Transcriptomic profiling of the mussel Mytilus trossulus with a special emphasis on integrin-like genes during development

doi: info:doi/10.1080/07924259.2019.1626774

Transcriptomic profiling of the mussel *Mytilus trossulus* with a special emphasis on integrin-like genes during development

Mariia A. Maiorova[a*], Noriyuki Satoh[b], Konstantin Khalturin[b], Nelly A. Odintsova[a,c]

[a]*Laboratory of Cytotechnology, National Scientific Center of Marine Biology, The Far Eastern Branch of the Russian Academy of Sciences, Vladivostok 690041, Russia;*

[b]*Marine Genomcs Unit, Okinawa Institute Science & Technology, Okinawa 9040495, Japan;*

[c] *Far Eastern Federal University, Vladivostok 690922, Russia.*

***Corresponding author:** Mariia A. Maiorova

National Scientific Center of Marine Biology, FEB RAS, Russia

Tel: (+7) 423-231-1161

Fax: (+7) 423-231-0905

E-mail: maiorovamariya@gmail.com

ORCID:

Mariia Maiorova http://orcid.org/0000-0001-5175-3641

Noriyuki Satoh http://orcid.org/0000-0003-4560-9250

Konstantin Khalturin http://orcid.org/ 0000-0003-4359-2993

Nelly Odintsova http://orcid.org/0000-0003-4041-6409

Abstract

This study is based on the Illumina RNA-sequencing data obtained for a *de novo* assembly of the transcriptome from early developmental stages and some tissues and cells of the adult mussel *Mytilus trossulus* (Mytilidae, Mollusca) using Trinity program. A total of 200079 contigs were obtained, and compared to the NCBI database using BLAST to search for sequence similarity. The number of annotated contigs under the GO term 3 categories was estimated to reach 19.96%. The BUSCO analysis determined a level of 99.2% completeness for the assembled transcriptome. The main findings include evidence that the mussel β integrin-like protein sequences are high similar to the β integrin-like proteins so far sequenced for all classes of Mollusca, while the highest similarity is observed between mussel and oyster (*Crasostrea gigas*) β integrin-like proteins. Our transcriptome dataset contributes to the genetic databases of non-model animals such as Bivalves and represents the first characterization of expressed sequences during early development of the mollusk *M. trossulus* from the Sea of Japan including the identification of candidate genes involved in cell adhesion.

Keywords: integrin-like proteins, cell adhesion, Mollusca, next-generation sequencing, stage-specific expression

Introduction

When cells began to organize into multicellular organisms, the appearance of extracellular matrix (ECM) is an important stage in this process (Wainright et al., 1993). A prerequisite for the emergence of long-lived multicellular organisms was the evolution of intercellular adhesion mechanisms (Abedin, King, 2010). Unicellular ancestors of Metazoa had a variety of protein components of cell adhesion complexes and receptor-type signalling molecules, such as integrins and cadherins, involved in metazoan multicellularity and development (Suga et al., 2013). Sequencing of the genome of one of the most primitive multicellular organisms, the sponge *Amphimedon queenslandica,* has revealed a set of genes associated with the emergence of metazoan multicellularity, among them genes involved in cell adhesion (Srivastava et al., 2010). In lower metazoans, Placozoans, genome includes genes for many proteins of the ECM found in the basement membranes of other animals, despite the lack of a basement membrane in Placozoans (Srivastava et al., 2008). In this study, we present an initial survey of candidate genes involved in cell adhesion, such as genes encoding integrin-like proteins, during early development of the bivalve mollusk *Mytilus trossulus,* because the integrin gene families that are particularly interesting in the context of the evolution of multicellularity. Unfortunately, the evolutionary history of integrins is largely unclear (Humphries, 2000; Hughes, 2001). It is known that one or more alpha integrin-subunits associate with specific beta integrin subunits, and □ and □ integrin subunits are encoded by evolutionarily unrelated gene families (Hynes, 1992).There is evidence of the divergence of alpha integrin-subunits until the separation of Protostomes from Deuterostomes (Hynes, 2012). Alpha integrin-subunit in mussels is located near alpha integrin-subunits of sponges, cnidarians, nematodes, arthropods, echinoderms and ascidians (Miyazawa et al., 2001; Takada et al., 2007). A phylogenetic tree constructed for these alpha subunits also indicates their ancestral position (Miyazawa et al., 2001).

The genome-sequencing data of the snail *Biomphalaria glabrata* and the Pacific oyster *Crassostrea gigas* confirmed the presence of genes encoding integrin-like proteins in mollusks

(Lockyer et al., 2007; Zhang et al., 2012). Moreover, shell formation in oysters was found to depend on genes associated with the ECM (Zhang et al., 2012). To date, *Mytilus* transcriptome sequencing has been reported only for adult *M. edulis* and *M. galloprovincialis* (Venier et al.,2003; Craft et al., 2010; Philipp et al., 2012; Tanguy et al., 2013; Toubiana et al., 2014; Moreira et al., 2015). One study deals with the transcription profiles of molluscan larvae (Bassim et al., 2014), where the active transcription of *M. edulis* genes encoding proteins of the ECM has been revealed.

The integrin homologs have been previously identified in some marine invertebrates from corals to mollusks (Burke, 1999; Grasso et al., 2008; Knack et al., 2008; Jia et al., 2015) and are highly conserved in the Metazoa kingdom (Burke 1999; Takada et al. 2007). The integrin β1 subunit is likely to be the common ancestor for all β integrin subunits in vertebrates (Hughes, 2001; Ewan et al., 2005). It should be borne in mind that, similar to integrins in vertebrates, studying the integrin-like proteins in invertebrates shows that they play an important role in developmental processes (Burke, 1999; Zhang et al., 2012; Bassim et al., 2014; Maiorova, Odintsova, 2015). Nevertheless, there is a few information about the participation of integrin-like proteins in the development of mollusks (Dyachuk et al., 2015).

Here, a set of positively selected genes related to integrin complex was identified in the transcription profiles of one of the representatives from the lophotrochozoan clade, the mussel *M. trossulus*. We have also conducted the quantitative expression analysis of mussel integrin homologs during early development. Additionally, an analysis of the mussel transcriptome revealed four novel full-length sequences, orthologous to mammalian transcripts, that seemed to be isoforms of two genes encoding β integrin-like proteins. The *β-A* transcript had isoforms *β-A1* and *β-A2* that differed by the insertion of 24 base pairs (bps) in a protein-coding region of the extracellular domain; the *β-B* transcript also had two isoforms (*β-B1* and *β-B2*) that differed by a small insertion in a non-coding region. We compared our results with the well-known sequences of β integrin-like proteins from other animals and found that the predicted β integrin-like proteins of mussel were most closely related to the integrins of mollusks (two resembling the

oyster β1-integrin and two more similar to the oyster β3-integrin (Zhang et al., 2012)). There are a few molecular studies for deep sea mussels but all of them are based only on mitochondrial or nuclear genes (Distel et al., 2000; Owada, 2007; Samadi et al., 2007; Liu et al., 2018), thus, we are forced to compare sequencing information with oyster genome. The present study provides a transcriptome that can serve as a reference for future studies of functional development of this important bivalve group in the marine ecosystems.

Methods

Marine farmed mussels (*M. trossulus,* Bivalvia: Mytilidae) used as a model were collected from the Vostok Bay of the Sea of Japan and stored in tubs filled with aerated running seawater (SW) at 5–10°C before performing the experiments. The spawning of sexually mature specimens was induced by thermal shock, as previously described (Dyachuk, Odintsova, 2009). Our studies did not involve endangered or protected species.

