Profiling of microRNAs and mRNAs in marine mussel *Mytilus galloprovincialis*

Deliang Yu, Huifeng Wu, Xiao Peng, Chenglong Ji, Xiaoying Zhang, Jun Song, Junle Qu

**Abstract**

MicroRNAs (miRNAs) are a class of noncoding RNA molecules containing 18–24 nucleotides, and those with conserved structures are able to regulate the expression of eukaryotic genes by inhibition or enhancement of mRNA translation. However, miRNAs of the blue mussel, *Mytilus galloprovincialis* have not been reported. *M. galloprovincialis* is a primary species distributed along coastal zones worldwide. To reveal the repertoire of miRNAs in *M. galloprovincialis*, we constructed small RNA libraries prepared from three different mussels, which were then sequenced by Solexa deep sequencing technology. A total of 32,836,817, 33,359,113 and 33,093,562 clean reads from the tissues of the three *M. galloprovincialis* were obtained. Based on sequence similarities and hairpin structure predictions, 137 *M. galloprovincialis* miRNAs (mg-miRNA) were identified. Among the mg-miRNAs, 104 were conserved across species, whereas 33 might be novel and specific for *M. galloprovincialis*. Some of the mg-miRNAs, such as let-7 and the miR-100 family are playing key roles in many metabolic pathways and are worthy of further study. By performing a whole genome-scale characterization of mg-miRNAs and proposing their potential functions, these results provide a foundation for understanding the biological processes of the blue mussel, *M. galloprovincialis*.

1. Introduction

The blue mussel (*Mytilus galloprovincialis*) is one of the main species cultured for seafood along the coastal areas of China and many coastal zones around the world. Mussels are also widely used marine organisms for environmental monitoring and assessment programs because of their sessile filter-feeding lifestyle, and high accumulation of pollutants from marine environments (Alimba and Faggio, 2019; Faggio et al., 2016; Ji et al., 2013; Yu et al., 2016). However, the functions of most molecules remain unclear.

MicroRNAs (miRNAs) are a class of small noncoding endogenous RNA molecules containing 18–24 nucleotides that function as negative regulators of target genes. miRNAs regulate gene expression by directing their target mRNAs for degradation or translational repression (Ambros, 2004; Bartel, 2004). Increasing evidence have shown that miRNAs are primary regulators in cell differentiation and growth, mobility, apoptosis (programmed cell death) and immune responses (Aboobaker et al., 2005; Bueno and Malumbres, 2011). Thus far, ~48,885 mature miRNAs have been identified from 271 species (www.mirbase.org, release 22.0), with many others yet to be identified. In mollusks, only 245 miRNAs have been identified, from *Haliotis rufescens*, *Lottia gigantea*, and *Melibe leonina*. MiRNAs have key roles in multiple biological processes in almost all species (Burgos-Aceves et al., 2018a). For many marine species, some novel miRNAs in marine medaka *Oryzias melastigma* target genes in a tissue-specific manner, such as neuron development and synaptic transmission in the brain, glucose and fat metabolism in the liver, and steroid genesis in the gonads. Such studies have led to further biomarker development for the assessment of environmental stresses and pollution in marine environments (Burgos-Aceves et al., 2018b, 2018c; Lai et al., 2015). In the marine oyster *Crassostrea gigas*, an invertebrate-specific miRNA was identified to have a negative role in both
synthesis and release of acetyl choline (ACh) and choline uptake in hemocytes during the early stages of pathogen infection (Chen et al., 2016). In another study, miRNA expression profiles in scallops (Chlamys farreri) were characterized in response to acute viral necrosis virus (AVNV) infection (Chen et al., 2014). The discovery of miRNAs in these studies has added a new dimension to our understanding of complex genetic regulatory networks in marine animals. However, the functions of miRNAs in marine mussels remain unclear.

