Charybdis japonica genome provides insights into desiccation adaptation and sex-determining region

Fangrui Lou[1], Tianxiang Gao[2], Qi Liu[3], and Zhiqiang Han[4]

[1]Yantai University
[2]Fishery College
[3]Wuhan Gooalgene Technology Company
[4]Affiliation not available

April 16, 2024

Abstract

Charybdis japonica predominantly inhabits the intertidal zones and has high desiccation tolerance. We present the first chromosome-level *C. japonica* genome, which contains 51 chromosomes, and the revised genome is 1431.02 Mb in length and has a contig N50 size of 29.67 Mb. Among the contigs, 91.42% were anchored to 51 chromosomes. Additionally, 824.02 Mb repeat elements, 30,900 coding genes, 474 miRNAs, 15,570 tRNAs, 309 rRNAs, and 157 miRNAs were identified in the *C. japonica* genome. The whole-genome resequencing data can contribute to the identification of sex-related single-nucleotide polymorphisms and insertion–deletion mutations. The 0-10,120,000 bp of chromosome 37 is the sex-determining region of *C. japonica*. Comparative genome analysis identified 1,138 *C. japonica*-specific gene families. Phylogenetic analysis showed that *C. japonica* has a close relationship with *Portunus trituberculatus*, which also belonged to family Portunidae, and differentiated 42.1-135.5 million years ago. Demographic history analysis suggested that the maximum effective population size of *C. japonica* was maintained until 0.5×105 years ago. Relative evolution rate showed that *C. japonica* evolved slower than *Daphnia magna*, *Penaeus vannamei*, and *P. trituberculatus*. Compared with other species, metabolism rate, oxygen supply, oxidative stress, and various transporter-related genes were expanded or underwent positive selection in *C. japonica*, which might contribute to *C. japonica*'s ability to overcome diverse stresses in drought environment. Decoding the present genome provides valuable information for revealing the desiccation-adaptive and sex-determining mechanisms of *C. japonica* and also enriches the genetic information to explore the evolutionary history and environmental adaptation strategies of other Portunidae crabs.

1. Introduction

The Portunidae family, which comprises approximately 300 identified species, has successfully settled in a variety of habitats from ocean trenches to hydrothermal vents, intertidal mud flats, and terrenes (Cui et al., 2021). Portunidae crabs have high ecological and economic value. Although genomic information is believed to reveal the regulatory mechanism of the biological characteristics of Portunidae members, only a handful of Portunidae genomes have been published, such as those of*Scylla paramamosain* (Zhao et al., 2021) and *Portunus trituberculatus* (Tang et al., 2020). Hence, completing the genome database of Portunidae crabs is urgently needed.

Charybdis japonica (H. Milne Edwards, 1861; Figure 1) belongs to genus *Charybdis* of family Portunidae (Decapoda, Crustacea). This species is a eurythermic and euryhaline marine crab (Yu et al., 2005) and predominantly inhabits the intertidal zones of China, Japan, the Korean Peninsula, and Southeast Asia countries. *C. japonica* is a delicious aquatic food with high amino acid and unsaturated fatty acid contents; thus, it has high economic and medicinal values (Yu et al., 2004). This fishery resource was further developed after the 1990s and gradually became an important fishery resource in China's coastal areas (Zheng, 2015).

Meanwhile, artificial *C. japonica* farming is active in China. The economic traits of *C. japonica* is gender specific during the farming process. Specifically, male *C. japonica* has a faster growth rate, but females show a higher economic value (Kolpakov and Kolpakov, 2011). Additionally, *C. japonica*has a higher desiccation tolerance than other crabs (Yu et al., 2004; Yu et al., 2005; Zheng et al., 2013), which seems to help improve survival rates for intertidal living and during farming and transportation processes.

Dry transportation is one of the important segments of artificial crustacean culture, which is beneficial to improve crustacean welfare and reduce economic loss. However, dry transportation can still more easily break the balance between crustacean and water environment and lead to the decrease of crustacean physique, vitality, and survival rate (Lorenzon et al., 2008; Paital, 2013). The tissues of crustaceans under desiccation are always subjected to water deficit stress and hypoxia stress (Ridgway et al., 2006; Paital, 2013). The gills are the first tissue to suffer from desiccation because it is the main respiratory organ of many crustaceans. The gill prefers to use oxygen in water rather than that in air. The absence of water media leads to the aggregation of gill lamellae, which reduces the gas exchange area of the gill and ultimately leads to gill dysfunction syndrome (Greenaway et al., 1996; Taylor and Wheatley, 1989; Levin, 2003). Insufficient oxygen supply will further hinder oxidative metabolism, reduce ATP production, cause imbalance in cell homeostasis, and even cause body death (Li et al., 2010). Therefore, desiccation tolerance is an important factor that influences the biological processes and survival of many marine crustaceans, whether living in the natural intertidal zone or in the dry transport process of artificial farming.

Many aquatic organisms have evolved and obtained some adaptive strategies to establish higher tolerance to desiccation. For example, some organisms in drought conditions can increase oxygen supply by beating their hearts faster (Morris et al., 1999). Reducing oxygen consumption rate is also an effective strategy to deal with the hypoxia stress caused by desiccation (Urbina et al., 2013). Aquatic organisms exposed to desiccation can increase the synthesis of respiratory proteins (such hemoglobin and hemocyanin) to enhance the oxygen-carrying rate of proteins and oxygen transport (Pascual, 2003; Urbina et al., 2013). Crustaceans living in intertidal zones may be exposed to air and suffer from drought stress for several hours during low tide. These crustaceans have evolved the hard chitin shell to keep their gills moist for long periods of time to adapt to the periodic dryness of the intertidal zone. Meanwhile, adjustments in metabolic mechanisms are also valuable for crustaceans living in the intertidal zone (Taylor and Greenaway, 1984; Lu et al., 2016). However, the molecular mechanism of desiccation tolerance in crustaceans has not yet been satisfactorily elucidated.

Considering the stronger desiccation tolerance of *C. japonica* (Yu et al., 2004; Yu et al., 2005; Zheng et al., 2013), this species can be used as a model crustacean to understand the regulation mechanism of desiccation tolerance. Whole-genome genetic information can provide a microperspective for revealing the desiccation-adaptive plasticity of*C. japonica* , that is, its flexibility in terms of the capability to cope with water deficit stress and hypoxia stress. Meanwhile, the sex differentiation of *C. japonica* may be similar to that of many crustaceans and controlled by androgenic or estrogen gland; thus, it is susceptible to the influence of the external environment (Cui et al., 2021). Therefore, decoding the whole-genome genetic information can also enhance our understanding of the molecular mechanisms of the sex differentiation of *C. japonica* . In fact, the publication of the whole-genome information of *C. japonica* will help enrich the genetic resources of family Portunidae and provide many insights into their evolutionary history and environmental adaptation strategies. Therefore, the whole-genome sequencing of *C. japonica* is necessary to carry out.

Here, we characterized a high-quality chromosome-anchored reference genome of*C. japonica* for the first time by combining short Illumina reads, long PacBio reads, and Hi-C reads. We tried to correlate the reported genomic data with the biology, evolutionary history, and desiccation tolerance mechanisms of this species. Additionally, the genome regions associated with underlying sex determination in *C. japonica* were predicted based on whole-genome resequencing data. In-depth study of this extensive data will provide groundbreaking mechanical insights into the genetic mechanisms associated with sex differentiation and desiccation tolerance in *C. japonica* and other crustaceans.

