



Full-Length Transcriptome Sequencing Reveals Tissue-Specific Gene Expression Profile of Mangrove Clam *Geloina erosa*

Xin Liao^{1*†}, Yunqing Liu^{1†}, Tingyu Han², Mingliu Yang¹, Wenai Liu¹, Yadi Wang³, Chunpeng He^{2*} and Zuhong Lu^{2*}

OPEN ACCESS

Edited by:

Youji Wang,
Shanghai Ocean University, China

Reviewed by:

Hongxia Wang,
Institute of Oceanology (CAS), China
Hongtao Nie,
Dalian Ocean University, China

*Correspondence:

Xin Liao
liaoX@mangrove.org.cn
Chunpeng He
cphe@seu.edu.cn
Zuhong Lu
zhlu@seu.edu.cn

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Aquatic Physiology,
a section of the journal
Frontiers in Physiology

Received: 11 January 2022

Accepted: 18 March 2022

Published: 20 April 2022

Citation:

Liao X, Liu Y, Han T, Yang M, Liu W,
Wang Y, He C and Lu Z (2022) Full-
Length Transcriptome Sequencing
Reveals Tissue-Specific Gene
Expression Profile of Mangrove Clam
Geloina erosa.
Front. Physiol. 13:851957.
doi: 10.3389/fphys.2022.851957

¹Guangxi Key Lab of Mangrove Conservation and Utilization, Guangxi Mangrove Research Center, Beihai, China, ²State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing, China, ³Henan Key Laboratory of Big Data Analysis and Processing, Institute of Data and Knowledge Engineering, School of Computer and Information Engineering, Henan University, Kaifeng, China

Mollusca is the second largest animal phylum and represents one of the most evolutionarily successful animal groups. *Geloina erosa*, a species of Corbiculidae, plays an important role in mangrove ecology. It is highly adaptable and can withstand environmental pollution and microbial infections. However, there is no reference genome or full-length transcriptome available for *G. erosa*. This impedes the study of the biological functions of its different tissues because transcriptome research requires reference genome or full-length transcriptome as a reference to improve accuracy. In this study, we applied a combination of Illumina and PacBio single-molecule real-time sequencing technologies to sequence the full-length transcriptomes of *G. erosa* tissues. Transcriptomes of nine samples obtained from three tissues (hepatopancreas, gill, and muscle) were sequenced using Illumina. Furthermore, we obtained 87,310 full-length reads non-chimeric sequences. After removing redundancy, 22,749 transcripts were obtained. The average Q score of 30 was 94.48%. In total, 271 alternative splicing events were predicted. There were 14,496 complete regions and 3,870 lncRNAs. Differential expression analysis revealed tissue-specific physiological functions. The gills mainly express functions related to filtration, metabolism, identifying pathogens and activating immunity, and neural activity. The hepatopancreas is the main tissue related to metabolism, it also involved in the immune response. The muscle mainly express functions related to muscle movement and control, it contains more energy metabolites than gill and hepatopancreas. Our research provides an important reference for studying the gene expression of *G. erosa* under various environmental stresses. Moreover, we present a reliable sequence that will provide an excellent foundation for further research on *G. erosa*.

Keywords: mangrove, *Geloina erosa*, SMRT, full-length transcriptomics, gill, hepatopancreas, muscle

INTRODUCTION

The mangrove clam *Gelonia erosa* syn. *Polymesoda erosa*, syn. *Gelonia expansa* (Tizzard et al., 2017; Yahya et al., 2018), belongs to Mollusca, Bivalvia, Heterodonta, Veneroida, Corbiculidae, *Gelonia* (Ingole et al., 2002), and is a type of large clam (Morton, 1976; Yingya et al., 1995). It is widely distributed throughout the Indian ocean and pacific coast and mainly inhabits mangroves, as well as estuaries and brackish marshes (Morton, 1976; Yingya et al., 1995). *G. erosa* plays an important role in the macrobenthic community in the mangrove habitat of the estuary. It is also the main target of mangrove tidal flat fisher. Despite this, there is little research on this species. *G. erosa* can use its mantle margin to achieve aerial respiration and its mantle cavity to absorb subterranean water during long-term exposure to the shore (Morton, 1976). They also take up pollutants present in the surrounding environment, which makes them excellent bioindicators of heavy metals and organic pollutants in mangrove environments (Modassir, 2000; Lai et al., 2011; Xing et al., 2017; Bertucci et al., 2018; Ma et al., 2019; Almeida et al., 2020; Wang et al., 2020; Zhang et al., 2020). In addition, *G. erosa* has important economic and medicinal value. Studies have found that the extract of *G. erosa* has high antiviral activity against influenza virus strains type-A (A/Mississippi 1/85/H3N2) and type-B (B/Harbin 7/94) (Morton, 1985; Chatterji et al., 2002).

To fully understand the biology of *G. erosa*, it is necessary to study the underlying genes that contribute to its functions and physical characteristics. Liu et al. (2021) published the world's first comprehensive mollusc genome database, which provides a valuable resource for studying the genome of mollusks. Zhao et al. (2017) studied tissues, from the gills and the hepatopancreas during external infections in *Sinonovacula constricta*. They found that gills are involved in a variety of ion transport processes, and the hepatopancreas is involved in immune responses. Kim et al. (2019) used the Illumina MiSeq platform to generate the transcriptome of the digestive gland in *Saxidomus purpuratus* and discovered that it performed functions related to innate immunity, digestion, and metabolism. Bertucci et al. (2018) performed RNA sequencing in *Corbicula fluminea*, a species in the Corbiculidae family that spread in freshwater and estuaries, exposed to water pollution. They found that genes for energy metabolism and survival in hypoxic conditions were differentially expressed depending on pollutants present in the environment *Corbicula fluminea* was living in.

The gills and hepatopancreas of *G. erosa* play an important role in environmental tolerance, resistance to pathogen infection, and metabolism. It is the focus of *G. erosa* research (Kim et al., 2019). As control group, elucidating the gene expression profiles and functions of muscle tissue is also important (Cappello et al., 2018).

High-quality transcripts are important for studying the biological functions of *G. erosa*. Due to the alternative gene splicing and a lack of genomic data, it is difficult to obtain accurate transcripts using second-generation sequencing methods. Third-generation sequencing (single molecule real-

time sequencing, SMRT) has the advantage of longer sequencing reads, thus avoiding read errors caused by alternative splicing. PacBio's long sequencing read length makes it ideal for identifying genetic subtypes, as well as novel genes and isoforms. However, PacBio has a high error rate.

In this study, full-length transcriptomes obtained by PacBio and Illumina sequencing were combined to construct the full-length transcriptome of *G. erosa*. Differential expression analysis was also performed between tissues taken from the gills, hepatopancreas, and muscle.

MATERIALS AND METHODS

Collection and Preparation of Animal Material

In this study, animal samples for sequencing were obtained from mangrove wetlands in Beihai, Guangxi, China (21.57°N, 109.16°E), and animals for real-time quantitative PCR were obtained from estuary in Qinzhou, Guangxi, China (21.46°N, 108.37°E). No specific permissions were required for sample collection from these locations. Field studies did not involve endangered or protected species. Animal welfare and experimental procedures were carried out in accordance with the Guide for the Care and Use of Laboratory Animals and the ethical regulations of the Guangxi Mangrove Research Center. The shell length and weight of *G. erosa* samples were listed in **Supplementary Table S1**.

RNA Extraction and Detection

Nine samples (three tissues-hepatopancreas, gill, and muscle, each with three biological replicates) were quickly separated on ice. Total RNA was extracted using Tiangen RNA preparation kits (Tiangen Biotech, Beijing, China) following the manufacturer's protocol. The following methods were used to test the RNA quality. First, the extent of RNA degradation was estimated using agarose gel electrophoresis. Second, RNA purity (OD260/280), concentration, and absorption peaks were determined using a Nanodrop spectrophotometer. Third, RNA integrity was determined using Agilent 2,100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States). Detection indexes included the RIN value, 28S/18S, baseline for spectra, and 5S peak. RNA samples were used to construct cDNA libraries. One microgram of each RNA sample from all samples was added together in a 1:1 ratio and used for PacBio single-molecule long-read sequencing. All nine samples were used for Illumina sequencing.

Library Preparation and Sequencing

PacBio Library Preparation and Sequencing

A Clontech SMARTer[®] PCR cDNA Synthesis Kit (Clontech Laboratories, 634,926, Mountain View, CA, United States) and the BluePippin Size Selection System protocol, as described by Pacific Biosciences (PN 100-092-800-03), were used to prepare the Iso-Seq library according to the Isoform Sequencing protocol (Iso-Seq).

Illumina Library Preparation and Sequencing

The Illumina library was prepared according to the Illumina (NEB, Ipswich, MA, United States) official NEBNext Ultra™ RNA Library Prep Kit (E7530L) protocol. After the sample was qualified, the eukaryotic mRNA was enriched with magnetic beads with oligo (dT). Subsequently, fragmentation buffer was added to break the mRNA into short fragments. Using the mRNA as a template, single stranded cDNA was synthesized with six-base random primers (random hexamers). Buffer, dNTPs, DNA polymerase I, and RNase H were then added to synthesize double stranded cDNA. The double-stranded cDNA was purified with AMPure XP beads and poly(A) tails were added before sequencing adapters were selected by fragment size using AMPure XP beads. The final library was obtained by purification with AMPure XP beads and PCR amplification of the above sequence, after which the Illumina HiSeq 2000 platform was used to generate 150 bp paired-end reads.

Bioinformatics Analysis

Error Correction Using Illumina Reads

Additional nucleotide errors in the consensus reads were corrected using Illumina RNA-seq data with LorDEC software (Salmela and Rivals, 2014). The full-length transcriptome was calibrated using multiple methods to improve data accuracy. The methods used are as follows. 1) The same template was sequenced multiple times to perform in-hole correction for zero-mode waveguide holes to obtain a high-quality Circular Consensus (CCS). 2) Cluster multi-copy transcriptional sequencing data was used to ensure there was no redundancy, perform inter-hole correction of zero-mode waveguide holes, and obtain cluster consensus sequences. 3) Arrow software was used to calibrate the cluster consensus sequence using a non-full-length sequence, and thus establish a polished consensus sequence. 4) The polished consensus sequence derived from the second-generation data was refined using LorDEC software (Salmela and Rivals, 2014), thus further improving sequencing accuracy. LorDEC is a software which uses second generation data to correct the data generated by third generation PacBio sequencing. It employs mixed error correction, and is highly accurate, far more so than other correction software such as LSC (Au et al., 2012) and PacBioToCA (Koren et al., 2012). Any redundancy in the corrected consensus reads was removed using CD-HIT (Fu et al., 2012) (-c 0.95 -T 6 -G 0 -aL 0.00 -aS 0.99) before obtaining the final transcripts for subsequent analysis. This software uses a heuristic algorithm to quickly find highly similar fragments among sequences.

Gene Functional Annotation

Gene functions were annotated using the following databases: NCBI non-redundant protein sequences (NR) (Li et al., 2002), protein family (Pfam) (Finn et al., 2016), Clusters of Orthologous Groups of proteins (KOG/COG) (Tatusov et al., 2000), Swiss-Prot (Bairoch and Apweiler, 2000), Kyoto Encyclopedia of Genes and Genomes Ortholog database (KEGG) (Kanehisa et al., 2004), and Gene Ontology (GO) (Ashburner et al., 2000). We used Diamond BLASTX software and an e-value of “1e-10” when

performing NR, KOG, Swiss-Prot and KEGG database analysis. Hmmscan was used for family identification with Pfam database.

