

Full-length transcriptomes of brain and ganglia and sex-based differentially-expressed genes involved in growth dimorphism of giant river prawn *Macrobrachium rosenbergii*

Dong Liu¹, Zhenzhen Hong¹, Lang Gui¹, Li Zhao¹, Wenzong Zhou², Yude Wang³, Shengming Sun¹, and Mingyou Li¹

¹Shanghai Ocean University

²Shanghai Academy of Agricultural Sciences

³Hunan Normal University

December 2, 2022

Abstract

The giant river prawn *Macrobrachium rosenbergii* is an important aquaculture prawn, showing sexual dimorphism in growth, with males growing much faster than females. However, the mechanisms controlling these complex traits are not yet well understood. The nervous system plays an important role in regulating life functions. This study aimed to obtain and characterize the full-length transcriptomes of brain and ganglion in female and male prawns by PacBio RNA sequencing. Based on the result of PacBio sequencing, transcript's functional annotation, transcript factors, and simple sequence repeat analysis, long non-coding RNAs (LncRNAs) and transposable element predictions were accomplished. Total 159.1-Gb subreads were obtained and average length was 2,175 bp, with 93.2% completeness. After clustering and polishing, 84,627 high quality unigene sequences were produced and annotated by functional databases. 6,367 transcript factors and 6,287 LncRNAs were predicted. Illumina sequencing of brains and ganglia extracted from female and male prawns was carried out. A significant number of differentially expressed genes (DEGs) were found and confirmed by qRT-PCR analysis. Of the related 435 genes in protein processing pathways in the endoplasmic reticula, compared to females, 42 DEGs were detected, and 21/26 DEGs with up-regulated expression in male prawn brain/ganglion. DEGs in this pathway are likely to be regulated by multiple LncRNAs in polypeptide folding and misfolded protein degradation in the different organs and sexes of the prawn. Our study lays a foundation for understanding the growth dimorphism controlled by nervous system, and is a valuable resource for sex-controlled breeding of prawns in the future.

1 Introduction

The giant river prawn *Macrobrachium rosenbergii* lives in fresh water in the tropical regions of northern Australia and Southeast Asia. This species is the largest *Macrobrachium* species with complex traits including an omnivorous diet, faster growth, and a short reproductive cycle. It has become a favored species in prawn farms and is commercially cultured in many countries and regions (Chen *et al.* 2012). *M. rosenbergii* was first introduced to China (from Japan) in 1976, raised rapidly in large scale. Its cultivation has maintained an increasing trend, reaching 139.6 thousand tons of live weight aquaculture production in 2020, which accounted for 47.48% of the worldwide total production (294 thousand tons) (FAO 2022; Ministry of Agriculture and Rural Affairs *et al.* 2020).

One striking feature of the giant river prawn is that males grow faster and distinctly larger than mature females. Additionally, the male's second pair of chelipeds is larger and thicker than that of the female, exhibiting sexual dimorphism (Tan *et al.* 2020). Genetic sex-determination in *M. rosenbergii* follows the

ZW mode, and WW females can sex reverse into functional males via androgenic gland cell transplantation (Levy *et al.* 2019). The sex-linked genes on the ZW chromosomes were observed (Ma *et al.* 2019). The giant river prawn's physiological molting is a basic process of the post-larvae prawn, through which they realize metamorphosis, growth, and development. Successful molting is regulated by both the nervous system and the neuroendocrine system, exhibiting an initial rise and a coordinated decline in the circulating concentration of molt hormone and molt-inhibiting hormone (Yan *et al.* 2016). Those genes that present in the ecdysone signal pathway, underlying hormonal regulation, have been identified as being involved in molting and epidermis reconstruction. These include chitin binding proteins, crustacean hyperglycemic hormone, calcification-related cuticular proteins, and chitinase in Penaeid white shrimp (*Litopenaeus vannamei*) (Zhang *et al.* 2019). In fact, a comparison between higher growth and lower growth transcriptomes revealed genes potentially involved in superior growth performance based on family selected stocks of *Litopenaeus vannamei* (Santos *et al.* 2021). However, little attention has been given, so far, to the regulation of sexual dimorphism in growth by the nervous system, which could be useful in understand nervous regulation of life processes (Bobkova *et al.* 2020).

In evolutionary biology, the nervous system tends to concentrate, a progress from scattered and simple nervous net to a complex centralized nervous system. The evolution of the nervous system allows animals to regulate body functions by generating, modulating, and transmitting information (Arendt *et al.* 2016). The brain of *M. rosenbergii* is located in a mass of spongy tissue within the base of its eyestalks. It comprises the protocerebrum, deutocerebrum, tritocerebrum, nerve roots, commissures, and clusters of cell bodies working in concert. Its chained neural system is composed of cerebral ganglia, thoracic ganglia and abdominal ganglia (Liao 2001). Extracts from the thoracic ganglia stimulate oocyte development in the giant river prawn (Liao *et al.* 2001), which acts indirectly on the gonads by triggering the release of a putative gonad stimulating factor from the thoracic ganglion (Jayasankar *et al.* 2020). Secretions from the eyestalk, brain, and thoracic ganglia of decapod crustaceans have a stimulatory effect on ovarian maturation (Mohamed & Diwan 1991). The eyestalk's regulation of ovarian maturation has been well-studied by Illumina sequencing in pacific white shrimp *Litopenaeus vannamei* (Wang *et al.* 2019). Due to a lack of completely related-gene analysis, the possible molecular mechanisms by which the brain and thoracic ganglion regulate gonad development and growth remain unclear. Nanopore-based full-length transcriptome sequencing methods are capable of sequencing transcripts from end to end at a single molecule level and can be used to annotate transcriptome structures in a variety of organisms (Roach *et al.* 2020). However, comparative studies of the transcriptomes for the brain and thoracic ganglia between sexes of the giant river prawn have received little attention outside of miRNA (Xia *et al.* 2022).

To investigate and gain a better understanding of sex-linked growth dimorphism through regulation of the brain and thoracic ganglia in *M. rosenbergii*, we analyzed the full-length transcriptomic characterization of the brain and thoracic ganglion of female and male prawns using SMART single molecule sequencing and compared gene expression differences between the sexes. This allowed us to discover that differentially expressed genes were significantly enriched in the endoplasmic reticulum pathway's protein processing, a unique protein-synthesizing mechanism that plays a crucial role in efficient neural communication and is responsible to growth dimorphism in the giant river prawn. To our knowledge, this is the first full-length transcriptome wide gene expression profiling of the brain and thoracic ganglion in *M. rosenbergii*. This study will serve as a basic resource for further studies of growth regulation mechanisms, and be of special interest to exploring molecular breeding for targeting growth genes.

2. Materials and Methods

2.1. Tissue material

Six mature prawns (3 males and 3 females) in good health were randomly selected as experimental specimens from Huzhou Fengsheng Bay Farm, Zhejiang province in China, and subjected to an ice bath for immobilization. Brain and thoracic ganglia tissues of each prawn (approximately 20 g) were rapidly excised and flash-frozen in liquid nitrogen. The dissected tissues of the 3 males and 3 females were mixed with those of the same sex to eliminate individual differences. These were stored at -80 °C until the total RNA

was extracted. The other tissues were collected according to experimental requirements. Animal experiments were approved by Animal Ethics Committee of Shanghai Ocean University.

2.2. RNA extraction and compete transcript sequencing

Total RNA was extracted from the mixed samples using Phygene Total RNA Isolation Kit (Phygene, Fuzhou, Fujian, China) following the manufacturer’s instructions. RNA purity was determined firstly by agarose gel electrophoresis, followed by use of a NanoDrop instrument (IMPLEN, CA, USA). RNA quantity was measured using an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA). The total RNA from three samples was used to generate one library. The poly (A) mRNA was isolated via Oligo (dT), and the mRNA was reversed into full length first-strand cDNA using a SMARTer PCR cDNA Synthesis Kit. The synthesized cDNA was enriched using PCR amplification. Fragment screening of partial cDNAs were performed by the BluePippin Size Selection System, and the 5-10 kb size optional fragments were selected and subsequently enriched using PCR amplification. The amplicators were used to construct a SMRTbell Library, subsequently sequenced on the PacBio RS II sequencing platform (HiSeq 2500, San Diego, CA, USA), and final complete reads were generated.

