# A chromosome-level genome of the kuruma shrimp (Marsupenaeus japonicus) provides insights into its evolution and cold-resistance mechanism

Xianyun Ren[1], Jianjian Lv[1], Meng Liu[2], Qiong Wang[1], Huixin Shao[1], Ping Liu[3], and Jian Li[3]

[1]Chinese Academy of Fishery Science Yellow Sea Fisheries Research Institute
[2]Affiliation not available
[3]Chinese Academy of Fishery Sciences

April 16, 2024

**Abstract**

Marsupenaeus japonicus is an important marine crustacean species. However, a lack of genomic resources hinders the use of whole genome sequencing to explore their genetic basis and molecular mechanisms for genome-assisted breeding. Consequently, we determined the chromosome-level genome of M. japonicus. In total, 478.42 Gb of genomic sequencing data were obtained using the PacBio platform. A final genome assembly of 1.54 Gb was generated with a contig N50 of 229.97 kb. The 266 Contigs were categorized into 42 chromosomes using high-throughput chromosome conformation capture (Hi-C) technology, with a scaffold N50 of 38.27 Mb, occupying 95.9% of the genome. We identified 24,317 protein-coding genes in the M. japonicus genome, among which annotation was available for 23,986 genes relying on information regarding known proteins in public databases. M. japonicus is commercially valuable because of its coloration and capability for live transport. A gene involved in heme oxygenase (decyclizing) activity and heme oxidation was identified under positive selection in M. japonicas. The truncated protein had an altered helical structure, which might be responsible for resistance to low oxygen, and even the colorful shell of M. japonicus. The high-quality genome assembly enabled the identification of genes associated with cold-stress and cold tolerance in kuruma shrimp through the comparison of eyestalk transcriptomes between the low temperature stressed shrimp and normal temperature shrimp. The genome assembly presented here could be useful in future studies to determine the genetic changes underpinning the ecological traits of M. japonicus, and provide insights for genome assisted breeding.

## 1. Introduction

Penaeid shrimps are marine crustaceans in the Penaeidae family, which comprises a number of economical important species, including the kuruma prawn (*Marsupenaeus japonicus* ) the giant tiger prawn (*Penaeus monodon* ), and the Pacific whiteleg shrimp (*Litopenaeus vannamei* ) (Koyama et al., 2010; Wilson et al., 2000; Yuan et al., 2018). These crustaceans are farmed commercially, making them valuable aquaculture species internationally (FAO, 2019). Penaeid shrimps have complex body patterns and specific structures, including appendages, segments, and antennae with lateral line-like sense organs (Thornber et al., 2019). Therefore, research using these species provides insights into the developmental biology of crustaceans. High quality crustacean genomes assembled at the chromosome level are currently available; however, the presence of high numbers or repeat sequences and the large genome size mean that most of these assemblies are incomplete.

*Marsupenaeus japonicus* is distributed widely, ranging from the Red Sea and East Africa to Japan and South-East Asia (Tsoi et al., 2007). In 2016, the global annual production of *M. japonicus*reached 57, 351

tons, the yield of which represented over 5% of the world's shrimp output, with an output value of more than 860 million US dollars (Figure 1) (FAO, 2019). Among shrimp, *M. japonicus* is one of the most cultured, and has several popular features, such as a good taste, high economic value, rapid growth rate, and low oxygen tolerance, making it suitable for live transportation (Zhao et al., 2021). However, kuruma shrimp have been greatly overexploited in recent years, causing a significant decrease in its natural abundance and prompting the initiation of artificial breeding (Zheng et al., 2020, Ren et al., 2020). Uniquely, *M. japonicus* are able to survive without water for long periods, permitting live transportation of these shrimp to distant markets (Francis et al., 2021). In China, *M. japonicus* is a good choice for species diversification because of its capability for live transportation, export demand, and high price (Wang et al., 2020).

Kuruma shrimp engage in sand diving; therefore, their growth is promoted by including sand substrates in shrimp ponds during aquaculture (Wang et al., 2018; Zhao et al., 2021). These sand substrates also contain mud, and provide an environment that promotes the growth of microorganisms and benthic organisms, which in turn provide nutrients to the shrimp and maintain the water quality (Almeida et al., 2012; Silva & Martinelli-Lemos, 2012). *M. japonicus'* adaptive plasticity, i.e., its ability to cope with changing environmental conditions, could be revealed via transcriptome sequencing combined with a reference genome. In studies on prawns, the distinctions between the morphological and physiological characteristics have rarely been explored, and little is known about its genetic changes. There are two publicly available (short-read based) draft genome assemblies; however, because the kuruma shrimp genome contains a high proportion of repetitive sequences; therefore, these assemblies are highly fragmented, comprising N50 contig lengths of 912 bp (Yuan et al., 2018) and 234,949 bp (Kawato et al., 2021). These genome resources are helpful to understand the genetics of kuruma shrimp; however, they lack the required completeness and contiguity (Ren et al., 2020). Thus, we still lack high-quality genome information at the chromosome level.

In the present study, a chromosome-anchored reference genome of *M. japonicus* was constructed, which is an important addition to the high-quality genome assemblies of decapods, which are currently only available for the black tiger shrimp *Penaeus monodon*(Uengwetwanit et al., 2021), the Pacific white shrimp *Litopenaeus vannamei* (Zhang et al., 2019), the Chinese shrimp (*Fenneropenaeus chinensis* ) (Wang et al., 2021b), the marbled crayfish *Procambarus virginalis* (Gutekunst et al., 2018), and the swimming crab*Portunus trituberculatus* (Tang et al., 2020). Although two draft*M. japonicas* genomes are available, few efforts to relate the species' biology to the genome assembly have been made. We constructed a chromosome-level genome assembly of *M. japonicus* by combining Pacific Bioscience oriented single-molecule real-time (SMRT) sequencing with Illumina paired-end sequencing and high-throughput chromosome conformation capture (Hi-C) technology. By constructing this chromosome-based genome assembly, we not only aimed to promote research into the genetic changes underpinning the ecological traits of *M. japonicus* , but also to provide important resources for the protection and breeding management of *M. japonicus* .

## 2. Materials and Methods

### 2.1 Statement of Ethics

The relevant national and international guidelines were followed during the conductance of the animal experiments and the Yellow Sea Fisheries Research Institute approved the experiments. No specific permits are required to catch wild shrimp from seawater in China. Endangered or protected species were not involved in this study.

### 2.2 Sample collection and preparation

A healthy male *M. japonicus* from the off-shore area by Fujian Takifugu Breeding Station (Zhangzhou, Fujian, China) was selected to sequence and assemble the genome. A standard phenol/chloroform protocol (Sambrook et al., 1989) was used to extract high-quality genomic DNA (gDNA) from muscle tissue. Agarose gel electrophoresis and Qubit 2.0 Fluorimeter (Life Technologies, Carlsbad, CA, USA), were used to check the quality and quantity of the extracted DNA, respectively. Transcriptomic samples from seven tissues, including eyestalk, gills, hepatopancreas, hemocytes, intestines, muscle, and stomach, were collected for cDNA library preparation. Total RNA was extracted using the TRIzol Reagent (Invitrogen, Waltham, MA,

USA) according to the manufacturer's instructions and the quality of RNA was checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

### 2.3 Library construction and sequencing

For Illumina sequencing, a short-insert paired-end DNA library with an insert size of 350 bp was constructed using Illumina's DNA library preparation kits following standard protocols, and then sequenced on the Illumina HiSeq X Ten platform (Illumina Inc., San Diego, CA, USA). Reads containing more than 20% low-quality bases (Q-value [?] 5), the reads containing more than 10% ambiguous bases, as well as the reads with adapter contamination, were all removed. For PacBio sequencing, the high-quality gDNA was sheared into approximately 20 kb fragments using a g-TURE device. Sheared DNA was purified and concentrated using AMPure XP Beads, and then were used to construct the single-molecule real-time (SMRT) bell libraries. Two 20 kb SMRT bell libraries were prepared and sequenced on the PacBio RS II sequencing platform (Pacific Biosciences, Menlo Park, CA, USA. Hi-C technology was further used to obtain the chromosome-level genome assembly. A Hi-C library was constructed according to the standard protocol described previously (Belton et al., 2012) and sequenced on the Illumina Hiseq X Ten platform to generate 150-bp paired-end reads. For long-read RNA sequencing (Iso-Seq), RNA samples from seven tissues were mixed at an equal molarity and sequenced on the PacBio Sequel platform. For Illumina RNA sequencing, the libraries for seven tissues were constructed using the NEBNext mRNA Library Prep Master Mix Set following the manufacturer's instructions (NEB, Ipswich, MA, USA). All libraries were subjected to paired-end 150 bp sequencing on the Illumina Hiseq X Ten.