Transcriptomics includes the analysis of a comprehensive set of transcripts expressed in cells under different conditions (e.g., a specific developmental stage or under stress), whereas RNA sequencing (de novo RNA-seq) technologies provide cost-effective high-throughput sequencing to advance genomic studies of non-model species. In many marine species, comprehensive genomic databases have been successfully constructed using RNA-seq, such as the owl limpet Lottia gigantea (Simakov et al., 2013), apple snail Pomacea canaliculata and pearl oyster Pinctada martensii (Sun et al., 2012; Jiao et al., 2014). However, miRNAs in Mytilus galloprovincialis (mg-miRNAs) have remained unclear. In the current study, we identified multiple mg-miRNAs and miRNAs in M. galloprovincialis and explored their potential functions by constructing small RNA libraries prepared with Solexa deep sequencing technology using the soft tissues of M. galloprovincialis. The results serve as a basis for further research on multiple biological processes in this organism, and establish a more comprehensive database of the miRNA and mRNA transcriptome of soft tissues in M. galloprovincialis.

2. Methods

2.1. Experimental samples and RNA extraction

Mussels (shell length: 3.5–5 cm, n = 3) were purchased at random from a fishery market in Yantai. The mussels were then transported to the laboratory and acclimatized in aerated normal filtered seawater (FSW) (salinity 31 psu, temperature 20 °C) for 7 days. During the acclimation period, all mussels were cultured under a 12 h: 12 h light: dark photoperiod, and fed daily with Chlorella vulgaris at a ratio of 2% of tissue per dry weight. Soft tissues were pooled from the shells and quickly dropped into liquid nitrogen. Then total RNA of soft tissues was extracted with TRIzol reagent (Invitrogen) according to the manufacturer’s instructions. To identify the maximum number of miRNAs and mRNAs, the total RNAs from all of the tissue samples were pooled in equal fractions.

2.2. miRNA and mRNA sequencing

In this study, BGISEQ-500 deep sequencing technology (Huada Gene, Shenzhen, Guangdong, China) was applied to sequence short reads. Clean sequencing reads were obtained by removing reads that contained poly-N with 5’ adapter contaminants, without 3’ adapters or the insert tag, containing poly A, T, G or C, low-quality reads, and reads < 18 nt. Following read clean-up, the high-quality reads were mapped to a reference sequence using Bowtie (Hoen et al., 2008), without mismatch to analyze their expression and distribution on the reference sequence. To remove tags originating from protein-coding genes, repeat sequences, rRNA, transfer RNA (rRNA), small nuclear RNA (snRNA), and small nucleolar RNAs ( snoRNAs), sRNA tags were mapped to the RepeatMasker, Rfam database. The clean reads were compared with the miRNA precursor and/or mature miRNAs of all animals in miRBase (version 22.0) to identify the sequence and number of miRNA families (not species specific) in the samples. The characteristics of the hairpin structures of miRNA precursors were evaluated to predict novel miRNAs.

For mRNA sequencing, low-quality reads (> 20% of the bases qualities being < 10), reads with adapters, and reads with unknown bases (N bases > 5%) were filtered to obtain the clean reads. These were then assembled into unigenes, followed by unigene functional annotation, simple sequence repeat (SSR) detection, and unigene expression level calculations. Finally, differential expressed genes (DEGs) were identified between samples and clustering analysis and functional annotations were performed. Raw reads were defined as reads containing low-quality, adaptor-polluted, and a high content of unknown base (N) reads; such reads were also removed before downstream analyses. Internal software was used to filter reads, as follows: (1) remove reads with adaptors; (2) remove reads in which unknown bases (N) were > 5%; and (3) remove low-quality reads (i.e., reads in which the percentage of bases with a quality < 10 was > 20% in a read). After filtering, the remaining reads were called ‘Clean Reads’ and stored in FASTQ format.

In the de novo assembly, Trinity was used to perform a de novo assembly with clean reads (PCR duplication was removed to improve the efficiency) and then TGICL (TIGR Gene Indices clustering tools) was used to cluster transcripts to unigenes. Trinity contains three independent software modules: Inchworm, Chrysalis, and Butterfly. It partitions the sequence data into many individual de Bruijn graphs, each representing the transcriptional complexity at a given gene or locus, and then processes each graph independently to extract full-length splicing isoforms and to tease apart transcripts derived from paralogous genes.