2.3. Data processing and full-length transcript functional annotation

The raw data (subreads) were filtered and corrected to obtain circular consensus sequences, in which the adaptors, barcodes, polyA, and chimera were eliminated and then polished using isoseq3 software (Abdel-Ghany *et al.* 2016) to obtain isoform sequences. After removing low-quality isoform sequences, based on min passes = 2 and min predicted accuracy = 99%, the high-quality isoform sequences were obtained and clustered into a unigene sequence via CD-HIT (Fu *et al.* 2012) with an identification of 98% model. The completeness of the unigene sequence was assessed by BUSCO with the arthropoda_odb9 database (Simao *et al.* 2015), and annotated functionally, using diamond software, with an e value of $e < 1e-5$ based on six different databases: non-redundant sequences (NR), eukaryotic ortholog groups (KOG), Gene Ontology (GO), Swissprot, Evolutionary Genealogy of Genes Non-supervised Orthologous Groups (eggNOG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (Bairoch *et al.* 2010; Kanehisa *et al.* 2000). The protein families were assigned by the HMMER 3.1 package (<http://hmmer.org/download.html>) with the Pfam database (Mistry *et al.* 2021).

2.4 Identification of TF, CDS, lncRNAs, SSR, and transposable elements

The transcription factors (TF) were identified using blast, comparing against the AnimalTFDB database (Huet *et al.* 2019). Coding sequences (CDS) of the unigenes were annotated through blast against NR, Swiss-Prot, and KOG databases. The long non-coding RNAs (lncRNAs) were predicted from transcripts without coding potential using Coding Potential Calculator 2 (Kang *et al.* 2017), Coding Non-coding Index (Sun *et al.* 2013), Pfam (Mistry *et al.* 2021) and PLEK (Li *et al.* 2014), with min length 200 bp and min ORF 300 bp as the cut-off criterion. Simple sequence repeats (SSRs) were identified using MISA (Beier *et al.* 2017). Transposable elements were identified using RepeatMasker (<http://www.repeatmasker.org>).

2.5. Differential expression analysis via Illumina cDNA library sequencing

To prepare the Illumina library, total RNA was extracted from brain and ganglia tissues of three individuals as described above, and the mRNAs were enriched with Oligo (dT) mRNA magnetic beads following the manufacturer’s instructions. The cDNA libraries were synthesized using mRNA fragments as templates. The quality of four cDNA libraries were checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and 150 bp paired-end reads were generated by Illumina HiSeq 2000 platform. The unigene sequences generated by SMART sequencing were used as a reference, and the clean reads obtained from Illumina sequencing were aligned to the reference using bowtie2 (Langmead & Salzberg 2012), to analyze gene expression counts. The fragments per kilobase of transcript per million mapped reads (FPKM) value for the gene expression level were calculated using eXpress (Roberts & Pachter 2013). Based on gene expression counts obtained from each sample, differentially expressed genes were identified using DESeq (Anders S 2012). The thresholds for significant differential expression were set as a p-value < 0.05 and absolute of \log_2 (fold

change) >2. Significance was tested using hypergeometric distribution. Finally, differentially expressed genes were used for GO and KEGG enrichment analysis.

2.6. Differentially expressed gene validation via qRT-PCR

Ten differentially expressed genes (DEGs) were randomly selected to verify their differential expression. The qRT-PCR primers were designed using Primer 6 and are listed in Table 1. The *actin* gene was used as the control. The qRT-PCR was carried out on a real-time PCR system (Eppendorf, Hamburg, Germany). Reacted mixture consisted of 1 μ l cDNA (60 ng/ μ l), 10 μ l Bestar SYBR Green qPCR Mastermix (DBI, Bioscience Inc., Hamburg, Germany), 1 μ l of each primer, and 7 μ l mill i-Q water. Reactions were performed at 95 °C for 30 s; 30 cycles of 95 °C for 5 s, 60 °C for 30 s and 72 °C for 60 s. The qRT-PCR results were obtained from three biological replicates, and the gene relative expression levels were calculated using the 2- $\Delta\Delta$ Ct method.

Table 1. Sequences of primers used for gene differential expression analysis

Gene	Primer sequence (5->3)	Primer sequence (5->3)
	Forward	Reverse
SMCP6	TGCAGTGGTTGTTGCACTTG	ACTGCACAGTAGCTGTTTGC
IAGBP	AACATGGCTGACGGTTCTTC	TCCGGACGTTGATGTTTCATG
ASMP	ACAGATCAGTGCCACACATG	ACGGGAACAAACCATCAAGC
Fem1b	AATTGCATGTGGGGCTGATG	TGTCCATGTGTGCTCCATTC
transformer 2	ACATCATGGAAGCAGAGAGGAC	TGGCATCCAGAACAACCTTGC
female lethal d-like protein	GGTGAATTGGCCCTTCAAAGG	TTCGCCTGCCTTAATTGCTG
DNA damage-binding protein 1	TTCGAGTGGACGAACGAAAG	ACGTCTTGTACTGCAGCAAG
alpha-2-macroglobulin	TCAGTGAAGCAGCCTTTTGC	TGATGTCTCTGTGGCCAAAC
heat shock protein 90	TCCGCAAGAACTTGGTCAAG	AGCCAACCTTTCGCGGTTAG
ganglioside GM2 activator	TATCGGTTTCATGTGGGTGGAG	ATGGATCAGGGCAAGGTTTCG
actin	AATCGTGCGTGACATCAAGG	TCTCGTTACCGATGGTGATGAC

Abbreviated SMCP6: the structural maintenance of chromosomes protein 6-like gene; IAGBP: the insulin-like androgenic gland hormone-binding protein gene; ASMP: the abnormal spindle-like microcephaly-associated protein homolog gene.

2.7 Construction of LncRNA and gene co-expression networks

A network of LncRNA-gene co-expressions was constructed using the Pearson correlation coefficient between the differentially expressed LncRNAs and differentially expressed mRNAs using the FPKM values of four tissue samples via Illumina sequencing. The threshold for positive correlation was set to PCC > 0.95 and P-value < 0.05. The regulatory network of LncRNAs and mRNA was established via Cytoscape (Smoot *et al.* 2011), and the ZFLNC database (Hu *et al.* 2018) was used to conduct a conservative analysis of lncRNA.

3. Results

3.1 Full-length transcription sequences

Brain and thoracic ganglia samples of sexually mature giant river prawns were used for single molecular real time (SMART) sequencing. After removing adaptor sequences and low-quality sequences, a total of 73,146,314 subreads (159.1 Gb) were obtained, and the average length of subread was 2,175 bp (Figure 1A). Subread sequences were subjected to self-correction to produce 1,725,793 circular consensus sequence reads (CCS), and the CCS sequences were clustered into 1,509,283 full-length non-chimerics (FLNC). The FLNCs were polished to obtain 99,728 highly quality isoform numbers (288 Mb), with an average length of 2,891 bp (Figure 1B). Finally, 84,627 unigene sequences (246 Mb) with an average read length of 2,913 bp (Figure 1C), were obtained using high quality isoform sequence clustering (identify = 98%). 78,559 transcriptions

were found between 1 and 10 isoforms (Figure 1D). The transcript length extended from 170 bp to 14,287 bp, with an average length of 3,061 bp.

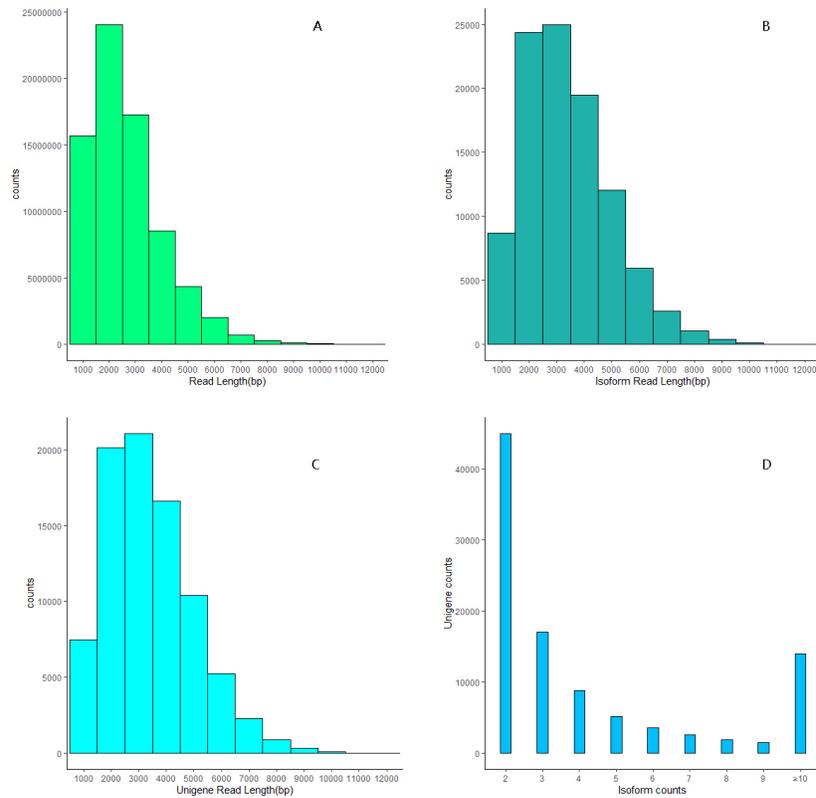


Figure 1 . Numbers and length distributions of subreads (A), isoform read (B), unigene read (C), and number of isoform (D) in unigene sequences with PacBio SMART sequencing method of *M. rosenbergii* .