We used TransDecoder (v3.0.0) (Altschul et al., 1997) to predict coding sequence (CDs). The Pfam database was used to identify reliable potential CDs from the transcript sequence.

Long non-coding RNAs (lncRNAs) are a class of RNA molecules whose transcripts exceed 200 nt in length and do not code for proteins. We used the Coding-Non-Coding-Index (CNCI) (Sun et al., 2013), coding potential calculator (CPC) (Kang et al., 2017), Pfam-scan (Finn et al., 2016), and coding potential assessment tool (CPAT) (Wang et al., 2013) as tools to predict the coding potential of the transcripts. We used the CNCI with default parameters. Using the NCBI eukaryotes protein database, we set the e-value to “1e-10” in CPC. We translated each transcript in all three possible frames and used Pfam Scan to crossmatch with the Pfam database and identify sequences coding for any known protein family domains. All Pfam searches were performed using the default parameters of -E 0.001 --domE 0.001. The transcripts not predicted to have coding potential by all four tools above were excluded, and those without coding potential were candidate sets of lncRNAs. We further used the LncTar (Haas et al., 2013) (version 1.0, parameters: ndG value = -0.1) as a tool to predict lncRNA-targeted transcripts.

Defining Alternative Splicing Events

We ran all-vs-all BLAST with high identity settings. BLAST alignments that met all the following criteria were classified as products of candidate alternative splicing (AS) events: 1) Two high-scoring segment pairs (HSPs) exist in the same alignment, with the same forward/reverse direction. 2) One sequence should be continuous or have only a small “overlap” size (smaller than 5 bp); the other one should be distinct, thus creating what is called an “AS Gap.” 3) The continuous sequence should almost completely align with the distinct sequence. 4) The AS gap should be longer than 100 bp and at least 100 bp away from the 3'/5' end.

Gene Expression Quantification

Bowtie2 was used to compare the reads sequenced from each sample with the transcript sequences. Fragments per kilobase of transcript per million mapped reads (FPKM) is the number of reads per million reads that are aligned to a gene per kilobase length. It is a commonly used method for estimating gene expression levels in transcriptome sequencing data analysis. FPKM can eliminate the influence of differences in transcript length and sequencing amount on calculated expression. The calculated level of transcript expression can be directly used to compare transcript expression differences between samples. Gene expression levels were estimated by RNA-Seq by Expectation-Maximization (RSEM) (Li and Dewey, 2011) for each sample, as follows: 1) Clean data was mapped back onto the transcript sequences, 2) The read count for each transcript was obtained from the mapping results and 3) The read count was transformed to fragments per kilobase million (FPKM) (Zhao et al., 2021). Bowtie2 of the comparison software in RSEM was set to enter an

end-to-end and sensitive mode, and the other parameters were set to default.

Gene Set Enrichment Analysis

We used Molecular Signatures DataBase (MSigDB) software (Liberzon et al., 2015) to measure the statistical enrichment of the genes. KEGG, GO, KOG, and Pfam databases were used to categorize transcripts into different gene classes. We then performed Fisher's exact test on each class in the ordered genes or the user-specified gene list. $p < 0.05$ was the threshold which defined significant enrichment.

Gene Co-Expression Analysis

The weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) approach is a popular method used to identify co-expressed and hub genes. WGCNA has been widely used to identify co-expressed gene modules and to find associations between modules and traits (Li et al., 2017). Based on the gene expression data, this software was used to analyze associations between modules and tissues (traits). In undirected networks, genes within the same module are highly related. After clustering genes into modules, each module was analyzed in two ways, which were described as follows: 1) Functional enrichment analysis was used to determine whether functional characteristics and traits that were identified were consistent with the aims of our research, and 2) Correlation analysis was conducted to determine the module that had the highest correlation with traits of interest. For all samples, we set power to 12, minModuleSize to 30, and maxBlockSize to the number of genes.

Quantitative Real-Time PCR Analysis

Twelve genes were randomly chosen to validate the RNA-seq data. The animal samples used for Real-Time PCR analysis were collected from another location distinct from the samples using for sequencing. Four tissues-hepatopancreas, gill, muscle and the female gonad were quickly separated on ice. Total RNA for qRT-PCR were isolated from four tissues using the Tiangen RNA preparation kits (Tiangen Biotech, Beijing, China) as previously described. First-strand cDNA was reverse transcribed using the PrimeScript™ RT reagent Kit with gDNA Eraser (Takara, Dalian, China). qRT-PCR was performed using TB Green Premix Ex Taq (Tli RNaseH Plus; Takara, Dalian, China). In the past experiment of our lab, to selected the suitable reference genes of different tissues in *G. erosa* by using qRT-PCR method, we compared four candidate genes, which were *β-actin*, *18S rRNA*, *GAPDH* and *α-tubulin*, and the *α-tubulin* was the most stable expressed gene in different tissues. Thus, we use *α-tubulin* as the reference gene in this study. All primers of selected genes were designed using Primer3Plus and are listed in **Supplementary Table S2**. The conditions for qRT-PCR amplification were as follows: incubation at 95°C for 30 s, 40 cycles of 95°C for 5 s and 60°C for 34 s. The specificity of primer amplicons was tested using a melting curve analysis and PCR products were verified by sequencing. The amplification was carried out in a QuantStudio3 (ABI, United States) containing three biological replicates and three technical replicates.

Transcriptome Coverage Assessment

The completeness of the transcriptome is difficult to assess due to the lack of a reference genome. The current common solution is to use the coverage of single-copy genes necessary for life to check the coverage of the transcriptome such as BUSCO (Benchmarking Universal Single-Copy Orthologs) (Manni et al., 2021). BUSCO (v5.3.0) is an open-source python software that evaluates genome assembly or transcripts based on gene evolution (with reference alignment). We use eukaryotes and mollusks as references, respectively, to evaluate our assembly results.

RESULTS

Polymerase Read Statistics

The PacBio RS II used for SMRT sequencing contained 150,000 ZMWs (zero-mode waveguides) per cell. The reads were provided to the ZMW well for sequencing. One read (P1) per ZMW is valid. A total of 450,876 polymerase read sequences were obtained. Polymerase read fragments less than 50 bp and sequence accuracy less than 0.75 were excluded. The remaining reads were removed from the linker and the linker sequence was filtered out to obtain subreads (**Table 1**). Using this process, 3,847,562 subreads were obtained.

Transcript Classification

Transcripts were classified as CDS or lncRNA sequences, based on their predicted protein coding ability. All transcript sequences are listed in **Supplementary Table S5**. A total of 19,411 ORFs were obtained, of which 14,496 were complete (**Figure 1**, **Supplementary Table S3**). A total of 3,870 lncRNA transcripts were identified, of which 978 contained target genes (**Supplementary Table S3**).

Alternative Splicing

The pre-mRNA (pre-mRNA) generated by gene transcription has a variety of splice forms. Different exons are selected to produce different mature mRNAs, which are translated into different proteins, which constitute the diversity of biological traits. This post-transcriptional mRNA processing is called alternative splicing. A total of 31,881 consensus isoforms were obtained. 271 alternative splicing events were identified (**Supplementary Table S3**).

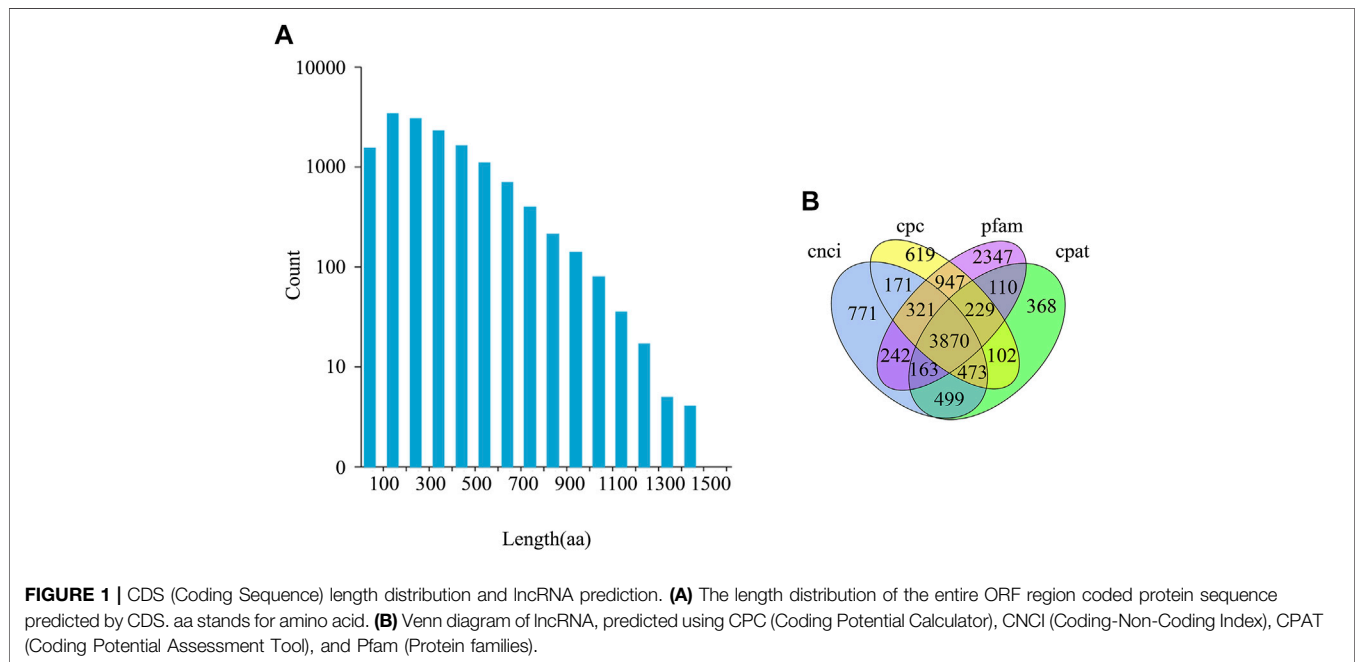
Transcriptome Coverage

In eukaryotes we annotated 183 (71.8%) Complete BUSCOs (C), of which 118 (46.3%) Complete and single-copy BUSCOs (S) and 65 (25.5%) Complete and duplicated BUSCOs (D). In addition, there are 13 (5.1%) Fragmented BUSCOs (F) and 59 (23.1%) Missing BUSCOs (M).

In the Mollusca phylum we annotated 2,298 (43.4%) Complete BUSCOs (C), of which, 1,526 (28.8%) Complete and single-copy BUSCOs (S) and 772 (14.6%) Complete and duplicated BUSCOs (D). In addition, there are 118 (2.2%) Fragmented BUSCOs (F) and 2,879 (54.4%) Missing BUSCOs (M).

TABLE 1 | Data statistics from Pacbio transcriptome sequencing.

cDNA Size	1–2K	2–3K	3–6K
Polymerase Reads	150,292	150,292	150,292
Polymerase Reads	107,731	115,408	108,719
Total Number of Subread Bases	3,099,865,656	3,208,845,527	3,050,228,963
Number of Subread	1,848,503	1,096,181	902,878
Subreads N50	1,696	3,050	3,835
Mean Subread length	1,676	2,927	3,378

**TABLE 2** | Annotated transcript number statistics table.

Database	Number of Annotated Transcripts
All	17,541
GO	6,002
KEGG	8,329
KOG	11,384
Pfam	14,207
Swissprot	9,666
COG	5,421
NR	16,985

"Database" indicates the database used for functional annotations. "Annotated number" refers to the number of transcripts used to obtain this annotation from the corresponding database.