Developing embryos were cultivated in 5-L tanks at 17°C, harvested at different developmental stages, including fertilized eggs, cleaving embryos, and larvae, up to the early veliger (Odintsova et al., 2010). We have collected the larvae after the hatching at the stage of swimming blastula (11 h after fertilization) at the mesh gauze (35 μm), then we have added SW in the tank. In due time, the trochophore larvae were collected at the new mesh gauze, and for following development (up to the early veliger) the trohophore larvae were transferred to a new tank with the fresh SW. Also, some cells and tissues (hemocytes, adductor, gills, digestive gland, mantle, and testis) of adult mollusks were used for constructing sequencing libraries. Total RNAs from the mussel embryonic or larval suspensions (about 0.05–0.1 ml of 300000 embryos or larvae) or cells and tissues of adult mollusks (50–100 mg) were lysed in 15–20 volumes of TRIzol Reagent (Thermo Fisher Scientific, USA), briefly frozen in liquid nitrogen, and stored at -80°C until further analysis.

RNA isolation

Total RNAs were extracted using the RNeasy Mini kit (Qiagen, USA). Contaminating DNA was removed with DNAase I (Fermentas, USA). RNA concentration was determined using a NanoDrop spectrophotometer (Thermo Scientific, USA). The RNA quality was checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA) according to the manufacturer's protocol.

Preparation of the mRNA fragment library and Sequencing

Sequencing libraries were prepared *using* the reagents provided in the *Illumina*® TruSeq® RNA *Sample Preparation Kit* based on 5 µg of total RNA in each sample. Libraries were quantified by real-time PCR using the KAPA Library Quantification Kit for the Illumina platform (Kapa Biosystems, USA). The DNA fragment size was verified using the Agilent 2100 Bioanalyzer DNA High Sensitivity Assay (Agilent Technologies, USA). The average fragment size was within the range from 360 to 490 bps. The libraries were sequenced using the information obtaining from two sequencers MiSeq (Illumina, USA) and HiSeq2500 (Illumina) in the Marine Genomics Unit (OIST, Japan) (see Supplementary table S1 for details of libraries and sample sets used for the different transcriptome reads). Longer sequence reads (2x280 bp) obtained with MiSeq were used for transcriptome assembly and for preliminary gene expression estimation. Short sequence reads (2x134 bp) obtained with HiSeq had higher coverage and were used for expression estimation of all genes.

The assembly of de novo transcriptome of M. trossulus

Raw reads were processed for adapter sequences removal and trimming using the Trimmomatic tool (Bolger et al*.,* 2014). Sequencing fragments quality was monitored using the FAST-QC program (www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). After removing ambiguous nucleotides and low-quality reads (quality scores <20), Illumina Miseq high-quality raw reads with Phred score ≥20 and length ≥ 36 bps were kept for further assembly and quantitative expression analysis. The assembly of *de novo* transcriptome was performed using the Trinity program (version r20140413p1) (Haas et al., 2013). A transcriptome assembly quality was accessed using QUAST (Gurevich et al., 2013). Transcriptome completeness was assessed

using the Benchmarking Universal Single-Copy Ortholog (BUSCO) assessment tool (Simão et al., 2015), software version 3.1.0 with default setting. The metazoa_odb9 dataset consisting of 978 single-copy genes was used as a reference.

Sequence Annotation

For the annotation steps, we searched for potential contaminant sequences in our assembly. We looked for sequences with bacterial signatures using CLARK-*l* (Ounit et al., 2015) and $BLAST_X$. Protein coding sequences of transcripts were analyzed via Transdecoder. The standard Trinotate annotation pipeline (https://trinotate.github.io/) was performed to annotate the assembled transcripts against the UniProt, NR, Pfam, and GO databases using the BLASTx program.

Quantitative gene expression

To assess the level of gene expression in the data of massive parallel sequencing between all tested early developmental stages, we used alignment readings to the received transcripts by the RSEM algorithm, and then analyzed values of FPKM (fragments per kilobase of transcript length per million mapped reads). For the quantitative expression analysis, the high-quality reads from each sample were mapped.

Availability of Supporting Data

Illumina read sequences from this study were deposited in the NCBI's SRA database (Sequence Read Archive) under the accession number SRP 137045. Assembled sequences have been submitted to the NCBI under the submission code SUB (3854015).

Amino acid sequence analysis of β-integrin-like proteins

A search for integrin homologs was performed using the algorithms BLASTp and BLASTx in the NCBI NR and UniProt/TrEMBLE databases. We compared the amino acid sequences of suspected mussel β- integrin-like proteins with the known sequences of β-integrins from some other organisms from different taxonomic groups using the ClustalW algorithm (see Supplementary Figure, S2). Then, sequence alignments were manually corrected by the provisions of 56 conserved cysteine residues using the program MEGA 7 (Kumar et al., 2016),

alignment was visualized using JalView (Waterhouse et al., 2009). Based on the alignment data, a phylogenetic analysis of the sequences was conducted using the method of maximum likelihood (ML) and the method of nearest neighbors (NJ) in the program MEGA 7 (Kumar et al., 2016). As the best model of the test data set, the WAG model was chosen.

Results

The *M. trossulus* transcriptome analysis included almost 40 million assembled raw RNA-Seq reads with the median length of 392 bp. Total reads were compiled into 200079 contigs ranging from 500–31400 bp, with an average length of contig of 728 bp. Sixteen sequencing libraries were constructed. Specific details related to the *M. trossulus* assembly statistics and annotation are included in Table 1 (also, see Supplementary table S1). The BUSCO analysis determined a level of 99.2% completeness for the assembled transcriptome. There are only two missing orthologs and six fragmented orthologs from the metazoa_odb9 datase in the *M. trossulus* transcriptome.

Using the NR database, the number of annotated sequences was estimated to reach 29.24%. Contigs were annotated using the Trinotate annotation pipeline (version r20140708, www.trinotate.github.io) with an E-value threshold of 1×10^{-5}. This pipeline assesses BLAST homologies between the assembly and SwissProt using BLASTx of the contig sequence, and BLASTp of TransDecoder Predicted Proteins. Among 200079 transcripts, 39934 (19.96%) were classified under the GO term three categories: biological process, molecular function, and cellular component (Fig. 1 A). Based on selected GO terms, related to integrins and extracellular matrix, the analysis was performed at the second level for all transcripts of three main categories, and the number of transcripts was shown on top of each column (Fig. 1 B).

A Venn diagram demonstrates the distribution of transcriptional active contigs during early development (Fig. 2). In the development process, the number of active transcripts increased in comparison with that at the early stages, reaching maximum values at the trochophore and veliger stages (about 70000).

To determine primary expression trends in the mussel transcriptome profile, we compared the gene expression patterns of each transcript of proteins involved in cell adhesion (□ and □ integrin-like proteins) between all tested developmental stages and found on a heat map that significant differences occurred in the expression of genes encoding integrin-like proteins during development (the expression level of these genes is marked by a color key) (Fig. 3). Expression of the *β-A1* gene (c163688_g2_i1) was decreased when comparing egg to the following stages, indicating the reduction of maternal expression of over 6–9-fold, whereas expression of the *β-A2* gene (c163688_g2_i2) began to increase gradually from the early embryonic stages (blastula stage) to early larval stages (veliger stage). In contrast, we observed low levels in the expression of both *β-B* genes (c164133_g1_i1 and c164133_g1_i2) during early mussel development (Fig. 3). The level of expression of α integrin-like genes was very low (in comparison of that of some β integrin-like genes) showing the absence of maternal α integrin-like mRNAs, with the exception of only one α integrin-like gene, c168669_g1_i1 (Fig. 3). At each time point, the top (annotated) differentially expressed contigs indicated different biological processes were likely to be important at different time. Specifically, at the trochophore stage, mussle development, cytoskeleton processes and mitosis were the dominant differential processes (Dyachuk et al., 2015; Maiorova, Odintsova, 2015). Later, at the veliger stage, cell adhesion, DNA repair, immune response and RNA processing were most prominent (this study). Qualitative analysis of differentially expressed transcripts over the time course showed the highest number occurred at the veliger stage (70057 contigs), whereas the minimal number of transcripts occurred at the blastula stage (40583 contigs).