3.2 Assessment of unigene completeness

The completeness and accuracy of the obtained unigenes were assessed via BUSCO with the arthropoda-odb9 database. Compared to 60 species with 1,066 available gene sequences in the database, 994 (93.2%) sequences were completely homologous to matches within the BUSCO database (Figure 2A). Our SMART data were determined to be of high quality and could be used for subsequent analysis.

3.3 Functional annotation of unigenes

All 84,627 SMRT sequencing from these unigenes were functionally annotated by searching seven databases: NR, Swiss-Prot, KEGG, KOG, eggNOG, GO, and Pfam. A total of 62,612 (73.99%) unigenes were annotated in at least one database. The most, 57,846 (68.35%) unigenes, were functionally annotated in the NR database. The least, 5,192 (6.14 %) unigenes, were functionally annotated in the KEGG database (Figure 2B). The functional annotations of all 84,627 unigenes are listed in Table S1.

M. rosenbergii

A total of 5,192 unigene sequences were annotated by the KEGG data and plotted to 87 operative catalogs. Genetic information processing was the largest transcript category including translation, transcription, folding, sorting, and degradation (Figure 4). The first four transcript-related pathways in the genetic information processing were spliceosome (505, 9.7%), ribosome (471, 9.1%), protein processing in endoplasmic reticulum (435, 8.4%), and RNA transport (4335, 8.3%) (Table S2). Gene function was classified by KOG data and most were defined specifically in signal transduction mechanisms, posttranslational modification, protein turnover, chaperones, cytoskeleton, intracellular trafficking, secretion, and vesicular transport (Figure S2). A large number of KOG-annotated genes were related to neuron cell commitment and sensation, such as neuropeptide signal transduction (122), innervation activity (50), sensory perception of sound (357), sensory perception of light stimulus (22), and neurotransmitter biosynthesis including gamma-aminobutyric acid (144), dopamine secretion (77), G protein-coupled serotonin receptor (10), and histamine secretion (7) (Table S1).

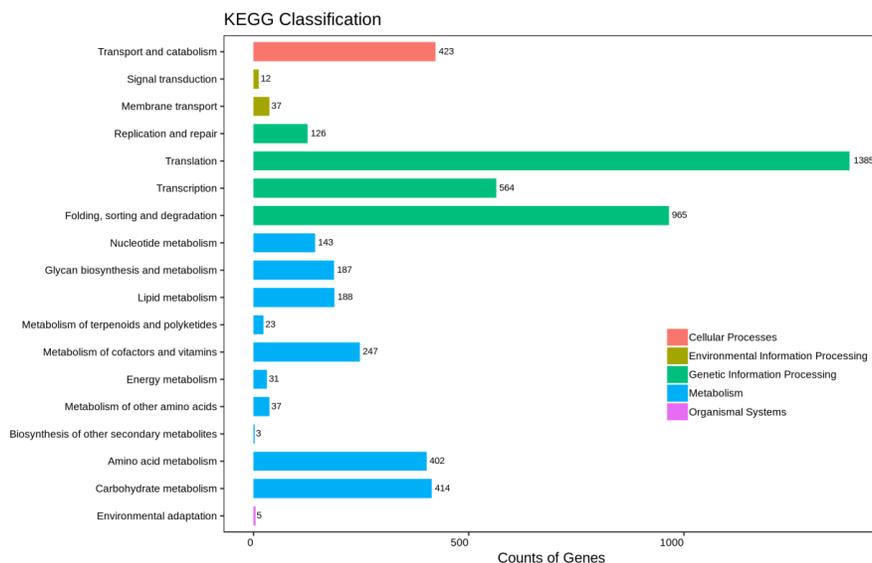


Figure 4. KEGG annotation of unique transcripts of *M. rosenbergii*

3.4 Results of TF, CDS, LncRNAs, SSR, and transposable elements

Transcription Factors (TFs) regulate gene expression at transcription of DNA into RNA by binding to special regions of DNA. The unigene sequences of giant river prawn were blasted against AnimalTFDB, and the top 10 species distribution included 945 (14.84%) from *Drosophila melanogaster*, 641 (10.07%) from *Homo sapiens*, and 448 (7.04%) from *Danio rerio* (Figure S3). A total of 6,367 putative TFs from 58 families, were predicted from these transcripts (Table S3). The top 27 annotated families are shown in Figure 5.

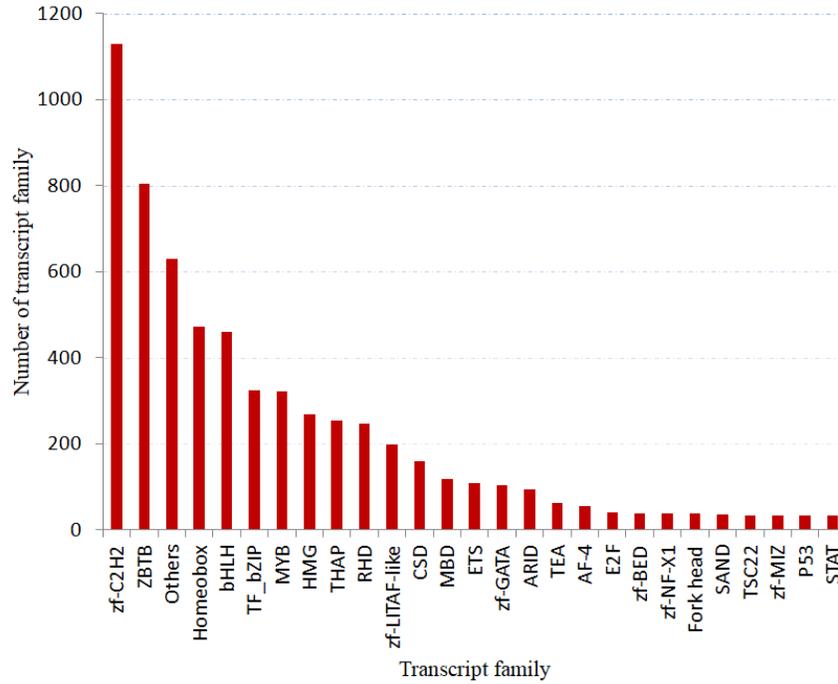


Figure 5 . Numbers and families of the top 27 transcription factors in *M. rosenbergii*

CDSs are critically important features of mRNA transcripts that contain coding exons. Of the 70,536 transcripts, 57,963 CDS were identified by blast against the NR, Swiss-Prot, and KOG databases. 63.4% of these transcript CDSs' length ranged between 201 bp and 1400 bp (Figure 6A).

LncRNAs were predicted by CPC2, CNCI, Pfam, and PLEK. Overall 6,287 LncRNA were identified (Figure 6B). Their length varied from 200 bp to 6,482 bp. A majority of lncRNA lengths (>71%) ranged between 1,000 and 3,000 bp, and the average length was 1,557.03 bp, shorter than that of mRNA's transcripts (Figure 6C).