### 2.4 Genome size estimation and assembly

The genome size of *M. japonicus* was estimated using $k$ -mer analysis. High-quality short-insert size reads were used to calculate the 17-mer frequency distribution and to estimate the genome size according to the formula: genome size = (total number of 17-mer)/(position of peak depth) (Marçais & Kingsford, 2011).

*De novo* assembly of the PacBio reads was performed using the wtdbg2 software (version 2.5) with the parameters of "–node-drop 0.20 –node-len 1536 –node-max 600 -s 0.05 -e 3" (Ruan & Li, 2020). Three rounds of consensus correction was performed using racon (version 1.3.1) with default parameters (Vaser et al., 2017), and then pilon (version 1.22) was used to polish the resulting assembly using the Illumina short paired-end read (Walker et al., 2014) Next, the high-quality paired-end reads generated by the Hi-C method were mapped onto the *M. japonicus* draft genome followed by filtering using HICUP (version 0.7.4) (Wingett et al., 2015) to generate a chromosome-level genome. Briefly, HICUP_TRUNCATER was used to truncate the Hi-C reads at the enzyme digestion ligation site (ˆGATC) and then the resulting trimmed forward and reverse reads were aligned to the genome by bowtie2 (version 2.2.5) (Langmead & Salzberg, 2012), yielding an alignment BAM file. Only unique high-quality and valid alignment results were used to build the raw intra- or inter-chromosomal interaction maps. Lastly, contigs were clustered and anchored into 42 pseudo-chromosomes using ALLHIC (Zhang et al., 2019). Juicebox (v1.18) (https://github.com/aidenlab/Juicebox) was used for manual fine-tuning in a graphic and inter-active fashion to obtain the final chromosome-level assembly.

To evaluate the accuracy of the genome assembly, Burrows-Wheeler Aligner (BWA) (Li & Durbin, 2009) incorporating parameters comprising '-k 32 -w 10 -B 3 -O 11 -E 4' was utilized to align the short-insert paired-end reads to the *M. japonicus* genome. To assess the completeness of the genome, BUSCO (Benchmarking Universal Single Copy Orthologs) (version 4.1.2) (Simão et al., 2015) was performed by searching against the arthropoda_odb10 datasets. In addition, genome completeness was assessed using CEGMA (Core Eukaryotic Genes Mapping Approach) (version 2.5) based on 248 conserved core eukaryotic gene sets (Parra et al., 2007).

### 2.6 Genome annotation

To detect transposable elements (TEs) in the *M. japonicus* genome, two approaches were used: *de novo* prediction and homology-based alignment. RepeatMasker and RepeatProteinMask (version 4.0.7) (http://www.repeatmasker.org/) were used in the homology-based alignment to screen the *M. japoni-*

*cus* genome against the Repbase library (Jurka et al., 2005). For *de novo* prediction, RepeatModeler (v1.0.5, http://repeatmasker.org/RepeatModeler.html), RepeatScout (v1.0.5) (Price et al., 2005), and LTR_FINDER (v1.07) (Xu and Wang, 2007) were used to build a *de novo* library of non-redundant repeats using default settings. Based on the constructed *de novo*library, RepeatMasker (v4.0.7) (Chen, 2009) was then run on the *M. japonicus* genome. In addition, the program Tandem Repeats Finder (TRF, v4.07b) (Benson, 1999) was used to predict tandem repeats, with default settings.

Three approaches were used to predict protein-coding genes in the*M. japonicus* genome, including homology-based prediction,*ab initio* prediction, and transcriptome-based prediction. For the homology-based prediction, TBLASTN (version 2.2.26; E-value [?] 1e-5) (Camacho et al, 2009) was used to align protein sequences from*Homo sapiens* (GCF_000001405.38), *Tetranychus urticae*(GCF_000239435.1), *Caenorhabditis elegans* (GCF_000002985.6),*Crassostrea gigas* (GCF_000297895.1), *Drosophila melanogaster* (GCF_000001215.4), *Daphnia pulex*(GCA_000187875.1), *Ixodes scapularis* (GCF_016920785.1),*Parasteatoda tepidariorum* (GCF_000365465.2), *Litopenaeus vannamei* (GCA_003789085.1), *Tribolium castaneum*(GCF_000002335.3), *Strongylocentrotus purpuratus*(GCF_000002235.5), *Cherax quadricarinatus* , and*Fenneropenaeus chinensis* onto the *M. japonicus* genome. Then, the BLAST hits were concatenated using the software Solar (Yu et al., 2006). GeneWise (version 2.4.1) (Stamatakis, 2014) was used to determine the accurate gene structure of the corresponding genomic region on each BLAST hit. Homology predictions were denoted as the "Homology-set". For *ab initio* prediction, Augustus (version 3.2.3) (Stamatakis, 2014), GlimmerHMM (version 3.0.4) (Majoros et al., 2004), Genscan (version 1.0) (Burge and Karlin, 1997), Geneid (version 1.4.4) (Burge and Karlin, 1997), and SNAP (version 2013-11-29) (Korf, 2004) were used to predict coding regions in the repeat-masked genome sequences. Trinity (version 2.0) (Grabherr et al, 2011) was used to assemble the RNA-seq data from the seven tissues. Program to Assemble Spliced Alignment (PASA) (Haas et al., 2003) then aligned these assembled transcript sequences, together with the full-length transcript sequences generated from PacBio, against the *M. japonicus* genome. Valid alignments were clustered according to genome mapping location and assembled into gene structures. Gene models created by PASA were denoted as the PASA-T-set (PASA transcript set). Besides, RNA-seq reads generated from Illumina were directly aligned onto the genome using Tophat (v2.0.13) (Trapnell et al., 2009), and the Cufflinks (v2.1.1) (Trapnell et al., 2012) was used to predict gene models (Cufflinks-set). Gene models obtained from all the methods were integrated into a comprehensive and non-redundant gene set using the software EvidenceModeler (EVM, v1.1.1) (Hssa et al., 2008). Weights for each type of evidence were set as follows: PASA-T-set > Homology-set > Cufflinks-set > Augustus = GlimmerHMM = Genscan = Geneid = SNAP.

Functional annotations were performed using BLASTP searches against the SwissProt (Boeckmann et al., 2003) and NCBI non-redundant protein (NR) databases (Pruitt et al., 2007) with e-value less than 1e-5. In addition, InterProScan (version 4.8) (Quevillon et al., 2005) was used to screen proteins against five databases: Pfam, PRINTS, PROSITE, ProDom, and SMART, to determine protein domains and motifs. Gene Ontology (GO) terms were retrieved from the corresponding InterPro entry (Apweiler et al., 2001). The Kyoto Encyclopedia of Genes and Genomes (KEGG) databases was also searched to identify enriched pathways (Kanehisa & Goto, 2000).

### 2.7 Gene family construction

To carry out gene family analysis, the whole protein-coding gene repertoires from 21 published genomes, including *Danio rerio* ,*Ctenopharyngodon idellus* , *Oreochromis niloticus* ,*Paralichthys olivaceus* , *Cynoglossus semilaevis* ,*Litopenaeus vannamei* , *Penaeus monodon* ,*Fenneropenaeus chinensis* , *Portunus trituberculatus* ,*Procambarus virginalis* , *Cherax quadricarinatus* ,*Daphnia magna* , *Locusta migratoria* , *Aedes aegypti* ,*Drosophila melanogaster* , *Bombyx mori* , *Daphnia pulex* , *Caenorhabditis elegans* , *Argopecten purpuratus* ,*Crassostrea gigas* , and *Branchiostoma floridae* , were retrieved. For genes with multiple alternative isoforms, the longest transcript of each gene (encoding more than 30 amino acids) was retained. All-against-all BLASTP (v2.2.26) with an e-value threshold of 1e-7 was performed to assess the similarities among the retained protein sequences. OrthoMCL software (Li et al., 2003) was used to construct gene families with

the parameter of '-inflation 1.5'.