2. Materials and methods

2.1 Sampling, DNA extraction, and RNA extraction

Wild adult female *C. japonica* (approximately 210 g) individuals were collected in December 2019 from Zhoushan. The animal study and procedure were approved by the Animal Ethics Committee of Yantai University. Fresh gill, muscle, heart, and intestinal tissues were collected using sterilized dissection scissor and scalpel and then snap-frozen in liquid nitrogen. High-quality genomic DNA was extracted from the muscle tissues of *C. japonica* using the Blood & Cell Culture DNA Mini Kit (QIAGEN, GER), and high-quality RNA was extracted from *C. japonica* tissues using the TRIzol Reagent Kit (Invitrogen, USA). The integrity, purity, and concentration of the genomic DNA were assessed using 1% agarose gel electrophoresis, NanoDrop 2000 analyzer (Thermo Fisher Scientific, USA), and Qubit 3.0 analyzer (Thermo Fisher Scientific, USA). RNA integrity and concentration were measured using the Agilent Bioanalyzer 2100 System (Agilent Technologies, USA) and NanoDrop 2000 analyzer (Thermo Fisher Scientific, USA).

2.2 Library construction and sequencing

Three libraries, including Illumina library, SMRTbell library, and Hi-C library, were combined to construct the high-quality *C. japonica*genome information. A high-quality short-insert (300–350 bp) paired-end (PE) Illumina library was constructed in accordance with the Illumina standard protocol (Illumina, USA), and the Illumina library was then sequenced on the Illumina NovaSeq-6000 platform. A high-quality SMRTbell library with a fragment size of 20 Kb was prepared using the SMRTBell Template Preparation Kit 1.0 (PacBio, USA) according to the manufacturer's protocol to obtain long genomics reads for promoting genome assembly. The constructed SMRTBell library was added to one SMRT cell and then transferred to the PacBio Sequel II sequencing platform for long-read genomic sequencing. A high-quality Hi-C library was constructed to obtain a chromosome-scale genome assembly, and the Hi-C library was sequenced using the Illumina NovaSeq-6000 platform. Additionally, a high-quality 150 bp PE RNA-sequencing (RNA-seq) library was constructed according to the Illumina standard protocol (Illumina, USA) and then sequenced on an Illumina NovaSeq-6000 platform.

Twenty *C. japonica* individuals (11 females and 9 males) were used for whole-genome resequencing. High-quality short-insert (300–350 bp) PE Illumina libraries were constructed in accordance with the Illumina standard protocol (Illumina, USA) and then sequenced on the Illumina NovaSeq-6000 platform.

2.3 Data filtering

All raw Illumina reads were filtered by removing reads that included adapter sequences, duplicated sequences, unknown nucleotides greater than 10%, and low-quality bases (quality scores [?] 5) greater than 50%. Hi-C reads that contain adapter sequences or less than 50 bp in length were removed, and only PE Hi-C reads were retained. Bases with a quality score of less than 20 at both ends of the reads were eliminated. All RNA-seq reads were filtered by removing reads with sequencing adaptors, unknown nucleotides (N ratio $>$ 10%), and low quality (quality scores [?] 5).

2.4 K-mer analyses of clean Illumina reads

The remaining clean Illumina reads can be used to estimate the genomic characteristics of*C. japonica* before its genome was assembled. In the present study, K-mer-based analysis was used to estimate the size, heterozygosity rate, and repeat sequence of the *C. japonica* genome (Liu et al., 2013). The 17-mer was selected for K-mer analysis to ensure that enough K-mers (4^{17}) were produced to cover the entire *C. japonica* genome.

2.5 *C. japonica* genome assembly and evaluation

The Wtdbg2 software (Ruan and Li, 2020) was applied to assemble the*C. japonica* genome with PacBio long sequencing reads, and the parameters as follows: best depth from input reads, 50.0; Kmer psize, 21; readCutoff, 1k. Although the PacBio long sequencing reads are reliable, they still need to reach a certain sequencing depth to ensure accuracy. First, long PacBio reads were applied to polish the consensus sequence output from the Wtdbg2 software. More specifically, the pbmm2 (Chaisson and Tesler, 2012) and minimap2 software (Li, 2018) were used to align the long PacBio reads to the consensus sequences, and the alignment

results were then corrected using the Arrow and Racon methods (Walker et al., 2014). Furthermore, clean Illumina reads were compared with the abovementioned long PacBio read-based polished genome sequences using the BWA software (version 0.7.10-r789; Li and Durbin, 2009) and then corrected using the Plion software (Walker et al., 2014). Finally, the de-redundancy of the corrected *C. japonica* genome was performed according to the depth distribution and sequence similarity of the reads. The filtered Hi-C reads were mapped to the polished *C. japonica* genome to detect the positional and directional errors in contigs during 3D DNA assembly (Dudchenko et al., 2017). The Juicerbox software (Durand et al., 2016) was used to modify the order and directions of some contigs and to help in the determination of chromosome boundaries. Genomic overlap was identified based on sequence homology and long-distance interaction patterns, and the chromosome-level *C. japonica* genome was obtained.

Three methods were performed to evaluate the assembly effect of the *C. japonica* genome. First, the genome sequence was interrupted with a step length of 1000 bp, and the interrupted sequences were compared with the nucleotide sequence (NT) database by the Basic Local Alignment Search Tool (BLAST) software to evaluate the accuracy of the genome sequences. Second, the BWA (version 0.7.10-r789; Li and Durbin, 2009) and minimap2 software (Li, 2018) were used to compare the short Illumina reads and long PacBio reads with the genome sequence, respectively. The consistency of sequencing reads and genome sequence was evaluated according to comparison rate. Additionally, the completeness of conserved *C. japonica* genes was evaluated using the Benchmarking Universal Single-copy Orthologs (BUSCO, version 2.0; Simao et al., 2015) and based on the orthologous gene database of arthropods. Meanwhile, RNA-seq reads were compared with the genome using the hisat2 software (Vaser et al., 2017) to assess genomic integrity.

2.6 *C. japonica* genome repeat element, coding gene and NcRNA annotation

C. japonica genome annotation was carried out from three perspectives, including repeat recognition, non-coding RNA (ncRNA) and gene structure prediction, and functional annotation.

As an important part of the genome, repeat elements, including tandem repeat and interpersed repeat (also known as transposon element [TE]). In the present study, the Tandem Repeat Finder software (Benson, 1999) was applied to find the tandem repeats in the *C. japonica* genome sequences. Meanwhile, the RepeatMasker and RepeatProteinMask software (http://www.repeatmasker.org) were executed to annotate the interpersed repeats of the *C. japonica* genome sequence based on the Repbase database (Jurka et al., 2005). Furthermore, the RepeatMasker software (Bedell et al., 2000) was used to compare the genome sequence to the repeat element database obtained by the abovementioned methods to obtain a set of repeat elements. The ultimate *C. japonica* genome repeat elements were obtained by removing the redundant repeat elements in the three methods.

Coding gene annotation includes structural prediction and functional annotation. First, three prediction strategies, including homologs, ab initio, and RNA-seq reads, were applied to predict the coding genes. In the present study, *Eriocheir sinensis* , *Penaeus monodon* ,*Penaeus vanamei* , and *P. trituberculatus* were selected as they are closely related to *C. japonica,* and the protein sequences of these species were downloaded for the structural prediction of *C. japonica* coding genes. Ab initio coding gene prediction was performed using Augustus (version 2.7) (Stanke et al., 2006) and the GenScan software (Burge and Karlin, 1997) with default settings. The filtered RNA-seq reads were mapped to the *C. japonica* genome sequences for transcript assembly using the TopHat software, and the Cufflinks software (Ghosh and Chan) was then used to predict the coding genes. The MAKER2 software (Carson and Mark, 2011) was used to remove the redundancy of coding genes predicted by the abovementioned methods, and the HiCESAP process was used to obtain more complete and accurate coding gene datasets. Predicted coding genes were then functionally annotated using InterPro (Zdobnov and Apweiler, 2001), Gene Ontology (GO) (Ashburner et al., 2000), Kyoto Encyclopedia of Genes and Genomes (KEGG)_ALL (Kanehisa and Goto, 2000), KEGG Orthology (KEGG_KO) (Kanehisa and Goto, 2000), Swiss-Prot (Bairoch and Apweiler, 2000), Translation of European Molecular Biology Laboratory nucleotide sequence (TrEMBL) (Boeckmann et al., 2003), TF, Pfam (Griffiths-Jones et al., 2005), NR, and Eukaryotic Orthologous Groups (KOG) (Tatusov et al., 2003) databases to determine the biological function and metabolic pathways involved in the coding gene products.