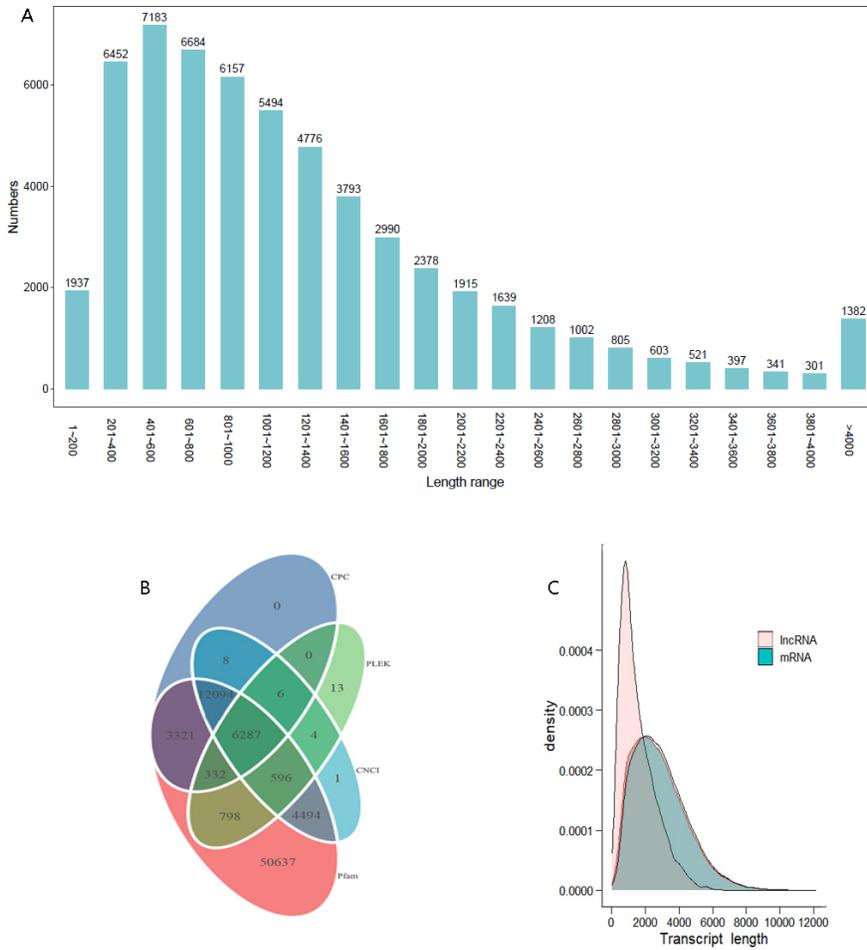


Figure 6 . Length and number distribution of CDS from unigenes (A), Venn diagram of LncRNAs predicted by CPC, PLEK, CNCI, Pfam methods (B), and density and length distributions of LncRNAs and mRNAs (C) in *M. rosenbergii* .

SSR has abundant polymorphisms as microsatellite markers. A total of 84,627 sequences were analyzed to identify 160,010 SSRs. The number of SSRs present in compound formation was 13,996. Most of the SSRs detected were mono-nucleotide repeats (80.10%), followed by di-nucleotide repeats (9.84%), tri-nucleotide repeats (9.56%), tetra-nucleotide repeats (0.44%), hexa-nucleotide repeats (0.04%) and penta-nucleotide (0.02%) (Table S4).

Transposable elements are DNA sequences able to move around within the genome and impact genome evolution. 37 SINEs, 245 LINEs, 42 LTR RNA types, and 950 DNA types of transposons were identified, and transposon content accounted for 0.11% of the total transcript length.

3.5. Differentially expressed genes in brain and thoracic ganglion between sexes based on Illumina sequencing results

Illumina sequencing was performed on brain and thoracic ganglion tissues of the female and male prawns, separately. After processing the raw reads, clean reads were obtained. The percentage of valid bases in all reads exceeded 92%, and the Q30 value of all sequences per library was as high as 94% (Table S5). Using the annotated unigenes as reference sequences, Illumina sequences were mapped to the reference sequences, and the mapped rates were 83.92% (male thoracic ganglion) to 90.49% (male brain) for all Illumina sequences in

each library (far exceeding the standard of 70% non-pollution conditions). Further, 79.06%-85.76% of the Illumina sequences were matched in proper pairs (Table S6).

Volcano plots showed that the sex-based gene expression levels were distinguishable and statistically significant (p value < 0.05 and absolute value of \log_2 -fold change > 2) (Figure 7). Compared with female prawns, male prawn brain and thoracic ganglia displayed 2,416 and 3,280 upregulated differentially expressed genes (DEGs), 2,871 and 2,931 DEGs were down-regulated, and 3,575 and 4,499 DEGs were exclusively expressed. 1,712 DEGs were shared by male and female prawns. GO annotation of DEGs revealed the top up-regulated genes in biological process, cellular component, molecular function categories in proper sequence to be hemolymph coagulation, extracellular region, lipid transporter activity for brain (Figure S4), as well as cell surface pattern recognition receptor signaling pathway, 4-aminobutyrate transaminase complex and its activity for thoracic ganglion (Figure S5). Growth-related genes in GO molecular function were actin binding (GO: 0003779), transforming growth factor beta receptor activity (GO: 0005026) etc. (Table 2). KEGG pathway annotation showed that DEGs were sorted into translation, transcript, protein folding, sorting, and degradation for the brain (Figure S6) and thoracic ganglion (Figure S7).

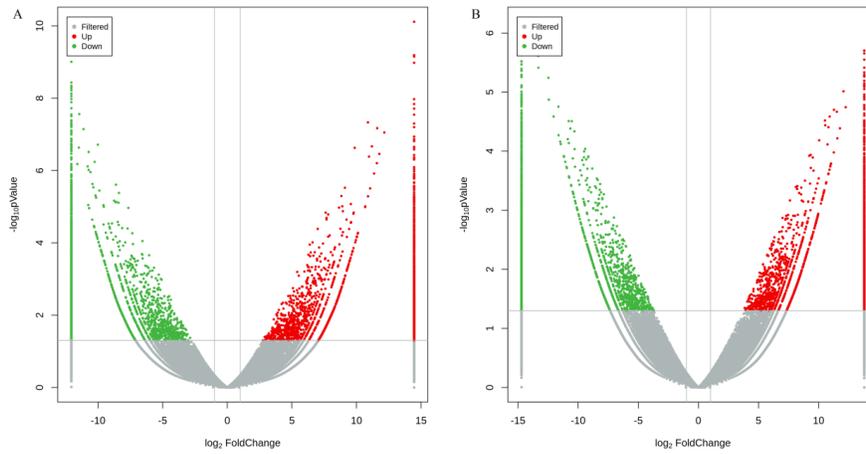


Figure 7. Volcanic diagrams of DEGs calculated based on raw counts from brains (A) and thoracic ganglia (B) of *M. rosenbergii*. Statistically significant differentially expressed genes are shown in the red dots (up-regulation in male prawn) and green dots (down-regulation in female prawn).

Table 2 Genes involved in growth based on GO term.

GO id	Molecular function	p-Value	q-Value	tissues	regulation	DEGs number
0003779	actin binding	0.0009	0.0182	brain	Up	80
0003779	actin binding	0.0173	0.0854	brain	Down	77
0005026	TGF- beta	0.0514	0.1443	brain	Down	1
0005026	TGF- beta	0.0043	0.0446	brain	Up	2
0019838	growth factor binding	0.126	0.2351	brain	Down	2
0019838	growth factor binding	0.0274	0.1083	brain	Up	3
0005520	IGF binding	0.2259	0.3368	brain	Down	1
0005520	IGF binding	0.2027	0.3232	brain	Up	1
0048273	p38-MAPK binding	0.0588	0.1627	brain	Up	2
0003779	actin binding	0.4687	0.5553	ganglion	Up	66
0003779	actin binding	0.0888	0.1977	ganglion	Down	81
0005026	TGF- beta	0.0065	0.0437	ganglion	Up	2
0005026	TGF- beta	0.0006	0.0117	ganglion	Down	3

GO id	Molecular function	p-Value	q-Value	tissues	regulation	DEGs number
0048273	p38-MAPK binding	0.3041	0.4302	ganglion	Down	1
0048273	p38-MAPK binding	0.0838	0.1896	ganglion	Up	2

Abbreviated TGF-beta: transforming growth factor beta receptor activity, type II; MAPK: mitogen activated protein kinase; IGF: insulin-like growth factor; IAG: insulin-like androgenic gland hormone.