### 2.8 Phylogenetic construction and divergence time estimation

The phylogenetic relationships between *M. japonicus* and other species were determined using 174 single-copy orthologous families. Protein sequences in each single-copy family were aligned using MUSCLE (version 3.8.31) with default parameters (Edgar et al., 2004) and a super alignment matrix was constructed by concatenating the alignments. Then, RAxML (version 8.0.19) (Stamatakis, 2014) was used to construct the phylogenetic tree using the maximum-likelihood (ML) method with the PROTGAMMAAUTO model for amino acid substitution. The MCMCTree program implemented in the PAML package was used to estimate the divergence times between the 21 species. Several calibration points applied in the present study were retrieved from the TimeTree website (*http://timetree.org/* ), including *D. rerio* and *C. semilaevis* (206-252 Mya), *A. purpuratus* and *C. gigas*(395-551 Mya), *C. elegans* and *P. trituberculatus* (623–877 Mya), *D. melanogaster* and *P. trituberculatus* (452-557 Mya), *D. melanogaster* and *L. migratoria* , *P. trituberculatus* and *C. quadricarinatus* (281-410 Mya), *P. trituberculatus* and *P. monodon* (271-430 Mya), *F. chinensis* , and *L. vannamei* (55-108 Mya).

### 2.9 Positive selection analysis

The single-copy gene families of *M. japonicus* , *L. vannamei* , *P. monodon* and *F. chinensis* , were used for the positive selection analysis. MUSCLE (Edgar, 2004) was used to perform multiple sequence alignment for protein sequences. A branch-site model of CODEML in the PAML package (Yang, 2007) was used to detect positive selection. The chi-squared test was used to calculate the P-value, which was corrected using the false discovery rate (FDR).

### 2.10 Analysis of stress response genes

To reveal RNA expression changes in response to stress, we acclimated 100 adult shrimp for seven days at 28 degC in oxygenated water in individual 120 L buckets. Shrimp were fed with clams at 5% of the total shrimp weight daily. Every day, 50% of the water was changed. To evaluate the transcriptional response of the shrimp to cold stress, 20 shrimp were maintained at 28 degC (CT, controls) and 20 shrimps were maintained at 10 degC (LT) chiller, and the eyestalks from both groups were sampled for RNA extraction at 12 h and 48 h, at which time the shrimps began to die according a pilot experiment conducted under the same conditions. Eyestalk mRNA was subjected to transcriptome sequencing using the Illumina Hiseq X Ten system (carried out by Novogene Biotechnology, Tianjin, China). We filtered the raw data to remove those reads containing ambiguous bases ('N') or large numbers of low-quality positions ($>$ 10 positions with quality scores $<$ 10). Hisat2 (v2.0.4) (Kim et al., 2015) was then used to map the high quality reads to the *M. japonicus* genome. Read counts were calculated using HTseq software (Anders et al., 2015) and the normalized read counts (fragments per kilobase of transcript per million mapped reads; FPKM) was used to estimate the expression level of the genes. Differentially expressed genes (DEGs) were identified using DESeq2 (Love et al., 2014) and genes with a log2 (fold-change (FC)) [?]1 and adjusted$p$ -value $<$0.05 were defined as significant DEGs.

## 3. Results

### 3.1 Library construction and sequencing

The genome of an adult male shrimp was sequenced using the Illumina and PacBio Sequel platforms. PacBio sequencing generated 244 Gb of data and Illumina sequencing generated 130 Gb data (Table 1). We estimated the genome size 1.39 Gb based on the Illumina sequenced short reads k-mer distribution. The genome sequencing depths were 175.54x (PacBio) and 92.78x (Illumina). The final assembly of the *M. japonicus* genome comprised 1.53 Gb, for which the contig N50 length was 229.974 kb( Table 2).

The Hi-C method successfully clustered the sequences into 42 groups (Figure 2). Finally, a chromosome-level assembly of 1.54 Gb, with a scaffold N50 length of 38.27 Mb, was obtained (Table 2).

### 3.2 Completeness of the assembly

The quality of assembly was evaluated by Illumina short reads mapping, for which the 88.53% coverage showed that the relative completeness of the genome assembly (Table S1). We identified 9,679,458 homozygous SNPs, which indicated a low homozygous rate (0.0099%) and demonstrated that the genome assembly was highly accurate at the single-base level (Table S2).

The BUSCO assessment revealed 87.6% of the 1,066 Arthropoda BUSCOs were identified in the assembled genome (Table S3). Meanwhile, according to the CEGMA analysis, 221 (89.11%) conserved genes were identified in the*M. japonicas* genome (Table S4). These results indicated that the genome assembly had good coverage and completeness.

### 3.3 Gene prediction and annotation

Next we identified TEs and tandem repeats in the *M. japonicus*genome assembly. Approximately 56.07% of the assembly represented repeats, among which the most abundant TEs were DNA transposons (44.76% of the genome), followed by simple repeats (16.88 %), long terminal repeats (LTR, 13.61 %), long interspersed elements (LINE, 7.07%), and short interspersed nuclear elements (SINE, 0.01%) (Table 3 and Figure S1). In the genome, 24,317 protein-coding genes were predicted, with an average of 5.5 exons and a 1,237.45 bp average CDS length (Figure 3 and Table 4). Pathway assignment was successful for 23,986 (98.6%) of the predicted protein-coding genes in almost one of six data pools( Table 5) .

### 3.4 Genome Evolution and Estimation of Evolutionary Rate

To clarify *M. japonicus* genome evolution, 46,975 gene families were clustered from 21 species. To reveal the phylogenetic relationships among *M. japonicus* and other species, 174 single-copy ortholog families derived from the 21 species were identified and the protein sequences of these orthologs were aligned. The maximum-likelihood method was used to performed phylogenetic analysis. According to the phylogenetic analysis, *M. japonicus* might have started to diverge from other species of shrimp approximately 119 million years ago (Figure 5).

### 3.5 Gene family comparison between the *M. japonicus* and other species

Ortholog analysis of genes from 21 species including *Danio rerio* ,*Ctenopharyngodon idellus* , *Oreochromis niloticus* ,*Paralichthys olivaceus* , *Cynoglossus semilaevis* ,*Litopenaeus vannamei* , *Penaeus monodon* ,*Fenneropenaeus chinensis* , *Portunus trituberculatus* ,*Procambarus virginalis* , *Cherax quadricarinatus* ,*Daphnia magna* , *Locusta migratoria* , *Aedes aegypti* ,*Drosophila melanogaster* , *Bombyx mori* , *Daphnia pulex* , *Caenorhabditis elegans* , *Argopecten purpuratus* ,*Crassostrea gigas* , and *Branchiostoma floridae* identified 46,954 clusters of gene families. 174 single-copy orthologs were used to construct the phylogeny and then the expansion and contraction of all ortholog families were analyzed at each node (Figure S4). In the*M. japonicus* lineage, we observed that 126 gene families had undergone expansion. We detected significant enrichment of neuronal function-associated genes (52 families), cuticle structure and remodeling genes (34 families), and transposons (25 families) (Figure S4A and Table S5). For the most highly expanded families, the associated gene ontology (GO) terms included binding, heterocyclic compound binding, and organic cyclic compound binding (Figure S4B and Table S6). We identified 24 contracted gene families in *M. japonicus* , for which the three most frequently associated GO terms were transition metal ion binding, oxidation-reduction process, and electron transfer activity (Figure S4C and Table S7).

There were 4,960 single-copy homologous gene families identified in the four Penaeus shrimps. Compared with those in *L. vannamei* ,*P. monodon* and *F. chinensis* , we observed 586 genes that were subjected to positive selection in *M. japonicus* , which were mainly associated with substance metabolic process. Among them, there was a gene annotated as heme oxygenase (HO)-like, which was clustered into the GO term of heme oxygenase (decyclizing) activity and heme oxidation. The HO-like protein of *M. japonicus* was truncated compared with the same gene in the other three Penaeus shrimps (Figure 6A). Positive selection sites were located at the 131st and 144th amino acid, based on the likelihood-ratio test. Variations in this gene resulted in the disappearance of the kinked helix structure (Figure 6B). Moreover, the reduction in hydrophobic amino acids caused the loss of a transmembrane domain (Figure 6C).