Since this is transcriptome research and not all tissues, not all genes are expressed. The BUSCO completeness is relatively low compared to the genome (Schaarschmidt et al., 2020). The BUSCO completeness of our transcriptome is around 71%, which is similar to the European starling research (around 63%) (Stuart et al., 2021) by ISO-seq. The proportion of fragmented genes is very small (5.1%), indicating that our assembly results are relatively reliable.

Gene Function Annotation

Due to the lack of corresponding genome annotation information for *G. erosa*, we annotated each transcript in seven databases. Genes were annotated by information of homologous proteins (NR, Swissprot), orthologous proteins (COG/KOG), protein domain information (Pfam), gene functions and involved pathways, biological processes, etc. (GO, KEGG). Eight databases were used to annotate gene functions: GO, KEGG, COG/KOG, Swissprot, NR, and Pfam. A total of 17,541 genes were annotated (Table 2; Supplementary Table S3). Among them, the NR database annotated the most transcripts (16,985), while the transcripts annotated by GO were relatively few (6,002).

Gene Expression Analysis

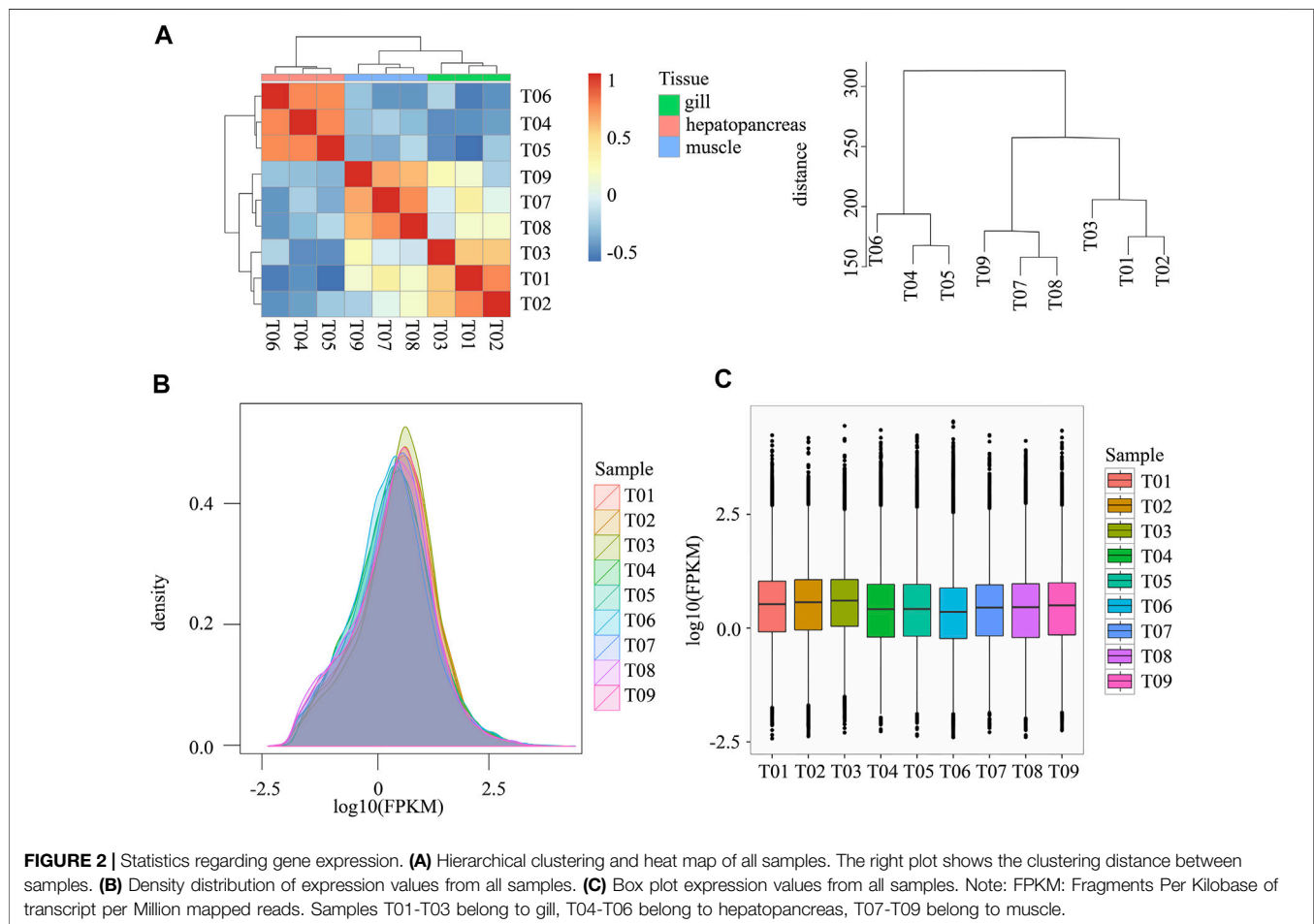
Bowtie2 was used to compare reads obtained by sequencing each sample with the transcript sequence (Table 3). Final quantitation of transcript expression was obtained and converted to FPKM (Figure 2; Supplementary Table S4). The expression levels of the randomly selected genes using qRT-PCR analysis matched these of high throughput sequencing data.

We adopted the Pearson Correlation Coefficient r as the correlation evaluation index of biological replicate

TABLE 3 | Comparison of results derived from second-generation sequencing vs. full-length transcripts.

BMK-ID	Total reads	Mapped Reads (%)	Unique Mapped Reads (%)	Multi Mapped Reads (%)
T01	31,520,554	18,157,222 (57.60%)	5,793,284 (31.91%)	12,363,938 (68.09%)
T02	32,752,380	17,857,615 (54.52%)	5,660,251 (31.70%)	12,197,364 (68.30%)
T03	27,667,629	16,465,587 (59.51%)	5,251,710 (31.90%)	11,213,877 (68.10%)
T04	32,933,308	18,811,133 (57.12%)	6,634,051 (35.27%)	12,177,082 (64.73%)
T05	30,115,694	17,211,031 (57.15%)	6,131,141 (35.62%)	11,079,890 (64.38%)
T06	29,407,809	18,630,763 (63.35%)	6,329,039 (33.97%)	12,301,724 (66.03%)
T07	28,991,037	17,665,871 (60.94%)	4,681,671 (26.50%)	12,984,200 (73.50%)
T08	31,697,911	18,775,882 (59.23%)	5,462,242 (29.09%)	13,313,640 (70.91%)
T09	30,691,848	18,716,281 (60.98%)	5,519,848 (29.49%)	13,196,433 (70.51%)

BMK-ID, refers to sample id. Total reads refer to total reads for each sample. Mapped reads refer to the percentage of reads that aligned to the reference transcript. Unique mapped reads refer to the percentage of mapped reads that were unique. Multi-mapped reads refer to the percentage of mapped reads that were not unique.



correlation. The value range of r is between -1 and 1 , greater than 0 indicates a positive correlation, and less than 0 indicates a negative correlation. Larger absolute values indicate stronger correlations. The gene expression abundances of the nine samples had similar distributions (**Figure 2B**). Different tissues expressed significantly different genes (**Figures 2A,C**). **Figure 2A** shows that differences between tissues are significantly greater than

differences within tissues. The difference in expression of hepatopancreas was greater than that of other tissues.

Differential Expression Statistics

From the gene expression correlation analysis (**Figure 2A**), we found significant differences between different tissues. In the process of differential expression genes (DEGs) analysis, we can do differential analysis between any two tissues. In this paper, we

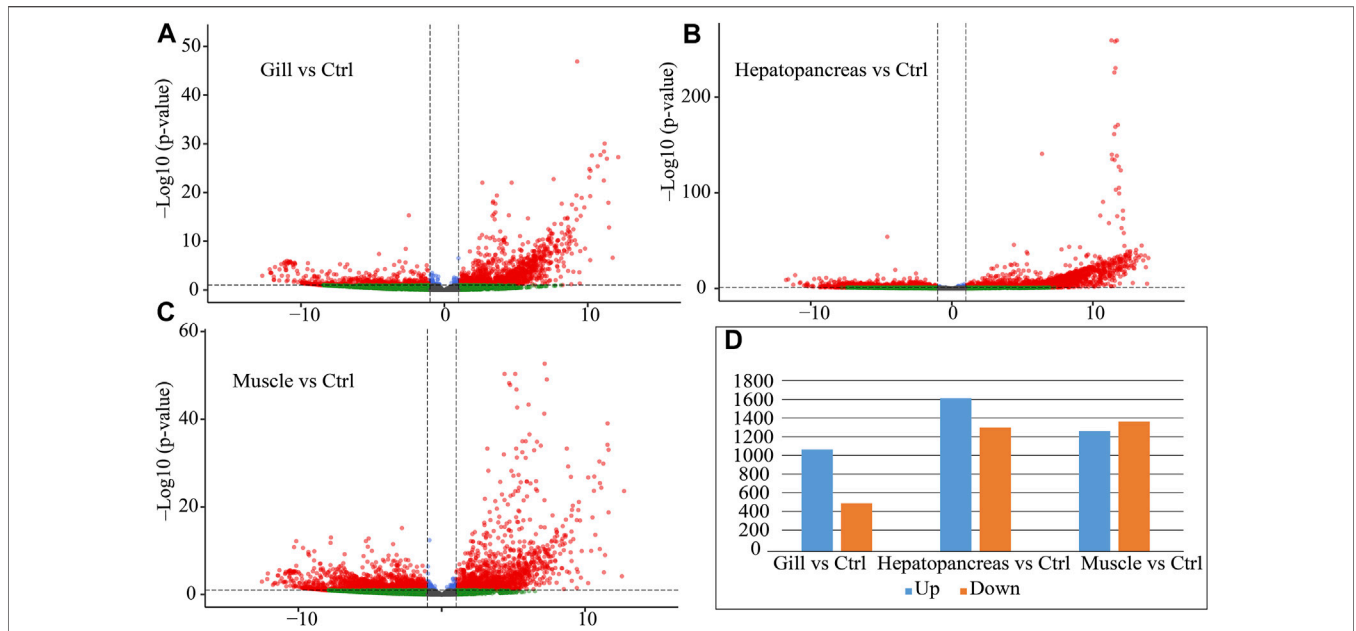


FIGURE 3 | Differential expression statistics for muscle, hepatopancreas and gill tissue. **(A)** Volcano plot of differential expression between gill and control (hepatopancreas and muscle). **(B)** Volcano plot of differential expression between hepatopancreas and control (gill and muscle). **(C)** Volcano plot of differential expression between muscle and control (gill and hepatopancreas). **(D)** Differential expression statistics for each tissue. Note: Ctrl: control; Up: upregulated genes; Down: downregulated genes.