3.6 Validation of differentially expressed genes

To validate the DEGs, ten genes were selected using a qRT-PCR method, to determine their relative expression levels. These genes included *Smc6* (structural maintenance of chromosomes protein 6) in ubiquitin protein ligase binding activity, *Iagbp* (insulin-like androgen gland hormone binding protein) in feedback IAG signaling, *Aspm* (abnormal spindle-like microcephaly associated protein homolog) in the functional regulation of neurogenesis and brain growth, *fem1b*, *transformer2*, and *female-lethal d* in sex determination and differentiation, *Ddb1* (DNA damage-binding protein 1) in protein processing-related E3 ubiquitin ligase complex, *Hsp90b1* (heat shock protein 90 beta 1) in stabilizing and folding other proteins, *Gsg2* (ganglioside GM2 activator) in reductions of GM2 gangliosidosis (mutations in this gene result in Alzheimer's disease in humans), and *alpha 2 macroglobulin* gene. The gene expression results from qRT-PCR are consistent with the values of FPKM from Illumina sequencing under the same conditions (Figure 8), indicating that the DEGs from the RNA-seq data were reliable.

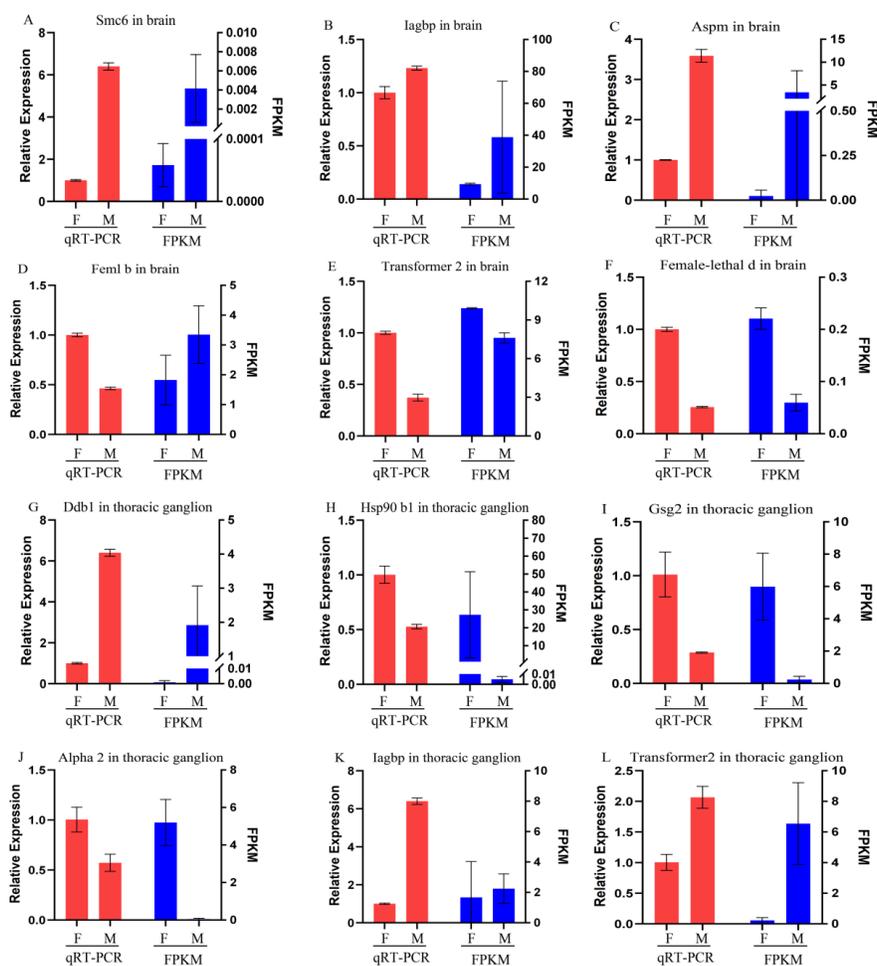


Figure 8 . Validation of ten DEG profiles by qRT-PCR. F: female prawn, M: male prawn. The FPKM (Fragments per kilobase of transcript per million mapped reads) value was calculated by eXpress (3). The abbreviated gene names are given above.

3.7 Gene expression analysis in protein processing signaling pathways

In DEG-enriched pathways, up-regulated expression of male prawn genes associated with protein processing in the endoplasmic reticulum play an important role in protein folding and misfolded protein elimination in both the brain (Figure 9A) and thoracic ganglion (Figure 9B). In fact, male prawns showed faster growth than females, and more proteins needed to be assembled. In turn, more misfolded protein needed to be eliminated in order to escape concentrations exceeding to the threshold of proteotoxic stress. 435 genes were identified as participating in pathways of protein processing in the endoplasmic reticulum (Table S1). Principle component analysis (PCA) separated these genes expressed in the brain and thoracic ganglion of males and females into different principal components. The first and second principal components accounted for 40.9% and 36.2% of the variation between the sexes (Figure 9C). Of 42 DEGs between the prawn sexes, 21 male genes had up-regulated expression in the brain and 26 had up-regulated expression in the thoracic ganglion compared with the female genes. A total of 38 DEGs were obtained from male prawns, and 39 DEGs from female prawns (Figure 9D). These DEGs, shown in a heat map, supported the results of the PCA analysis.

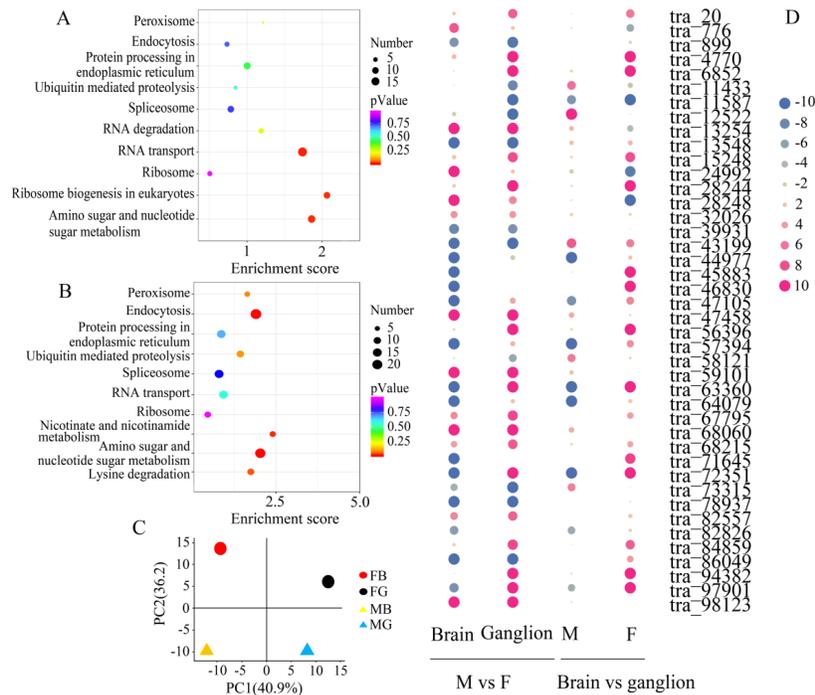


Figure 9 . Annotation and distribution of DEGs ($|\log_2FC| > 2$, p value < 0.05) in the pathway of protein processing in the endoplasmic reticulum of *M. rosenbergii* . KEGG enrichment of DEGs in up-regulated expression in the brain (A) and up-regulated expression in the thoracic ganglion (B). The DEGs were separated into different principal components (C) the distribution is shown (D) according to tissues and sexes of *M. rosenbergii* . FB: female brain; FG: female thoracic ganglion; MB: male brain; MG: male thoracic ganglion; M: male; F: female.

The DEGs were involved in protein recognition by luminal chaperones (k09486, 3 genes), polypeptide folding, and correction (k1718, 3 genes). Most of the DEGs presented in the ER-associated degradation pathway

(ERAD), a process which starts with misfolded protein recognition, followed by ubiquitination and retrotranslocation into the cytosol for degradation in proteasome (Figure 10). Protein disulfide- isomerases (PDIs, K09580, 10 genes) and Osteosarcoma-9 (OS-9, k10088, 1 gene) delivered misfolded polypeptide to the ubiquitin ligase complex, which is composed of ubiquitin-protein ligases such as Dao10 (k10661, 3 genes), Sel1L (k14026, 4 genes), and Hsp40 (k09502, 5 genes) in ERAD. A translocon-associated protein, TRAP (k13251, 3 genes), showed up-regulated expression in the male brain and thoracic ganglion, which plays a role in accelerating ER degradation of unfolded proteins (Figure 10).

