession number PRJNA419475. The CA and CHIT sequences described here are accessible via GenBank accession numbers MG827120–MG827134 and are included in Supplemental_Data_File_1. Other sequences are included in Table S6.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.6245#supplemental-information.

REFERENCES

Aguilera F, McDougall C, Degnan BM. 2014. Evolution of the tyrosinase gene family in bivalve molluscs: independent expansion of the mantle gene repertoire. *Acta Biomaterialia* **10(9)**:3855–3865 DOI 10.1016/j.actbio.2014.03.031.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215(3)**:403–410 DOI 10.1016/S0022-2836(05)80360-2.

Arivalagan J, Marie B, Sleight VA, Clark MS, Berland S, Marie A. 2016. Shell matrix proteins of the clam, *Mya truncata*: Roles beyond shell formation through proteomic study. *Marine Genomics* **27**:69–74 DOI 10.1016/j.margen.2016.03.005.

Aspatwar A, Tolvanen MEE, Parkkila S. 2010. Phylogeny and expression of carbonic anhydrase-related proteins. *BMC Molecular Biology* **11(1)**:25 DOI 10.1186/1471-2199-11-25.

Bjärnmark NA, Yarra T, Churcher AM, Felix RC, Clark MS, Power DM. 2016. Transcriptomics provides insight into *Mytilus galloprovincialis* (Mollusca: Bivalvia) mantle function and its role in biomineralisation. *Marine Genomics* **27**:37–45 DOI 10.1016/j.margen.2016.03.004.

Blank S, Arnoldi M, Khoshnavaz S, Treccani L, Kuntz M, Mann K, Grathwohl G, Fritz M. 2003. The nacre protein perlucin nucleates growth of calcium carbonate crystals. *Journal of Microscopy* **212(3)**:280–291 DOI 10.1111/j.1365-2818.2003.01263.x.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30(15)**:2114–2120 DOI 10.1093/bioinformatics/btu170.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25(15)**:1972–1973 DOI 10.1093/bioinformatics/btp348.

Clark MS, Thorne MAS, Vieira FA, Cardoso JCR, Power DM, Peck LS. 2010. Insights into shell deposition in the Antarctic bivalve *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing. *BMC Genomics* **11(1)**:362 DOI 10.1186/1471-2164-11-362.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21(18)**:3674–3676 DOI 10.1093/bioinformatics/bti610.

De Zwaan A, Eertman RHM. 1996. Anoxic or aerial survival of bivalves and other euryoxic invertebrates as a useful response to environmental stress—a comprehensive review. *Comparative Biochemistry and Physiology—Part C: Toxicology & Pharmacology* **113(2)**:299–312 DOI 10.1016/0742-8413(95)02101-9.

Du L, Li Y, Zhang X, Yue B. 2013. MSDB: a user-friendly program for reporting distribution and building databases of microsatellites from genome sequences. *Journal of Heredity* **104(1)**:154–157 DOI 10.1093/jhered/ess082.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32(5)**:1792–1797 DOI 10.1093/nar/gkh340.

Fang D, Xu G, Hu Y, Pan C, Xie L, Zhang R. 2011. Identification of genes directly involved in shell formation and their functions in pearl oyster, *Pinctada fucata*. *PLOS ONE* **6(7)**:e21860 DOI 10.1371/journal.pone.0021860.

Feng D, Li Q, Yu H, Kong L, Du S. 2017. Identification of conserved proteins from diverse shell matrix proteome in *Crassostrea gigas*: characterization of genetic bases regulating shell formation. *Scientific Reports* **7(1)**:45754 DOI 10.1038/srep45754.

Feng D, Li Q, Yu H, Zhao X, Kong L. 2015. Comparative transcriptome analysis of the pacific oyster *crassostrea gigas* characterized by shell colors: identification of genetic bases potentially involved in pigmentation. *PLOS ONE* **10(12)**:e0145257 DOI 10.1371/journal.pone.0145257.