Database comparisons revealed that the mussel β-integrin-like proteins belong to different sub-groups (Table 2). Mussel *β-A*-integrin-like proteins were identified as possible orthologs of integrin β1 from of the oyster *C. gigas*, whereas mussel *β-B*-integrin-like proteins were most closely related to oyster β3-integrins. Among identified contigs, four full-length transcripts (2834, 2858, 3534 and 3649 bps) corresponding β-integrin-like proteins were found in embryos and larvae at the different developmental stages. All hallmarks of β-integrins are conserved in all

sequences of the mussel β-integrin-like proteins: the cysteine-rich stalk of the extracellular domain containing 56 conserved cysteine residues, transmembrane and cytoplasmic domains (see Supplementary figure S2).

To assess the divergence of β-integrin-like subunits in Protostomes and Deuterostomes, a cladogram was constructed, showing the relationships among β integrin-like proteins of various animals (Fig. 4). The difference of the topologies of phylogenetic trees of amino acid sequences of β-integrin-like proteins of various animals constructed from the method of ML and the method of NJ was minimal; thus, we used the ML method. β-integrins of Protostomes clustered into one clade in which a plurality of sub-groups can be distinguished. The sequence of mussel *β-A* integrin-like proteins showed a high degree of similarity to the sequence of oyster β1-integrin (XP_011419533.1) and the sequences of β integrin-like proteins in Gastropods, *Lottia gigantea* (V4AU86_LOTGI), and *B. glabrata* (O96444_BIOGL). β integrin-like proteins of other representatives of Lophotrochozoa, such as the lingulid brachiopod *Lingula anatina* (XP_013415842.1), the polychaete *Capitella teleata* (X2BBV3_CAPTE), and the freshwater leech *Helobdella robusta* (tr|T1EIY7|T1EIY7_HELRO) were situated in the same group. The sequence of β integrin-like protein β-B of the mussel *M. trossulus* (c164133_g1) together with β3-integrin sequences from the pearl oyster *Pinctada fucata* (G9JKY4_PINFU) and the Pacific oyster *C. gigas* (XP_011453738.1) formed a more distantly related group to the mussel β-A integrin-like proteins (c163688_g2). It resulted in high bootstrap support values.

Amino acid sequence analysis of α-integrin-like proteins revealed the presence of some conservative domains, which are characteristic of integrins: integrin α2 superfamily domain and beta-propellor repeats.

The invertebrate species distribution of the top BLAST hits is given in Supplementary figure S3. Only species appearing in more than 0.5% of the BLAST hits are represented.

Discussion

Today, the study of newly discovered protein begins by searching for previously characterized proteins that have similar amino acid sequences. These homologous proteins can be identified in different organisms, despite the large phylogenetic distances.

More than 2000 mussel genes have two isoforms, and near 1000 mussel genes have five isoforms (Gerdol et al., 2015). This is not surprising as mussel populations have been previously reported to be characterized by an extremely high level of heterozygosity (Skibinski et al., 1980; Koehn, Gaffney, 1984; Mosquera et al., 2003). The mussel genome is presently estimated to contain 25000 genes (Murgarella et al., 2016), which is close in the size to the oyster genome (29353 genes) (Takeuchi et al., 2012; Takeuchi et al., 2016). This fact coincides with the data of other authors for marine molluscan genomes (Yoshida et al., 2011).

Typical of β integrin-like proteins consist of three different domains, which include the highly conserved extracellular domain. The presence of 56 conserved cysteine residues in the extracellular domain of all tested β integrin-like proteins (Hynes, 2012; our data) confirms the well-known fact that disulfide bonds in polypeptide chains of different proteins are mainly located on the outside of the membrane, providing the native structure of polypeptides and being critical for interactions between polypeptide chains (Bretscher, 1973). *β-A* transcripts corresponding predicted β integrin-like proteins appear to be the products of alternative splicing of the same gene *β-A*, whereas *β-B* transcripts resembling the oyster β3-integrin seem to be a result of the allelic diversity of the gene *β-B*.

β-integrin subunits in Spongia and Corals have been supported to be formed independently of each other, and a division into the integrin β-subunit classes in vertebrates occurred late in Evolution – likely in the Deuterostome subtaxon line only, and perhaps only within Chordates (Brower et al., 1997; Satoh, 2016). In contrast, other analyses resolve the divergence of β integrin-like subunits in invertebrates as occurs independently from the ancestral form in several lines of bilateral animals, and there are no orthologs of vertebrate β-integrins (Burke, 1999; Knack et al., 2008). Nevertheless, β integrin-like proteins of *Drosophila,* the nematode *Caenorhabditis elegans,* and the sea urchin *Strongylocentrotus purpuratus* show a

high degree of sequence similarity to the β-integrin subunit in vertebrates, and β integrin-like proteins of these invertebrates are also involved in the development processes (Brabant, Brower, 1993; Gettner et al., 1995; Marsden, Burke, 1998).

To understand variations in differential gene expression patterns between larvae and adults, transcriptional activity has been studied in just a few animals, including some marine invertebrates (Woods et al., 2004; Azumi et al., 2007; Williams et al., 2009; Conaco et al., 2012). Our results have also demonstrated various expression patterns of genes associated with cell adhesion and encoded integrin-like proteins throughout early development in Bivalves, both α and β integrin-like subunits. In this study we examined the expression levels of some transcripts of α integrin-like subunits (of about 14) and only four of β integrin-like subunits. Most of α integrin chains were expressed at a relatively low level during early mussel development, whereas the expression levels of some β integrin-like subunits were found to be significantly higher. A low expression level of α integrin-like genes shows the absence of maternal α integrin-like mRNAs (with the exception of transcript c168669_g2_i1) and a high expression level of β integrin-like genes shows of the presence of maternal β integrin-like mRNAs. We assume that these are, mainly, *β-A1* transcripts (c163688_g2_i1) (based on the results presented in Fig. 3). The previously obtained data on the presence of maternal integrin mRNAs was reported for hydroid polyps (Reber-Müller et al., 2001; Knack et al., 2008), as well as for higher animals. They indicate the importance of maternally programmed behavior of some cell adhesion molecules.

Genes participating in signal transduction and stimulus response, related to defense pathways, have been previously reported in the oyster genome (Zhang et al., 2012). Their active expression was observed both in adult mollusks and during oyster development. In pathogen recognition and elimination of gram-negative bacteria, an integrin from the oyster *C. gigas* has been showed to enhance the phagocytosis of oyster hemocytes. The full-length cDNA of this oyster integrin was 2571 bp (Jia et al., 2015), that is close in the size to two full-length integrin-like transcripts which we detected in the genome of the mussel larvae (2834 and 2858 bp). In

adult mussels, only genes associated with multixenobiotic resistance, glutamate biosynthetic process, and the maintenance of ciliary structures were identified (Murgarella et al., 2016). Moreover, it was previously reported that antioxidant enzymes and antimicrobial peptides were discovered in adult bivalves: *M. edulis* (Philipp et al., 2012; Tanguy et al., 2013) and *C. gigas (*Gonzalez et al., 2007). In larval *M. edulis,* genes encoding proteins of the ECM have been additionally revealed (Bassim et al., 2014). These authors have found an end of the development of larval structures and the beginning of the exponential increase of adult structures already in post-metamorphic pediveligers.