2.3. miRNA and mRNA expression analysis

To identify the miRNAs differentially expressed between three different mussels, expression data were log2-transformed and plotted on a scatter plot. Briefly, miRNA expression data from the three libraries were normalized to obtain the number of transcripts per million reads (TPM), using the normalization formula:

\[
\text{Normalized expression} = \frac{\text{mapped readcount}}{\text{Total reads}} \times 1 \times 10^6
\]

TPM analysis step could eliminate the influence of sequencing discrepancies on the calculation of miRNA expression. The fold-change and p-values were calculated from the normalized expression and the p-values were then adjusted using the q-value. The q-value < 0.01 and \(| \log_2 (\text{foldchange}) | > 1 \) was set as the threshold for significantly different expression (default). Finally, the \( \log_2 \)-ratio figure and scatter plot were generated. When the normalized expression of a miRNA was zero between three libraries, its expression value was adjusted to 0.01 (because 0 cannot be plotted on a log plot). If the normalized expression of a specific miRNA in three libraries was < 1, further differential expression analyses were conducted without that miRNA.

2.4. Transcriptome annotation

Gene Ontology (GO; www.geneontology.org) is an international standard classification system for gene function. The distribution of those selected target genes (miRNA and mRNA) among biological pathways/functions in gene ontologies will be identified to characterize the biological differences between samples based on gene function. Using this method, candidate target genes were mapped to the GO terms (biological functions) in the database (www.geneontology.org). The number of genes in every term was calculated, and a hypergeometric test was performed to identify significantly enriched GO terms in the target gene candidate list from the background reference gene list.

The Kyoto Encyclopedia of Genes and Genomes (KEGG; www.kegg.jp) database is a public database containing biological pathway data, and is a resource for understanding high-level functions and processes active in a biological system (Kanehisa et al., 2008). In our work, KEGG pathway analysis helps us identify significantly enriched metabolic pathways or signal transduction pathways enriched in target gene
candidates, compared with a reference gene background, using the hypergeometric test.

2.5. Statistical analysis

The R statistical software package (R, Auckland, New Zealand) was used to analyze for the RNA-seq data. R was used for graphical representations, as well as to correct for multiple testing and p-value corrections. The data were expressed as the mean ± standard error, with \( p < 0.05 \) indicating a significant difference. Graphpad 5.0 (San Diego, CA, USA) was used to generate figures.

3. Results

3.1. Overview of miRNA and mRNA-seq data

To identify miRNAs that were differentially expressed in three mussels, three miRNA libraries were constructed by Solexa sequencing. The error rate of the sequencing data from the three libraries was 0.01%, and Q30 was > 97.8%, indicating that the sequencing data were of high quality and suitable for this study. A total of 32,836,817, 33,359,113, and 33,093,562 reads were acquired from the *M. galloprovincialis* soft tissues of the three mussels. After discarding the sequences that were below the background, 76.56%, 81.71%, and 68.08% of the miRNAs were mapped to the mussel (*Bathymodiolus platifrons*) genome, for which all genomic information has been reported and have a close genetic relationship with *M. galloprovincialis* (Sun et al., 2017). Subsequently, all identical reads were classified by group, providing 7,697,619, 6,100,958, and 10,564,320 unique sequences, respectively. The specific miRNA length ranges from the clean reads were subtracted from these sequences and the length distribution of miRNAs is shown in Fig. 3. Most of the miRNAs were 21–24 nucleotides in length. Therefore, the calculated gene expressions were directly used to compare the differences in gene expression between samples. A total of 32,836,817, 33,359,113, and 33,093,562 miRNAs were obtained for each of the three samples. The remaining reads were further compared by using the software RepeatMasker and Rfam database, to remove possible mRNA, rRNA, tRNA, snoRNA, snRNA, and repeat sequences. However, because some sRNA tags might map to more than one category, the following priority rule was applied: rRNA etc. (Genbank > Rfam) > known miRNA > repeat > exon > intron. All the clean reads were then divided into the following categories: exon_sense, exon_antisense, intron_sense, intron_antisense, miRNA, rRNA, repeat, scRNA, snRNA, snoRNA, srpRNA, tRNA, and unknown (sequences not mapped to any known reference databases). The composition of the RNA classes in each library is shown in Fig. 1. Among the clean reads, 25,139,198 (76.56%), 27,258,155 reads (81.71%) and 22,529,242 (68.08%) were mapped to the mussel genome (Sun et al., 2017). Known miRNAs accounted for 77.61%, 77.61% and 77.78% of the total clean reads. The analyses of these three libraries confirmed that these miRNA sequences were enriched among the miRNA libraries.