NcRNAs, such as ribosomal RNA (rRNA), microRNA (miRNA), transfer RNA (tRNA), and small nuclear RNA (snRNA), are RNAs that do not translate proteins but have important biological functions. MiRNA functions in gene silencing and can degrade its target gene or inhibit the translation of the target gene into protein. TRNA and rRNA are directly involved in protein synthesis. SnRNA is involved in the processing of RNA precursors and is the main component of RNA spliceosomes. The tRNAscan-SE software (version 1.3.1) (Lowe and Eddy, 1997) can be used to search for tRNA sequences in the *C. japonica* genome according to the structural characteristics of tRNA. Considering the high conservation of rRNA, the BLASTN software (Altschul et al., 1990) can be used to search for rRNA in the *C. japonica* genome based on the rRNA sequences of closely related species. Additionally, miRNA and snRNA were predicted using the INFERNAL software (version 1.1) (Nawrocki, 2014).

2.7 Whole-genome resequencing analyses

The BWA software (version 0.7.10-r789) (Li and Durbin, 2009) was used to compare high-quality whole-genome resequencing reads with the assembled *C. japonica* genomes, and reads with low mapping efficiency were further removed. The parameters were as follows: mem –M –t –K 10000000. Filtered reads in "sam" format was sorted using the Picard software (https://github.com/broadinstitute/picard) to remove polymerase chain reaction (PCR) duplication. Single-nucleotide polymorphism (SNP) and short insertion/deletion (InDel) were called using a Bayesian approach as implemented in the SAMtools software (Li et al., 2009). Finally, population differentiation index (F_{ST}) and genotype frequency were used to locate the sex-determining regions of *C. japonica* .

2.8 Comparative genomic analyses and testing for genomic selection

We performed an extensive orthologous gene comparison of *C. japonica* with eight other model species with genome datasets, including*Chionoecetes opilio* (GCA_016584305.1), *Hyalella azteca*(GCA_000764305.3), *Penaeus vannamei* (GCA_003789085.1), *P. trituberculatus* (GCA_017591435.1),*Drosophila melanogaster* (GCA_003401745.1),*Amphibalanus amphitrite* (GCA_019059575.1), and *Daphnia magna*(GCA_003990815.1). We downloaded the protein sequences from the National Center for Biotechnology Information database. Subsequently, we extracted the orthologous groups using the ORTHOMCL software (version 2.0.9) (Chen et al., 2006) and filtered the BLASTP results with default parameters. The single-copy orthologous genes shared by all nine species were further aligned using the MUSCLE software (version 3.8.31) (Edgar, 2004), and conserved sequences were extracted from each concatenated nucleotide sequence using the Gblocks software with parameter -t=c. We performed 1000 nonparametric bootstrap replicates for the optimal GTRGAMMA substitution model of all concatenated nucleotide sequences, and then a phylogenetic tree of the nine species was constructed using the RaxML software (version 8) (Stamatakis, 2014). The divergence times of the nine species was estimated using the r8s software, and fossil evidence was used to calibrate the divergence time. Furthermore, the CAFE software (version 3.1) (De Bie et al., 2006) was applied to analyze the expansion and contraction of the gene family, and $P < 0.05$ was used to indicate remarkably changed gene families. Gene enrichment analysis was performed for the extended and contracted gene families based on GO and KEGG databases, respectively.

2.9 Demographic history of *C. japonica*

The demographic history of *C. japonica* was estimated using pairwise sequentially markovian coalescent (PSMC) method (Li and Durbin, 2011) based on a mutation rate of 2e-9 and generation year of 2.

2.10 Relative evolution rate

Two methods, namely, Tajima's relative rate test model in the MEGA7 software (Kumar et al., 2016) and tpcv model in the LINTRE software (Takezaki et al., 1995), were applied to evaluate the relative evolution rate between *C. japonica* and other species. Meanwhile, *D. melanogaster* was used as an outgroup in both analyses.

2.11 Desiccation-adaptive mechanisms of *C. japonica*

Two desiccation-tolerant species (including *C. japonica* and*A. amphitrite*) were selected as foreground species, and then the CAFE software (version 3.1) (De Bie et al., 2006) was applied to analyze the expansion and contraction of the gene family to explore the desiccation adaptive mechanisms of *C. japonica* . Multiple tree files were constructed for the nine species using all single-copy orthologous genes The codeml program in the PAML software (version 4.9) (Yang, 2007) was applied to estimate the non-synonymous/synonymous ratio (w) to determine the positively selected genes (PSGs) of *C. japonica* and *A. amphitrite* . Finally, gene enrichment analysis was performed for the expanded gene families and PSGs based on GO and KEGG databases, respectively.

3. Results

3.1 Evaluation of *C. japonica* genomic characteristics

A total of 164.57 Gb of Illumina data, which correspond to 1,101,954,864 Illumina reads, were generated from the Illumina NovaSeq-6000 platform. The Q20, Q30, and GC content of the Illumina reads were 96.70%, 91.57%, and 42.29%, respectively. After 17-mer was set, 123,219,056,032 K-mers with a depth of 84 were obtained. The*C. japonica* genome size assessed by K-mer analysis was 1,398 Mb. After the negative effect caused by incorrect K-mer was eliminated, the revised*C. japonica* genome size was 1,384 Mb. Meanwhile, the heterozygous ratio and repeat of Illumina reads were 1.09% and 61.24%, respectively.

3.2 Genome assembly and completeness of the assembled genome

The PacBio Sequel II platform generated 219.37 Gb (156.92-fold coverage) long sequencing reads, which correspond to 18,433,125 subreads. Of all the subreads, 16,799,937 subreads were larger than 2 Kb in length. The maximum length, mean length, N50, N90, and GC content of the long PacBio reads were 264,033 bp, 11,900 bp, 17,148 bp, 6,872 bp, and 43.60%, respectively. The long reads were assembled into 27,518 contigs, which correspond to 1,507,921,137 bp. Among all the contigs, 27,517 contigs were larger than 2 Kb in length. The maximum length, N50, N90, and GC content of the assembled contigs were 1,759,181 bp, 107,376 bp, 23,463 bp, and 40.90%, respectively. After the contigs were polished with short Illumina reads and long PacBio reads, 27,223 corrected contigs (1,529,349,659 bp) were obtained, and 27,220 contigs were larger than 2 Kb in length. The maximum length, N50, N90, and GC content of the assembled corrected contigs were 1,758,737 bp, 109,775 bp, 24,031 bp, and 41.20%, respectively. The genome was subsequently processed to remove redundancies and produced a 1429.38 Mb genome with a contig N50, N90, and GC content of 118,954 bp, 30,162 bp, and 41.30%, respectively. Approximately 240.85 Gb clean Hi-C data were generated, and 91.42% of the contigs were anchored to 51 chromosomes to assemble the genome into chromosome level. The final chromosome-level *C. japonica* genome after Hi-C data-assisted assembly was 1431.02 Mb with a contig N50 size of 29.67 Mb (Figure.2).

The short Illumina reads and long PacBio reads were further compared with the assembled*C. japonica* genome sequences to assess the consistency of sequences. A comparison of the results showed that 94.11% and 87.67% of short Illumina reads and long PacBio reads were successfully mapped to the assembled genome, respectively (Table.1). Based on Arthropoda gene set, 86.40% complete BUSCOs were found in the *C. japonica* genome, including 82.27% of the complete and single-copy BUSCOs and 4.13% complete and duplicated BUSCOs (Table.2).