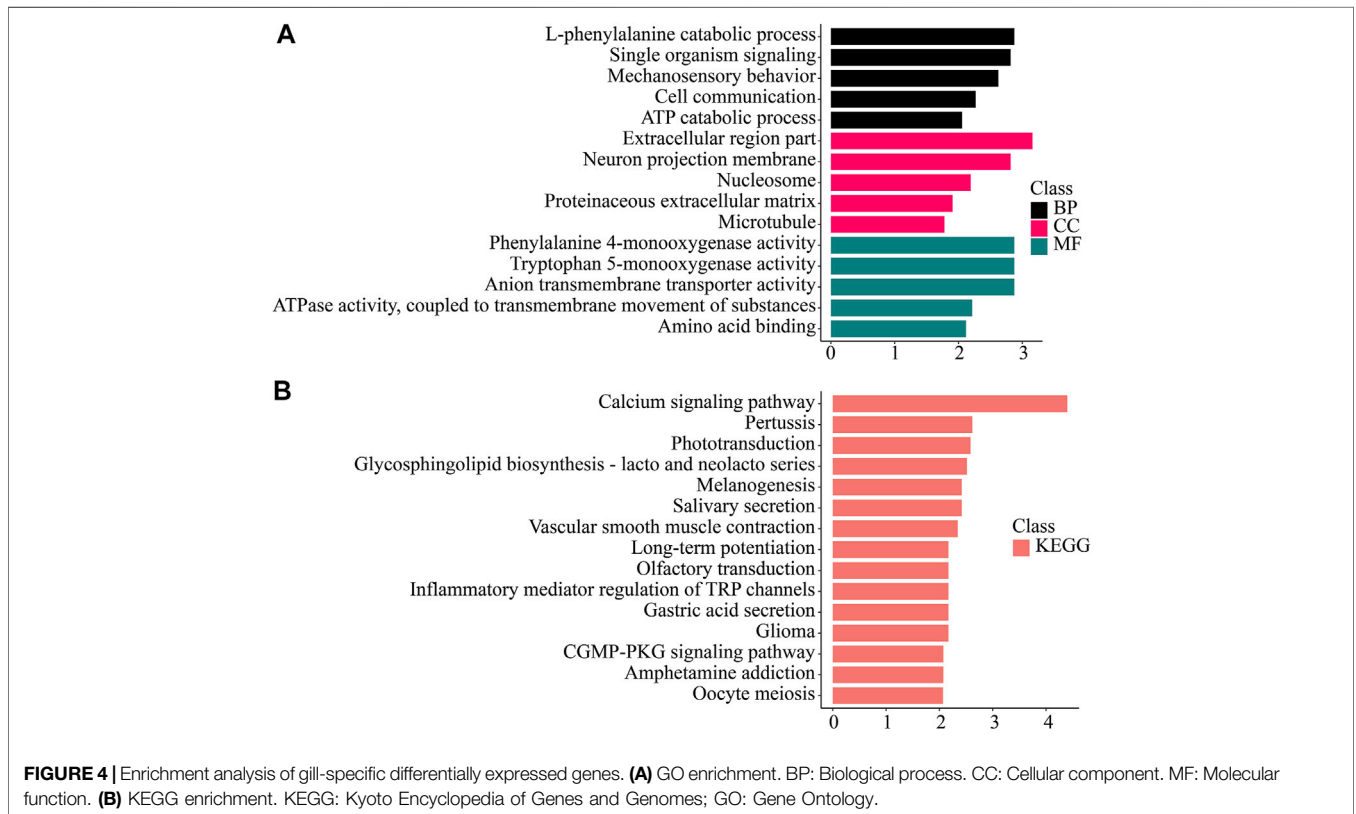
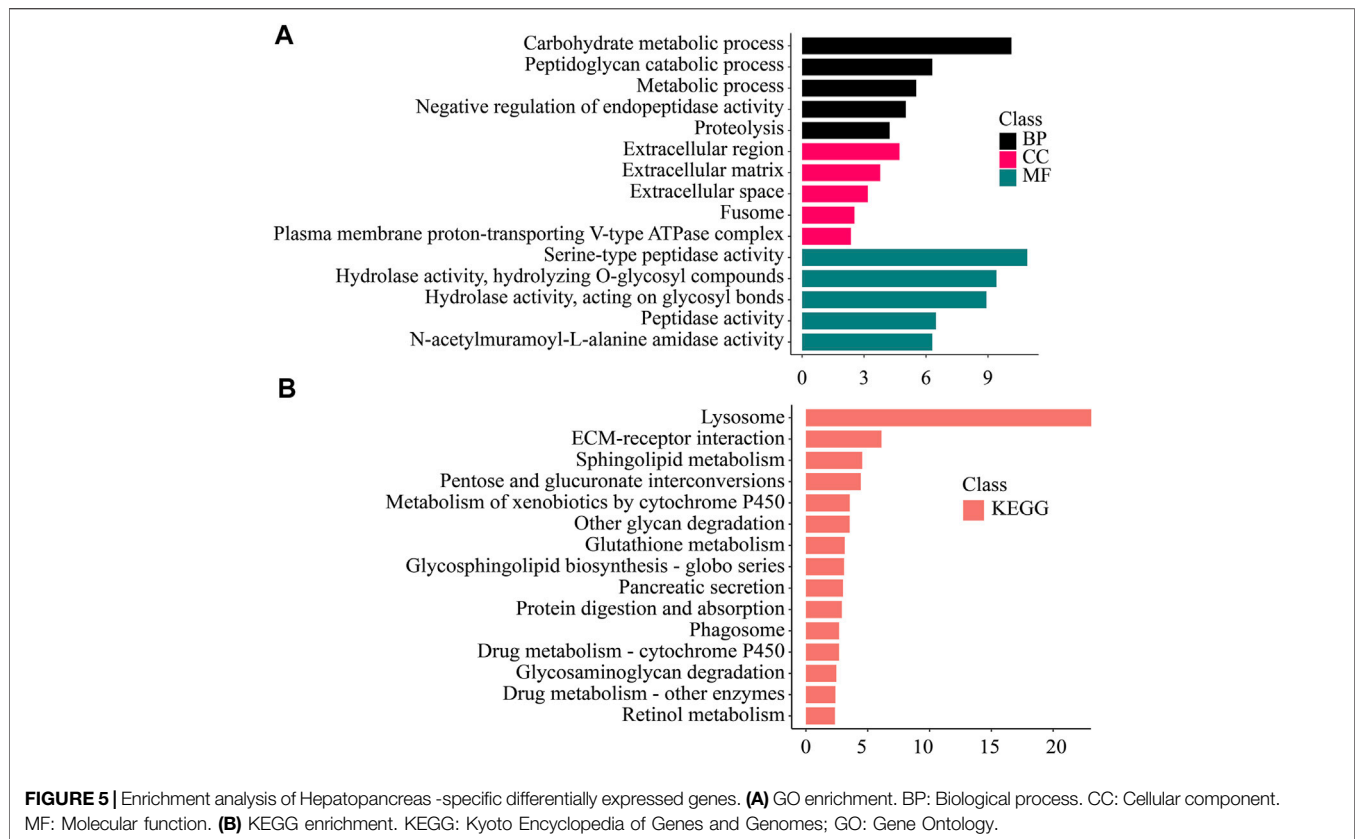


FIGURE 4 | Enrichment analysis of gill-specific differentially expressed genes. **(A)** GO enrichment. BP: Biological process. CC: Cellular component. MF: Molecular function. **(B)** KEGG enrichment. KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene Ontology.



focus on tissue-specific expression. We take the gill, hepatopancreas, and muscle as case groups, respectively, other tissues were used as the control group for DEGs analysis between tissues. We adopted Deseq2 as a differential expression analysis tool.

A total of 22,749 transcripts were quantitatively expressed (**Supplementary Table S5**). Among them, 2,981 genes were differentially expressed between the hepatopancreas and other tissues (gill and muscle), 1,614 genes were differentially expressed between gills and other tissues (hepatopancreas and muscle), and 2,692 genes were differentially expressed between muscle and other tissues (hepatopancreas and gill) (**Figure 3; Supplementary Table S5**).

We found that DEGs have the following characteristics. 1. Hepatopancreas had more DEGs, and their fold changes were also higher (**Figure 3B**, UP: 1,649, DOWN: 1,332). 2. Most of the DEGs in gills were up-regulated (UP: 1,096, DOWN: 518). 3. The number of up-regulated and down-regulated genes in muscle are roughly the same (UP: 1,295, DOWN: 1,397). These results suggest that hepatopancreas and gills may perform specific functions.

Enrichment Analysis and Analysis of Gene Co-Expression Between Tissues

To study the function of the differentially expressed genes (DEGs), we performed GO, KEGG, KOG/COG, and Pfam enrichment analysis on the differentially expressed genes in each tissue (**Figures 4–6; Supplementary Table S5**).

We found that genes significantly up-regulated in gills were enriched in Mechanosensory behavior, L-phenylalanine catabolic process, Phenylalanine 4-monooxygenase activity and other processes and functions. In the KEGG database, Calcium signaling pathway, Phototransduction, Glycosphingolipid biosynthesis-lacto and neolacto series pathway and other pathways were enriched.

Genes that were significantly up-regulated in the hepatopancreas were enriched in various metabolic processes (Carbohydrate metabolic process, Peptidoglycan catabolic process, Proteolysis, Hydrolase activity, Hydrolyzing O-glycosyl compounds). Variety of digestive and metabolic activities (Sphingolipid metabolism, Pancreatic secretion, Protein digestion and absorption), detoxification (Drug metabolism—cytochrome P450, Drug metabolism—other enzymes etc.), xenobiotic degradation (Lysosome, Metabolism of xenobiotics by cytochrome P450), and other pathways were significantly enriched.

In muscle tissue, significantly up-regulated genes were enriched in functions related to muscle such as muscle cell cellular homeostasis, motor activity, calcium ion binding, actin filament binding. In the KEGG pathway, it was related to carbohydrate metabolism (Carbon metabolism, Glycolysis/Gluconeogenesis), muscle activity (Vascular smooth muscle contraction), nerve conduction (Adrenergic signaling in cardiomyocytes, Dopaminergic synapse), and intracellular signal transduction (Calcium signaling pathway,