### 3.6 Identification of cold-associated genes

Compared with the control group (28 degC), the eyestalk tissue exhibited 720 (LT *vs.* CT, 343 upregulated and 377 downregulated) and 2,347 (LT *vs.* CT, 1,275 upregulated and 1,072 downregulated) DEGs at 12 h and 48 h under 10 degC exposure, respectively (|Log2 fold-change| >1 and $p$ -value <0.05; Table S8). According to GO analysis, the DEGs from 12 h LT *vs.* CT were enriched mainly in chitin metabolic process (GO:0006030), extracellular region (GO:0005576), and chitin binding (GO:0008061) in the biological process, cellular component, molecular function GO categories, respectively. We observed the enrichment of certain GO terms involved in the neuroendocrine system, such as endopeptidase activity (GO:0004175), endopeptidase inhibitor activity (GO:0004866), and endopeptidase regulator activity (GO:0061135) (Table S9, Figure S5A). In the DEGs from 48 h LT *vs* . CT group, GO term analysis showed that enrichment mainly in organo-nitrogen compound catabolic process (GO:1901565), proteasome core complex (GO:0005839), and threonine-type endopeptidase activity (GO:0004298), in the biological process, cellular component, and molecular function GO categories respectively. We observed the enrichment of certain GO terms involved in the neuroendocrine system, such as cellular catabolic process (GO:0044248), heterocycle catabolic process (GO:0046700), threonine-type endopeptidase activity (GO:0004298), threonine-type peptidase activity (GO:0070003), and endopeptidase activity (GO:0004175) (Table S9, Figure S5B).

We identified the top 20 significantly enriched KEGG pathways in the eyestalk. In the 12 h LT *vs* . CT group comparison, spliceosome (dme03040), proteasome (dme03050), and lysosome (dme04142) were the main clusters. We also observed enrichment of certain KEGG pathways associated with the neuroendocrine system, such as longevity regulating pathway-multiple species (dme04213), phosphatidylinositol signaling system (dme04070), and endocytosis (dme04144) (Table S10, Figure S6A). As for the 48 h LT *vs.* CT group, proteasome (dme03050), drug metabolism-other enzymes (dme00983), and drug metabolism-cytochrome P450 (dme00982) were the main clusters. We also observed enrichment of certain KEGG pathways associated with the neuroendocrine system, such as ubiquitin mediated proteolysis (dme04120), phosphatidylinositol signaling system (dme04070), neuroactive ligand-receptor interaction (dme04080), and SNARE interactions in vesicular transport (dme04130) (Table S10, Figure S6B).

We observed that under low temperature stress, many DEGs were upregulated significantly, most of which were associated closely with the neuroendocrine system, amino acid transport and metabolism, carbohydrate transport, lipid transport and metabolism, HSPs, and hemocyanin. Thus, these DEGs might participate in the *M. japonicus* cold tolerance mechanism (Fig. 8).

### 3.6 Data Records

The sequencing datasets and genomic assembly are maintained in a common repository. The genomic Illumina sequencing data, genomic PacBio sequencing data, and the Hi-C genomic sequencing data are deposited in the NCBI Sequence Read Archive (SRA) database with Bioproject accession PRJNA677851.

### 3.7 Technical Validation

To assess the quality of the assembled genome, its accuracy and completeness were evaluated using BUSCO analysis and the short-read mapping method (Simao et al., 2015). The integrity and completeness of the sampled *M. japonicus* genome was evaluated using BUSCO v4.1.2 based on the arthropoda_odb10 database. We found that 89.4% and 1.0% of 1,013 BUSCO genes are monitored, while BUSCO type genes were partially revealed in the genomic database. Burrows-Wheeler Aligner (BWA) (version 0.7.8-r455) (Li & Durbin, 2009) was used to align the next generation sequencing short reads to the genome and more than 89.44% of the reads were reliably assigned, representing a relatively high proportion of the short-view sequencing data.

### 3.8 Usage Notes

During assembly of the Hi-C sequencing data contig sequences into chromosomes; 100 bp were added to show the unknown gap size between different contigs in the chromosome sequences.

### 3.9 Code validity

The bioinformatic tools and versions used in this work have been described in the main text. Default parameters were applied if no parameters were mentioned. The following is some additional software information used in this work: 1. Trinity version v2.1.1 was used to assemble the RNA sequencing reads, *https://github.com/ trinityrnaseq/trinityrnaseq* . 2. RepeatMasker version 4.0.5 was used to mask the repeat sequences in the genome,*http://repeatmasker.org/* .

## 4. Discussion

*Litopenaeus vannamei* , *Penaeus monodon* , *M. japonicus* , and *Penaeus chinensis* are the four most cultured prawns worldwide. The genomes of *L. vannamei* , *P. monodon* , and *F. chinensis* have been published (Zhang et al., 2019; Uengwetwanit et al., 2021; Wang et al., 2021b); however, so far, no genome of *M. japonicus* has been reported, despite the economic and ecological significance of this species. *M. japonicus* is a major research object due to its position as the most important cultured shrimp. Previously, this species was believed to show adaptive plasticity, and this view was supported by the observed expansion of stress response-related gene families in the present study. Genomic information regarding this species could help to gain a better understanding of its habit of sand diving as well as its environmental adaptation. The present study reports the whole-genome sequence of*M. japonicus* , which was assembled using PacBio long-read data and Hi-C techniques. The genome was assembled based on a 10-Mb benchmark (Reference standard for genome biology, 2018), and is considered to be a high-quality reference genome because of its N50 scaffold length of 38.26 Mb. We believe that it is one of the highest quality crustacean genomes available currently.

Previously, the *M. japonicus* genome was estimated to be 1.94 Gb (Yuan et al., 2018). In the present study, the genome size was 1.54 Gb, which is smaller than that of two other penaeid shrimps, *L. vannam* ei (2.60 Gb) (Zhang et al., 2019) and *P. monodon* (2.59 Gb) (Uengwetwanit et al., 2017), and larger than that of *F. chinensis* (1.38 Gb) (Wang et al., 2021b). Compared with the genome of *L. vannam* ei ( Zhang et al., 2019), the *M. japonicus genom* e contained slightly fewer genes (24,317*vs* . 25,596). The *M. japonicus* and *L. vannamei*genomes contain a similar gene numbers. However, the *M. japonicus*genome contains a markedly lower proportion of repetitive sequences compared with that of *L. vannamei* (56.07% *vs* . 78%, respectively) (Zhang et al., 2019), which might have contributed to genome contraction in *M. japonicus.*

*M. japonicus* is a benthic species that needs sand to survive, while *L. vannamei* is planktonic. The microenvironment of the sediment is very complex, comprising both water and soil. In addition, benthic bivalves have become adapted to extreme environments comprising enriched pathogens and ions, and a low oxygen content. Therefore, benthic crustaceans are likely to have specific molecular mechanisms endowing them with tolerance to extreme environments. *M. japonicus* ' adaption to a lifestyle comprising burial in sediment is likely to involve specific gene families associated with complex signaling systems, ion binding systems, and the immune system.

*M. japonicus* is commercially valuable because of its consumer-appreciated coloration and its capability of being transported live without water (Cheng & Chen, 2002). An HO-like, gene, whose encoded protein is involved in heme oxygenase (decyclizing) activity and heme oxidation, was identified under positive selection in *M. japonicus* . Heme is an important cofactor for oxygen transfer, oxygen storage, and oxygen activation (Shimizu et al., 2019; Tsiftsoglou et al., 2006). It exists in the form of hemocyanin in crustacea. The main function of hemocyanin is oxygen transport (Burmester, 2004; Zhang et al., 2020). *M. japonicus* can tolerate an extremely hypoxic environment; therefore, we speculated that the positive selection on this heme-related gene could benefit the resistance to low oxygen of *M. japonicus* .

Meanwhile, the HO-like protein might participate in shell color regulation. Shell color plays a significant role in consumer acceptability of crustacean species. True green pigments in animals are mostly porphyrinoids. Endogenous porphyrins resulting from the breakdown of heme are usually known as bile pigments, especially biliverdin (Kikuchi et al., 2005; Martins et al., 2019; Schmid & McDonagh, 1975). In mammals, heme oxygenase (HO) is a universal enzyme that degrades heme to biliverdin-IX alpha (BV-IXalpha), liberating ferrous iron ($Fe^{2+}$) and carbon monoxide (CO) as by-products (Mahawar & Shekhawat, 2018; Shekhawat & Verma, 2010). Porphyrins and metalloporphyrins show different colors when they coordinate with different

metal ions (Bonkovsky et al., 2013). A previous study on*F. chinensis* reported that porphyrin metabolism participates in body color formation (Wang et al., 2019).