The matrix completeness of Illumina data was shown to be superior to that of other data (Smith et al., 2011) and has increased our basic knowledge of the genomes of Bivalves. Using the next-generation sequencing techniques and the Trinity program, we generated a large *M. trossulus* transcriptome database. The protein domains identified may provide more relevant information of cell adhesive function. Our findings suggested that the predicted *β* integrin-like genes have different functions during the mussel development.

Acknowledgments

The work presented in this paper was carried out during M. Maiorova's Ph.D. research at the Okinawa Institute Science & Technology (Japan) and in the National Scientific center of Marine Biology, Far Eastern Branch of the Russian Academy of Sciences (Russia). This research was supported through computational resources provided by the Shared Facility Center 'Data Center of FEB RAS' (Khabarovsk, Russia). Financial support was partially provided by the Russian Foundation for Basic Research (grant no. 18-34-00064/18). The specimens were sequenced in the Marine Genomics Unit (OIST, Okinawa, Japan). The authors are grateful to the staff of OIST, who helped in sequencing samples for their technical support.

Disclosure statement

The authors declare that they have no competing interests.

Author Contributions

M.M., N.O. and N.S. designed and coordinated the research; M.M. and N.O. took part in the experiments with mussel embryos and larvae; M.M. and K.K. performed the Illumina RNA-sequencing in the Marine Genomics Unit (OIST, Japan); M.M. analyzed the obtained data and conducted a bioinformatics' analysis. All authors read and approved the final manuscript.

Figure legends:

Figure 1. Functional annotation of assembled contigs in the mussel *M. trossulus* based on gene ontology (GO) categorization. GO analysis was performed at the second level for three main categories (cellular component, molecular function and biological process) (A) and for selected GO terms, related to integrins and extracellular matrix, in all three categories (B). The number of transcripts is shown on top of each column.

Figure 2. Venn diagram showing the overlap of transcriptionally active contigs (whose FPKMs ≥1 at least in one from the libraries) comparing different early developmental stages in the mussel *Mytilus trossulus*.

Figure 3. Heat map of quantitative gene expression of α and β integrin-like transcripts during mussel development (in eggs, zygotes, embryos and early larvae). The level of expression is marked by a color key. Data obtained using HiSeq Sequencing.

Figure 4. Cladogram showing the relationships among β integrin-like proteins in Protostomes and Deuterostomes. Amino acid sequences were aligned, and a phylogenetic tree was made by the ML method using amino acid sequences for β integrin-like proteins of the species identified. The bootstrap values above 50% are marked on the tree branches. The branches tested in branch-sites tests of selection for β integrin-like proteins are indicated.

Supplementary material

Supplementary Table S1. Details of the assembled libraries produced by the different sequencing runs. Quantitative evaluation of the data obtained MiSeq/HiSeq sequencing.

Supplementary figure S2. Amino acid sequences of mussel β integrin-like proteins (β A and β B) aligned with representative β integrin sequences: Human β1 (HsItgβ1; P05556); *Drosophila melanogaster* β-PS (P11584/ITBX_DROME), oyster *C. gigas* β1-integrin (XP_011419533.1); oyster *C. gigas* β3-integrin (XP_011453738.1).

Supplementary figure S3. Invertebrate species distribution of the top BLAST hits. Only species appearing in more than 0.5% of the BLAST hits are represented.

References

Abedin M, King N. 2010. Diverse evolutionary paths to cell adhesion. Trends Cell Biology. 20 (12):734–742.

Azumi K, Sabau SV, Fujie M, Usami T, Koyanagi R, Kawashima T, Fujiwara S, Ogasawara M, Satake M, Nonaka M, Wang HG, Satou Y, Satoh N. 2007. Gene expression profile during the life cycle of the urochordate *Ciona intestinalis*. Developmental Biology. 308(2):572–582.

Bassim S, Tanguy A, Genard B, Moraga D, Tremblay R. 2014. Identification of *Mytilus edulis* genetic regulators during early development. Gene. 551:65–78.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Brabant MC, Brower DL. 1993. PS2 integrin requirements in *Drosophila* embryo and wing morphogenesis. Developmental Biology. 157(1):49–59.

Bretscher *MS. 1973. Membrane structure*: some general principles. *Science*. 181(4100):622–629.

Brower DL, Brower SM, Hayward DC, Ball EE. 1997. Molecular evolution of integrins: genes encoding integrin beta subunits from a coral and a sponge. PNAS of the USA. 94(17): 9182–9187.

Burke RD. 1999. Invertebrate integrins: structure, function, and evolution. International Review of Cytology. 191:257–284.

Conaco C, Neveu P, Zhou H, Arcila M Luz, Degnan SM, Degnan BM, Kosik KS. 2012. Transcriptome profiling of the demosponge *Amphimedon queenslandica* reveals genome-wide events that accompany major life cycle transitions. BMC Genomics. 13:209–228.

Craft JA, Gilbert JA, Temperton B, Dempsey KE, Ashelford K, Tiwari B, Hutchinson TH, Chipman JK. 2010. Pyrosequencing of *Mytilus galloprovincialis* cDNAs: tissue-specific expression patterns. PloS one. 5:e8875.

Distel DL, Baco AR, Chuang E, Morrill W, Cavanaugh C, Smith CR. 2000. Do mussels take wooden steps to deep-sea vents? Nature. 403:725–726.

Dyachuk V, Odintsova N. 2009. Development of the larval muscle system in the mussel *Mytilus trossulus* (Mollusca, Bivalvia). Development, growth & differentiation. 51:69–79.

Dyachuk VA, Maiorova MA, Odintsova NA. 2015. Identification of beta integrin-like- and fibronectin-like proteins in the bivalve mollusk *Mytilus trossulus*. Development, growth & differentiation. 57:515–528.

Ewan R, Huxley-Jones J, Mould AP, Humphries MJ, Robertson DL, Boot-Handford RP. 2005. The integrins of the urochordate *Ciona intestinalis* provide novel insights into the molecular evolution of the vertebrate integrin family. BMC Evolutionary Biology. 5:31–49.

Gerdol M, Venier P, Pallavicini A. 2015. The genome of the Pacific oyster *Crassostrea gigas* brings new insights on the massive expansion of the C1q gene family in Bivalvia. Developmental & Comparative Immunology. 49(1):59–71.

Gettner SN, Kenyon C, Reichardt LF. 1995. Characterization of beta pat-3 heterodimers, a family of essential integrin receptors in *C. elegans*. The Journal of Cell Biology. 129:1127–1141.

Gonzalez M., Gueguena Y., Desserrea G., Lorgerila J. de, Romestanda B., Bachèrea E. 2007. Molecular characterization of two isoforms of defensin from hemocytes of the oyster *Crassostrea gigas.* Developmental & Comparative Immunology. 31(4):332–339.

Grasso LC, Maindonald J, Rudd S, Hayward DC, Saint R, Miller DJ, Ball EE. 2008. Microarray analysis identifies candidate genes for key roles in coral development. BMC Genomics. 9:540–557.

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

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc R.D, Friedman N, Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols. 8(8):1494–1512.

Hughes AL. 2001. Evolution of the integrin alpha and beta protein families. The Journal of Molecular Evolution. 52:63–72.