3.3 Repeat element annotation, gene prediction, and gene annotation

Four methods were applied to annotate the repeat elements of *C. japonica* genome and 824.02 Mb repeat elements were identified, which account for 57.65% of the assembly genome. The top 3 repeat elements included DNA transposable element (DNA, 318.62 Mb; 22.29% of the genome), long interspersed nuclear elements (236.31 Mb, 16.53% of the genome), and long terminal repeats (179.30 Mb, 12.54% of the genome, Table.3).

3.4 Gene structure prediction and function annotation

A total of 30,900 coding genes were eventually predicted using the three strategies. The average gene length,

average coding gene length, average exon per gene, average exon length, and average intron length were 11,027, 1,386, 5.12, 341.12, and 2,255 bp, respectively (Table 4). We also functionally annotated all coding genes based on 10 publicly available protein databases. The results showed that 21,979, 16,540, 21,220, 11,310, 15,903, 23,894, 1,458, 20,810, 23,010, and 16,956 coding genes were successfully mapped to the InterPro, GO, KEGG_ALL, KEGG_KO, Swissprot, TrEMBL, TF, Pfam, NR, and KOG databases, respectively. Among all the coding genes, 25,325 were annotated in at least one database, and 5,575 were unannotated.

In the present study, 474 miRNAs, 15,570 tRNAs, 309 rRNAs, and 157 snRNAs were identified from the *C. japonica* genome. The total lengths of the miRNAs, tRNAs, rRNAs, and snRNAs were 57,561 bp (0.004027% of genome), 1,135,923 bp (0.079470% of genome), 49,795 bp (0.003484% of genome), and 28,110 bp (0.001967% of genome), respectively (Table 5).

3.5 Sex-determining region identification of *C. japonica*

The whole-genome resequencing of 20 *C. japonica* individuals from different genders provided a genome-wide scan of SNPs and InDels. Results showed 9 SNP variants (Number: 1,811,614; Table 6) and 12 InDel variants (Number: 1,671,995; Table 7). The allele in chromosome 37 of *C. japonica* had the highest F_{ST} aggregation, which means that chromosome 37 is the sex-determining chromosome of *C. japonica* (Figure 3). Furthermore, we calculated the genotype frequencies using bulk segregant analysis (BSA), and the distribution pattern of Euclidean distance (ED) showed that a difference between 0 and 10,120,000 bp in the chromosome 37 of female and male *C. japonica* (Figure 4).

3.6 Phylogenetic relationship and evolutionary history of *C. japonica*

We compared the *C. japonica* genome with seven other model species, including *A. amphitrite* , which is also a desiccation-tolerant species to analyze the desiccation-adaptive evolution of *C. japonica* . Of all the genes (30,900) of *C. japonica* , 23,684 genes were assigned to 12,396 gene families (Figure.5). Compared with the seven other species, 1,138 gene families were unique to *C. japonica* (Table 8, Figure 6). Using ORTHOMCL software, we identified a set of 968 single-copy orthologous genes shared by eight species (Figure 6). Phylogenetic analysis based on single-copy orthologous genes clustered *C. japonica* and *P. trituberculatus* into one branch, and the differentiation time between *C. japonica* and *P. trituberculatus* was 42.1–135.5 Mya (Figure 7). Compared with seven other species, we identified 2,417 expanded and 2,709 contracted gene families in *C. japonica* (Figure 7). The expanded gene families of *C. japonica* were significantly enriched in 1,375 GO terms and 85 KEGG pathways (corrected $P < 0.05$).

We applied the PSMC method to estimate the demographic history of *C. japonica* . The historical demographic pattern (Figure 8) shows that *C. japonica* maintained a relatively stable population size, followed by an obvious expansion, and reached their maximum effective population size about 0.5×10^5 years ago.

3.7 Relative evolution rate of *C. japonica*

Relative evolution rate can explore the historical status and molecular evolution rate of species. The present study analyzed the relative evolution rates of *C. japonica* and other species. The results showed that *C. japonica* evolved faster than *A. amphitrite* ,*H. azteca* , and *C. opilio* but evolved slower than *D. magna* , *P. vannamei* , and *P. trituberculatus* (Figure 9, Table 9, and Table 10).

3.8 Desiccation-adaptive evolution of *C. japonica*

Changes in gene copy number affect the desiccation-adaptive evolution of *C. japonica.* The present study examined the expansion and contraction of gene families and PSGs in *C. japonica* and *A. amphitrite* genomes to uncover the underlying desiccation-adaptive mechanisms. Compared with other species, the expanded gene families or PSGs were predominantly involved in metabolism rate (i.e., glycogen phosphorylase gene), oxygen supply (i.e., inositol 1,4,5-trisphosphate receptor [*IP3*] gene, ribose-phosphate pyrophosphokinase A [*PRPS1*] gene, serine/threonine-protein phosphatase [*STK*] gene, glutathione reductase [*GSR*] gene), oxidative stress (i.e., heat shock protein [*HSP*] gene, zinc finger [*ZNF*] gene), and various transporters

(annotation information of expanded gene families and PSGs, which might contribute to *C. japonica* 's ability to overcome diverse stresses in drought environment.

4. Discussion

Despite their ecological and economic importance, few crab genomes have been reported. *C. japonica* is of major interest, because it is the most important near-shore economic crab. Moreover, *C. japonica*has long tolerance to desiccation. In the present study, we assembled the first chromosome-level *C. japonica* genome with a genome size of 1431.02 Mb and a contig N50 of 29.67 Mb, which is larger than the genome size of *P. trituberculatus* (1.00 Gb) (Tang et al., 2020) of family Portunidae but smaller than that of *S. paramamosain*(1.55 Gb) (Zhao et al., 2021). The genome size after short Illumina read correction and long PacBio read correction was also confirmed by survey sequence results (1,398 Mb). Notably, the number (n = 51) of chromosomes assembled based on Hi-C reads differs from the karyotype analysis result (n = 57) (Yan et al., 2009). Some *C. japonica* chromosomes have extremely small morphology and are rich in repeat elements, and most chromosomes in the mitosis metaphase are spot-shaped and difficult to distinguish; therefore, identifying all *C. japonica* chromosomes through Hi-C interaction is very difficult (Lou et al., 2004). In fact, we did find some unmanageable repeat sequences through Hi-C interaction. A large number of sequencing reads (short Illumina reads and long PacBio reads) can be successfully mapped to the assembled genome sequence, which means that these reads have high consistency. Meanwhile, 86.40% complete BUSCOs were found in the *C. japonica* genome; thus, the genome that we assembled is more complete. A total of 824.02 Mb of repeat elements were identified and account for 57.65% of the *C. japonica* genome. We speculate that the high proportion of repeat elements is one of the critical reasons for the large genome size of*C. japonica* . In fact, previous association analysis between the number of whole-genome repetitive sequences and genome size in 44 plants and 68 vertebrates confirmed that the proportion of repetitive sequences is positively correlated with genome size (Gao et al., 2018). Additionally, we obtained accurate genome-annotated information through many methods. In conclusion, we presented a basic resource for exploring*C. japonica* biological processes at the genomic level.