The body of crustaceans is covered by a rigid exoskeleton, the cuticle, which protects the inner organs against outer environmental factors. Immunological detection showed a diffuse distribution of hemocyanin over the cuticle of *M. japonicus* (Adachi et al., 2005). A red color-related protein was purified from the shell of *M. japonicus* , which was identified belonging to the hemocyanin family (Pan et al., 2020). In this study, the change to the structure of the HO-like protein could result in a change in its function. The loss of the kinked helix structure and transmembrane domain possibly make it easier for the protein to dissociate in the hemolymph to exert its effects. In summary, we considered that the positive selection on the HO-like gene is at least partly responsible for colorful appearance of *M. japonicus* .

The reported reference genome assembly is of high quality, with an improved gene set with increased contiguity and an improved annotation rate. Indeed, 98.64% of the predicted genes could be annotated functionally. This suggested that the genome assembly could better facilitate transcriptomic studies in kuruma shrimp compared with RNA-seq mapping based on *de novo* transcriptome assembly. We used RNA-seq mapping based on the reference genome to identify genes associated with the shrimp's cold-resistant performance. Profitable shrimp production relies on the low-temperature tolerance traits of this species (Jiang et al., 2019). *M. japonicus* is a warm water species that can thrive at temperatures between 18 and 30 degC (optimum = 28 degC), although it can survive at temperatures as low as 6 degC (Ren et al., 2020). However,*M. japonicus* 's nursery temperature should not be lower than 16 degC. Thus, the breeding season and geographical distribution of *M. japonicus* are limited by low temperature, which also affects production and cultivation efficiency. Therefore, the identification or development of a low-temperature tolerant variety of *M. japonicus* remains necessary. Despite its economic and ecological importance, that lack of a high-quality reference genome has limited our knowledge of genes related to cold tolerance in this species.

In the present study, we explored the effects of acute cold stress on the eyestalk transcriptome of *M. japonicus* to identify candidate cold tolerance-related and cold change-related genes. We further studied DEGs associated with the nervous and endocrine systems because of their importance in the regulation a variety of physiological functions and the maintenance of system-wide homeostasis under normal and stressful conditions (Wu et al., 2019; TsuTsui et al., 2020; Ye et al., 2021). A previous study found that increases in biogenic amines enhanced insect cold tolerance (Lubawy et al., 2013). In agreement with that study, in the endocrine system, we observed marked increases in the expression levels of genes encoding Gs-α, PC2, G-protein coupled receptors (GPCRs), and the dopamine receptor 1 (DAR1) after low temperature stress, while the *DAR2* gene showed the opposite trend. As an important biogenic amine, dopamine (DA) binds to GPCRs, and depending on the target tissue and receptor type, stimulates different secondary messengers, mostly $Ca^{2+}$ or cAMP (Farooqui, 2012). The regulation of many behaviors involves the participation of DA, including feeding, locomotion, and evoking systemic responses to different stressful conditions (Aparicio-Simón et al., 2018; Tong et al., 2021). G proteins, important signal transduction factors located on the cell membrane, comprise of three kinds of subunits α, β and γ, which are involved in the regulation of multiple processes (Wei et al., 2020). For example, a GPCR binds to GnRH, which activates intracellular signaling to promotes the synthesis and secretion of follicle stimulating hormone and luteinizing hormone (Ibuchi & Nagayama, 2021). It has been reported that glucose metabolism is modulated by the mature Crustacean hyperglycemic hormone (CHH) protein, produced by PC2-like protein cleavage of the CHH propeptide (Tangprasittipap et al., 2012). These results suggested that DA enhances CHH expression by promoting PC2, resulting in regulation of the hemolymph glucose concentration. Energy mobilization is one of the most important functions of the release of biogenic amines into the insect hemolymph (Lorenz & Gäde, 2009). The mammalian DAR family comprises two classes (D1-like and D2-like), classified according to their intracellular signaling pathways and pharmacological properties. D1-like receptors bind to Gs/Golf class of $G_\alpha$ proteins, resulting in the activation of adenylyl cyclase, which increases the intracellular cAMP level. By contrast, D2-like receptors bind to $G_\alpha$i/Go proteins, resulting in inhibition of adenylyl cyclase, which decreases intracellular cAMP levels (Chen et al., 2017). Therefore, the comparison of eyestalk transcriptomes identified genes associated with cold-stress, which could lead to better management of shrimp farming, with a consequent reduction in

the exploitation of natural populations.

That cold tolerance is a complex regulatory process was demonstrated by the association of DEGs with several pathways. We further studied those DEGs related to metabolism, because they might be a strategy used by*M. japonicus* to resist low temperature stress. Notably, "Glycine, serine and threonine metabolism", "Glutathione metabolism", "Ubiquitin mediated proteolysis", "Glyoxylate and dicarboxylate metabolism", "Thiamine metabolism", "Glycerophospholipid metabolism", "Alanine, aspartate and glutamate metabolism", and "Glycolysis/Gluconeogenesis" were enriched among the DEGs of *M. japonicus* exposed to low temperature for 48 h (Tab. S10). Cold stress upregulated the genes encoding glucose-6-phosphatase and ATP-generating enzymes, which suggested that heat stress induced the rapid production of ATP. This might be related to molecular chaperones requiring high levels of ATP for their functions (Chu et al., 2020). In lipid metabolism, the cold-stress upregulated DEGs were mainly associated with unsaturated fatty acids (UFA) biosynthesis. UFAs are key components of cellular membranes and affect energy metabolism. From the transcriptomic results, genes encoding fatty acid synthesis-related proteins, such as fatty aldehyde dehydrogenase, long chain fatty acid CoA ligase, acetyl-CoA carboxylase, and stearoyl-CoA desaturase, showed high expression after acute cold stress. Fatty aldehyde dehydrogenase, which has a vital function in biomembrane structure and function, catalyzes the conversion of palmitic acid to UFAs by introducing unsaturated bonds. Accumulation of UFAs effectively restores biomembrane fluidity and membrane-related enzyme activity (Los & Murata, 1998). In the *de novo* synthesis of fatty acids, acetyl-CoA carboxylase is the rate-limiting enzyme, catalyzing the conversion of acetyl-CoA to malonyl-CoA (Wang et al., 2021a). As key enzymes for fatty acid synthesis, long chain fatty acid CoA ligase and stearoyl-CoA desaturase, combined with NADPH, produce palmitic acid, which is required for immune defense, biological growth, and biomembranes (Zuo et al., 2017). The observed higher expression of genes encoding fatty acid metabolism regulators in shrimp is not only important for shrimp biology, but also provides a guide for the construction of feed formulations that promote the endurance of shrimp under low temperatures. Taken together, these data indicated that in *M. japonicus* low temperature resistance, metabolic response pathways might have vital functions.

In addition, temperature change and temperature stress-related immune-related genes were identified among the DEGs, such as cold upregulated *hsp70* , *hsp90* , and *hemocyanin C chain-like* . *Hsp70* is conserved gene that responds to various stresses, including pathogen invasion and temperature changes by promoting recovery of the cells from damaged proteins, thereby increasing the probability of survival and life span (Valenzuela-Castillo et al., 2019; Zininga et al., 2018). Hsp90 participates in various cell regulation pathways, and a remarkable proportion of its target proteins are kinases involved in signal transduction, transcriptional, cell cycle regulation, and steroid hormone receptors, as well as mediating the refolding of stress-denatured proteins (Terasawa et al., 2005). In crustaceans, hemocyanin is a respiratory protein in the hemolymph, which is mainly responsible for oxygen binding and carbon dioxide transport (Cheng et al., 2002; Zheng et al., 2019). Overall, HSPs and hemocyanin might have important functions in low temperature adaptation. Kuruma shrimp show sand diving behavior. The sand substrates at the bottom of the sea contain an abundance of oxygen-consuming organic matter and a low oxygen content in the buried sediment (Wei et al., 2020). Resistance to low temperature is an energy consuming process. With less energy production capacity and more energy consumption, *M. japonicus* might not be particularly resistant to low temperature.