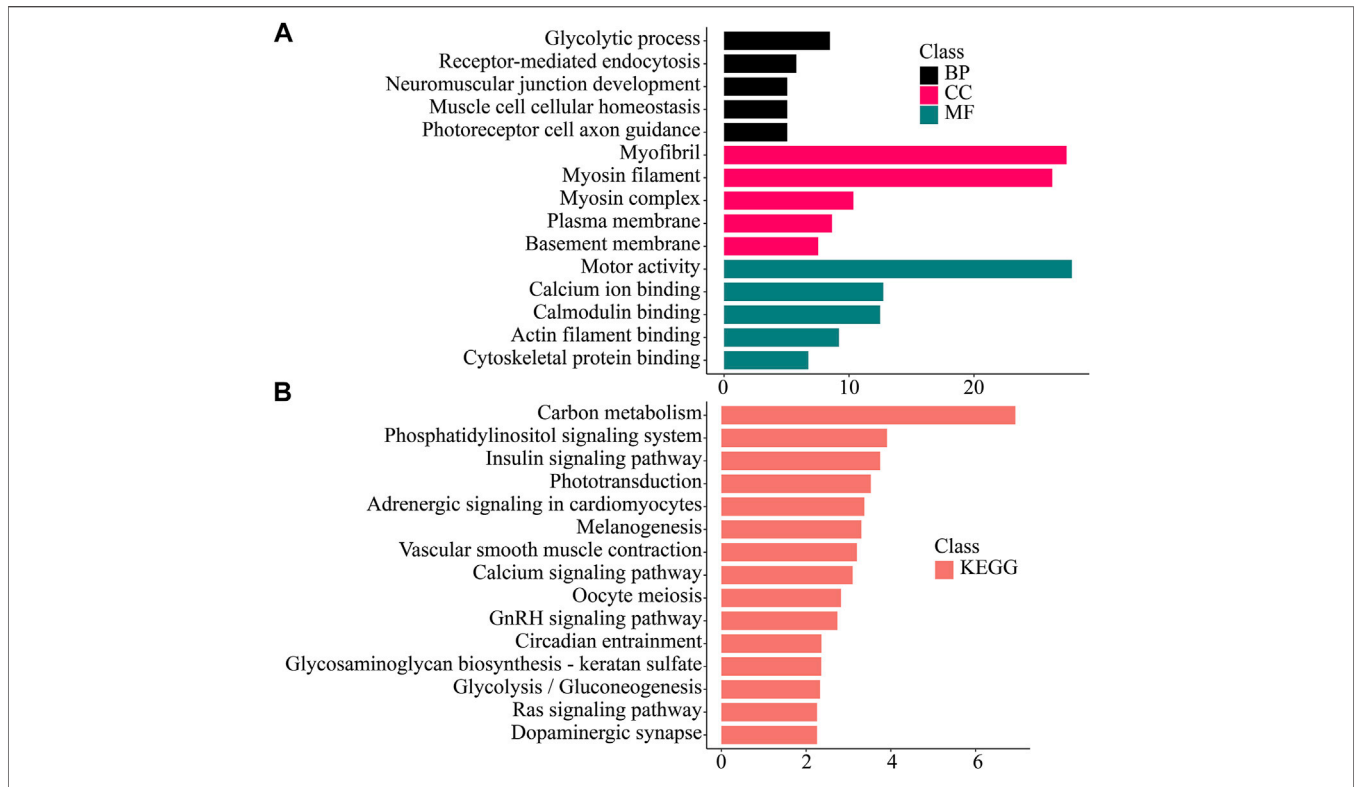


FIGURE 6 | Enrichment analysis of Muscle-specific differentially expressed genes. **(A)** GO enrichment. BP: Biological process. CC: Cellular component. MF: Molecular function. **(B)** KEGG enrichment. KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene Ontology.

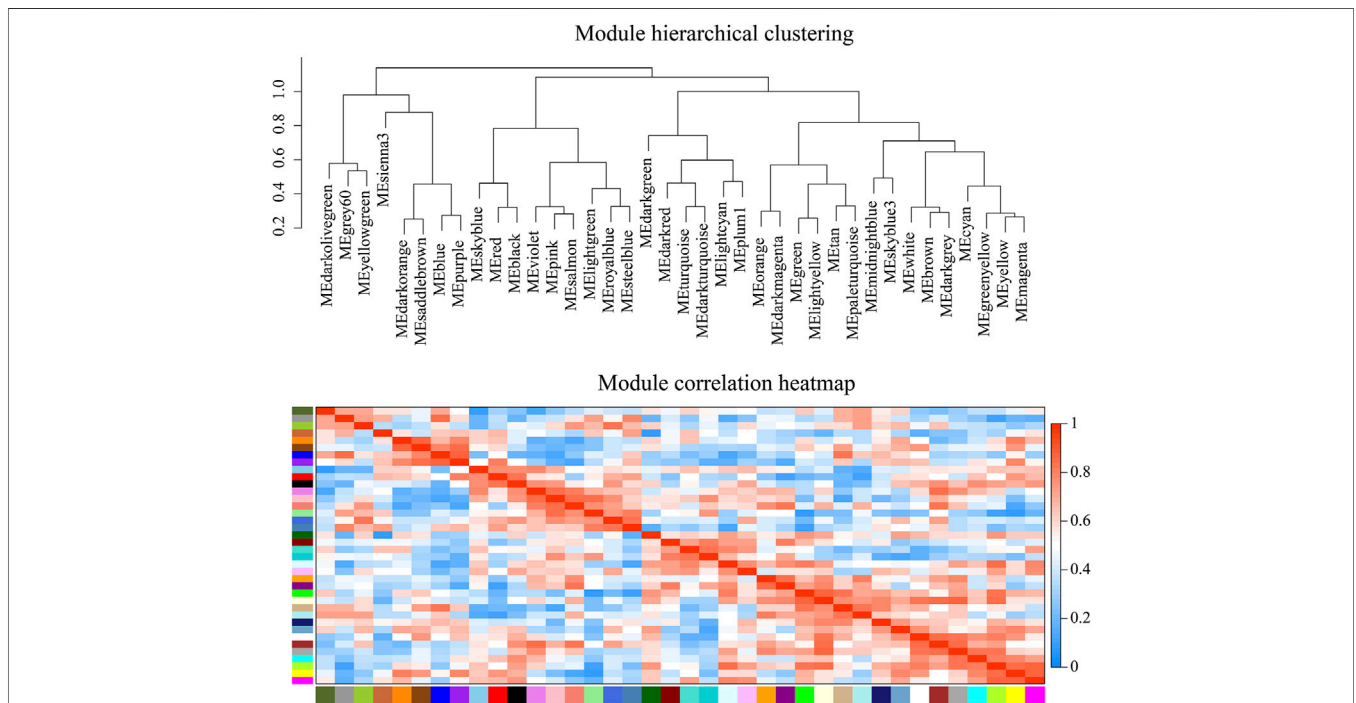
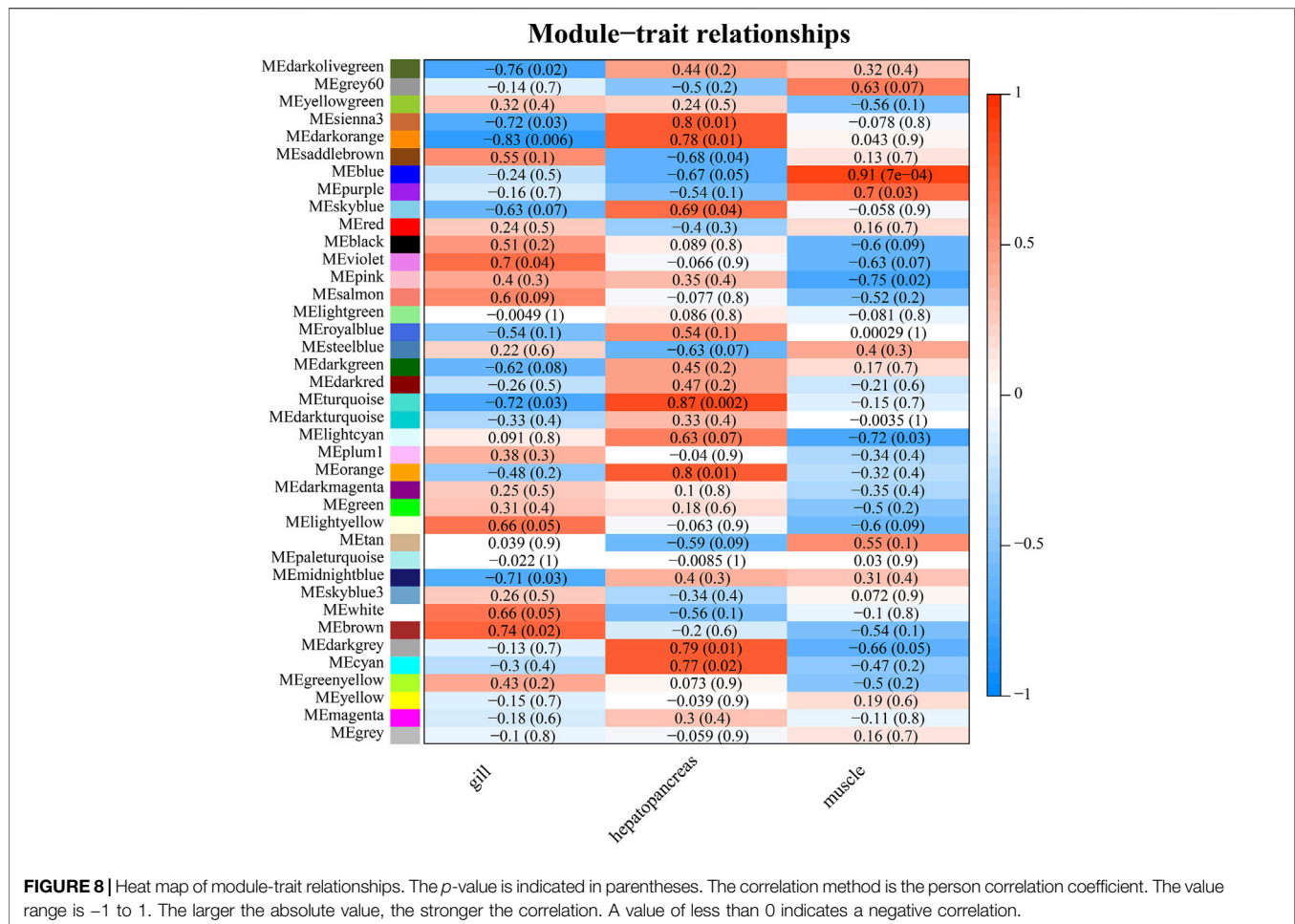


FIGURE 7 | Weighted gene co-expression network analysis (WGCNA) module hierarchical clustering. The value range of correlation is 0–1.



Phosphatidylinositol signaling system, Ras signaling pathway) related pathways were significantly enriched.

The detailed functions involved in the gills, hepatopancreas, muscle, were shown in the Discussion section.

WGCNA analysis can discover gene modules with similar expression patterns, as well as correlations between gene modules and phenotypic properties. By this approach, we identified 38 modules. There were 3,897 genes in the largest module, and 37 genes in the smallest module. The average amount of genes within a module was 588. (Figure 7; Supplementary Table S6). Module-tissue correlation analysis revealed that two modules were significantly positively correlated with gill tissue ($p < 0.05$); eight with hepatopancreas tissue, and two with muscle tissue (Figure 8; Supplementary Table S6).