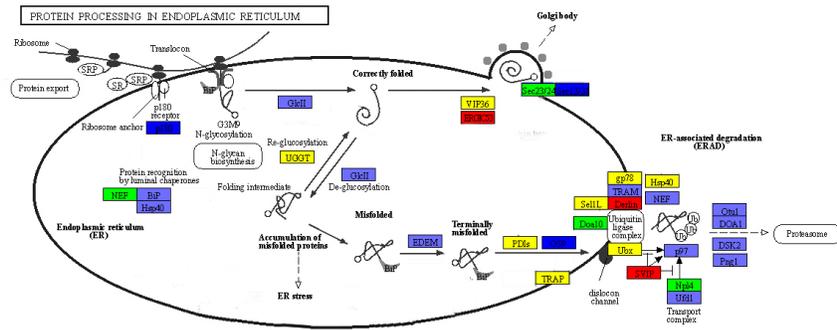


Figure 10 . DEGs involved in the pathway of protein processing in the endoplasmic reticulum. DEGs with up-regulated expression colored in red; DEGs with down-regulated expression colored in green, and DEGs with up/down-regulated expression colored in yellow (male prawn compared to female prawn). Genes colored in blue denoted that these genes were identified from the full-length transcripts, without differential expression in the nervous system between sexes of *M. rosenbergii*.

3.8 The network of differently expressed lncRNA-gene coexpression in ERAD

In the brain and nerve ganglion, the collaboration between the differentially expressed (DE-) lncRNAs and the differentially expressed (DE-) mRNAs showed 5,086 pairs with co-expression which were composed of 26 DE-lncRNAs and 1,328 DE-mRNAs. In these DE-lncRNAs and DE-mRNAs, significantly correlated lncRNA-mRNA couplings were obtained from 19 DE-lncRNAs and 6 DE-mRNAs in ERDA, based on a Pearson correlation coefficient with $|R| > 0.95$ and P-value < 0.05 , and PDIs.2 (degree=17) indicating the highest degrees (Figure 11A). The 6 DE-mRNAs showed a similar model of expression in the brain (Figure 11B) and ganglion (Figure 11C). The lncRNA-mRNA coupling suggested that the regulation of PDIs.2 by multiple lncRNAs was likely to occur in the nervous system. The mRNAs in the co-expression network enriched in ERDA may play a critical role in the growth of male *M. rosenbergii* which is worth further study.

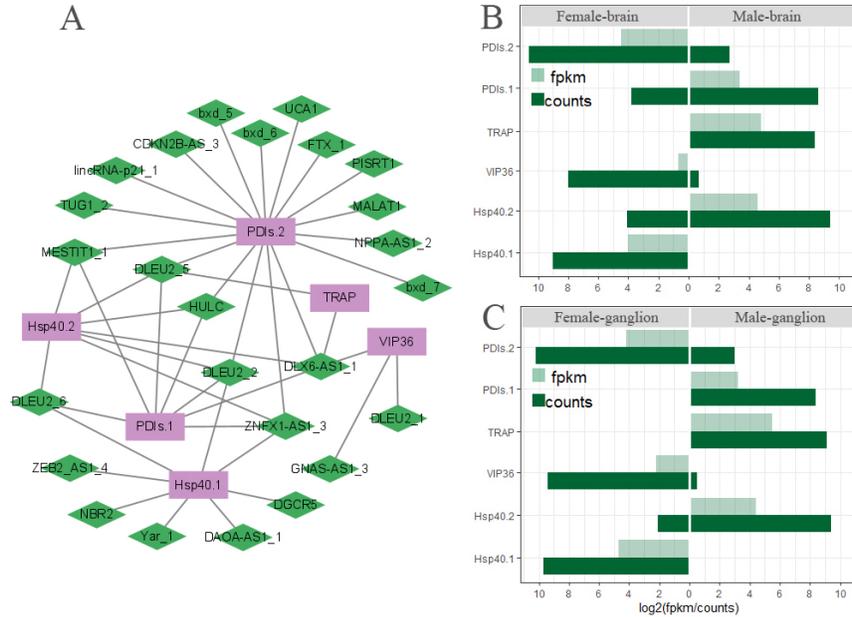


Figure 11 . Differently expressed lncRNA and DEG co-expression networks (A) and gene expression in tissues vs sex (B & C) of *M. rosenbergii* . The purple rectangles represent genes; green diamonds represent lncRNAs. FPKM value was calculated using eXpress (3), and counts was calculated using bowtie2 (2).

4. Discussion

Using the PacBio RS II sequencing platform, a 159.1 GB subread base with 73,146,314 subreads and 2,175 average reads were generated. 84,627 unigenes were detected in *M. rosenbergii* . All unigene sequences were assessed using BUSCO and 93.2% unigenes showed high homology and complete matches, suggesting that our full-length cDNA sequences are a rich resource for further functional genomics analysis in *M. rosenbergii* .

All unigene sequences were functionally annotated using the NR database. 37.6% of the unigenes were aligned with *Hyalella azteca* which is a widespread species of amphipod crustacean, followed by decapod crustaceans with lower match rates (4.67% for *Procambarus clarkii*, and 3.02% for *M. rosenbergii*). These results suggest that our full-length transcripts provide additional novelty and complexity to functionally important proteins previously unannotated in the *M. rosenbergii* transcriptome. Transcripts annotated using the GO database were enriched in categories associated with biological process, cellular components, and molecular function, and subcategories including metabolic processes, biological regulation, catalytic activity, and transporter activity. According to KEGG annotations, genetic information processing was the dominant enriched transcript group, especially in pathways for spliceosomes, ribosomes, protein processing in the endoplasmic reticulum, and RNA transport. According to the gene function interpretation of SMRT transcripts using the KOG database, a large number of genes were functionally involved in neuron cell commitment, neuropeptide signal transduction, and neurotransmitter biosynthesis. This might provide clues for further study of brain- and nerve ganglia-regulation in sexual dimorphism pertaining to growth, development, and morphologies of *M. rosenbergii* . Neurosecretory cells in the brain produce various neuropeptides with regulatory effects on biological functions, and allatostatin, crustacean female sex hormone, crustacean hyperglycemic hormone, and eclosion hormone were identified from transcriptome of the brain of the Chinese mitten crab (Liu *et al.* 2019).

Transcription factors are proteins that modulate the transcription of specific genes by binding to DNA

which regulates the proliferation, migration and differentiation of neural tissue (Feng *et al.* 2021; Qian *et al.* 2022). 6,367 putative transcription factors were identified in *M. rosenbergii*, and their function requires further investigation. LncRNAs constitute an important layer of regulation in gene expression at either the transcriptional or post-transcriptional levels, during fundamental processes (Necsulea *et al.* 2014). Transcriptome profiles of LncRNAs in the brain of zebrafish showed significant sex differentiation between male and female individuals (Yuan *et al.* 2019). A total of 6,287 LncRNAs have been obtained by four analytical methods from *M. rosenbergii*, and the number are similar to that seen in the Platypus, but significantly lower than in other mammals (Necsulea *et al.* 2014). LncRNA evolution and biological function are driven by transposable elements, particularly transposable element inserts at transcription starting sites, governing LncRNA transcription (Kelley & Rinn 2012). It is interesting that the number of transposable elements identified from the transcription of *M. rosenbergii* is lower than what is seen in vertebrates (Liu *et al.* 2020).

According to the Illumina sequences from the brain and nerve ganglion of *M. rosenbergii*, a large number of DEGs were found and annotated using the GO database. Growth function was a prevalent term in the GO database. In the giant freshwater prawn, the periodic shedding of the exoskeleton is one of the most important physiological processes essential for growth, including molting and muscle development. Molting-related genes, for instance, molt-inhibiting hormone, crustacean hyperglycemic hormone, ecdysteroid receptor, and retinoic acid X receptor (Zhang *et al.* 2019) were not found in DEGs in the present study. In contrast, growth-associated gene expressions were observed in DEGs. In particular, *actin* is one of the most abundant intracellular proteins and a preferential binding of actin-beta to myosin which contains actin binding sites, plays a key role in many essential biological processes for cell adhesion, migration, and contractility in muscle (Vedula & Kashina 2018). Increasing myofiber numbers (hyperplasia) is important in body growth of aquatic animals. For example, sea bream *Sparus aurata*'s muscle hyperplasia contributes greatly to its adult size, while zebrafish *Brachydanio rerio* shows little hyperplasia and reaches only a small adult size (Rowlerson *et al.* 1997). In the present study, GO enrichment results from 80 and 66 up-regulated genes involved a common actin binding process in the male brain and nerve ganglion respectively, indicating that the faster growth males compared to females in the giant freshwater prawn is under nervous system control. In the DEGs, the insulin-like androgenic gland hormone binding protein (IAGBP) gene showed up-regulated expressions in male brains. It has demonstrated that the IAGBP, as a binding partner of the insulin-like androgenic gland hormone (IAG), directly increases IAG transcripts, which play an important role in the growth and development of male *M. rosenbergii* (Yang *et al.* 2020). Meanwhile, certain growth factors, such as transforming growth factor beta (TGF-beta), and p38-MAPK binding proteins (Wyganowska-Swiatkowska *et al.* 2015), were found to be up-regulated in the male brain and nerve ganglion. TGF-beta is a central mediator of fibroblast activation, and, in skeletal muscle, TGF-beta signals by binding to tissue-specific combinations of receptor subtypes, triggering the activation of a muscle hypertrophy program, promoting muscle growth (Forouhan *et al.* 2022). In the brain/nerve ganglion, differentially expressed genes between male and female prawns lay the foundation for the further study of gene expression and manipulation in growth and development in giant freshwater prawn.