## 5. Conclusions

In the present study, we report a high-quality chromosome-scale genome assembly of *M. japonicus* . Examination of the characteristics and functional features of the genome identified some of the molecular mechanisms regulating the shrimp's low temperature tolerance. This reference genome allows downstream biological and industrial investigations that were previously limited by a lack of genomic resources. This genome assembly will be useful for *M. japonicus*research and that of related shrimp and crustacean species, permitting detailed ecological studies.

## Acknowledgements

This research was supported by the project of the National Key R&D Program of China (grant num-

ber 2018YFD0901304), the agricultural improved variety project in Shandong Province-Breeding of breakthrough new prawn varieties with high quality and resistance (grant number 2019LZGC014), the China Agriculture Research System (grant number CARS-48), and the Basic scientific research business expenses of Chinese Academy of Fishery Sciences of "Innovation team project of ecological aquaculture in seawater pond" (2020td46). The authors thank the contribution of several anonymous reviewers of earlier versions, who have greatly improved the quality of this manuscript.

**Author contributions**X.R. wrote the manuscript. P.L. and J.L. conceived the project. J.L. and M.L. performed the bioinformatics analysis. H.S. collected the samples. Q.W. supervised this work. All authors contributed to the final manuscript editing.

**Competing interests**There are no competing interests.

**References**

Adachi, K., Endo, H., Watanabe, T., Nishioka, T., & Hirata, T. (2005). Hemocyanin in the exoskeleton of crustaceans: enzymatic properties and immunolocalization. *Pigment Cell Research,* 18(2), 136–143.

*http://doi.org/10.1111/j.1600-0749.2005. 00217.x*

Almeida, A. C., Fransozo, A., Teixeira, G. M., Hiroki, K. A. N., Furlan, M., & Bertini, G. (2012). Ecological distribution of the shrimp*Nematopalaemon schmitti* (Crustacea: Decapoda: Caridea) in three bays on the south-eastern coast of Brazil. *African Journal of Marine Science* , *34* (1), 93–102.

*https://doi.org/10.2989/ 1814232X.2012.673292*

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* , *31* (2), 166–169.

Aparicio-Simón, B., Piñón, M., Racotta, R., & Racotta, I. S. (2018). Neuroendocrine and metabolic responses of Pacific whiteleg shrimp*Penaeus vannamei* exposed to hypoxia stress. *Latin american journal of aquatic research* , *46* (2), 364–376.

*http://dx.doi.org/10.3856/vol46-issue2-fulltext-12*

Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., et al. (2001). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic acids research* , *29* (1), 37–40.*https://doi.org/10.1093/nar/29.1.37*

Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., & Dekker, J. (2012). Hi–C: a comprehensive technique to capture the conformation of genomes. *Methods* , *58* (3), 268–276.*https://doi.org/10.1016/j.ymeth.2012.05.001*

Bergman, C. M., & Quesneville, H. (2007). Discovering and detecting transposable elements in genome sequences. *Briefings in Bioinformatics* , *8* (6), 382–392.

*https://doi.org/10.1093/bib/bbm048*

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* , *27* (2), 573–580.*https://doi.org/10.1093/nar/27.2.573*

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic acids research* , *31* (1), 365–370.*https://doi.org/10. 1093/nar/gkg095*

Bonkovsky, H. L., Guo, J. T., Hou, W., Li, T., Narang, T., & Thapar, M. (2013). Porphyrin and heme metabolism and the porphyrias.*Comparative Physiology* , *3* (1), 365–401. *https: //doi.org/10.1002/cphy.c120006*

Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology* , *268* (1), 78-94.

Burmester, T. (2004). Evolutionary history and diversity of arthropod hemocyanins. *Micron, 35* (1-2), 121–122.

*https://doi.org/10.1016/j.micron.2003.10.034*

Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* ,*268* (1), 78–94. *https://doi.org/10. 1006/jmbi.1997.0951*

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* , *10* (1), 1–9.

*https://doi.org/10.1186/1471-2105-10-421*

Cheng, S. Y., & Chen, J. C. (2002). Study on the oxyhemocyanin, deoxyhemocyanin, oxygen affinity and acid-base balance of*Marsupenaeus japonicus* following exposure to combined elevated nitrite and nitrate. *Aquatic Toxicology* , *61* (3-4), 181–193. *https://doi.org.10.1016/s0166-445x(02)00053-x*

Cheng, W., Liu, C. H., Yan, D. F., & Chen, J. C. (2002). Hemolymph oxy-hemocyanin, protein, osmolality and electrolyte levels of whiteleg shrimp *Litopenaeus vannamei* in relation to size and molt stage.*Aquaculture* , *211* , 325–339.*https://doi.org/10.1016/S0044-8486(01)00768-2*

Chen, N. S. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics* ,*25* (1), 4–10.*https://doi.org/10. 1002/0471250953.bi0410s05*

Chen, P., Chen, P., Li, T., Shen, Q., Yan, D. F., & Zhang, L., (2017). Two dopamine D2-like receptor genes from the silkworm (*Bombyx mori* ) and their evolutionary history in metazoan. *Scientific reports* , *7* (1), 1–12.*https://doi.org/10.1038/s41598 -017-07055-5*

Chu, T., Liu, F., Qin, G., Zhan, W., Wang, M., & Lou, B. (2020). Transcriptome analysis of the *Larimichthys polyactis* under heat and cold stress. *Cryobiology* , *96* , 175–183.*https://doi.org/10.1016/j.cryobiol.2020.06.014*

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* ,*32* (5), 1792–1797.*https://doi.org/10. 1093/nar/gkh340*

Farooqui, T. (2012). Review of octopamine in insect nervous systems.

*Open Access*

*Insect Physiology* 4, 1. *http://doi.org/10.2147/oaip.s20911*

Food and Agriculture Organization of the United Nations Global Production Statistics, 2019.*http://www.fao.org/fishery/statistics/global-aquaculture-production/ query/en* (accessed 17 February 2019)

Francis, B., Antony, J., Aravind, R., Shyne Anand, P., Balasubramanian, C. P., & Vijayan, K. K. (2021). Reproductive performance, salinity tolerance, growth and production performance of a cryptic species*Penaeus* (*Marsupenaeus* ) *japonicus* . *Aquaculture Research* . *https://doi.org/10.1111/are.15424*

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* ,*29* (7), 644–652.*https://doi.org/ 10.1038/nbt.1883*

Gutekunst, J., Andriantsoa, R., Falckenhayn, C., Hanna, K., Stein, W., Rasamy, J., & Lyko, F. (2018). Clonal genome evolution and rapid invasive spread of the marbled crayfish. *Nature Ecology & Evolution* , *2* (3), 567–573.

*https://doi.org/10.1038/ s4155 9-018-0467-9*

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith Jr, R. K., Hannick, L. I., & White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies.*Nucleic Acids Research* , *31* (19), 5654–5666.*https://doi.org/10.1093/nar/gkg770*

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using Evidence Modeler and the Program to Assemble Spliced Alignments. *Genome Biology* ,*9* (1), 1–22.*https://doi.org/10.1186/gb-2008-9-1-r7*

Ibuchi, K., & Nagayama, T. (2021). Opposing effects of dopamine on agonistic behaviour in crayfish. *Journal of Experimental Biology* ,*224* (12), jeb242057. *https://doi.org/10.1242/jeb.242057*

Jiang, S., Zhou, F., Yang, Q., Huang, J., Yang, L., & Jiang, S. (2019). Impact of temperature stress on oxygen and energy metabolism in the hepatopancreas of the black tiger shrimp, *Penaeus monodon*(Crustacea: Decapoda: Penaeidae). *Pakistan Journal of Zoology* , *51* (1), 141–148.