Fitzer SC, Phoenix VR, Cusack M, Kamenos NA. 2014. Ocean acidification impacts mussel control on biomineralisation. *Scientific Reports* **4(1)**:6218 DOI 10.1038/srep06218.

Freer A, Bridgett S, Jiang JH, Cusack M. 2014. Biomineral proteins from *Mytilus edulis* mantle tissue transcriptome. *Marine Biotechnology* **16(1)**:34–45 DOI 10.1007/s10126-013-9516-1.

Gaitán-Espitia JD, Quintero-Galvis JF, Mesas A, D'Elía G. 2016. Mitogenomics of southern hemisphere blue mussels (Bivalvia: Pteriomorphia): insights into the evolutionary characteristics of the *Mytilus edulis* complex. *Scientific Reports* **6(1)**:26853 DOI 10.1038/srep26853.

Gao P, Liao Z, Wang X-X, Bao LF, Fan MH, Li XM, Wu CW, Xia SW. 2015. Layer-by-Layer Proteomic Analysis of Mytilus galloprovincialis shell. *PLOS ONE* **10(7)**:e0137487 DOI 10.1371/journal.pone.0133913.

Gardner JPA, Zbawicka M, Westfall KM, Wenne R. 2016. Invasive blue mussels threaten regional scale genetic diversity in mainland and remote offshore locations: the need for baseline data and enhanced protection in the Southern Ocean. *Global Change Biology* **22(9)**:3182–3195 DOI 10.1111/gcb.13332.

Gesualdo I, Aniello F, Branno M, Palumbo A. 1997. Molecular cloning of a peroxidase mRNA specifically expressed in the ink gland of *Sepia officinalis*. *Biochimica et Biophysica Acta (BBA)—Gene Structure and Expression* **1353(2)**:111–117 DOI 10.1016/S0167-4781(97)00088-2.

Goffredo S, Prada F, Caroselli E, Capaccioni B, Zaccanti F, Pasquini L, Fantazzini P, Fermani S, Reggi M, Levy O, Fabricius KE, Dubinsky Z, Falini G. 2014. Biomineralization control related to population density under ocean acidification. *Nature Climate Change* **4(7)**:593–597 DOI 10.1038/nclimate2241.

Gosling EM. 2003. *Bivalve molluscs biology, ecology and culture*. Oxford: Blackwell Publishing.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **29(7)**:644–652 DOI 10.1038/nbt.1883.

Henikoff S, Henikoff JG. 1992. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America* **89(22)**:10915–10919 DOI 10.1073/pnas.89.22.10915.

Herlitze I, Marie B, Marin F, Jackson DJ. 2018. Molecular modularity and asymmetry of the molluscan mantle revealed by a gene expression atlas. *Gigascience* **7(6)**:giy056 DOI 10.1093/gigascience/giy056.

Hüning AK, Melzner F, Thomsen J, Gutowska MA, Krämer L, Frickenhaus S, Rosenstiel P, Pörtner H-O, Philipp EER, Lucassen M. 2013. Impacts of seawater acidifcation on mantle gene expression patterns of the Baltic Sea blue mussel: implications for shell formation and energy metabolism. *Marine Biology* **160(8)**:1845–1861 DOI 10.1007/s00227-012-1930-9.

Hüning AK, Lange SM, Ramesh K, Jacob DE, Jackson DJ, Panknin U, Gutowska MA, Philipp EE, Rosenstiel P, Lucassen M, Melzner F. 2016. A shell regeneration assay to identify biomineralization candidate genes in mytilid mussels. *Marine Genomics* **27**:57–67 DOI 10.1016/j.margen.2016.03.011.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23(2)**:254–267 DOI 10.1093/molbev/msj030.

Jiao Y, Wang H, Du X, Zhao X, Wang Q, Huang R, Deng Y. 2012. Dermatopontin, a shell matrix protein gene from pearl oyster *Pinctada martensii*, participates in nacre formation. *Biochemical and Biophysical Research Communications* **425(3)**:679–683 DOI 10.1016/j.bbrc.2012.07.099.