Humphries MJ. 2000. Integrin structure. Biochem. Soc. Trans. 28:311–339.

Hynes RO. 1992. Integrins: versatility, modulation, and signaling in cell adhesion. Cell. 69:11–25.

Hynes RO. 2012. The evolution of metazoan extracellular matrix. The Journal of Cell Biology. 196(6):671–679.

Jia Z, Zhang T, Jiang S, Wang M, Cheng Q, Sun M, Wang L, Song L. 2015. An integrin from oyster *Crassostrea gigas* mediates the phagocytosis toward *Vibrio splendidus* through LPS binding activity. Developmental & Comparative Immunology. 53:253–264.

Knack BA, Iguchi A, Shinzato Ch, Hayward DC, Ball EE, Miller DJ. 2008. Unexpected diversity of cnidarian integrins: expression during coral gastrulation. BMC Evolutionary Biology. 8:136–147.

Koehn RK, Gaffney PM. 1984. Genetic heterozygosity and growth rate in *Mytilus edulis*. Marine Biology. 82:1–7.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution. 33(7):1870–1874.

Liu J, Liu H, Zhang H. 2018. Phylogeny and evolutionary radiation of the marine mussels (Bivalvia:Mytilidae) based on mitochondrial and nuclear genes. Molecular Phylogenetics and Evolution. 126:233–240.

Lockyer AE, Spinks JN, Walker AJ, Kane RA, Noble LR, Rollinson D, Dias-Neto E, Jones CS. 2007. *Biomphalaria glabrata* transcriptome: identification of cell-signalling, transcriptional control and immune-related genes from open reading frame expressed sequence tags (ORESTES). Developmental & Comparative Immunology. 31(8):763–782.

Maiorova MA, Odintsova NA. 2015. Beta integrin–like protein-mediated adhesion and its disturbances during cell cultivation of the mussel *Mytilus trossulus*. Cell Tissue Research. 361 (2):581–592.

Marsden M, Burke RD. 1998.The beta integrin subunit is necessary for gastrulation in sea urchin embryos. Developmental Biology. 203(1):134–148.

Miyazawa S, Azumi K, Nonaka M. 2001. Cloning and characterization of integrin alpha subunits from the solitary ascidian, *Halocynthia roretzi.* Journal of immunology. 166(3):1710–1715.

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. 16:728.

Mosquera E, Lopez JL, Alvarez G. 2003. Genetic variability of the marine mussel *Mytilus galloprovincialis* assessed using two-dimensional electrophoresis. Heredity. 90(6):432–442.

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(7):e0160081.

Odintsova NA, Dyachuk VA, Nezlin LP. 2010. Muscle and neuronal differentiation in primary cell culture of larval *Mytilus trossulus* (Mollusca: Bivalvia). Cell and Tissue Research. 339(3):625–637.

Owada M. 2007. Functional morphology and phylogeny of the rock-boring bivalves Leiosolenus and Lithophaga (Bivalvia: Mytilidae): a third functional clade. Marine Biology 150:853–860.

Ounit R, Wanamaker S, Close TJ, Lonardi S. 2015. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. BMC genomics. 16:236.

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

Reber-Muller S, Studer R, Muller P, Yanze N, Schmid V. 2001. Integrin and talin in the jellyfish *Podocoryne carnea*. Cell Biology International. 25:753–769.

Samadi S, Quéméré E, Lorion J, Tillier A, von Cosel R, Lopez P, Cruaud C, Couloux A, Boisselier-Dubayle M.-C. 2007. Molecular phylogeny in mytilids supports the wooden steps to deep-sea vents hypothesis. C. R. Biology. 330:446–456.

Satoh N. 2016. Chordate Origins and Evolution: The Molecular Evolutionary Road to Vertebrates. London NW1 7DX, Academic Press Ltd-Elsevier Science Ltd. 1–206.

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

Skibinski DOF, Cross TF, Ahmad M. 1980. Electrophoretic investigation of systematic relationships in the marine mussels *Modiolus modiolus* L., *Mytilus edulis* L., and *Mytilus galloprovincialis* Lmk. (Mytilidae; Mollusca). Biological Journal of the Linnean Society. London. Blackwell Publishing Ltd. 13:65–73.

Smith S, Wilson NG, Goetz F, Feehery C, Andrade SCS, Rouse GW, Gilbert G, Dunn CW. 2011. Resolving the evolutionary relationships of molluscs with phylogenomics tools. Nature. 480 (7377):364–367.

Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML, Signorovitch AY, Moreno MA, Kamm K, Grimwood J, Schmutz J, Shapiro H, Grigoriev IV, Buss LW, Schierwater B, Dellaporta SL, Rokhsar DS. 2008. The *Trichoplax* genome and the nature of placozoans. Nature. 454 (7207):955–960.

Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, Larroux C, Putnam NH, Stanke M, Adamska M, Darling A, Degnan SM, Oakley TH, Plachetzki DC, Zhai Y, Adamski M, Calcino A, Cummins SF, Goodstein DM, Harris C, Jackson DJ, Leys SP, Shu S, Woodcroft BJ, Vervoort M, Kosik KS, Manning G, Degnan BM, Rokhsar DS. 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature. 466 (7307):720–726.

Suga H, Chen Z, de Mendoza A, Seb´e-Pedr´os A, Brown MW, Kramer E, Carr M, Kerner P, Vervoort M, Sánchez-Pons N, Torruella G, Derelle R, Manning G, Lang BF, Russ C, Haas BJ, Roger AJ, Nusbaum C, Ruiz-Trillo I. 2013. The *Capsaspora* genome reveals a complex unicellular prehistory of animals. Nature Communications. 4:2325–2334.

Takada Y, Ye X, Simon S. 2007. The integrins. Genome Biology. 8:215–222.

Takeuchi T, Kawashima T, Koyanagi R, Gyoja F, Tanaka M, Ikuta T, Shoguchi E, Fujiwara M, Shinzato C, Hisata K, Fujie M, Usami T, Nagai K, Maeyama K, Okamoto K, Aoki H, Ishikawa T, Masaoka T, Fujiwara A, Endo K, Endo H, Nagasawa H, Kinoshita S, Asakawa S, Watabe S, Satoh N. 2012. Draft genome of the pearl oyster *Pinctada fucata*: A platform for understanding bivalve biology. DNA Research. 19(2):117–130.

Takeuchi T, Koyanagi R, Gyoja F, Kanda M, Hisata K, Fujie M, Goto H, Yamasaki S, Nagai K, Morino Y. 2016. Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. Zoological letters. 2:3.

Tanguy M, McKenna P, Gauthier-Clerc S, Pellerin J, Danger J-M, Siah A. 2013. Sequence analysis of a normalized cDNA library of *Mytilus edulis* hemocytes exposed to *Vibrio splendidus* LGP32 strain. Results in Immunology. 3:40–50.

Toubiana M, Rosani U, Giambelluca S, Cammarata M, Gerdol M, Pallavicini A, Venier P, Roch P. 2014.Toll signal transduction pathway in bivalves: Complete cds of intermediate elements and related gene transcription levels in hemocytes of immune stimulated *Mytilus galloprovincialis*. Developmental & Comparative Immunology. 45 (2):300–312.

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

Wainright PO, Hinkle G, Sogin ML, Stickel SK. 1993. Monophyletic origins of the metazoa: an evolutionary link with fungi. Science. 260:340–342.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics. 25: 1189–1191.

Williams EA, Degnan BM, Gunter H, Jackson DJ, Woodcroft BJ, Degnan SM. 2009. Widespread transcriptional changes pre-empt the critical pelagic-benthic transition in the vetigastropod *Haliotis asinina*. Molecular Ecology. 18(5):1006–1025.