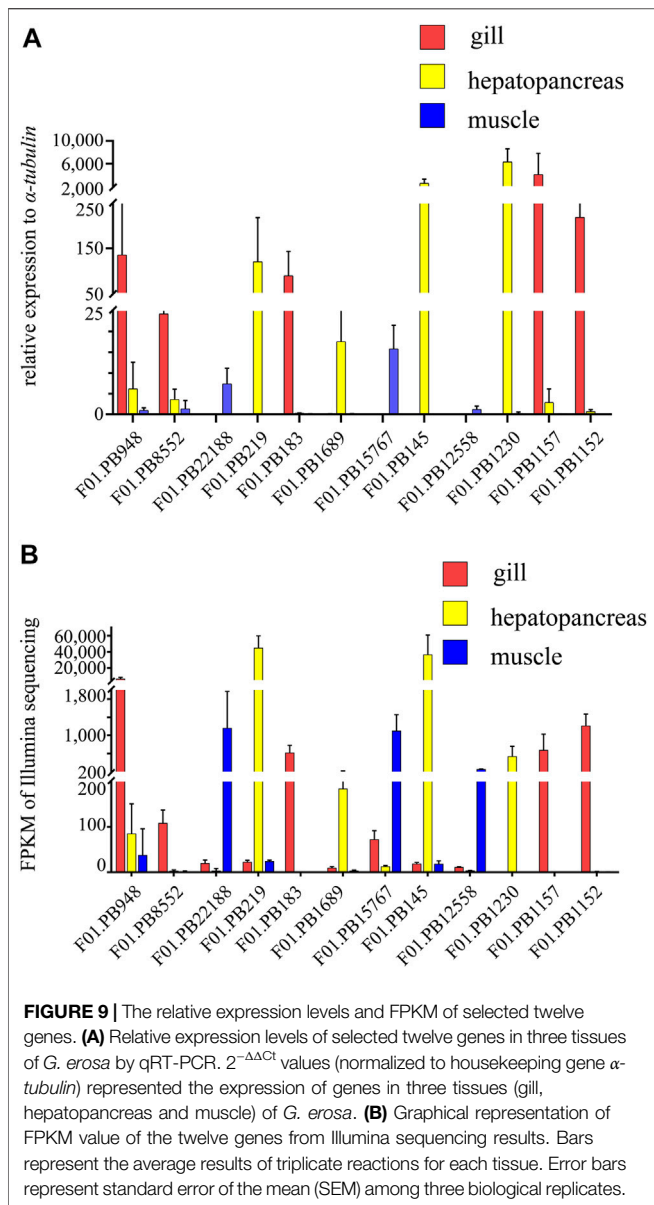
Validation of Differentially Expressed Transcripts of Tissues

Twelve genes among annotated or un-annotated DEGs were randomly selected, and real-time PCR (qRT-PCR) analysis was conducted to validate the differentially expressed transcripts. The total RNA extracted from four tissues of *G. erosa* samples was used for Real-Time PCR analysis. Data were quantified using the

$2^{-\Delta\Delta Ct}$ method based on Ct Values. All the samples were normalized to the housekeeping gene *α -tubulin*, and the concerned three tissues were normalized to the female gonad. The relative expression levels of twelve genes in the three tissues and the FPKM of these genes from RNA-seq results were showed in (Figure 9). All the twelve genes showed consistent qRT-PCR expression patterns as the high throughput sequencing data. The results showed that the results of RNA-seq technique were credible.

DISCUSSION

Geloina erosa plays an important role in mangrove ecology, however, compared to well-studied economic mussels, little attention has been paid to it. Full-length cDNA sequences are basic resources for molecular studies. In this study, full-length transcriptomes of *G. erosa* was sequenced by SMRT and Illumina sequencing were applied for studying the tissue-specific physiological functions. Additionally, the expression levels of randomly selected genes using qRT-PCR analysis validated the credibility of high throughput sequencing data.



Gill-Specific Genes and Functions

The gills evolve from the endoderm (Green et al., 2015; Janvier, 2015; He et al., 2020). The gill is the first barrier in the body of bivalves and is the main point of interaction between the organism and the environment (Cannuel et al., 2009). It contains many filaments that increase its surface area to allow constant exchange of substances between the environment and the body. Because of this, it is exposed to many environmental stresses such as microorganisms, pollutants, pH and changes in salinity (Moreira et al., 2015). In most bivalves, gills are also involved in respiration and cilia suspension feeding. Mussels use the cilia on their gills to pump water through the inhalant siphon into the mantle cavity, where ciliary action along the gills draws water over these filter-feeding organs (Galbraith et al., 2009). Biological processes enriched in the gill may be related to this filtering function, including mechanosensory behavior (Bezares-

Calderón et al., 2020) and cell communication. Cellular components necessary for filtering include a proteinaceous extracellular matrix, microtubules, and kinesin complexes. Molecular functions necessary for filtering include structural components of the cytoskeleton and microtubule motor activity (Gäde, 1981; Dobrzanskaya et al., 2010).

We found that biological processes related to metabolism are enriched in the gills of *G. erosa*, such as inositol biosynthetic and phospholipid biosynthetic processes. In gene co-expression (WGCNA) analysis, we found that the MEbrown module was positively correlated with the gills. After enrichment analysis of all genes in this module, the GO items that were significantly enriched included glycosaminoglycan biosynthesis-keratan sulfate, glycosphingolipid biosynthesis-lacto, and neolacto series.

Gills may play an important role in identifying pathogens and activating immunity (Zhao et al., 2017; Saco et al., 2020). Immune activation leads to activation of phagocytes, generation of reactive oxygen species (ROS), and secretion of immune effectors and cytokines (Soudant et al., 2013). Previous research has shown that gills express many transcripts related to the recognition of antigens (Wang et al., 2008; Gerdol et al., 2011; Romero et al., 2011; Moreira et al., 2015). Monooxygenase is involved in immune responses in *Caenorhabditis elegans* (Wani et al., 2021). In the transcriptome of gill, we found active monooxygenase (phenylalanine 4-monooxygenase activity, tryptophan 5-monooxygenase activity) enriched in GO Molecular function (MF) gene set. We also found that cytochrome P450 CYP2 subfamily were enriched in the KOG database, and cytochrome P450 and multicopper oxidase were enriched in the Pfam database. The glutaredoxin protein is involved in the detoxification of ROS, and in this study, gene F01. PB8852 is homologous to the *glutaredoxin* of a kind of abalone, *Haliotis diversicolor supertexta* (Nam et al., 2016). The result of qRT-PCR showed that, expression of gene F01. PB8852 was abundant in gills. Taken together, these results indicate that gills play an important role in immunity and are likely the first line of defense against infection in *G. erosa*.

Furthermore, we found genes involved in multiple pathways and functions related to neural activity. Using the KEGG database, we identified genes involved in the neurotrophin signaling pathway, long-term potentiation, and dopaminergic. Using the Pfam database, we identified genes related to neurotransmitter gated transmembrane ion channels. Compared with the hepatopancreas and muscle, gill metabolite analysis revealed the presence of more neurotransmitters, such as acetylcholine (Cappello et al., 2018). By subcellular localization, we discovered genes related to neuron projection membranes, glial cell migration, and other neural pathways. In the Pfam database, the region corresponding to transmembrane neurotransmitter-gated ion channels was significantly enriched. The acetylcholine receptor region was enriched in the KOG database (**Supplementary Table S3**) (Jørgensen, 1974; Cappello et al., 2015). Together, these results suggest that gill activity is under the control of the nervous system (Carroll and Catapano, 2007).

Hepatopancreatic Specific Genes and Functions

The hepatopancreas is an organ with combined functions of the liver and pancreas (Xu et al., 2021). The vertebrate hepatopancreas is also called the mid-intestinal gland (Rószler, 2014). Research suggests that the invertebrate hepatopancreas is involved in biological processes such as glycogen, calcium, and lipid storage, as well as excretion of nitrogen metabolites. (Sumner, 1965; Rószler, 2014). However, the invertebrate hepatopancreas is not homologous to the vertebrate hepatopancreas, but rather to the vertebrate bilaterian midguts (Hashimshony et al., 2015; Steinmetz et al., 2017). The digestive gland-like system of invertebrates is controlled by the Pdx gene, which belongs to the ParaHox cluster (Fritsch et al., 2016; He et al., 2020).

A variety of metabolic processes, such as metabolism of carbohydrates and lipids, and peptidoglycan catabolism, were enriched in this study. In molecular function (MF), a variety of hydrolase activities were enriched, such as hydrolyzation of O-glycosyl compounds and glycosyl bonds. Oxidoreductase activity and peptidase activity was also enriched. In the KEGG database, we identified genes related to pancreatic secretion, pathways involving the lysosome and phagosome, drug metabolism, cytochrome P450, lipid metabolism (sphingolipid metabolism, glycosphingolipid biosynthesis-globo series), amino acid metabolism (glycine, serine, and threonine metabolism), sugar metabolism (pentose and glucuronate interconversions, other glycan degradation, starch and sucrose metabolism), and vitamin ligand metabolism (folate biosynthesis, arachidonic acid metabolism, pantothenate and CoA biosynthesis). A variety of enzymes related to protein metabolism were found in the KOG database: trypsin, zinc carboxypeptidase, cysteine proteinase cathepsin L, hydrolytic enzymes of the alpha/beta hydrolase fold, glutamate/aspartate, and neutral acid transporters. In the Pfam database, we found that the papain family cysteine protease, trypsin, glycosyl hydrolase family 9, peptidase C1-like family, and peptidase family M13 were related to protein metabolism. Lipase is involved in lipid metabolism. After WGCNA modular analysis, the MEturquoise module was found to be positively correlated with the hepatopancreas. This module was enriched in pentose and glucuronate interconversions, glycine, serine, and threonine metabolism, pancreatic secretion, protein digestion, and absorption. GO analysis also revealed pathways related to carbohydrate, protein, and lipid metabolism. We also found functions and pathways related to chitin metabolism, such as the chitin-binding peritrophin-A domain, chitinase.

The hepatopancreas is an important organ involved in the immune response, as well as responses to oxidative and heat stress in mollusks and crustaceans (Rószler, 2014). Among the differentially expressed genes in the hepatopancreas, many belong to the monooxygenase family (Moore et al., 1980), such as peptidylglycine alpha-amidating monooxygenase, beta, beta-carotene 15,15'-dioxygenase and related enzymes, and copper type II ascorbate-dependent monooxygenase. The result of qRT-PCR showed that, expression of gene F01.PB219 and F01.PB145 were abundant in hepatopancreas.

Gene F01.PB219 is homologous to *lysozyme* of a clam, *Mactra quadrangularis*, and the lysozyme protein has antibacterial activity against a number of bacterial species. Gene F01.PB145 is homologous to the *defensin* of a kind of mussel, *Dreissena polymorpha*, and the defensins are members of a large family of cationic antimicrobial peptides that form an essential element of innate immunity (Gerdol et al., 2020). Taken together, these results suggest that the hepatopancreas plays an important role in metabolism, immune response, and detoxification (Rószler, 2014).

Muscle-Specific Genes and Functions

Bivalve is an important fishery commodity, and muscle is the main edible part of bivalve organisms. Bivalve muscle has a high collagen and low fat content, and is rich in mineral elements (Tabakaeva et al., 2018).

Previous analysis of metabolite content in *G. erosa* has revealed that muscle tissue is far richer in energy metabolites than hepatopancreas or gill tissue. (Cappello et al., 2018). We found that biological processes related to energy metabolism include GTP catabolism, oxidation-reduction reactions, glycolytic processes, and the tricarboxylic acid cycle. Cellular components related to energy metabolism, include mitochondrial respiratory chain complex II. Molecular functions related to energy metabolism include cytochrome-b5 reductase activity, NAD(P)H involvement and malate dehydrogenase activity. We found KEGG pathway related energy metabolism including carbon metabolism, insulin signaling pathway, and glycolysis/gluconeogenesis. In WGCNA analysis, the MEblue module, which is enriched in functions related to glucose metabolism and glycolytic processes, was positively correlated with muscle tissue.