KEGG pathway analysis showed that DEGs were significantly enriched in the protein processing in the endoplasmic reticulum pathway, which not only would be a unique protein-synthesizing mechanism, but also is a crucial player in efficient neural communications (Khan 2022). Endoplasmic reticulum (ER)-to-Golgi trafficking is an essential cellular process in the secretory pathway, and selective export of proteins from the ER requires transport signals for efficient recruitment of a transmembrane intracellular lectin ER-Golgi intermediate compartment protein 53 (ERGIC-53) into budding vesicles (Nufer *et al.* 2003). Overexpression of exogenous ERGIC-53 has been shown to increase protein-secreted production in batch cultivation (Yutaka *et al.* 2022). ERGIC-53 used as an ER export determinant was only highly expressed in the male nerve ganglion of giant freshwater prawn, suggesting that it facilitates secretion of protein including neuropeptides and hormones transported to the Golgi for delivery to their final destinations. Rainbow trout that are transferred from freshwater to seawater experience growth stunting and down-regulation in growth factor IGF-I transcription has revealed endoplasmic reticulum stress to be a key mechanism underlying this growth

stunting phenotype (Morro *et al.* 2021). The maintenance of intracellular hormones can be controlled by correctly folded models, and misfolded proteins in the endoplasmic reticulum are eliminated by a highly effective protein-degradation process known as ERAD. We found multiple genes involved in terminally misfolded protein elimination in ERAD, and members of two gene families were markedly upregulated in male giant freshwater prawns, Dailin-2 in the brain, and SVIP in the nerve ganglion. Derlin-2 is homologous with Derlin-1 and capable of recognizing the misfolded protein. It physically interacts with p97, SEL1L, and SVIP to form a retrotranslocated complex, thereby, using the Derlin-1 channel, the misfolded protein is transported into the proteasome for degradation via ERAD (Ballar *et al.* 2007; van de Weijer *et al.* 2014). Therefore, it is conceivable that male prawns are more likely than females to remove misfolded proteins, since the male grows faster than the female giant freshwater prawn.

To explain the regulation mechanism underlying DEGs in protein processing in the endoplasmic reticulum, a LncRNA and mRNA co-expression network analysis was performed to find differentially expressed LncRNAs regulating transcription of DEGs in the nervous system, such as LncRNAs DLX6-AS1, ZNFX1-AS1, and HULC which are reported to play critical roles in promoting oncogenic phenotypes of cancer cell lines (Ghafouri-Fard *et al.* 2022; Xian *et al.* 2018). LncRNAs are highly expressed in the nervous system and control gene expression in the developing and adult brain, mediating neural differentiation (Chen & Zhou 2017). However, these conclusions have some limitations due to the comparative analysis of tissue next-generation sequencing data. Further *in vivo* studies will help to understand these LncRNAs' roles in the nervous system of the giant freshwater prawn.

Acknowledgments

This study was supported by State Key Laboratory of Developmental Biology of Freshwater (2020KF002), the National Key R&D Program of China (2018YFD0901205), and National Natural Science Foundation of China (31672700).

References

- Abdel-Ghany SE, Hamilton M, Jacobi JL, *et al.* (2016) A survey of the sorghum transcriptome using single-molecule long reads. *Nature Communications* **7**.
- Anders S HW (2012) Differential expression of RNA-Seq data at the gene level – the DESeq package. *Heidelberg, Germany: European Molecular Biology Laboratory (EMBL)*, 1-24.
- Arendt D, Tosches MA, Marlow H (2016) From nerve net to nerve ring, nerve cord and brain - evolution of the nervous system. *Nature Reviews Neuroscience* **17**, 61-72.
- Bairoch A, Bougueleret L, Altairac S, Amendolia V, Zhang J (2010) The Universal Protein Resource (UniProt) 2009. *Nucleic acids research* **37**, D169-D174.
- Ballar P, Zhong Y, Nagahama M, *et al.* (2007) Identification of SVIP as an endogenous inhibitor of endoplasmic reticulum-associated degradation. *Journal of Biological Chemistry* **282**, 33908-33914.
- Beier S, Thiel T, Munch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* **33**, 2583-2585.
- Bobkova NV, Poltavtseva RA, Leonov SV, Sukhikh GT (2020) Neuroregeneration: Regulation in Neurodegenerative Diseases and Aging. *Biochemistry-Moscow* **85**, 108-130.
- Chen XF, Yang GL, Kong J, *et al.* (2012) Effect of artificial culture and selective breeding on the genetic diversity of *Macrobrachium rosenbergii*. *Acta Hydrobiologica Sinica* **36**, 866-873.
- Chen Y, Zhou J (2017) LncRNAs: macromolecules with big roles in neurobiology and neurological diseases. *Metabolic Brain Disease* **32**, 281-291.
- FAO (2022) *The State of World Fisheries and Aquaculture 2022, Towards Blue Transformation* FAO, Rome.

- Feng X, Guan N, Xu ES, Miao Y, Li CG (2021) Transcription Factors Leading to High Expression of Neuropeptide L1CAM in Brain Metastases from Lung Adenocarcinoma and Clinical Prognostic Analysis. *Disease Markers* **2021** .
- Forouhan M, Lim WF, Zanetti-Domingues LC, *et al.* (2022) AR cooperates with SMAD4 to maintain skeletal muscle homeostasis. *Acta Neuropathologica* **143** , 713-731.
- Fu LM, Niu BF, Zhu ZW, Wu ST, Li WZ (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28** , 3150-3152.
- Ghafouri-Fard S, Najafi S, Hussen BM, *et al.* (2022) DLX6-AS1: A Long Non-coding RNA With Oncogenic Features. *Frontiers in Cell and Developmental Biology* **10** .
- Hu H, Miao YR, Jia LH, *et al.*(2019) AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic acids research* **47** , D33-D38.
- Hu X, Chen W, Li J, *et al.*(2018) ZFLNC: a comprehensive and well-annotated database for zebrafish lncRNA. *Database-the Journal of Biological Databases and Curation* .
- Jayasankar V, Tomy S, Wilder MN (2020) Insights on Molecular Mechanisms of Ovarian Development in Decapod Crustacea: Focus on Vitellogenesis-Stimulating Factors and Pathways. *Frontiers in Endocrinology* **11** .
- Kanehisa M, Goto S, Kawashima S, Nakaya A, S., S. K. (2000) KEGG: kyoto encyclopaedia of genes and genomes. *Nucleic acids research volume* **28** , 27-30.
- Kang YJ, Yang DC, Kong L, *et al.* (2017) CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic acids research* **45** , W12-W16.
- Kelley D, Rinn J (2012) Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biology* **13** .
- Khan S (2022) Endoplasmic Reticulum in Metaplasticity: From Information Processing to Synaptic Proteostasis. *Molecular Neurobiology* **59** , 5630-5655.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9** , 357-U354.
- Levy T, Rosen O, Manor R, *et al.* (2019) Production of WW males lacking the masculine Z chromosome and mining the *Macrobrachium rosenbergii* genome for sex-chromosomes. *Scientific reports* **9** .
- Li AM, Zhang JY, Zhou ZY (2014) PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics* **15** .
- Liao JY (2001) The Morphology and Structure of the Brain of the Prawn, *Macrobrachium rosenbergii* (Crustacea:Decapoda). *Acta Scientiarum Naturalium Universitatis Sunyatseni* **40** , 85-88.
- Liao JY, Zhang YF, Sun JX, Lin YA (2001) Partial isolation and bioassay of gonad-stimulating hormone from the brain of *Macrobrachium rosenbergii*. *JOURNAL OF FISHERIES OF CHINA* **25** , 5-10.
- Liu D, Yang JQ, Tang WQ, *et al.* (2020) SINE Retrotransposon variation drives Ecotypic disparity in natural populations of *Coilia nasus*. *Mobile Dna* **11** .
- Liu X, Ma KY, Liu ZQ, *et al.*(2019) Transcriptome analysis of the brain of the Chinese mitten crab, *Eriocheir sinensis*, for neuropeptide abundance profiles during ovarian development. *Animal Reproduction Science* **201** , 63-70.
- Ma KY, Yu SH, Du YX, *et al.*(2019) Construction of a Genomic Bacterial Artificial Chromosome (BAC) Library for the Prawn *Macrobrachium rosenbergii* and Initial Analysis of ZW Chromosome-Derived BAC Inserts. *Marine Biotechnology* **21** , 206-216.