Woods RG, Roper KE, Gauthier M, Bebell LM, Sung K, Degnan BM, Lavin MF. 2004. Gene expression during early ascidian metamorphosis requires signalling by Hemps, an EGF-like protein. Development. 131(12):2921–2933.

Yoshida M-A, Ishikura Y, Moritaki T, Shoguchi E, Shimizu KK, Sese J, Ogura A. Genome structure analysis of molluscs revealed whole genome duplication and lineage specific repeat variation. Gene 483:63–71.

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-Loso 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 P.M., 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.

Description	Statistics
Total number of contigs	200079
Number of "genes" excluding isoforms	155959
Number of contigs (>1000 bp)	35270
Number of contigs (>5000 bp)	1709
Number of contigs (>10000 bp)	114
Number of contigs (>25000 bp)	1
Total length of contigs (bp)	5656866
Total length of contigs (>500 bp)	104471164
Average length of contig (bp)	728
Median length of contig (bp)	392
Largest contig (bp)	31400
GC (%) of contig	34.03
N50 (bp)	1126
N50 (subject only to the contigs >500 bp)	1804
L50 (subject only to the contigs >500 bp)	16377
% annotated contigs (NCBI/NR)	29,24
% annotated contigs (UniProt/TrEMBL)	29,99
% annotated contigs (Gene Ontology)	19,96

Species, degree of identity / similarity	*M. trossulus,* β-A-integrin-like protein	*M. trossulus,* *β-B*-integrin-like protein	*C. gigas,* β 1-like integrin	*C. gigas,* β 3- like integrin	*Lottia gigantea,* β 1-integrin	*Lingula anatina,* β-PS-like integrin	*Strongylocentrotus purpuratus,* β-integrin	*Branchiostoma lanceolatum,* β 1-integrin
M. trossulus, β-A-integrin-like protein	100	38/54	54/70	33/52	54/71	52/68	39/57	41/60
M. trossulus, *β-B*-integrin-like protein	38/54	100	35/52	46/62	37/55	39/57	34/51	35/52

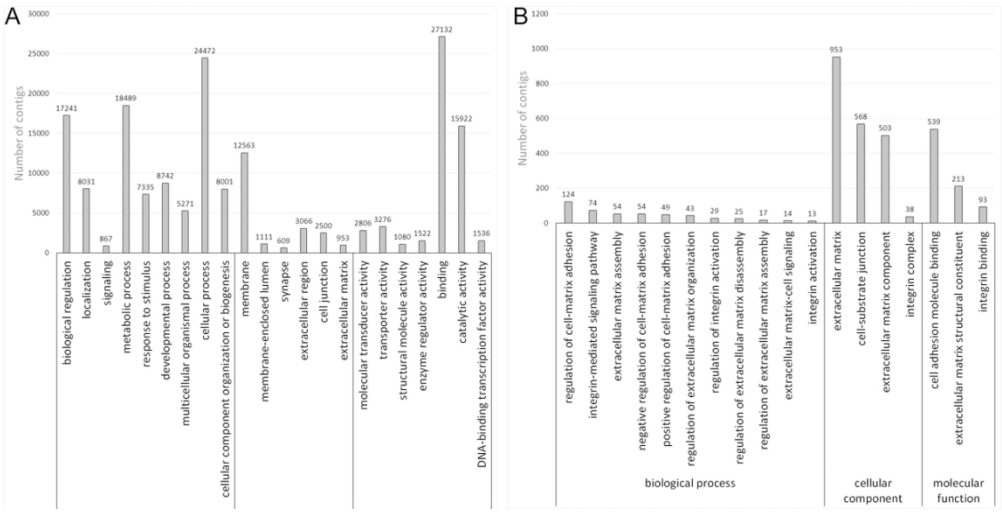


Figure 1. Functional annotation of assembled contigs in the mussel M. trossulus based on gene ontology (GO) categorization. GO analysis was performed at the second level for three main categories (cellular component, molecular function and biological process) (A) and for selected GO terms, related to integrins and extracellular matrix, in all three categories (B). The number of transcripts is shown on top of each column.

107x54mm (300 x 300 DPI)

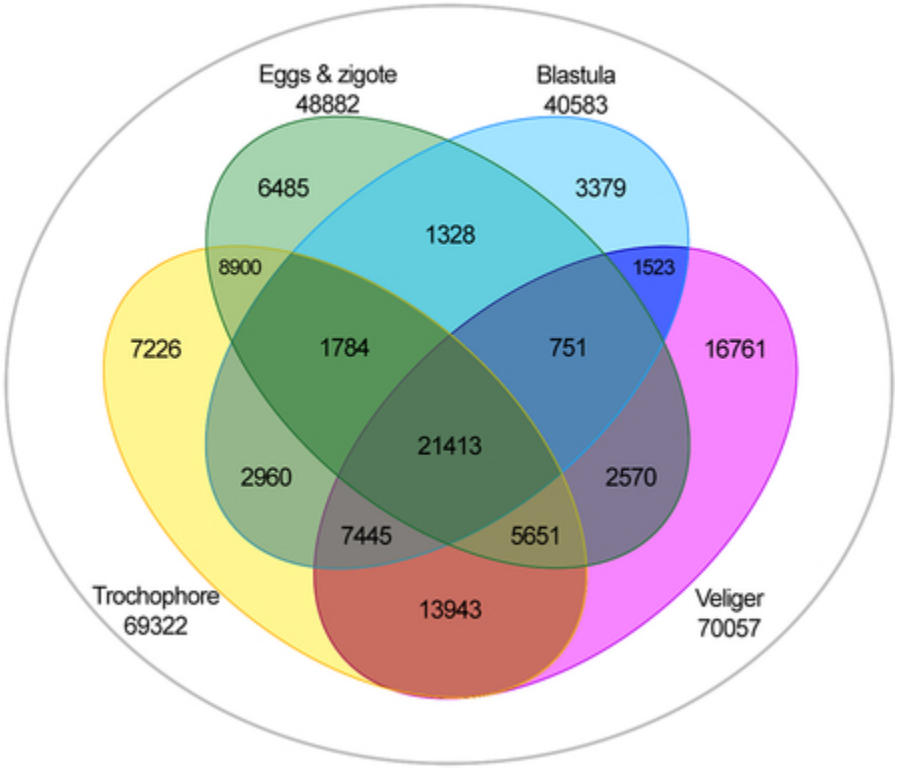


Figure 2. Venn diagram showing the overlap of transcriptionally active contigs (whose FPKMs ≥1 at least in one from the libraries) comparing different early developmental stages in the mussel Mytilus trossulus.