4.1 Whole-genome resequencing reads predicted a potentially sex-determining region

Based on the whole-genome resequencing, we predicted a sex-determining region (approximately 10 Mb) in chromosome 37 of *C. japonica* . Interestingly, we found some genes in this region that seem to be involved in sex differentiation in *C. japonica* , namely, the*F-box* gene and WD repeat domain (*WDR*) gene. A previous study considered that the F-box protein encoded by recombinant *FBXW* gene (containing*F-box* gene and *WDR* gene) is specific for substrate recognition in ubiquitin-mediated proteolysis (Risseeuw et al., 2003). It can bind S-phase kinase-associated protein 1 (SKP1) and Cullin to form an evolutionary conserved Skp–Cullin–F-box (SCF) ubiquitin ligase complex (Figure 10) (Zhang et al., 2002; Welcker and Clurman, 2008; Roberts et al., 2020), which can selectively bind to degraded proteins and participate in the ubiquitin–proteasome pathway (UPP) (Li et al., 2010). The UPP system in plants causes functional loss of male germ cells by regulating the conversion of functional proteins (Kipreos and Pagano, 2000). The sex-determining mechanism of crustaceans living in the open environment is still in the early stage of evolution, and the intensity of genetic factors in determining the gender of crustaceans is lower than that of the external environment (Lou et al., 2004; Wu et al., 2002). Therefore, we have every reason to infer that the UPP system regulated by recombinant *FBXW* gene may potentially limit the testis development of male *C. japonica* , reduce sperm production, and even promote vitellus formation. However, this mechanism is still speculative and needs further verification.

4.2 More accurate evolutionary history of *C. japonica*

The presented genomic resource can help us better understand the evolutionary history of *C. japonica* . The differentiation of conserved single-copy orthologous genes always leads to species differentiation; thus, we believe that the evolutionary history of*C. japonica* can be more accurately deetermined using widespread single-copy orthologous genes (Liang et al., 2013; Fitch, 1970). In the present study, 968 single-copy orthologous genes were obtained and then applied to construct the phylogenetic relationship of *C. japonica* .

Expectedly, the constructed phylogenetic structure is considerably congruent with the prevailing morphological and molecular biological views that crabs (*P. trituberculatus* and *C. japonica*) from family Portunidae eventually cluster into one branch. Moreover, the evolutionary sequence of *C. japonica* was identified based on differentiation times, and these results provided evidence to prove the existence of "carcinization" (Keiler et al., 2017; Cui et al., 2021). In fact, 1,138 unique gene families were found in the genome, and these gene families can contribute to *C. japonica* -specific adaptation (Wissler et al., 2013). Additionally, the relatively evolution rates of different species may also mean differences in environmental stability (Tang et al., 2020).

4.3 Desiccation-adaptive mechanisms of *C. japonica*

Comparative genomics can provide insights into the unique adaptive plasticity of marine species (Han et al., 2021; Xu et al., 2021). In the present study, we hypothesized the adaptive mechanisms of desiccation tolerance based on comparative genomics. Previous studies suggested that crustaceans may continuously undergo the double stress of dehydration and hypoxia in a drought environment. Specifically, desiccation disrupts the internal osmotic pressure and metabolic capacity of crustaceans (Haupt et al., 2006; Carlson and Rowe, 2009). Moreover, desiccation can reduce the oxygen-binding ability of hemoglobin and result in hypoxic stress in crustaceans (Omori et al., 1998; Allen et al., 2012). The prolonged disruption of the respiratory mechanism may further impair the immune function of crustaceans and eventually lead to death (Madenjian et al., 1987). Additionally, the activity of the antioxidant enzyme system directly determines the desiccation tolerance of crustaceans (Jiang et al., 2014). Based on the annotation information, the desiccation-adaptive mechanisms of*C. japonica* was supported by the marked expansion of gene families or the positive selection of genes related to metabolism rate, oxygen supply, oxidative stress, and various transporters.

4.3.1 Glycometabolism provides carbon skeleton and energy for *C. japonica* in drought environment

Carbohydrates can provide carbon skeleton and energy for the growth and metabolism of organisms and enhance stress resistance. Glycometabolism is also the center of biological metabolism, which connects the metabolism of proteins, lipids, nucleic acids, and secondary substances (Sun et al., 2015). Based on the comparative genome, abundant glycometabolism-related genes were expanded or positively selected in *C. japonica* and further contributed to the desiccation tolerance of *C. japonica* . The reasons for the improvement in the desiccation tolerance of *C. japonica* by carbohydrate may be as follows: (i) carbohydrates can accumulate and regulate osmotic pressure and enhance the water retention of cells (Savitch et al., 2000; Wilson et al., 2001); (ii) carbohydrates can protect biological substances, such as biofilms and biological macromolecules, to maintain cytoskeletal integrity in drought environments (Hincha et al., 2002; Saadeldin et al., 2020); (iii) carbohydrates can produce energy for resistance to desiccation (Savitch et al., 2000); and (iv) carbohydrates can also be used as a signal molecule involved in some desiccation resistance regulation mechanisms (Paul et al., 2008).

4.3.2 High oxygen transport ability improves the hypoxia tolerance of *C. japonica*

Although hemocyanin has a low affinity for oxygen, it is still an essential oxidizing medium for crustacean tissues in the absence of oxygen (Figure 11) (Charlotte, 1980; Fujieda et al., 2010). Desiccation can inhibit oxygen supply and induce respiratory acidosis in crustaceans (Defur, 1988). Respiratory acidosis further reduces the ability of oxygen and hemoglobin to bind, which exacerbates oxygen deficiency and tissue hypoxia (Allen et al., 2012). However, desiccation-tolerant crustaceans can regulate ion (including Ca^{2+} and Mg^{2+}) concentration to improve the oxygen-binding ability of hemocyanin and promote oxygen transfer between tissues (Charlotte, 1980). The expansion or positive selection of the *IP3*and *PRPS1* genes means that *C. japonica* can maintain Ca^{2+} and Mg^{2+} homeostasis to ensure the oxygen-binding ability of hemocyanin to resist hypoxia in drought environments. Hypoxia-inducible factors (HIFs) also play an important role in hypoxia response by binding with hypoxia response elements near the downstream gene promoter for transcriptional activation (Hirota, 2002; Martens et al., 2007). The *STK* gene in *C. japonica*can enhance HIF phosphorylation and increase the molecular weight of HIFs (Richard et al., 1999). In addition, the *GSR* gene, which prevents the oxidative decomposition of hemocyanin, was also positively selected or expanded.

Glutathione reductase encoded by the *GSR*gene can effectively maintain reduced glutathione in cells, which plays an important role in preventing the oxidative decomposition of hemocyanin (Kanzok et al., 2001). In conclusion, the maintenance of hemocyanin level and oxygen-carrying capacity will help *C. japonica* cope with hypoxia in drought environment.

4.3.3 *C. japonica* seem to have a greater capacity for oxidative stress

We detected the remarkable expansion or positive selection of *HSP*genes (including *HSP70* and *HSP90*), which probably encode prototypical chaperonin to prevent the accumulation of unfolding proteins and/or assist in the refolding or degradation of denatured proteins (Haslbeck et al., 2005; Beck et al., 2009). The *ZNF* gene family was also enriched in the *C. japonica* genome. Zinc finger proteins are the most abundant protein in eukaryotes and have a wide range of structures and functions (Laity et al., 2011). Interestingly, the positive selection or expansion of RING-type zinc finger proteins occurred in *C. japonica* . As a zinc finger protein widely found in plants and animals, RING-type zinc finger protein confers a strong drought tolerance in *Arabidopsis thaliana* (Moon et al., 2010), because RING-type zinc finger protein can regulate the ubiquitin–proteasome system to degrade misfolded proteins and thus improve the defense of an organism against external environmental stress (Ryota et al., 2002).

In conclusion, *C. japonica* 's metabolism, oxygen supply, and antioxidative stress abilities may help this species to survive longer in drought environments. Due to the high desiccation-tolerance capability of *C. japonica* , this information will be valuable in the transportation and farming strategies.