- Ministry of Agriculture and Rural Affairs, National Fisheries Technology Extension Center, China Society of Fisheries (2020) *China Fishery Statistical Yearbook* China Agriculture Press, Beijing.
- Mistry J, Chuguransky S, Williams L, *et al.* (2021) Pfam: The protein families database in 2021. *Nucleic acids research* **49** , D412-D419.
- Mohamed KS, Diwan AD (1991) Neuroendocrine regulation of ovarian maturation in the Indian white prawn *Penaeus indicus* H. Milne Edwards. *Aquaculture* **98** , 381-393.
- Morro B, Broughton R, Balseiro P, *et al.* (2021) Endoplasmic reticulum stress as a key mechanism in stunted growth of seawater rainbow trout (*Oncorhynchus mykiss*). *BMC genomics* **22** .
- Necsulea A, Soumillon M, Warnefors M, *et al.* (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505** , 635-+.
- Nufer O, Kappeler F, Guldbrandsen S, Hauri HP (2003) ER export of ERGIC-53 is controlled by cooperation of targeting determinants in all three of its domains. *Journal of Cell Science* **116** , 4429-4440.
- Qian TM, Qiao PP, Lu YN, Wang HK (2022) Transcription factor SS18L1 regulates the proliferation, migration and differentiation of Schwann cells in peripheral nerve injury. *Frontiers in Veterinary Science* **9** .
- Roach NP, Sadowski N, Alessi AF, *et al.* (2020) The full-length transcriptome of *C. elegans* using direct RNA sequencing. *Genome research* **30** , 299-312.
- Roberts A, Pachter L (2013) Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods* **10** , 71-U99.
- Rowlerson A, Radaelli G, Mascarello F, Veggetti A (1997) Regeneration of skeletal muscle in two teleost fish: *Sparus aurata* and *Brachydanio rerio*. *Cell and Tissue Research* **289** , 311-322.
- Santos CA, Andrade SCS, Teixeira AK, *et al.* (2021) Transcriptome differential expression analysis reveals the activated genes in *Litopenaeus vannamei* shrimp families of superior growth performance. *Aquaculture* **531** .
- Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31** , 3210-3212.
- Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27** , 431-432.
- Sun L, Luo HT, Bu DC, *et al.* (2013) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic acids research* **41** .
- Tan K, Jiang HG, Jiang D, Wang W (2020) Sex reversal and the androgenic gland (AG) in *Macrobrachium rosenbergii* : A review. *Aquaculture and Fisheries* **5** , 283-288.
- van de Weijer ML, Bassik MC, Luteijn RD, *et al.* (2014) A high-coverage shRNA screen identifies TMEM129 as an E3 ligase involved in ER-associated protein degradation. *Nature Communications* **5** .
- Vedula P, Kashina A (2018) The makings of the 'actin code': regulation of actin's biological function at the amino acid and nucleotide level. *Journal of Cell Science* **131** .
- Wang ZK, Luan S, Meng XH, *et al.* (2019) Comparative transcriptomic characterization of the eyestalk in Pacific white shrimp (*Litopenaeus vannamei*) during ovarian maturation. *General and Comparative Endocrinology* **274** , 60-72.
- Wyganowska-Swiatkowska M, Urbaniak P, Nohawica MM, Kotwicka M, Jankun J (2015) Enamel matrix proteins exhibit growth factor activity: A review of evidence at the cellular and molecular levels. *Experimental and Therapeutic Medicine* **9** , 2025-2033.

Xia J, Liu D, Zhou WZ, *et al.*(2022) Comparative transcriptome analysis of brain and gonad reveals reproduction-related miRNAs in the giant prawn, *Macrobrachium rosenbergii*. *Frontiers in Genetics* **13** .

Xian HP, Zhuo ZL, Sun YJ, Liang B, Zhao XT (2018) Circulating long non-coding RNAs HULC and ZNF1-AS1 are potential biomarkers in patients with gastric cancer. *Oncology Letters* **16** , 4689-4698.

Yan BL, Xue B, Zhao L, Gao H (2016) Progress on the research of the growth-related genes in crustaceans. *Transactions of Oceanology and Limnology* **4** , 97-104.

Yang G, Lu ZJ, Qin ZD, *et al.*(2020) Insight into the Regulatory Relationships between the Insulin-Like Androgenic Gland Hormone Gene and the Insulin-Like Androgenic Gland Hormone-binding Protein Gene in Giant Freshwater Prawns (*Macrobrachium rosenbergii*). *International journal of molecular sciences* **21** .

Yuan WL, Jiang SW, Sun D, *et al.* (2019) Transcriptome profiling analysis of sex-based differentially expressed mRNAs and lncRNAs in the brains of mature zebrafish (*Danio rerio*). *BMC genomics* **20** .

Yutaka K, Noriko Y, Yuichi K, Takeshi O (2022) Effect of co-overexpression of the cargo receptor ERGIC-53/MCFD2 on antibody production and intracellular IgG secretion in recombinant Chinese hamster ovary cells. *Journal of bioscience and bioengineering* **22** , 190-196.

Zhang XJ, Yuan JB, Sun YM, *et al.* (2019) Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nature Communications* **10** .

Data Accessibility

All data generated and analyzed in this study are included in this article and its additional files.

Ethics Statement

All experiments were conducted in strict accordance with the guidelines of the Committee for Laboratory Animal Research at Shanghai Ocean University.

Author Contributions

Conceptualization, experimental design, and project administration: LG, ML; Methodology: DL, ZH, LZ; Supervision: LG, ML; Writing of the original draft: DL, LG, and ML. Writing—review and editing: WZ, YW and SS. All authors read and agreed to the published version of the manuscript.

Figure Legends

Figure S1 Homologous top 10 species distribution of *Macrobrachium rosenbergii*'s unigenes annotated in the NR database

Figure S2 KOG function classification of the unigenes from full-length transcripts of *Macrobrachium rosenbergii*

Figure S3 Homologous top 10 species distribution of transcription factors from *Macrobrachium rosenbergii*

Figure S4 GO annotation of the up-regulated differentially expressed genes in brains of male prawns compared to female prawn

Figure S5 Top 30 GO terms of up-regulated differentially expressed genes in the thoracic ganglion of male prawns compared to female prawns

Figure S6 KEGG annotation of differentially expressed genes in the brain of male prawns compared to female prawns

Figure S7 KEGG annotation of differentially expressed genes in the thoracic ganglion of male prawns compared to female prawns

Tables Legends

Table S1 Annotation of the unigenes from full-length transcripts of *Macrobrachium rosenbergii* using NR, Swissprot, KEGG, KOG, eggNOG, GO, and Pfam databases

Table S2 KEGG pathway of the unigenes from full-length transcripts of *Macrobrachium rosenbergii*

Table S3 Transcription factor annotation of the unigenes from full-length transcripts of *Macrobrachium rosenbergii* by AnimalTFDB database

Table S4 Type of simple sequence repeats of unigenes from full-length transcripts of *Macrobrachium rosenbergii*

Table S5 Statistics of transcriptome from brain and thoracic ganglion of *Macrobrachium rosenbergii* by Illumina sequencing

Table S6 Statistics of Illumina transcriptome from brain and thoracic ganglion mapping with unigenes of *Macrobrachium rosenbergii*