37x32mm (300 x 300 DPI)

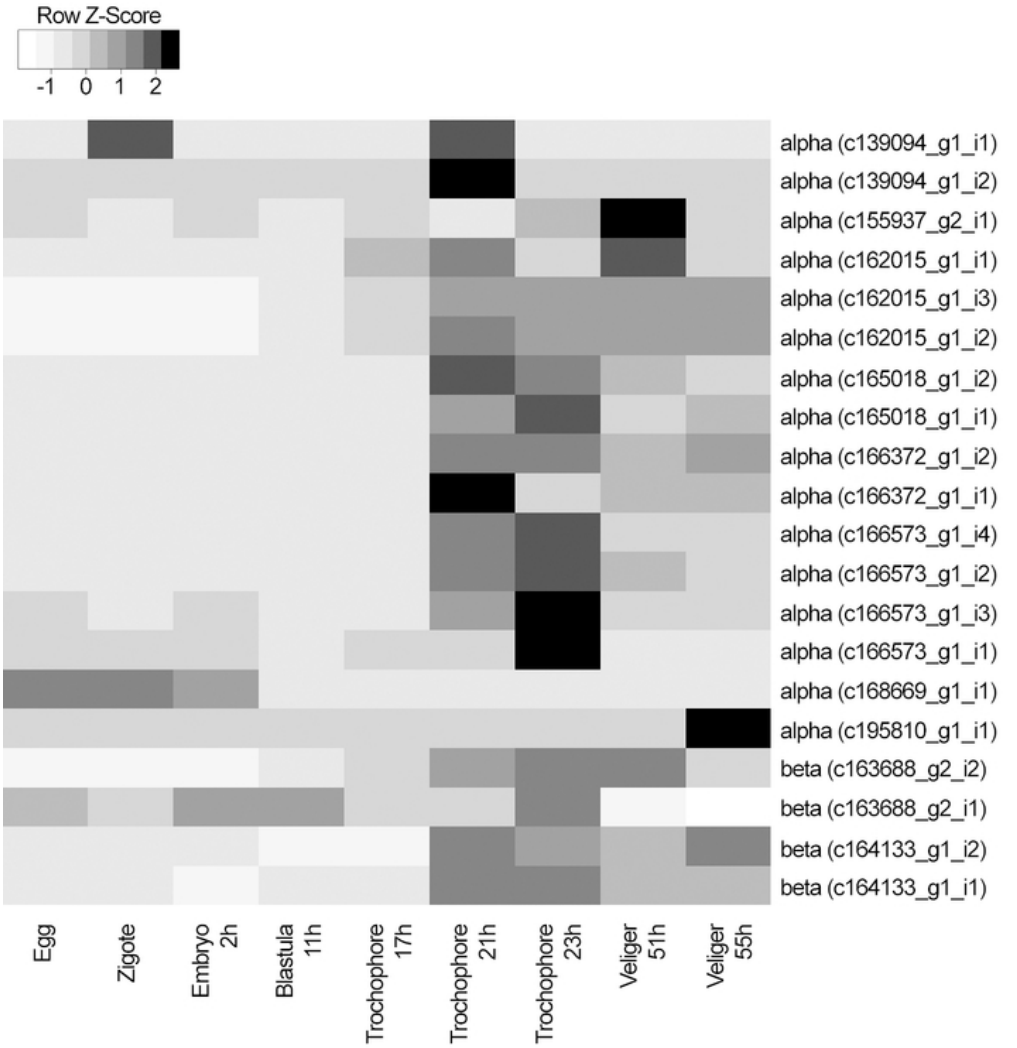


Figure 3. Heat map of quantitative gene expression of □ and □ integrin-like transcripts during mussel development (in eggs, zygotes, embryos and early larvae). The level of expression is marked by a color key. Data obtained using HiSeq Sequencing.

59x62mm (300 x 300 DPI)