As a result of mRNA sequencing, 19.85 Gb bases were generated in total on the BGISEQ-500 sequencing platform. Assembling all the samples and filtering the abundance resulted in 126,858 unigenes; the total length, average length, N50, and GC content of these unigenes were 139,638,363 bp, 1100 bp, 2142 bp, and 34.32% respectively. The unigenes were annotated by aligning them with seven functional databases, resulting in 55,037 (NR: 43.38%), 13,094 (NT: 10.32%), 38,282 (Swissprot: 30.18%), 34,413 (KOG: 27.13%), 40,065 (KEGG: 31.58%), 8055 (GO: 6.35%), and 46,648 (InterPro: 36.77%) annotated unigenes (Fig. 2). For functional annotation results, 61,949 CDS (Coding sequences) were detected by Transdecoder, 9061 SSR were detected distributed on 7847 unigenes, and 7683 transcription factor (TF)-coding unigenes were predicted.

3.2. Identification of miRNAs and potential novel miRNAs

To identify known miRNAs in *M. galloprovincialis*, the data set was compared with miRNAs (miRNA precursors and mature miRNAs) in the miRBase (version 22.0). The three samples provided a total of
25,139,198, 27,258,155, and 22,529,242 unique sequences, respectively. Among these, 104 were mature miRNAs, as listed in Table 1. A cDNA library of mg-miRNAs was sequenced using the BGISEQ-500 sequencing platform. After removal of the low-quality and adaptor sequences, 27,479,838 reads representing 317,630 distinct sequences were obtained. Given that miRNAs with known functions are typically 18–24 bp in length, the unique size distribution patterns of the miRNAs were analyzed (Fig. 1). The results showed that the size of miRNAs was not evenly distributed, with the number of 22-bp sequences being significantly higher than the number of sequences with < 22 or > 22 bp. The same result was observed for 34.48% of all sequences, which is consistent with the distribution of the most common sizes of miRNAs in other species (Ambros, 2001). Therefore, this analysis suggests that miRNAs in the mg-miRNA library are enriched as designed.

The presence of a hairpin RNA structure is a typical characteristic of miRNA precursor, which can be applied to predict novel miRNAs. Therefore, miRDeep2 miRNA prediction software was applied to identify potential novel miRNAs. Briefly, the precursor sequences were mapped to the mollusk genome by exploring their secondary structures, Dicer cleavage sites, and binding energies, leading to the identification of 33 potential novel miRNAs (Table 2).

### 3.3. Transcriptome annotation analysis

GO enrichment assessments were applied to analyze the candidate target genes for differentially expressed miRNAs to predict biological functions. Statistical analyses of the significantly enriched number of genes in each term are summarized in Fig. 4. KEGG pathway analyses showed that the candidate genes of miRNAs in the soft tissue of *M. galloprovincialis* were involved in 325 pathways. Moreover, the results indicated that several enriched terms were involved in biological processes, including cancer pathways, focal adhesion, PI3K-Akt signaling pathway, endocytosis, and neuroactive ligand–receptor interactions (Fig. 4).

### 4. Discussion

miRNAs are important molecules in a series of mammalian biological processes (Bartel, 2018). An increasing number of miRNAs have been identified in various species, including marine animals (Chen et al., 2014; He et al., 2017; Huo et al., 2017). The blue mussel (*M. galloprovincialis*) is a popular sea food worldwide, produced in huge quantities, and is also used for monitoring the marine environment (Freitas et al., 2019; Pagano et al., 2016; Pagano et al., 2017). However, the functions of most molecules remain unclear, especially of non-coding RNA, for instance, miRNA. Lists of miRNAs have been reported for a few molluscan species (Chen et al., 2017; Martín-Gómez et al., 2014a; Zheng et al., 2016; Zhou et al., 2014) and miRNA families have been investigated in the limpet genome (Kenny et al., 2015). Given that a general overview of *M. galloprovincialis* miRNA was lacking, the current study used transcriptomic data sets available for *M. galloprovincialis* to identify and characterize core elements involved in the miRNA formation pathway in this mussel.