*http://dx.doi.org/10.17582/journal.pjz/2019. 51.1.141.148*

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research* , *110* (1-4), 462–467.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* , *28* (1), 27–30.*https://doi.org/10.1093/nar/28.1.27*

Kawato, S., Nishitsuji, K., Arimoto, A., Hisata, K., Kawamitsu, M., & Hirono, I. (2021). Genome and transcriptome assemblies of the kuruma shrimp, *Marsupenaeus japonicus* . *G3 Genes| Genomes| Genetics* . *https://doi.org/10.1093/ g3journal/jkab268*

Kim, D., Langmead, B. & Salzberg, S. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods 12*(4), 357–360.*https://doi.org/10.1038/ nmeth.3317*

Kikuchi, G., Yoshida, T., & Noguchi, M. (2005). Heme oxygenase and heme degradation. *Biochemical Biophysical Research Communication, 338* (1), 558–567. *https://doi.org/10. 1016/j.bbrc.2005.08.020*

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* , *5* (1), 59.*https://doi.org/10.1186/1471-2105-5-59*

Koyama, T., Asakawa, S., Katagiri, T., Shimizu, A., Fagutao, F.F., & Mavichak, R. (2010). Hyper-expansion of large DNA segments in the genome of kuruma shrimp, *Marsupenaeus japonicus* . *BMC Genomics* ,*11* (1), 1–12.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods* , *9* (4), 357–359.*https://doi.org/10.1038/nmeth.1923*

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* , *25* (14), 1754–1760. *https://doi.org/10.1093/ bioinformatics/btp324*

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research* , *13* (9), 2178–2189.

*https://doi.org/10.1101/gr.1224503*

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., & Wang, J. (2010). The sequence and de novo assembly of the giant panda genome.*Nature* , *463* (7279), 311–317.*https://doi.org/10.1038/nature08696*

Lorenz, M. W., & Gäde, G. (2009). Hormonal regulation of energy metabolism in insects as a driving force for performance. *Integrative and Comparative Biology* , *49* (4), 380-392. *https://doi.org/10.1093/icb/icp019*

Los, D. A., & Murata, N. (1998). Structure and expression of fatty acid desaturases. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* , *1394* (1), 3–15.*https://doi.org/10.1016/S0005-2760(98)00091-5*

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* , *15* (12), 1–21.*https://doi.org/10.1186/s13059-014-0550-8*

Lubawy, J., Urbański, A., Colinet, H., Pflueger, H. J., & Marciniak, P. (2020). Role of the insect neuroendocrine system in the response to cold stress. *Frontiers in physiology* , *11* , 376.

*https://doi.org/10.3389/fphys.2020.00376*

Mahawar, L., & Shekhawat, G. S. (2018). Haem oxygenase: A functionally diverse enzyme of photosynthetic organisms and its role in phytochrome chromophore biosynthesis, cellular signalling and defence mechanisms.*Plant Cell and Environment, 41* (3), 483–500.

*https://doi.org/10.1111/pce.13116*

Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004) TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders.*Bioinformatics* , *20* (16), 2878–2879.

*https://doi.org/10.1093/bioinformatics/bth315*

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers.*Bioinformatics* , *27* (6), 764–770.

Martins, C., Rodrigo, A. P., Cabrita, L., Henriques, P., Parola, A. J., & Costa, P. M. (2019). The complexity of porphyrin-like pigments in a marine annelid sheds new light on haem metabolism in aquatic invertebrates. *Scientific Report, 9* (1), 12930.

*http://doi.org/10.1038/s41598-019-49433-1*

Pan, C., Ishizaki, S., Chen, S., Hao, S., Zhou, J., & Yang, X. (2020). Purification, characterization and antibacterial activities of red color-related protein found in the shell of kuruma shrimp,*Marsupenaeus japonicus* . *Food Chemistry, 310* , 125819.

*http://doi.org/10.1016/j.foodchem.2019.125819*

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* , *23* (9), 1061–1067.

Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes.*Bioinformatics* , *21* (suppl_1), i351–i358.*https://doi.org/ 10.1093/bioinformatics/bti1018*

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated nonredundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* , *35* (suppl_1), D61–D65.

*https://doi.org/10.1093/ nar/gkl842*

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., & Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Research* , *33* (suppl_2), W116–W120. *https://doi.org/10.1093/nar/gki442*

Ren, X.Y., Xu, Y., Zhang, Y.B., Wang, X., Liu, P., & Li, J. (2020). Comparative accumulation and transcriptomic analysis of juvenile *Marsupenaeus japonicus* under cadmium or copper exposure. *Chemosphere* , *249* , 126157.*https://doi.org/ 10.1016/j.chemosphere.2020.126157*

Ruan, J., & Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature Methods, 17* (2), 155–158.*https://doi.org/10.1038/s41592-019-0669-3*

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Molecular cloning: a laboratory manual (No. Ed. 2). Cold spring harbor laboratory press.

Schmid, R., & McDonagh, A. F. (1975). The enzymatic formation of bilirubin. *Annals of the New York Academy of Sciences, 244* , 533–552. *https://doi.org/10.1111/j. 1749-6632.1975.tb41553.x*

Shekhawat, G. S., & Verma, K. (2010). Haem oxygenase (HO): an overlooked enzyme of plant metabolism and defence. *Journal of the Experimental Botany* , *61* (9), 2255–2270.

*http://doi.org/10.1093/jxb/erq074*

Shimizu, T., Lengalova, A., Martinek, V., & Martinkova, M. (2019). Heme: emergent roles of heme in signal transduction, functional regulation and as catalytic centres. *Chemical Society Reviews ,48* (24), 5624–5657. *http://doi.org/10.1039/c9cs00268e*

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* , *31* (19), 3210–3212.*https://doi.org/10. 1093/bioinformatics/btv351*

Silva, D. C., & Martinelli-Lemos, J. M. (2012). Species composition and abundance of the benthic community of Axiidea and Gebiidea (Crustacea: Decapoda) in the Marapanim Bay, Amazon estuary, northern Brazil.*Zoologia-Curitiba* , *29* (2), 144–158.

*https://doi.org/10.1590/S1984-46702012000200007*

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics ,30* (9), 1312–1313. *10.1093/bioinformatics/btu033*

Tang, B., Zhang, D., Li, H., Jiang, S., Zhang, H., & Ren, Y. (2020). Chromosome-level genome assembly reveals the unique genome evolution of the swimming crab (*Portunus trituberculatus* ). *GigaScience* ,*9* (1), giz161. *gigascience/giz161*

Tangprasittipap, A., Chouwdee, S., Phiwsaiya, K., Laiphrom, S., Senapin, S., Flegel, T. W., & Sritunyalucksana, K. (2012). Structure and expression of a shrimp prohormone convertase 2. *General and comparative endocrinology* , *178* (2), 185–193.*https://doi.org/10.1016/j.ygcen.2012.05.005*

Terasawa, K., Minami, M., & Minami, Y. (2005). Constantly updated knowledge of Hsp90. *Journal of Biochemistry* , *137* (4), 443–447.*https://doi.org/10.1093/jb/ mvi056*

Thornber, K., Verner-Jeffreys, D., Hinchliffe, S., Rahman, M. M., Bass, D., & Tyler, C. R. (2019). Evaluating antimicrobial resistance in the global shrimp industry. *Reviews in Aquaculture* , *12* (2), 966–986. *https://doi.org/10.1111/raq.12367*

Tong, R., Pan, L., Zhang, X., & Li, Y. (2021). Neuroendocrine-immune regulation mechanism in crustaceans: A review. *Reviews in Aquaculture* . *https://doi.org/10. 1111/raq.12603*

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics ,25* (9), 1105–1111. *https://doi.org/10. 1093/bioinformatics/ btp120*

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* , *7* (3), 562–578.*https://doi.org/10. 1038/nprot.2012.016*

Tsiftsoglou, A. S., Tsamadou, A. I., & Papadopoulou, L. C. (2006). Heme as key regulator of major mammalian cellular functions: molecular, cellular, and pharmacological aspects. *Pharmacology Therapeutics, 111* (2), 327–345. *https://doi. org.10.1016/j.pharmthera.2005.10.017*

Tsoi, K. H., Chan, T. Y., & Chu, K. H. (2007). Molecular population structure of the kuruma shrimp *Penaeus japonicus* species complex in western Pacific. *Marine Biology* , *150* (6), 1345–1364.*https://doi.org/10.1007/s00227-006-0426-x*

Tsutsui, N., Kobayashi, Y., Izumikawa, K., & Sakamoto, T. (2020). Transcriptomic analysis of the kuruma prawn *Marsupenaeus japonicus* reveals possible peripheral regulation of the ovary. *Frontiers in Endocrinology* , *11* , 541.*https://doi.org/10. 3389/fendo.2020.00541*

Uengwetwanit, T., Pootakham, W., Nookaew, I., Sonthirod, C., Angthong, P., Sittikankaew, K., & Karoonuthaisiri, N. (2021). A chromosome-level assembly of the black tiger shrimp (*Penaeus monodon* ) genome facilitates the identification of novel growth-associated genes. *Molecular Ecology Resources. https://doi.org/ 10. 1111/1755-0998.12952.*

Valenzuela-Castillo, A., Sanchez-Paz, A., Castro-Longoria, R., Lopez-Torres, M. A., & Grijalva-Chon, J. M. (2019). Hsp70 function and polymorphism, its implications for mollusk aquaculture: a review.*Latin American Journal of Aquatic Research* , *47* (2), 224–231. *http://dx.doi.org/10.3856/vol47-issue2 -fulltext-2*

Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads.*Genome Research* , *27* (5), 737–746.