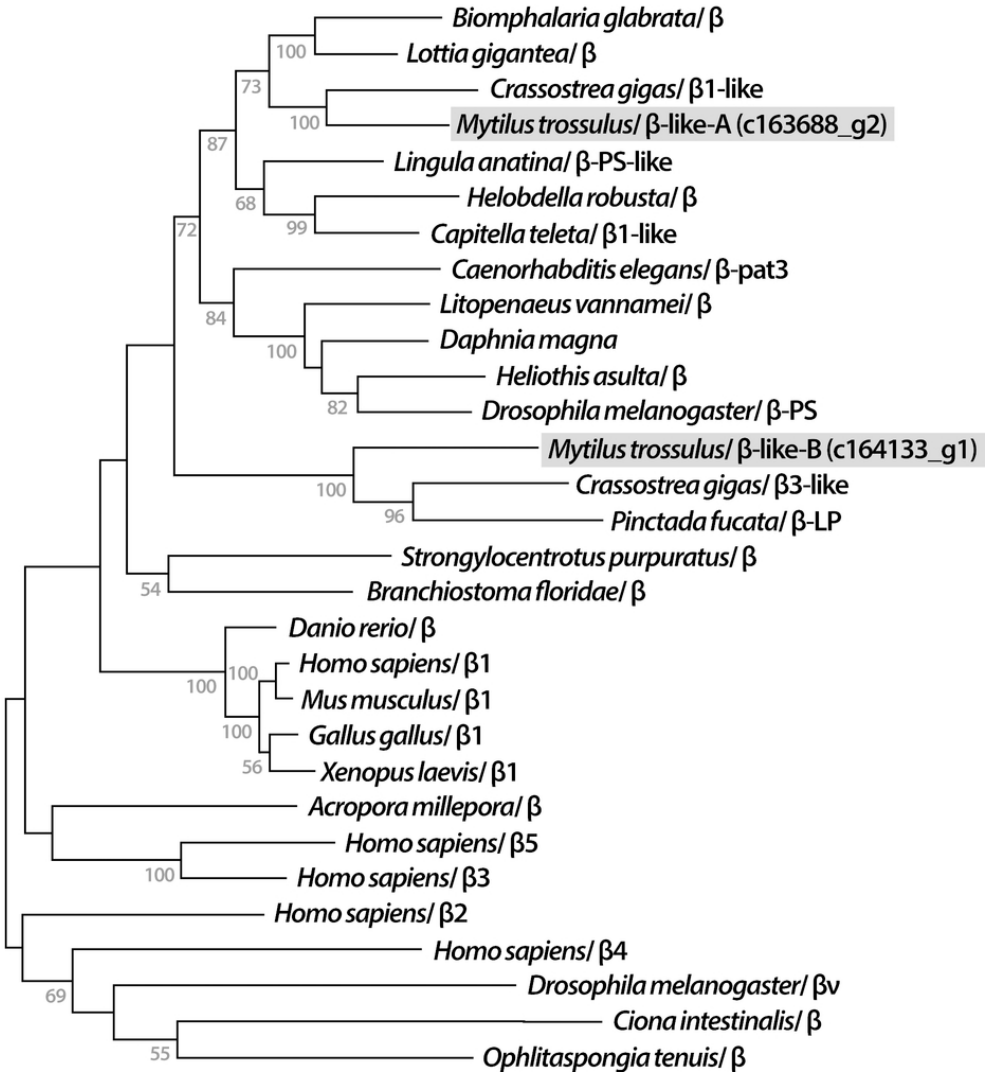


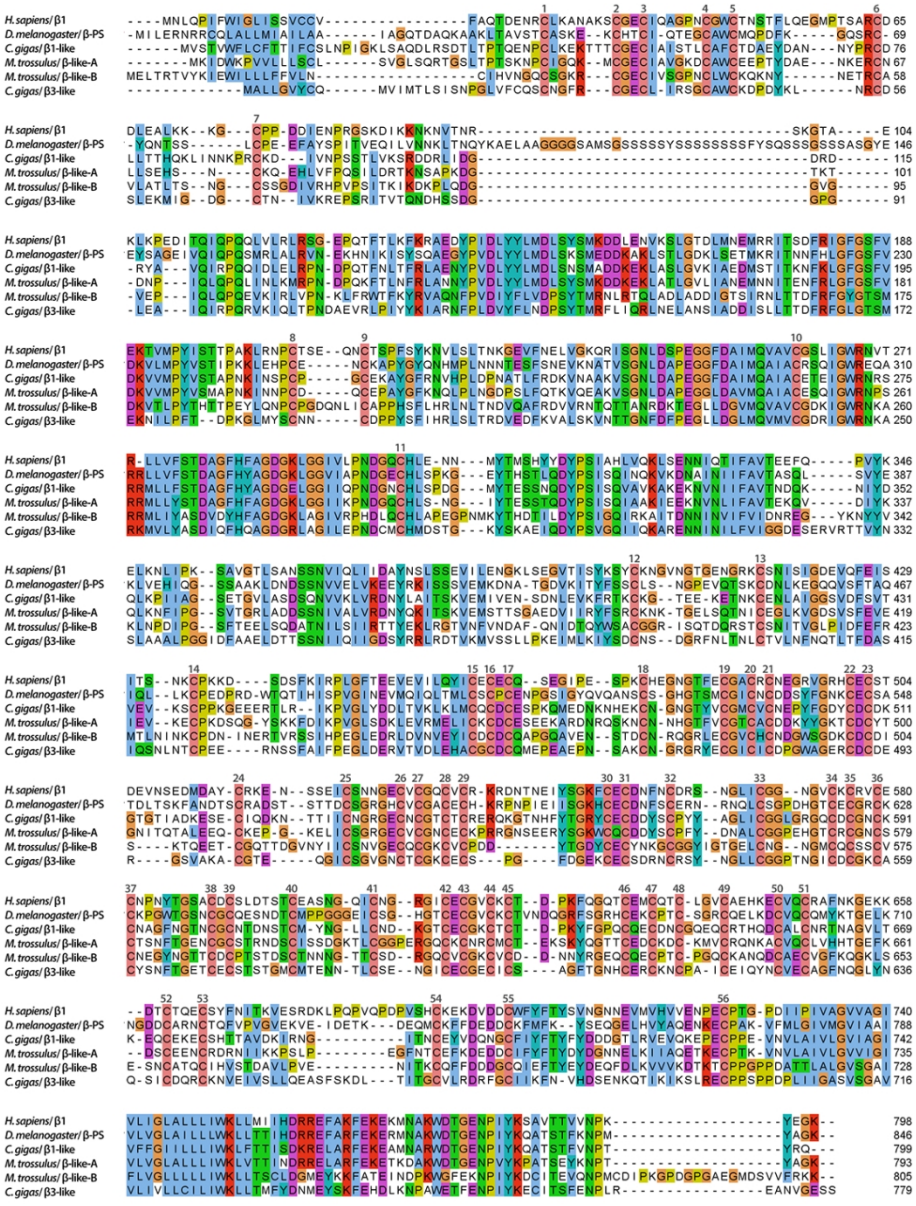
Figure 4. Cladogram showing the relationships among β integrin-like proteins in Protostomes and Deuterostomes. Amino acid sequences were aligned, and a phylogenetic tree was made by the ML method using amino acid sequences for β integrin-like proteins of the species identified. The bootstrap values above 50% are marked on the tree branches. The branches tested in branch-sites tests of selection for β integrin-like proteins are indicated.

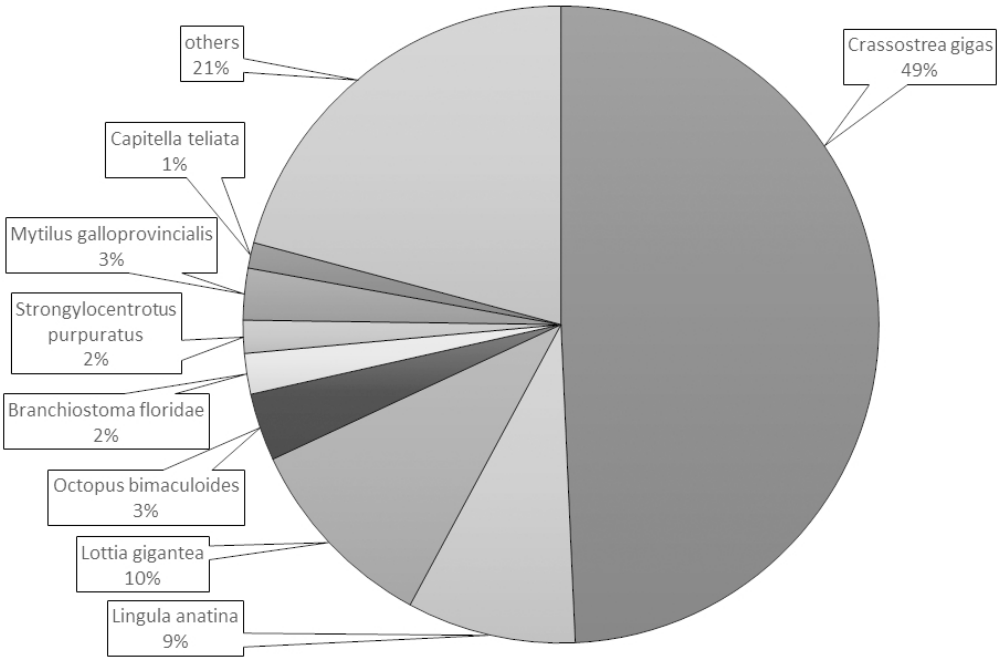
80x87mm (300 x 300 DPI)

Table S1. Details of the assembled libraries produced by the different sequencing runs. Quantitative evaluation of the data obtained MiSeq/HiSeq sequencing

Sample	concentration of cDNA libraries, nM	the average fragment length, bp	Number of reads		The percentage of reads used for the assembly of the transcriptome, %	The percentage of reads used for the analysis of quantitative expression,%		The percentage of reads successfully mapped at the transcriptome,%	
			MiSeq	HiSeq	MiSeq	MiSeq	HiSeq	MiSeq	HiSeq
Eggs	270,80	447	3088489	24463500	79,5	79,3	90,5	58,6	58,5
Fertilization eggs	35,10	409	2869307	27053170	77,4	77,3	90,0	62,2	62,2
Cleaving embryos, 2 h	350,82	485	4267044	25170927	91,0	90,8	89,5	57,6	56,6
Blastula, 11 h	110,42	410	2692405	25734753	78,0	78,0	90,3	62,7	62,7
Trochophore, 17 h	94,17	417	2289399	23263384	78,0	78,0	89,8	62,2	62,1
Trochophore, 21 h	64,94	426	2929434	32987911	76,4	76,3	89,6	62,9	62,7
Trochophore, 23 h	432,93	463	4371898	18494295	89,8	89,6	88,9	61,6	61,0
Trochophore, 24 h	343,86	490	2853912	20215770	79,6	79,5	92,4	59,9	59,7
Veliger, 51 h	82,80	445	3793360	17114374	77,6	77,6	93,3	62,4	62,2
Veliger, 55 h	440,75	481	5644674	30675844	81,7	81,5	92,6	61,9	62,1
Adductor	57,01	480	1062211	25895022	84,9	84,9	92,0	73,4	72,9
Gills	60,23	487	1082172	18020820	84,1	84,1	91,8	64,7	62,3

Testis	77,07	464	847069	42820964	85,7	85,7	93,4	56,1	55,0
Digestive gland	46,91	404	802017	19894536	85,5	85,5	93,9	66,2	64,6
Mantle	28,41	436	989591	28220848	83,0	83,0	93,1	72,9	70,8
Hemocytes	34,44	401	793843	15570065	85,9	85,8	94,5	62,6	61,1


H.sapiens/β1
D.melanogaster/β-PS
C.gigas/β1-like
M.trossulus/β-like-A
M.trossulus/β-like-B
C.gigas/β3-like


others
21%
Crassostrea gigas
49%
Capitella teliata
1%
Mytilus galloprovincialis
3%
Strongylocentrotus
purpuratus
2%
Branchiostoma floridae
2%
Octopus bimaculoides
3%
Lottia gigantea
10%
Lingula anatina
9%