BGISEQ-500 deep sequencing technology is used to sequence shorter reads. The specified size of mRNA and miRNA molecules can be directly sequenced in the sample by this technology without any miRNA sequence information to provide a more comprehensive miRNA expression profile and expedite the discovery of new miRNAs (Huang
This technology has been widely utilized to identify conserved and novel miRNAs in various species (Chen et al., 2009; Ge et al., 2013; Hackl et al., 2011; Lv et al., 2012; Song et al., 2010). In the current study, 104 conserved mg-miRNAs and 33 novel miRNAs were identified by deep sequencing techniques. Some miRNAs revealed to have key roles in multiple biological processes in marine species were also identified in M. galloprovincialis. For example, miR-31 is a conserved and critical regulator of gene expression in many pathogenic processes in vertebrates. It modulates respiratory bursts via targeting Aip105 during pathologival development of the sea cucumber, Apostichopus japonicas (Lu et al., 2015). In the present study, two members of the miR-31 family (miR-31_2, miR-31_3) were identified with high expression levels in M. galloprovincialis. In addition, miR-92a is a disease related regulator that is involved not only in tumorogenesis, but also in multiple host–pathogen interactions in vertebrates. Exposure to bacteria (Vibrio splendidus) and lipopolysaccharides (LPS) significantly increased the expression of A. japonicus miR-92a at different time points (Zhang et al., 2014). In marine shrimp (Marsupenaeus japonicus), miR-7 was upregulated in response to white spot syndrome virus (WSSV) infection and might have an evolutionarily conserved role (Huang and Zhang, 2012). In the current study, both the miR-92a and miR-7 families (miR-7, miR-7-5p_3, and miR-7-5p_5) were expressed at a high level, suggesting that these miRNAs participate in many disease defense pathways in mussels.

Among the conserved miRNAs, five miRNAs (let-7, miR-100, miR-184, miR-279, and miR-34) were present in > 10 million copies in the mg-miRNA libraries. Studies have shown that both the let-7 and miR-100 families are highly conserved in different sequences and functions (Christodoulou et al., 2010; Roush and Slack, 2008). The sequences of the mature let-7 families are 100% conserved among bilaterians (Caygill and Johnston, 2008; Pasquinelli et al., 2000). Temporal regulation of let-7 is also conserved, and its expression is largely limited to developmental stages corresponding to the transitional period between the juvenile and adult individuals in multiple species, such as nematode (Caenorhabditis elegans), fly (Drosophila melanogaster), zebrafish, and rat (Schulman et al., 2005). Drosophila let-7 co-expresses with other miRNA families in numerous tissues late in the third larval instar (L3), peaking in pupae during metamorphosis (Hutvágner et al., 2001; Semper et al., 2013).
Table 2

The information of potential novel miRNAs in the mussel *Mytilus galloprovincialis*.