*https://doi.org/10.1101/gr.214270.116*

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one* , *9* (11), e112963.*https://doi.org/10.1371/journal.pone.0112963*

Wang, L., Wang, H., Wu, N., Zhu, C.K., Wu, X. J., Ding, C. L., & Li, J. L. (2021a). Gene expression associated with acute low thermal stress of*Macrobrachium nipponense* . *Aquaculture Research* , *52* , 4325–4335. *https://doi.org/10.1111/are. 15270*

Wang, P. P., Wang, J., Su, Y. Q., Liu, Z. X., & Mao, Y. (2020). Air exposure affects physiological responses, innate immunity, apoptosis and DNA methylation of kuruma shrimp, *Marsupenaeus japonicus* . *Frontiers in Physiology* , *11* , 223.*https://doi.org/10.3389/fphys.2020.00223*

Wang, Q., He, Y., & Li, J. (2019). Conjoint Analysis of SMRT- and Illumina-Based RNA-Sequencing Data of *Fenneropenaeus chinensis*Provides Insight Into Sex-Biased Expression Genes Involved in Sexual Dimorphism. *Frontiers in Genetics* , *10* , 1175.

*https://doi.org/10.3389/fgene.2019.01175*

Wang, Q., Ren, X., Liu, P., Li, J., Lv, J., & Wang, J. (2021b). Improved genome assembly of Chinese shrimp (*Fenneropenaeus chinensis* ) suggests adaptation to the environment during evolution and domestication. *Molecular Ecology Resources* .

*https://doi.org/10.1111/1755-0998.13463*

Wang, W. L., Ishikawa, M., Koshio, S., Yokoyama, S., Hossain, M. S., & Moss, A. S. (2018). Effects of dietary astaxanthin supplementation on juvenile kuruma shrimp, *Marsupenaeus japonicus* .*Aquaculture* , *491* , 197–204.*https://doi.org/10.1016/j.*

*aquaculture.2018.03.025*

Wei, C., Pan, L., Zhang, X., & Tong, R. (2020). Comparative transcriptome analysis of eyestalk from the white shrimp*Litopenaeus vannamei* after the injection of dopamine. *Gene* , *763* , 145115.

*https://doi.org/10.1016/j.gene.2020.145115*

Wei, M., Ge, H., Shao, C., Yan, X., Nie, H., & Duan, H. (2020). Chromosome-level clam genome helps elucidate the molecular basis of adaptation to a buried lifestyle. *iScience* , *23* (6), 101148.*https://doi.org/10.1016/j.isci.2020.101148*

Wilson, K., Cahill, V., Ballment, E., & Benzie, J. (2000). The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon* : are malacostracan crustaceans more closely related to insects than to

branchiopods? *Molecular Biology Evolution* , *17* , 863–874.*https://doi.org/10.1093/oxfordjournals.molbev.a026366*

Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., & Andrews, S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research* , *4* . *https://doi.org/10.12688/f1000research.7334.1*

Wu, D., Huang, Y., Chen, Q., Jiang, Q., Li, Y., & Zhao, Y. (2019). Effects and transcriptional responses in the hepatopancreas of red claw crayfish *Cherax quadricarinatus* under cold stress. *Journal of Thermal Biology* , *85* , 102404.

Xu, Z., & Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* , *35* (suppl_2), W265–W268.

*https://doi.org/10.1093/nar/gkm286*

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood.*Molecular Biological Evolution, 24* (8), 1586–1591.

*https://doi.org/10.1093/molbev/msm088*

Yang, Z., & Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages.*Molecular Biological Evolution, 19* (6), 908–917.

*https://doi.org/10.1093/oxfordjournals.molbev.a00 4148*

Ye, H., Wilder, M. N., Dircksen, H., & Chung, J. S. (2021). Recent Advances in Crustacean Endocrinology. *Frontiers in Endocrinology* ,*12* .

*https://doi.org/10. 3389/fendo.2021.730642*

Yu, X. J., Zheng, H. K., Wang, J., Wang, W., & Su, B. (2006). Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. *Genomics* , 88(6), 745–751.*https://doi.org/10.1016/ j.ygeno.2006.05.008*

Yuan, J., Zhang, X., Liu, C., Yu, Y., Wei, J., & Xiang, J. (2018). Genomic resources and comparative analyses of two economical penaeid shrimp species, *Marsupenaeus japonicus* and *Penaeus monodon* . *Marine Genomics* , *39* , 22–25.*https://doi.org/10.1016/j.margen.2017.12.006*

Zhang, J., Nielsen, R., & Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biological Evolution* , *22* (12), 2472–2479. *https://doi.org/10.1093/ molbev/msi237*

Zhang, X., Zhang, S., Zhao, Q., Ming, R., & Tang, H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nature plants* , *5* (8), 833–845.

*https://doi.org/10.1038/s41477-019-0487-8*

Zhang, Z., Yao, D., Yang, P., Zheng, Z., Aweya, J. J., & Lun, J. (2020). Nuclear receptor E75 is a transcription suppressor of the*Litopenaeus vannamei* small subunit hemocyanin gene.*Developmental & Comparative Immunology* , *107* , 103662.

*https: //doi.org/10.1016/j.dci.2020.103662*

Zhang, X., Yuan, J., Sun, Y., Li, S., Gao, Y. I., Yu, Y., et al. (2019). Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nature Communications* , *10* (1), 356.

*https://doi. org/10.1038/s41467-018-08197-4*

Zhao, J., Liu, X., Liao, X., He, Z., Chen, X., Sun, C., & Ni, Z. (2021). Transcriptome analysis reveals the effects of sand substrate removal and body colour change of kuruma shrimp, *Marsupenaeus japonicus* . *Aquaculture Research* , *52* , 577–588.

*https://doi.org/10.1111/are.14915*

Zheng, J., Cao, J., Mao, Y., Su, Y., & Wang, J. (2019). Effects of thermal stress on oxidative stress and antioxidant response, heat shock proteins expression profiles and histological changes in*Marsupenaeus japonicus* . *Ecological Indicators* , *101* , 780–791.*https://doi.org/10.1111/are.14915*

Zheng, J. B., Mao, Y., Su, Y. Q., & Wang, J. (2020). Cross talk between heat shock protein 10 and a heat shock factor identified from*Marsupenaeus japonicas. International Journal of Biological Macromolecules* , *147* , 1041–1052. *https://doi. org/10.1016/j.ijbiomac.2019.10.072*

Zininga, T., Ramatsui, L., & Shonhai, A. (2018). Heat shock proteins as immunomodulants. *Molecules* , *23* , E2846.*https://doi.org/10.3390/molecules2311 2846*

Zuo, H., Gao, J., Yuan, J., Deng, H., Yang, L., Weng, S., He, J., & Xu, X. (2017). Fatty acid synthase plays a positive role in shrimp immune responses against *Vibrio parahaemolyticus* infection. *Fish & Shellfish Immunology* , *60* , 282–288.

*https://doi.org/10.1016/j.fsi.2016.11.054*

**Hosted file**

`Figures.docx` available at https://authorea.com/users/406867/articles/712559-a-chromosome-level-genome-of-the-kuruma-shrimp-marsupenaeus-japonicus-provides-insights-into-its-evolution-and-cold-resistance-mechanism

**Hosted file**

`Tables.docx` available at https://authorea.com/users/406867/articles/712559-a-chromosome-level-genome-of-the-kuruma-shrimp-marsupenaeus-japonicus-provides-insights-into-its-evolution-and-cold-resistance-mechanism