Johnson KM, Hofmann GE. 2016. A transcriptome resource for the Antarctic pteropod *Limacina helicina antarctica*. *Marine Genomics* **28**:25–28 DOI 10.1016/j.margen.2016.04.002.
Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8(3)**:275–282 DOI 10.1093/bioinformatics/8.3.275.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods* **14(6)**:587–589 DOI 10.1038/nmeth.4285.

Kijewski T, Smietanka B, Zbawicka M, Gosling E, Hummel H, Wenne R. 2011. Distribution of *Mytilus* taxa in European coastal areas as inferred from molecular markers. *Journal of Sea Research* **65(2)**:224–234 DOI 10.1016/j.seares.2010.10.004.

Kijewski T, Wijsman JWM, Hummel H, Wenne R. 2009. Genetic composition of cultured and wild mussels *Mytilus* from The Netherlands and transfers from Ireland and Great Britain. *Aquaculture* **287(3–4)**:292–296 DOI 10.1016/j.aquaculture.2008.10.048.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35(6)**:1547–1549 DOI 10.1093/molbev/msy096.

Larraín MA, Zbawicka M, Araneda C, Gardner JPA, Wenne R. 2017. Native and invasive taxa on the Pacific coast of South America: Impacts on biodiversity and aquaculture of blue mussels (*Mytilus* spp.). *Evolutionary Applications* **11(3)**:298–311 DOI 10.1111/eva.12553.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Molecular Biology and Evolution* **25(7)**:1307–1320 DOI 10.1093/molbev/msn067.

Le Roy N, Jackson DJ, Marie B, Ramos-Silva P, Marin F. 2014. The evolution of metazoan α-carbonic anhydrases and their roles in calcium carbonate biomineralization. *Frontiers in Zoology* **11(1)**:75–91 DOI 10.1186/s12983-014-0075-8.

Letunic I, Doerks T, Bork P. 2015. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Research* **43(D1)**:D257–D260 DOI 10.1093/nar/gku949.

Li S, Liu C, Huang J, Liu Y, Zhang S, Zheng G, Xie L, Zhang R. 2016. Transcriptome and biomineralization responses of the pearl oyster *Pinctada fucata* to elevated CO_2 and temperature. *Scientific Reports* **6(1)**:18943 DOI 10.1038/srep18943.

Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **13(9)**:2178–2189 DOI 10.1101/gr.1224503.

Li H, Zhang B, Fan S, Liu B, Su J, Yu D. 2017a. Identification and differential expression of biomineralization genes in the mantle of pearl oyster *Pinctada fucata*. *Marine Biotechnology* **19(3)**:266–276 DOI 10.1007/s10126-017-9748-6.

Li J, Zhang Y, Mao F, Tong Y, Liu Y, Zhang Y, Yu Z. 2017b. Characterization and identification of differentially expressed genes involved in thermal adaptation of the Hong Kong Oyster *Crassostrea hongkongensis* by digital gene expression profiling. *Frontiers in Marine Science* **4**:112 DOI 10.3389/fmars.2017.00112.

Liao Z, Bao L-F, Fan M-H, Gao P, Wang X-X, Qin C-L, Li X-M. 2015. In-depth proteomic analysis of nacre, prism, and myostracum of *Mytilus* shell. *Journal of proteomics* **122**:26–40 DOI 10.1016/j.jprot.2015.03.027.

Liu J, Yang D, Liu S, Li S, Xu G, Zheng G, Xie L, Zhang R. 2015. Microarray: a global analysis of biomineralization-related gene expression profiles during larval development in the pearl oyster, *Pinctada fucata*. *BMC Genomics* **16(1)**:325 DOI 10.1186/s12864-015-1524-2.

Lockwood BL, Connor KM, Gracey AY. 2015. The environmentally tuned transcriptomes of *Mytilus* mussels. *Journal of Experimental Biology* **218(12)**:1822–1833 DOI 10.1242/jeb.118190.

Lowenstam H, Weiner S, Veis A. 1990. On Biomineralization. *Science* **247**:1129–1130.