<table>
<thead>
<tr>
<th>miRNA id</th>
<th>Sequence (mature)</th>
<th>Chromosome</th>
<th>Tag information</th>
</tr>
</thead>
<tbody>
<tr>
<td>novel_mir1</td>
<td>uucugugcugegaaacugcucu</td>
<td>Bpl_scaf_25767</td>
<td>tag25633,115,tag20972,1207,tgal15046,595921,t23312,182,-</td>
</tr>
<tr>
<td>novel_mir2</td>
<td>uaaacuacguacugccgsgaag</td>
<td>Bpl_scaf_44438</td>
<td>tag13189,208,tgal19664,15,tgal7216,555,tgal6570,952,-</td>
</tr>
<tr>
<td>novel_mir3</td>
<td>uacuacacccgguauaugugcgegg</td>
<td>Bpl_scaf_41094</td>
<td>tag18740,2356,tgal16776,35</td>
</tr>
<tr>
<td>novel_mir4</td>
<td>uacuacacacccgguauaugugcgegg</td>
<td>Bpl_scaf_30185</td>
<td>tag22836,1096,tgal18995,1416,tgal22244,141,tgal1762,158,-</td>
</tr>
<tr>
<td>novel_mir5</td>
<td>uacacacaacacagguauaugugcgegg</td>
<td>Bpl_scaf_35208</td>
<td>tag14889,157,tgal3628,3549,tgal15779,39,tgal21630,45264,-</td>
</tr>
<tr>
<td>novel_mir6</td>
<td>uacuacacccguacgguagcgegg</td>
<td>Bpl_scaf_61139</td>
<td>tag21279,2578,tgal7558,33,tgal24635,208,tgal11394,229,-</td>
</tr>
<tr>
<td>novel_mir7</td>
<td>caaguccacacauccgguauaugugcgegg</td>
<td>Bpl_scaf_8330</td>
<td>tag22378,10313,tgal25711,39,tgal23727,10,tgal20385,122,-</td>
</tr>
<tr>
<td>novel_mir8</td>
<td>caaguccacacauccgguauaugugcgegg</td>
<td>Bpl_scaf_11649</td>
<td>tag17170,29,tgal163,167,tgal25109,19,tgal43343,37,-</td>
</tr>
<tr>
<td>novel_mir9</td>
<td>caaguccacacauccgguauaugugcgegg</td>
<td>Bpl_scaf_681</td>
<td>tag17877,21035,tgal24464,64,tgal6007,25,tgal16303,59</td>
</tr>
<tr>
<td>novel_mir10</td>
<td>caaguccacacauccgguauaugugcgegg</td>
<td>Bpl_scaf_63585</td>
<td>tag1976,25,tgal793,43314,tgal1241,121,tgal981,2143,-</td>
</tr>
<tr>
<td>novel_mir11</td>
<td>caaguccacacauccgguauaugugcgegg</td>
<td>Bpl_scaf_2982</td>
<td>tag9994,17,tgal16275,219,tgal16367,121,tgal23335,37</td>
</tr>
<tr>
<td>novel_mir12</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_23</td>
<td>tag22643,229,tgal1046,49,tgal19822,28</td>
</tr>
<tr>
<td>novel_mir13</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_23</td>
<td>tag22643,229,tgal1046,49,tgal19822,28</td>
</tr>
<tr>
<td>novel_mir14</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_23</td>
<td>tag22643,229,tgal1046,49,tgal19822,28</td>
</tr>
<tr>
<td>novel_mir15</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_23</td>
<td>tag22643,229,tgal1046,49,tgal19822,28</td>
</tr>
<tr>
<td>novel_mir16</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_23</td>
<td>tag22643,229,tgal1046,49,tgal19822,28</td>
</tr>
<tr>
<td>novel_mir17</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_23</td>
<td>tag22643,229,tgal1046,49,tgal19822,28</td>
</tr>
<tr>
<td>novel_mir18</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_23</td>
<td>tag22643,229,tgal1046,49,tgal19822,28</td>
</tr>
<tr>
<td>novel_mir19</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_23</td>
<td>tag22643,229,tgal1046,49,tgal19822,28</td>
</tr>
<tr>
<td>novel_mir20</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_33037</td>
<td>tag5153,128,tgal2156,367,tgal4922,29</td>
</tr>
<tr>
<td>novel_mir21</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_13745</td>
<td>tag4035,110,tgal19513,752,tgal18346,551</td>
</tr>
<tr>
<td>novel_mir22</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_50342</td>
<td>tag10019,399,tgal25408,6958,tgal17477,39021,tgal10138,152</td>
</tr>
<tr>
<td>novel_mir23</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_52347</td>
<td>tag1824,709</td>
</tr>
<tr>
<td>novel_mir24</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_56410</td>
<td>tag1707,713749,615,tgal10519,72,tgal25619,78,-</td>
</tr>
<tr>
<td>novel_mir25</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_13745</td>
<td>tag1050,37,tgal3707,701,tgal19490,44,tgal10785,315,-</td>
</tr>
<tr>
<td>novel_mir26</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_46488</td>
<td>tag2371,101,tgal11108,5672,tgal23387,145,tgal1236,783</td>
</tr>
<tr>
<td>novel_mir27</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_63435</td>
<td>tag7457,32601,tgal960,6290,tgal2988,38,tgal960,432,-</td>
</tr>
<tr>
<td>novel_mir28</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_33596</td>
<td>tag7233,93,tgal15253,655,tgal20940,14,-</td>
</tr>
<tr>
<td>novel_mir29</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_27638</td>
<td>tag20164,10,tgal4534,228</td>
</tr>
<tr>
<td>novel_mir30</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_27638</td>
<td>tag20164,10,tgal4534,228</td>
</tr>
<tr>
<td>novel_mir31</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_46577</td>
<td>tag9851,161,tgal3626,31,tgal1739,139,tgal23933,10,-</td>
</tr>
<tr>
<td>novel_mir32</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_45214</td>
<td>tag1049,53215,tgal12305,433</td>
</tr>
<tr>
<td>novel_mir33</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_22813</td>
<td>tag22601,16</td>
</tr>
<tr>
<td>novel_mir34</td>
<td>gcaaucaacagcaacguagcgegg</td>
<td>Bpl_scaf_23153</td>
<td>tag22601,16</td>
</tr>
</tbody>
</table>