5. Conclusion

A chromosomal-level genome of *C. japonica* has been published for the first time by integrating the short Illumina reads, long PacBio reads, and Hi-C reads. This is a highly assembled and reliable genome and it enriches the crustacean genome database. Meanwhile, we successfully annotated the repeat elements, coding genes, miRNAs, tRNA, rRNAs, and snRNAs in the *C. japonica* genome. These information will be necessary for future research into *C. japonica* biology. Additionally, the whole-genome resequencing data of *C. japonica*with different gender provided a genome-wide scan of SNPs and InDels, which were used to predict the sex-determining chromosome of *C. japonica* . *C. japonica* is a crustacean with strong desiccation tolerance. This study used comparative genome analysis to study the evolutionary history of *C. japonica* , as well as a series of genes that may contribute to desiccation adaption in *C. japonica* . The present results were beneficial in elucidating the evolution process in*C. japonica* and other marine organisms under desiccation stress.

Acknowledgements

This work was supported by the Zhejiang Provincial Natural Science Foundation of China (LR21D060003) and the National Natural Science Foundation of China (32070513).

Author contributions

Z.Q.H. and F.R.L. conceived and managed the project. Z.Q.H. and T.X.G. collected the sequencing samples. Q.L. extracted the DNA/RNA and performed the genome sequencing. F.R.L. analyzed the data. Z.Q.H. and F.R.L. wrote the manuscript. All authors reviewed and approved the final manuscript.

Conflicts of interests

No

Data Accessibility Statement

Raw sequencing data (including PacBio, Hi-C, Survey, and RNA-seq reads) for*C. japonica* genome has been deposited at the Sequence Read Archive (SRA) (SRR16072374, SRR16072376, SRR16072375 and SRR16078881) under BioProject number PRJNA766329. Raw whole-genome resequencing data of 20 *C. japonica* has been deposited at the SRR16894624 to SRR16894643 under BioProject number PRJNA779101.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (2008) Basic Local Alignment Search Tool (BLAST). *J Mol Biol* , **215** , 403-410. DOI: 10.1016/S0022-2836(05)80360-2.

Allen BJ, Rodgers B, Tuan YH, Levinton JS (2012) Size-dependent temperature and desiccation constraints on performance capacity: implications for sexual selection in a fiddler crab. *J Exp Mar Biol Ecol* , **438** , 93-99. DOI: 10.1016/j.jembe.2012.09.009.

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* ,**268** , 78-94. DOI: 10.1111/j.1365-294X.2005.02509.x.

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* , **27** , 573-580. DOI: 10.1093/nar/27.2.573.

Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.*Nucl Acids Res* , **28** , 45-48. DOI: 10.1093/nar/28.1.45.

Bedell JA, Korf I, Gish W (2000) MaskerAid: A performance enhancement to RepeatMasker. *Bioinformatics* , **16** , 1040-1041. DOI: 10.1093/bioin forma tics/16.11.1040.

Boeckmann B, Bairoch A, Apweiler R, Blatter M, Estreicher A, Gasteiger E, Martin M, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003.*Nucleic Acids Res* , **31** , 365-370. DOI: 10.1093/nar/gkg095.

Beck M, Malmstrom JA, Lange V, Schmidt A, Deutsch EW, Aebersold R (2009) Visual proteomics of the human pathogen *Leptospira interrogans* .*Nat Methods* , **6** , 817-823. DOI: 10.1038/nmeth.1390.

Charlotte PM (1980) Respiration function of the hemocyanins. *Am Zool* , **20** , 19-38.

Chen F, Mackey AJ, Jr CJS, Roos DS (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res* , **34** , D363-D368. DOI: 10.1093/nar/gkj123.

Carlson BE, Rowe MP (2009) Temperature and desiccation effects on the antipredator behavior of*Centruroides vittatus* (Scorpiones: Buthidae).*J Arachnol* , **37** , 321-330. DOI: 10.1636/Hi09-06.1.

Chaisson MJ, Tesler G (2012) Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* , **13** , 238. DOI: 10.1186/1471-2105-13-238.

Cantarel BL, Korf L, Robb SMC, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Carson H, Mark Y (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* ,**12** , 491. DOI: 10.1186/1471-2105-12-491.

Cui ZX, Liu Y, Yuan JB, Zhang XJ, Ventura T, Ma KY, Sun S, Song CW, Zhan DL, Yang YN, Fan GY, Cai QL, Du J, Shi CC, Hao SJ, Fitzgibbon QP, Smith GG, Xiang JH, Chan TY, Hui M, Bao CC, Li FH, Chu KH (2021) The Chinese mitten crab genome provides insights into adaptive plasticity and developmental regulation. *Nat Commun* , **12** , 2395. DOI: 10.1038/s41467-021-22604-3.

Defur PL (1988) Systemic respiratory adaptations to air exposure in intertidal decapod crustaceans. *Am Zool* , 28, 115-124. DOI: 10.1093/icb/28.1.115.

De Bie T, Cristianini N, Demuth JP, Hahn MW (2006) CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* ,**22** , 1269-1271. DOI: 10.1093/bioinformatics/btl097.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Adien EL (2016) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst* , **3** , 99-101. DOI: 10.1016/j.cels.2015.07.012.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL (2017) *De novo*assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* , **356** , 92-95. DOI: 10.1126/science.aal3327.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* , **32** , 1792-1797. DOI: 10.1093/nar/gkh340.

Fitch WM (1970) Distinguishing homologous from analogous proteins. *Syst Zool* **19** , 99-113. DOI: 10.2307/2412448.

Fujieda N, Yakiyama A, Itoh S (2010) Catalytic oxygenation of phenols by arthropod hemocyanin, an oxygen carrier protein, from *Portunus trituberculatus* . *Dalton Trans* , **39** , 3083-3092. DOI: 10.1039/c000760a.

Greenaway P, Morris S, McMahon BR, Gallagher CAFL (1996) Air breathing by the purple shore crab, *Hemigrapsus nudus* (Dana). I. Morphology, behaviour, and respiratory gas exchange. *Physiol Zool* , **69** , 785-805. DOI: 10.2307/30164230.

Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Bateman A (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* , **33** , D121-D124. DOI: 10.1093/nar/gki081.

Ghosh S, Chan CKK (2016) Analysis of RNA-Seq data using TopHat and Cufflinks. *Methods Mol Biol* , **1374** , 339-361. DOI: 10.1007/978-1-4939-3167-5_18.

Gao SH, Yu HY, Wu SY, Wang S, Geng JN, Luo YF, Hu SN (2018) Advances of sequencing and assembling technologies for complex genomes.*Hereditas* , **40** , 944-963. DOI: 10.16288/j.yczz.18-255.

Hirota K (2002) Hypoxia-inducible factor 1, a master transcription factor of cellular hypoxic gene expression. *J Anesth* , **16** , 150-159. DOI: 10.1007/s005400200011.

Haslbeck M, Franzmann T, Weinfurtner D, Buchner J (2005) Some like it hot: the structure and function of small heat-shock proteins. *Nat Struct Mol Biol* , **12** , 842-846. DOI: 10.1038/nsmb993.

Haupt P, Brouwer SL, Branch GM, Gäde G (2006) Effects of exposure to air on the escape behaviour and haemolymph chemistry of the South African Cape lobster, *Jasus lalandii* . *Fish Res* ,**81** , 210-218. DOI: 10.1016/j.fishres.2006.07.004.

Han ZQ, Guo XY, Liu Q, Liu SS, Zhang ZX, Xiao SJ, Gao TX (2021) Whole-genome resequencing of Japanese whiting (*Sillago japonica*) provide insights into local adaptations. *Zool Res* , 42, 548-561. DOI: 10.24272/j.issn.2095-8137.2021.116.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* , **110** , 462-467. DOI: 10.1159/000084979。

Jiang N, Wang F, Lu YL, Dong SL (2014) Effect of air-exposure stress on antioxidant capacity and stress response of swimming crab (*Portunus trituberculatus*). *Journal of Fisheries of China* , **38** , 1996-2004. (In Chinese).

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* ,**28** , 27-30. DOI: 10.1093/nar/28.1.27.