Malachowicz M, Kijewska A, Wenne R. 2015. Transcriptome analysis of gill tissue of Atlantic cod *Gadus morhua* L. from the Baltic Sea. *Marine Genomics* **27**:37–40 DOI 10.1016/j.margen.2015.04.005.

Malachowicz M, Wenne R, Burzynski A. 2017. De novo assembly of the sea trout (*Salmo trutta* m. *trutta*) skin transcriptome to identify putative genes involved in the immune response and epidermal mucus secretion. *PLOS ONE* **12(2)**:e0172282 DOI 10.1371/journal.pone.0172282.

Mann K, Jackson DJ. 2014. Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. *BMC Genomics* **15(1)**:249 DOI 10.1186/1471-2164-15-249.

Marin F, Luquet G, Marie B, Medakovic D. 2008. Molluscan shell proteins: primary structure, origin, and evolution. *Current Topics in Developmental Biology* **80**:209–276 DOI 10.1016/S0070-2153(07)80006-8.

Min XJ, Butler G, Storms R, Tsang A. 2005. OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research* **33(Web Server)**:W677–W680 DOI 10.1093/nar/gki394.

Moreira R, Pereiro P, Canchaya C, Posada D, Figueras A, Novoa B. 2015. RNA-Seq in *Mytilus galloprovincialis*: comparative transcriptomics and expression profiles among different tissues. *BMC Genomics* **16(1)**:728 DOI 10.1186/s12864-015-1817-5.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **35(Web Server)**:W182–W185 DOI 10.1093/nar/gkm321.

Moschino V, Bressan M, Cavaleri L, Da Ros L. 2015. Shell-shape and morphometric variability in *Mytilus galloprovincialis* from micro-tidal environments: responses to different hydrodynamic drivers. *Marine Ecology* **36(4)**:1440–1453 DOI 10.1111/maec.12244.

Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* **32(1)**:268–274 DOI 10.1093/molbev/msu300.

Ramos-Silva P, Kaandorp J, Huisman L, Marie B, Zanella-Cléon I, Guichard N, Miller DJ, Marin F. 2013. The skeletal proteome of the coral *Acropora millepora*: the evolution of calcification by co-option and domain shuffling. *Molecular Biology and Evolution* **30(9)**:2099–2112 DOI 10.1093/molbev/mst109.

Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, Loire E, Lourenco JM, Nabholz B, Roux C, Tsagkogeorga G, Weber AA, Weinert LA, Belkhir K, Bierne N, Glémin S, Galtier N. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515(7526)**:261–263 DOI 10.1038/nature13685.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61(3)**:539–542 DOI 10.1093/sysbio/sys029.

Shi M, Lin Y, Xu G, Xie L, Hu X, Bao Z, Zhang R. 2013. Characterization of the Zhikong scallop (*Chlamys farreri*) mantle transcriptome and identification of biomineralization-related genes. *Marine Biotechnology* **15(6)**:706–715 DOI 10.1007/s10126-013-9517-0.

Sleight VA, Thorne MAS, Peck LS, Arivalagan J, Berland S, Marie A, Clark MS. 2016. Characterisation of the mantle transcriptome and biomineralisation genes in the blunt-gaper clam, *Mya truncata*. *Marine Genomics* **27**:47–55 DOI 10.1016/j.margen.2016.01.003.

Smietanka B, Burzynski A, Wenne R. 2009. Molecular population genetics of male and female mitochondrial genomes in European mussels *Mytilus*. *Marine Biology* **156(5)**:913–925 DOI 10.1007/s00227-009-1137-x.

Smietanka B, Zbawicka M, Sanko T, Wenne R, Burzynski A. 2013. Molecular population genetics of male and female mitochondrial genomes in subarctic *Mytilus trossulus*. *Marine Biology* **160(7)**:1709–1721 DOI 10.1007/s00227-013-2223-7.

Steffani CN, Branch GM. 2003. Growth rate, condition, and shell shape of *Mytilus galloprovincialis*: responses to wave exposure. *Marine Ecology Progress Series* **246**:197–209 DOI 10.3354/meps246197.