*Comparative Biochemistry and Physiology, Part C 230 (2020) 108609*
MiR-100s are considered to be ancient and conserved because they are observed in most metazoans. In the current study, four members of the miR-100 family (miR-100_2, miR-100-5p, miR-100a-5p, and miR-100-5p_4) were identified in *M. galloprovincialis*, based on conserved miRNA expression patterns (e.g. miR-100, miR-31, miR-34, miR-92, and miR-124) (Grimson et al., 2008; Lee et al., 2016; Martín-Gómez et al., 2014b; Mulder and Berezikov, 2010). Early studies demonstrated a central role for the miR-34 family in promoting cell cycle arrest and cell death following stress in human cells. However, the biological significance of this response is unclear in *M. galloprovincialis*.

Interestingly, a total of 33 novel mg-miRNAs were identified (Table 2) and further revealed to be *M. galloprovincialis* specific. The read numbers of each novel miRNA were lower than those of the majority of the conserved miRNAs. These results are consistent with previous reports stating that nonconserved miRNAs are usually expressed at lower levels than conserved miRNAs (Liu et al., 2010; Zhu et al., 2012). Such a finding might indicate development-specific functions of these novel miRNAs in specific species. However, further studies could be performed to determine whether scarce miRNAs are expressed at higher levels in different developmental stages. Future functional studies will also provide insights into the functions of novel miRNAs in *M. galloprovincialis*.

In this study, miRNAs and mRNAs from the soft tissues of *M. galloprovincialis* were sequenced using the BGISEQ-500 sequencing platform. A total of 104 known miRNAs were identified and 33 potential novel miRNAs were predicted. In addition, both GO enrichment and KEGG pathway analyses of target genes from the three miRNA libraries were carried out. The results of KEGG pathway analyses revealed several enriched terms involved in *M. galloprovincialis* soft tissues (e.g., metabolic pathways, focal adhesion, PI3K-Akt signaling pathway, endocytosis, and neuroactive ligand–receptor interactions), which might affect the balance of the internal environment. Thus, future studies should be performed to determine whether miRNAs and mRNAs are involved in the regulation of environmental adaptation in *M. galloprovincialis*.

### 5. Conclusion

This study provides the first indication of miRNA and mRNA expression in the soft tissues of *M. galloprovincialis*. These results will be helpful for furthering the study of miRNA regulation and function in *M. galloprovincialis*.

### Declaration of competing interest

The authors declare no conflict of interest.
Acknowledgments

This work has been partially supported by the National Key Research and Development Program of China (2018YFC0910602); the National Natural Science Foundation of China (31771584/41806142/61525503/6162010616/61835009/41976152); (Key) Project of Department of Education of Guangdong Province (2016KCXTD007); Guangdong Natural Science Foundation Innovation Team (2014A030312008); Shenzhen Basic Research Project (JCYJ20170818100153423/JCYJ20170412105003520); Natural Science Foundation of SZU (No. 2017000193); China Postdoctoral Science Foundation (No. 2018M643151); and Shandong Key Laboratory of Coastal Environmental Processes, YICCAS (No. 2019SDHADKFJJ03).

References


Fig. 4. GO functional classification of mussel Mytilus galloprovincialis. The x-axis indicates the number of DEGs (presented by its square root value) and the y-axis represents GO terms. All GO terms were grouped into three ontologies: blue indicates biological process, brown indicates cellular component, and orange indicates molecular function. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
oyster. Open Biol. 6 (8), 160059.
Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature 408 (6808), 86–89.