Kipreos ET, Pagano M (2000) The F-box protein family. *Genome Biol* , **1** , 30021-30027. DOI: 10.1186/gb-2000-1-5-reviews3002.

Kanzok SM, Fechner A, Bauer H, Ulschmid JK, Müller HM, Botella-Munoz J, Schneuwly S, Schirmer R, Becker K (2001) Substitution of the thioredoxin system for glutathione reductase in *Drosophila melanogaster* .*Science* , **291** , 643-646. DOI: 10.1126/science.291.5504.643.

Kolpakov N, Kolpakov E (2011) On the biology of the Japanese swimming crab *Charybdis japonica* (Portunidae) in waters of primorye at the northern boundary of their range. *Russ J Mar Biol* , **37** , 570-578. DOI: 10.1134/S1063074011070042.

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* , **33** , 1870. DOI: 10.1093/molbe v/msw054.

Keiler J, Wirkner CS, Richter S (2017) One hundred years of carcinization - the evolution of the crab-like habitus in Anomura (Arthropoda: Crustacea). *Biological Journal of the Linnean Society* , **121** , 200-222.

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* , **25** , 955-964. DOI: 10.1093/nar/25.5.955.

Levin LA (2003) Oxygen minimum zone benthos: adaptation and community response to hypoxia. Oceanogr Mar Biol, 41, 1-45. DOI: 10.1007/s00248-002-2023-y.

Lou YD, Liu YH, Qiu GF (2004) Advances in sex determination of shrimps (prwans) and crabs. *J Shanghai Ocean Uni* , 2, 157-163. (In Chinese)

Lorenzon S, Giulianini PG, Libralato S, Martinis M, Ferrero EA (2008) Stress effect of two different transport systems on the physiological profiles of the crab *Cancer pagurus* .*Aquaculture* ,**278** , 156-163. DOI: 10.1016/j.aquaculture.2008.03.011.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform.*Bioinformatics* , **25** , 1754-1760. DOI: 10.1093/bioinformatics/btp324.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* ,**25** , 2078-2079. DOI: 10.1093/bioinformatics/btp352.

Li L, Xia K, Fu YF, Tian Y (2010) Plant F-box protein and its biological function. *Agricultural science and technology* , **11** , 9-12. (In Chinese)

Li J, Chen P, Liu P, Gao B, Wang Q, Jian L (2010) The cytosolic manganese superoxide dismutase cDNA in swimming crab*Portunus trituberculatus* : Molecular cloning, characterization and expression. *Aquaculture* , **309** , 31-37. DOI: 10.1016/j.aquaculture.2010.09.008.

Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences.*Nature* , **475** , 493-496. DOI: 10.1038/nature10231.

Laity JH, Lee BM, Wright PE (2011) Zinc finger proteins: New insights into structural and functional diversity.*Curr Opin Struct Biol* , **11** , 39-46. DOI: 10.1016/S0959-440X(00)00167-6.

Liu BH, Shi YJ, Yuan JY, Hu XS, Zhang H, Li N, Li ZY, Chen YX, Mu DS, Fan W (2013) Estimation of genomic characteristics by analyzing k-mer frequency in *de novo* genome projects. *ArXiv* , 1308.2012. DOI: 10.1016/S0925-4005(96)02015-1.

Liang D, Shen X, Zhang P (2013) One thousand two hundred ninety nuclear genes from a genome-wide survey support lungfishes as the sister group of tetrapods. *Mol Biol Evol* , **30** , 1803-1807. DOI: 10.1093/molbev/mst072.

Lu Y, Zhang D, Wang F, Dong S (2016) Hypothermal effects on survival, energy homeostasis and expression of energy-related genes of swimming crabs *Portunus trituberculatus* during air exposure.*J Therm Biol* , **60** , 33-40. DOI: 10.1016/j.jtherbio.2016.06.007.

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences.*Bioinformatics* , **34** , 3094-3100. DOI: 10.1093/bioinformatics/bty191.

Madenjian CM, Rogers GL, Fast AW (1987) Predicting night time dissolved oxygen loss in prawn ponds of Hawaii: Part I. Evaluation of traditional methods.*Aquacult Eng* , **6** , 191-208. DOI: 10.1016/0144-8609(87)90004-5.

Morris S, Oliver S (1999) Circulatory, respiratory and metabolic response to emersion and low temperature of *Jasus edwardsii* : simulation studies of commercial shipping methods. *Comp Biochem Phys A* , **122** , 299-308. DOI: 10.2514/2.983.

Martens LK, Kirschner KM, Warnecke C, Scholz H (2007) Hypoxia-inducible factor-1 (HIF-1) is a transcriptional activator of the TrkB neurotrophin receptor gene. *J Biol Chem* , **282** , 14379-14388. DOI: 10.1074/jbc.M609857200.

Moon YR, Seok KC, Woo TK (2010) The Arabidopsis C3H2C3-type RING E3 ubiquitin ligase AtAIRP1 is a positive regulator of an abscisic acid-dependent response to drought stress. *Plant Physiol* , **154** , 1983-1997. DOI: 10.1104/pp.110.164749.

Markl J (2013) Evolution of molluscan hemocyanin structures. *BBA-Proteins Proteom* ,**1834** , 1840-1852. DOI: 10.1016/j.bbapap.2013.02.020.

Nawrocki EP (2014) Annotating functional RNAs in genomes using infernal. *Methods Mol Biol* ,**1097** , 163-197. DOI: 10.1007/978-1-62703-709-9_9.

Pascual C, Gaxiola G, Rosas C (2003) Blood metabolites and hemocyanin of the white shrimp, *Litopenaeus vannamei* : the effect of culture conditions and a comparison with other crustacean species. *Mar Biol* , **142** , 735-745. DOI: 10.1007/s00227-002-0995-2.

Paul MJ, Primavesi LF, Jhurreea D, Zhang Y (2008) Trehalose metabolism and signaling. *Annu Rev Plant Biol* , **59** , 417-441. DOI: 10.1146/annurev.arplant.59.032607.092945.

Paital B (2013) Antioxidant and oxidative stress parameters in brain of *Heteropneustes fossilis* under air exposure condition; role of mitochondrial electron transport chain. *Ecotoxicol Environ Saf* , **95** , 69-77. DOI: 10.1016/j.ecoenv.2013.05.016.

Richard DE, Berra E, Gothié E, Roux D, Pouysségur J (1999) p42/p44 mitogen-activated protein kinases phosphorylate hypoxia-inducible factor 1α (HIF-1α) and enhance the transcriptional activity of HIF-1. *J Biol Chem* , **274** , 32631-32637. DOI: 10.1074/jbc.274.46.32631.

Ryota T, Noriyuki M, Akihiko N, Koji H, Chiharu A, Naoto S, Eiichi M (2002) EL5, a rice N-acetylchitooligosaccharide elicitor-responsive RING-H2 finger protein, is a ubiquitin ligase which functions in vitro in co-operation with an elicitor-responsive ubiquitin-conjugating enzyme, OsuBC5b. *Plant J* , **30** , 447-455. DOI: 10.1046/j.1365-313x.2002.01299.x.

Risseeuw EP, Daskalchuk TE, Banks TW, Liu E, Cotelesage J, Hellmann H, Estelle M, Somers DE, Crosby WL (2003) Protein interaction analysis of SCF ubiquitin E3 ligase subunits from Arabidopsis. *Plant J* ,**34** , 753-767. DOI: 10.1046/j.1365-313x.2003.01768.x.

Ridgway ID, Taylor AC, Atkinson R, Stentiford GD, Chang ES, Chang SA, Neil DM (2006) Morbidity and mortality in Norway lobsters, *Nephrops norvegicus* : physiological, immunological and pathological effects of aerial exposure. *J Exp Mar Biol Ecol* , **328** , 251-264. DOI: 10.1016/j.jembe.2005.07.015.