Sun X, Yang A, Wu B, Zhou L, Liu Z. 2015. Characterization of the mantle transcriptome of yesso scallop (*Patinopecten yessoensis*): identification of genes potentially involved in biomineralization and pigmentation. *PLOS ONE* **10(4)**:e0122967 DOI 10.1371/journal.pone.0122967.

Telesca L, Michalek K, Sanders T, Peck LS, Thyrring J, Harper EM. 2018. Blue mussel shell shape plasticity and natural environments: a quantitative approach. *Scientific Reports* **8(1)**:2865 DOI 10.1038/s41598-018-20122-9.

Wang X, Li L, Zhu Y, Du Y, Song X, Chen Y, Huang R, Que H, Fang X, Zhang G. 2013. Oyster shell proteins originate from multiple organs and their probable transport pathway to the shell formation front. *PLOS ONE* **8(6)**:e66522 DOI 10.1371/journal.pone.0066522.

Wenne R, Bach L, Zbawicka M, Strand MJ, McDonald JH. 2016. A first report on coexistence and hybridization of *Mytilus trossulus* and *M. edulis* mussels in Greenland. *Polar Biology* **39(2)**:343–355 DOI 10.1007/s00300-015-1785-x.

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution* **18(5)**:691–699 DOI 10.1093/oxfordjournals.molbev.a003851.

Vendrami DL, Shah A, Telesca L, Hoffman JI. 2016. Mining the transcriptomes of four commercially important shellfish species for single nucleotide polymorphisms within biomineralization genes. *Marine Genomics* **27**:17–23 DOI 10.1016/j.margen.2015.12.009.

Vidal R, Peñaloza C, Urzúa R, Toro JE. 2009. Screening of ESTs from *Mytilus* for the detection of SSR markers in *Mytilus californianus*. *Molecular Ecology Resources* **9(5)**:1409–1411 DOI 10.1111/j.1755-0998.2009.02682.x.

Xu M, Jiang L, Shen K-N, Wu CH, He G, Hsiaod CH-D. 2016. Transcriptome response to copper heavy metal stress in hard-shelled mussel (*Mytilus coruscus*). *Genomics Data* **7**:152–154.

Yarra T, Gharbi K, Blaxter M, Peck LS, Clark MS. 2016. Characterization of the mantle transcriptome in bivalves: *Pecten maximus*, *Mytilus edulis* and *Crassostrea gigas*. *Marine Genomics* **27**:9–15 DOI 10.1016/j.margen.2016.04.003.

Zbawicka M, Drywa A, Smietanka B, Wenne R. 2012. Identification and validation of novel SNP markers in European populations of marine *Mytilus* mussels. *Marine Biology* **159(6)**:1347–1362 DOI 10.1007/s00227-012-1915-8.

Zbawicka M, Trucco MI, Wenne R. 2018. Single nucleotide polymorphisms in native South American Atlantic coast populations of smooth shelled mussels: hybridisation with invasive European *Mytilus galloprovincialis*. *Genetics Selection Evolution* **50(1)**:5 DOI 10.1186/s12711-018-0376-z.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PW, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A, Hedgecock D, Xu Z, Liu Y, Domazet-Lošo T, Du Y, Sun X, Zhang S, Liu B, Cheng P, Jiang X, Li J, Fan D, Wang W,

Fu W, Wang T, Wang B, Zhang J, Peng Z, Li Y, Li N, Wang J, Chen M, He Y, Tan F, Song X, Zheng Q, Huang R, Yang H, Du X, Chen L, Yang M, Gaffney PM, Wang S, Luo L, She Z, Ming Y, Huang W, Zhang S, Huang B, Zhang Y, Qu T, Ni P, Miao G, Wang J, Wang Q, Steinberg CE, Wang H, Li N, Qian L, Zhang G, Li Y, Yang H, Liu X, Wang J, Yin Y, Wang J. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490(7418)**:49–54 DOI 10.1038/nature11413.