Previous studies have shown that muscles express many myofibril and calcium-related proteins (Moreira et al., 2015). In the current study, many functions related to muscle movement and control have been found to be enriched in *G. erosa*, such as neuromuscular junction development, cellular calcium ion homeostasis, myofibril assembly, striated muscle myosin thick filament assembly, actin nucleation, actin filament severing and capping, and synaptic transmission. Cellular components related to muscle function include myofibrils, myosin filaments, myosin complexes, postsynaptic membranes, axons, Z discs, and myosin IV and V complexes. Molecular functions related to muscle function include motor activity, calcium ion binding, calmodulin binding, actin binding, and neurotransmitter sodium symporter activity. In WGCNA analysis, the MEblue module, which was positively correlated with muscle, was also enriched in adrenergic signaling in cardiomyocytes, vascular smooth muscle contraction, actin crosslink formation, and synaptic growth at the neuromuscular junction. The result of qRT-PCR verified the reliability of sequencing result. The genes F01.PB22188, F01.PB15767 and F01.PB12558 with highly relative expression were respectively homologous to *paramyosin-3*, *myosin heavy chain* and *Filamin-A*.

We also identified pathways and functions related to photosensitivity. Biological processes related to photosensitivity include photoreceptor cell axon guidance, metarhodopsin

inactivation, regulation of light-activated channel activity, and bioluminescence. The MEblue module was found to be enriched for functions related to photosensitivity, including phototransduction, photoreceptor activity, bioluminescence, and photoreceptor cell axon guidance.

The findings can be used to study the tissue-specific functions of *G. erosa* under natural conditions. The full-length transcript data provided by this study will provide a reliable and highly valuable resource for further *G. erosa* alternative splicing and protein structure research. Furthermore, this study may in the future act as a reference point for *G. erosa* transcriptomics. It could be used as a control for future studies which investigate gene expression in *G. erosa* under abnormal environmental stresses and other conditions. However, there are several disadvantages in our study. Firstly, our results are for gene expression in different tissues under normal conditions, which may be different for transcripts under other conditions such as environmental stress. Secondly, our results are based on bulk RNA-seq results with limited resolution. In the future, we will use single cell sequencing to quantify the gene expression of *G. erosa*.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the National Center for Biotechnology Information (NCBI) in Bioproject: PRJNA544778, with ID: SAMN21552671 to SAMN21552679 (T01 to T09). All annotation data (**Supplementary Table S3**), full-length sequence of transcripts (**Supplementary Data Sheet 1.ZIP**), expression data (**Supplementary Table S4**), and enrichment analysis results (**Supplementary Table S5**, **Supplementary Table S6**) are shown in the **Supplementary Materials**.

REFERENCES

- Almeida, Â., Silva, M. G., Soares, A. M. V. M., and Freitas, R. (2020). Concentrations Levels and Effects of 17alpha-Ethinylestradiol in Freshwater and marine Waters and Bivalves: A Review. *Environ. Res.* 185, 109316. doi:10.1016/j.envres.2020.109316
- Altschul, S., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* 25, 3389–3402. doi:10.1093/nar/25.17.3389
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: Tool for the Unification of Biology. *Nat. Genet.* 25, 25–29. doi:10.1038/75556
- Au, K. F., Underwood, J. G., Lee, L., and Wong, W. H. (2012). Improving PacBio Long Read Accuracy by Short Read Alignment. *PLoS One* 7, e46679. doi:10.1371/journal.pone.0046679
- Bairoch, A., and Apweiler, R. (2000). The SWISS-PROT Protein Sequence Database and its Supplement TrEMBL in 2000. *Nucleic Acids Res.* 28, 45–48. doi:10.1093/nar/28.1.45
- Bertucci, A., Pierron, F., Gourves, P.-Y., Klopp, C., Lagarde, G., Pereto, C., et al. (2018). Whole-transcriptome Response to Wastewater Treatment Plant and Stormwater Effluents in the Asian Clam, *Corbicula fluminea*. *Ecotoxicology Environ. Saf.* 165, 96–106. doi:10.1016/j.ecoenv.2018.08.090

AUTHOR CONTRIBUTIONS

XL, CH, and ZL contributed to conception and design of the study. YL, MY, WL, and TH organized the database. YL, and YW performed the statistical analysis. YL wrote the first draft of the manuscript. XL, and YL wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (32060282 and 62106066); Natural Science Foundation of Guangxi Province (2019GXNSFBA185027); Open Research Fund Program of Guangxi Key Lab of Mangrove Conservation and Utilization (GKLMC-202002).

ACKNOWLEDGMENTS

We acknowledge the Biomarker Corporation (Beijing, China) for the facilities and expertise of PacBio platform for libraries construction and sequencing. We would like to thank Editage (www.editage.com) for English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2022.851957/full#supplementary-material>

- Bezares-Calderón, L. A., Berger, J., and Jékely, G. (2020). Diversity of Cilia-Based Mechanosensory Systems and Their Functions in marine Animal Behaviour. *Phil. Trans. R. Soc. B* 375, 20190376. doi:10.1098/rstb.2019.0376
- Cannuel, R., Beninger, P. G., McCombie, H., and Boudry, P. (2009). Gill Development and Evolutionary Implications in the Blue Mussel *Mytilus Edulis* (Bivalvia: Mytilidae). *Biol. Bull.* 217, 173–188. doi:10.1086/BBLv217n2p173
- Cappello, T., Giannetto, A., Parrino, V., Maisano, M., Oliva, S., De Marco, G., et al. (2018). Baseline Levels of Metabolites in Different Tissues of Mussel *Mytilus galloprovincialis* (Bivalvia: Mytilidae). *Comp. Biochem. Physiol. D: Genomics Proteomics* 26, 32–39. doi:10.1016/j.cbd.2018.03.005
- Cappello, T., Maisano, M., Giannetto, A., Parrino, V., Mauceri, A., and Fasulo, S. (2015). Neurotoxicological Effects on marine Mussel *Mytilus galloprovincialis* Caged at Petrochemical Contaminated Areas (Eastern Sicily, Italy): 1H NMR and Immunohistochemical Assays. *Comp. Biochem. Physiol. C: Toxicol. Pharmacol.* 169, 7–15. doi:10.1016/j.cbpc.2014.12.006
- Carroll, M. A., and Catapane, E. J. (2007). The Nervous System Control of Lateral Ciliary Activity of the Gill of the Bivalve Mollusc, *Crassostrea virginica*. *Comp. Biochem. Physiol. A: Mol. Integr. Physiol.* 148, 445–450. doi:10.1016/j.cbpa.2007.06.003
- Chatterji, A., Ansari, Z. A., Ingole, B. S., Bichurina, M. A., Sovetova, M., and Boikov, Y. A. (2002). Indian marine Bivalves: Potential Source of Antiviral Drugs. *Curr. Sci.* 82.
- Dobrzanskaya, A. V., Matusovskaya, G. G., Matusovsky, O. S., and Shelud'ko, N. S. (2010). Thin Filaments of Bivalve Smooth Muscle May Contain a Calponin-like Protein. *Biophysic* 55, 703–706. doi:10.1134/S0006350910050039

- Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam Protein Families Database: Towards a More Sustainable Future. *Nucleic Acids Res.* 44, D279–D285. doi:10.1093/nar/gkv1344
- Fritsch, M., Wollesen, T., and Wanninger, A. (2016). Hox and ParaHox Gene Expression in Early Body Plan Patterning of Polyplacophoran Mollusks. *J. Exp. Zool. (Mol. Dev. Evol.)* 326, 89–104. doi:10.1002/jez.b.22671
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: Accelerated for Clustering the Next-Generation Sequencing Data. *Bioinformatics* 28, 3150–3152. doi:10.1093/bioinformatics/bts565
- Gäde, G. (1981). Energy Production during Swimming in the Adductor Muscle of the Bivalve Lima Hians: Comparison with the Data from Other Bivalve Mollusks. *Physiol. Zoolog.* 54, 400–406. doi:10.1086/physzool.54.4.30155832
- Galbraith, H. S., Frazier, S. E., Allison, B., and Vaughn, C. C. (2009). Comparison of Gill Surface Morphology across a Guild of Suspension-Feeding Unionid Bivalves. *J. Molluscan Stud.* 75, 103–107. doi:10.1093/mollusc/eyn045
- Gerdol, M., Manfrin, C., De Moro, G., Figueras, A., Novoa, B., Venier, P., et al. (2011). The C1q Domain Containing Proteins of the Mediterranean Mussel *Mytilus galloprovincialis*: A Widespread and Diverse Family of Immune-Related Molecules. *Dev. Comp. Immunol.* 35, 635–643. doi:10.1016/j.dci.2011.01.018
- Gerdol, M., Schmitt, P., Venier, P., Rocha, G., Rosa, R. D., and Destoumieux-Garzon, D. (2020). Functional Insights from the Evolutionary Diversification of Big Defensins. *Front. Immunol.* 11. doi:10.3389/fimmu.2020.00758
- Green, S. A., Simoes-Costa, M., and Bronner, M. E. (2015). Evolution of Vertebrates as Viewed from the Crest. *Nature* 520, 474–482. doi:10.1038/nature14436
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De Novo transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis. *Nat. Protoc.* 8, 1494–1512. doi:10.1038/nprot.2013.084
- Hashimshony, T., Feder, M., Levin, M., Hall, B. K., and Yanai, I. (2015). Spatiotemporal Transcriptomics Reveals the Evolutionary History of the Endoderm Germ Layer. *Nature* 519, 219–222. doi:10.1038/nature13996
- He, C., Han, T., Liao, X., Guan, R., Chen, J.-Y., Tremblay, K. D., et al. (2020). On the Origin of Vertebrate Body Plan: Insights from the Endoderm Using the Hourglass Model. *Gene Expr. Patterns* 37, 119125. doi:10.1016/j.gep.2020.119125
- Hu, M., Zheng, X., Fan, C.-M., and Zheng, Y. (2020). Lineage Dynamics of the Endosymbiotic Cell Type in the Soft Coral *Xenia*. *Nature* 582, 534–538. doi:10.1038/s41586-020-2385-7
- Ingle, B. S., Naik, S., Furtado, R., Ansari, Z. A., and Chatterji, A. (2002). *Population Characteristics of the Mangrove Clam Polymesoda (Geloína) Erosa (Solander, 1786) in the Chorao Mangrove, Goa, India*: Indian Counc. Agric. Res. Old Goa.
- Janvier, P. (2015). Facts and Fancies about Early Fossil Chordates and Vertebrates. *Nature* 520, 483–489. doi:10.1038/nature14437
- Jørgensen, C. B. (1974). On Gill Function in the mussel *Mytilus Edulis* L. *Ophelia* 13, 187–232. doi:10.1080/00785326.1974.10430595
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG Resource for Deciphering the Genome. *Nucleic Acids Res.* 32, 277D–280D. doi:10.1093/nar/gkh063
- Kang, Y.-J., Yang, D.-C., Kong, L., Hou, M., Meng, Y.-Q., Wei, L., et al. (2017). CPC2: A Fast and Accurate Coding Potential Calculator Based on Sequence Intrinsic Features. *Nucleic Acids Res.* 45, W12–W16. doi:10.1093/nar/gkx428
- Kim, B.-M., Ahn, D.-H., Kim, H., Lee, J. S., Rhee, J.-S., and Park, H. (2019). Transcriptome Profiling Suggests Roles of Innate Immunity and Digestion Metabolism in Purplish Washington Clam. *Genes Genom* 41, 183–191. doi:10.1007/s13258-018-0750-9
- Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., et al. (2012). Hybrid Error Correction and De Novo Assembly of Single-Molecule Sequencing Reads. *Nat. Biotechnol.* 30, 693–700. doi:10.1038/nbt.2280
- Lai, T., He, B., Fan, H., Zhou, R., and Yang, Y. (2011). Effects of Cadmium Stress on the Activities of Antioxidant Enzymes, Digestive Enzymes and the Membrane Lipid Peroxidation of the Mangrove Mud Clam *Geloina Coaxans* (Gmelin). *Shengtai Xuebao/Acta Ecol. Sin.* 31.
- Langfelder, P., and Horvath, S. (2008). WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9. doi:10.1186/1471-2105-9-559
- Levy, S., Elek, A., Grau-Bové, X., Menéndez-Bravo, S., Iglesias, M., Tanay, A., et al. (2021). A Stony Coral Cell Atlas Illuminates the Molecular and Cellular Basis of Coral Symbiosis, Calcification, and Immunity. *Cell* 184, 2973–2987. e18. doi:10.1016/j.cell.2021.04.005
- Li, B., and Dewey, C. N. (2011). RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinformatics* 12. doi:10.1186/1471-2105-12-323
- Li, W., Jaroszewski, L., and Godzik, A. (2002). Tolerating Some Redundancy Significantly Speeds up Clustering of Large Protein Databases. *Bioinformatics* 18, 77–82. doi:10.1093/bioinformatics/18.1.77
- Li, Z., Han, X., Song, X., Zhang, Y., Jiang, J., Han, Q., et al. (2017). Overexpressing the Sedum Alfredii Cu/Zn Superoxide Dismutase Increased Resistance to Oxidative Stress in Transgenic Arabidopsis. *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.01010
- Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* 1, 417–425. doi:10.1016/j.cels.2015.12.004
- Liu, F., Li, Y., Yu, H., Zhang, L., Hu, J., Bao, Z., et al. (2021). MolluscDB: An Integrated Functional and Evolutionary Genomics Database for the Hyper-Diverse Animal Phylum Mollusca. *Nucleic Acids Res.* 49, D988–D997. doi:10.1093/nar/gkaa918
- Ma, B., Ran, Z., Xu, X., Xu, J., Liao, K., Cao, J., et al. (2019). Comparative Transcriptome Analyses Provide Insights into the Adaptation Mechanisms to Acute Salt Stresses in Juvenile *Sinonovacula Constricta*. *Genes Genom* 41, 599–612. doi:10.1007/s13258-019-00805-x
- Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 38, 4647–4654. doi:10.1093/molbev/msab199
- Modassir, Y. (2000). Effect of Salinity on the Toxicity of Mercury in Mangrove Clam, *Polymesoda Erosa* (Lightfoot 1786). *Asian Fish. Sci.* 13.
- Moore, M. N., Livingstone, D. R., Donkin, P., Bayne, B. L., Widdows, J., and Lowe, D. M. (1980). Mixed Function Oxygenases and Xenobiotic Detoxication/toxication Systems in Bivalve Molluscs. *Helgolander Meeresunters* 33, 278–291. doi:10.1007/BF02414753
- Moreira, R., Pereira, P., Canchaya, C., Posada, D., Figueras, A., and Novoa, B. (2015). RNA-seq in *Mytilus galloprovincialis*: Comparative Transcriptomics and Expression Profiles Among Different Tissues. *BMC Genomics* 16. doi:10.1186/s12864-015-1817-5
- Morton, B. (1976). The Biology and Functional Morphology of the Southeast Asian Mangrove Bivalve, *Polymesoda (Geloína) Erosa* (Solander, 1786) (Bivalvia: Corbiculidae). *Can. J. Zool.* 54, 482–500. doi:10.1139/z76-055
- Morton, B. (1985). The Reproductive Strategy of the Mangrove Bivalve *Polymesoda-Erosa* Bivalvia Corbiculoidea in Hong-Kong. *Malacol. Rev.* 18.
- Nam, B.-H., Jung, M., Subramaniyam, S., Yoo, S.-i., Markkandan, K., Moon, J.-Y., et al. (2016). Transcriptome Analysis Revealed Changes of Multiple Genes Involved in *haliotis* Discus Hannai Innate Immunity during vibrio Parahemolyticus Infection. *PLoS One* 11, e0153474. doi:10.1371/journal.pone.0153474
- Romero, A., Dios, S., Poisa-Beiro, L., Costa, M. M., Posada, D., Figueras, A., et al. (2011). Individual Sequence Variability and Functional Activities of Fibrinogen-Related Proteins (FREPs) in the Mediterranean Mussel (*Mytilus galloprovincialis*) Suggest Ancient and Complex Immune Recognition Models in Invertebrates. *Dev. Comp. Immunol.* 35, 334–344. doi:10.1016/j.dci.2010.10.007
- Röszer, T. (2014). The Invertebrate Midintestinal Gland (“hepatopancreas”) Is an Evolutionary Forerunner in the Integration of Immunity and Metabolism. *Cell Tissue Res* 358, 685–695. doi:10.1007/s00441-014-1985-7
- Saco, A., Rey-Campos, M., Novoa, B., and Figueras, A. (2020). Transcriptomic Response of Mussel Gills after a *Vibrio Splendidus* Infection Demonstrates Their Role in the Immune Response. *Front. Immunol.* 11. doi:10.3389/fimmu.2020.615580
- Salmela, L., and Rivals, E. (2014). LoRDEC: Accurate and Efficient Long Read Error Correction. *Bioinformatics* 30, 3506–3514. doi:10.1093/bioinformatics/btu538

- Schaarschmidt, S., Fischer, A., Lawas, L. M. F., Alam, R., Septiningsih, E. M., Bailey-Serres, J., et al. (2020). Utilizing Pacbio Iso-Seq for Novel Transcript and Gene Discovery of Abiotic Stress Responses in *Oryza Sativa* L. *Ijms* 21, 8148. doi:10.3390/ijms21218148
- Soudant, P., E. Chu, F.-L., and Volety, A. (2013). Host-parasite Interactions: Marine Bivalve Molluscs and Protozoan Parasites, Perkinsus Species. *J. Invertebr. Pathol.* 114, 196–216. doi:10.1016/j.jip.2013.06.001
- Steinmetz, P. R. H., Aman, A., Kraus, J. E. M., and Technau, U. (2017). Gut-like Ectodermal Tissue in a Sea Anemone Challenges Germ Layer Homology. *Nat. Ecol. Evol.* 1, 1535–1542. doi:10.1038/s41559-017-0285-5
- Stuart, K. C., Edwards, R. J., Cheng, Y., Warren, W. C., Burt, D. W., Sherwin, W. B., et al. (2021). *Transcript- and Annotation-Guided Genome Assembly of the European Starling*. New York, NY: bioRxiv.
- Sumner, A. T. (1965). The Cytology and Histochemistry of the Digestive Gland Cells of Helix. *Q. J. Microsc. Sci.* s3-106, 173–192. doi:10.1242/jcs.s3-106.74.173
- Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing Sequence Intrinsic Composition to Classify Protein-Coding and Long Non-coding Transcripts. *Nucleic Acids Res.* 41, e166. doi:10.1093/nar/gkt646
- Tabakaeva, O. V., Tabakaev, A. V., and Piekoszewski, W. (2018). Nutritional Composition and Total Collagen Content of Two Commercially Important Edible Bivalve Molluscs from the Sea of Japan Coast. *J. Food Sci. Technol.* 55, 4877–4886. doi:10.1007/s13197-018-3422-5
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG Database: A Tool for Genome-Scale Analysis of Protein Functions and Evolution. *Nucleic Acids Res.* 28, 33–36. doi:10.1093/nar/28.1.33
- Twaddle, R. W., Wurster, C. M., Bird, M. I., and Ulm, S. (2017). Complexities in the Palaeoenvironmental and Archaeological Interpretation of Isotopic Analyses of the Mud Shell *Geloina Erosa* (Lightfoot, 1786). *J. Archaeological Sci. Rep.* 12, 613–624. doi:10.1016/j.jasrep.2017.03.010
- Wang, G., Zhang, C., and Huang, B. (2020). Transcriptome Analysis and Histopathological Observations of *Geloina Erosa* Gills upon Cr(VI) Exposure. *Comp. Biochem. Physiol. Part C: Toxicol. Pharmacol.* 231, 108706. doi:10.1016/J.CBPC.2020.108706
- Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J.-P., and Li, W. (2013). CPAT: Coding-Potential Assessment Tool Using an Alignment-free Logistic Regression Model. *Nucleic Acids Res.* 41, e74. doi:10.1093/nar/gkt006
- Wang, N., Lee, Y.-H., and Lee, J. (2008). Recombinant Perlucin Nucleates the Growth of Calcium Carbonate Crystals: Molecular Cloning and Characterization of Perlucin from Disk Abalone, *Haliotis Discus Discus*. *Comp. Biochem. Physiol. B: Biochem. Mol. Biol.* 149, 354–361. doi:10.1016/j.cbpb.2007.10.007
- Wani, K. A., Goswamy, D., Taubert, S., Ratnappan, R., Ghazi, A., and Irazoqui, J. E. (2021). NHR-49/PPAR- α and HLH-30/TFEB Cooperate for *C. elegans* Host Defense via a Flavin-Containing Monooxygenase. *Elife* 10. doi:10.7554/eLife.62775
- Xing, Y. Z., Nong, Y., Lu, Y. Z., Yang, M. L., and Yan, B. (2017). Response Characteristics of Oxidative Stress Biomarkers of *Polymesoda Erosa* to Exposure of SCCPs. *Zhongguo Huanjing Kexue/china Environ. Sci.* 37.
- Xu, F., Marlétaz, F., Gavriouchkina, D., Liu, X., Sauka-Spengler, T., Zhang, G., et al. (2021). Evidence from Oyster Suggests an Ancient Role for Pdx in Regulating Insulin Gene Expression in Animals. *Nat. Commun.* 12. doi:10.1038/s41467-021-23216-7
- Yahya, N., Idris, I., Rosli, N. S., and Bachok, Z. (2018). Population Dynamics of Mangrove Clam, *Geloina Expansa* (Mousson, 1849) (Mollusca, Bivalvia) in a Malaysian Mangrove System of South China Sea. *J. Sustain. Sci. Manag.* 13.
- Yingya, C., Xianhu, H., and Dongke, W. (1995). Studies on the Ecology of *Polymesoda Erosa* (Solander). *Trop. Oceanol.* 14, 94–98.
- Zhang, T., Yan, Z., Zheng, X., Wang, S., Fan, J., and Liu, Z. (2020). Effects of Acute Ammonia Toxicity on Oxidative Stress, DNA Damage and Apoptosis in Digestive Gland and Gill of Asian Clam (*Corbicula Fluminea*). *Fish Shellfish Immunol.* 99, 514–525. doi:10.1016/j.fsi.2020.02.046
- Zhao, X., Duan, X., Wang, Z., Zhang, W., Li, Y., Jin, C., et al. (2017). Comparative Transcriptome Analysis of *Sinonovacula Constricta* in Gills and Hepatopancreas in Response to *Vibrio Parahaemolyticus* Infection. *Fish Shellfish Immunol.* 67, 523–535. doi:10.1016/j.fsi.2017.06.040
- Zhao, Y., Li, M.-C., Konaté, M. M., Chen, L., Das, B., Karlovich, C., et al. (2021). TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-Seq Data from the NCI Patient-Derived Models Repository. *J. Transl. Med.* 19, 1–15. doi:10.1186/S12967-021-02936-W

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liao, Liu, Han, Yang, Liu, Wang, He and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.