Ruan J, Li H (2020) Fast and accurate long-read assembly with Wtdbg2. *Nat Methods* ,**17** , 155-158. DOI: 10.1038/s41592-019-0669-3.

Roberts JZ, Holohan C, Sessler T, Fox J, Crawford N, Riley JS, Khawaja H, Majkut J, Evergren E, Humphreys LM, Ferris J, Higgins C, Espona-Fiedler M, Moynagh P, McDade SS, Longley DB (2020) The SCF^{Skp2} ubiquitin ligase complex modulates TRAIL-R2-induced apoptosis by regulating FLIP(L). *Cell Death Differ* , **27** , 2726-2741. DOI: 10.1038/s41418-020-0539-7.

Savitch LV, Harney T, Huner NPA (2000) Sucrose metabolism in spring and winter wheat in response to high irradiance, cold stress and cold acclimation. *Physiol Plantarum* ,**108** , 270-278. DOI: 10.1034/j.1399-3054.2000.108003270.x.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B (2006) AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* , **34** , W435-439. DOI: 10.1093/nar/gkl200.

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

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. DOI: 10.1093/bioinformatics/btv351.

Sun YM, Liu LJ, Feng MF, Wang JH, Cang J, Li S, Bao YZ, Wang XT (2015) Research progress of sugar metabolism of plants under cold stress. *Journal of Northeast Agricultural University* , **46** , 95-102. (In Chinese)

Saadeldin IM, Swelum AA, Elsafadi M, Mahmood A, Osama A, Shikshaky H, Alfayez M, Alowaimer AN, Magdeldin S (2020) Thermotolerance and plasticity of camel somatic cells exposed to acute and chronic heat stress. *J Adv Res* , **22** , 105-118. DOI: 10.1016/j.jare.2019.11.009.

Taylor HH, Greenaway P (1984) The role of the gills and branchiostegites in gas exchange in a bimodally breathing crab, *Holthuisana transversa* : evidence for a facultative change in the distribution of the respiratory circulation.*J Exp Biol* , **111** , 103-121.

Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetic test of the molecular clock and linearized trees. *Mol Biol Evol* ,**12** , 823-833. DOI: 10.1093/oxfordjournals.molbev.a040259.

Taylor EW, Wheatley MG (1980) Ventilation, heart rate and respiratory gas exchange in the crayfish*Austropotamobius pallipes* (Lereboullet) submerged in normoxic water and following 3 h exposure in air at 15.*J Comp Physiol B* , **138** , 67-78. DOI: 10.1007/bf00688737.

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf AI, Yin JJ, Natale DA (2003) The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* , **4** , 41. DOI: 10.1186/1471-2105-4-41.

Tang BP, Zhang DZ, Li HR, Jiang SH, Zhang HB, Xuan FJ, Ge BM, Wang ZF, Liu Y, Sha ZL, Cheng YX, Jiang W, Jiang H, Wang ZK, Wang K, Li CF, Sun Y, She SS, Qiu Q, Wang W, Li XZ, Li YX, Liu QN, Ren YD (2020) Chromosome-level genome assembly reveals the unique genome evolution of the swimming crab (*Portunus trituberculatus*). *GigaScience* ,**9** , giz161. DOI: 10.1093/gigascience/giz161.

Urbina MA, Paschke K, Gebauer P, Cumillaf JP, Rosas C (2013) Physiological responses of the southern king crab, *Lithodes santolla* (Decapoda: Lithodidae), to aerial exposure. *Comp Biochem Phys A* ,**166** , 538-545. DOI: 10.1016/j.cbpa.2013.08.006.

Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly from long uncorrected reads.*Genome Res* , **27** , 737-746. DOI: 10.1101/gr.214270.116.

Wilson RG, Kachman SD, Martin AR (2001) Seasonal changes in glucose, fructose, sucrose, and fructans in the roots of dandelion. *Weed Sci* , **49** , 150-155. DOI: 10.1614/0043-1745(2001)049[0150:SCIGFS]2.0.CO;2.

Wu CG, Xiang JH (2002) Genetic determination and exogenous influence in sex differentiation in crustacean. *Developmental and Reproductive Biology* , **11** , 88-94.

Welcker M, Clurman BE (2008) FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation. *Nat Rev Cancer* , **8** , 83-93. DOI: 10.1038/nrc2290.

Wissler L, Gadau J, Simola DF, Helmkampf M, Bornberg-Bauer E (2013) Mechanisms and dynamics of orphan gene emergence in insect genomes. *Genome Biol Evol* , **5** , 439-455. DOI: 10.1093/gbe/evt009.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *Plos One* , **9** , e112963. DOI: 10.1371/journal.pone.0112963.

Xu SY, Lu ZC, Cao BB, Han DD, Cai SS, Han ZQ, Gao TX (2021) Chromosome-scale genome assembly of brownspotted flathead*Platycephalus* sp.1 provides insights into demersal adaptation in flathead fish. *Zool Res* , **42** , 660-665. DOI: 10.24272/j.issn.2095-8137.2021.236.

Yu CG, Song HT, Yao GZ (2004) Assessment of the crab stock biomass in the continental shelf waters of the East China Sea. *Journal of Fisheries of China* , 28, 41-46. (In Chinese)

Yu CG, Song HT, Yao GZ (2005) The quantity distribution and biological property of *Charybdis japonica* in the East China Sea. *J Shanghai Ocean Uni* , **14** , 40-45. (In Chinese)

Yang ZH (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol***24** , 1586-1591. DOI: 10.1093/molbev/msm088.

Yan BL, Xu XH, Xu JT, Xu GC (2009) Karyotype analysis on *Charybdis japonica* . *Marine Science* ,**33** , 41-45. DOI: 10.1360/972009-495. (In Chinese)

Zdobnov EM, Apweiler R (2001) InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* ,**17** , 847-848. DOI: 10.1093/bioinformatics/17.9.847.

Zheng N, Schulman BA, Song LZ, Miller JJ, Jeffrey PD, Wang P, Chu C, Koepp DM, Elledge SJ, Pagano M, Conaway RC, Conaway JW, Harper JW, Pavletich NP (2002) Structure of the Cul1-Rbx1-Skp1-F box^{Skp2} SCF ubiquitin ligase complex. *Nature* , **416** , 703-709. DOI: 10.1038/416703a.

Zheng W, Jia CH, Han ZQ (2013) Morphological variation among wild populations of *Charybdis japonica* from coastal waters of China.*Journal of Zhejiang Ocean University* , 32, 477-483. (In Chinese)

Zheng W (2015) Molecular phylogeography study of *Charybdis japonica* and *Charybdis bimaculata* in East China Sea and Yellow Sea. *Zhejiang Ocean University* , pp2. (In Chinese).

Zhao M, Wang W, Zhang FY, Ma CY, Liu ZQ, Yang MH, Chen W, Li QS, Cui MS, Jiang KJ, Feng CL, Li JT, Ma LB (2021) A chromosome-level genome of the mud crab (*Scylla paramamosain* estampador) provides insights into the evolution of chemical and light perception in this crustacean. *Mol Ecol Resour* , **21** , 1299-1317. DOI: 10.1111/1755-0998.13332.

Figure legend

Figure.1. *Charybdis japonica* .

Figure.2. Genome-wide Hi-C heatmap of *C. japonica* . From outer circle to inner circle: chromosome (a), gene distribution (b), GC content of the genome (c), short Illumina read sequencing depth (d), long PacBio read depth (e), DNA transposable element (f), and ;ong terminal repeat (g).

Figure 3. Manhattan plot of F_{ST}.

Figure 4. ED^2 distribution.

Figure 5. Venn diagram of shared and unique gene families in four (A) or five (B) species.

Figure 6. Number of homologous genes among the genomes of related species.

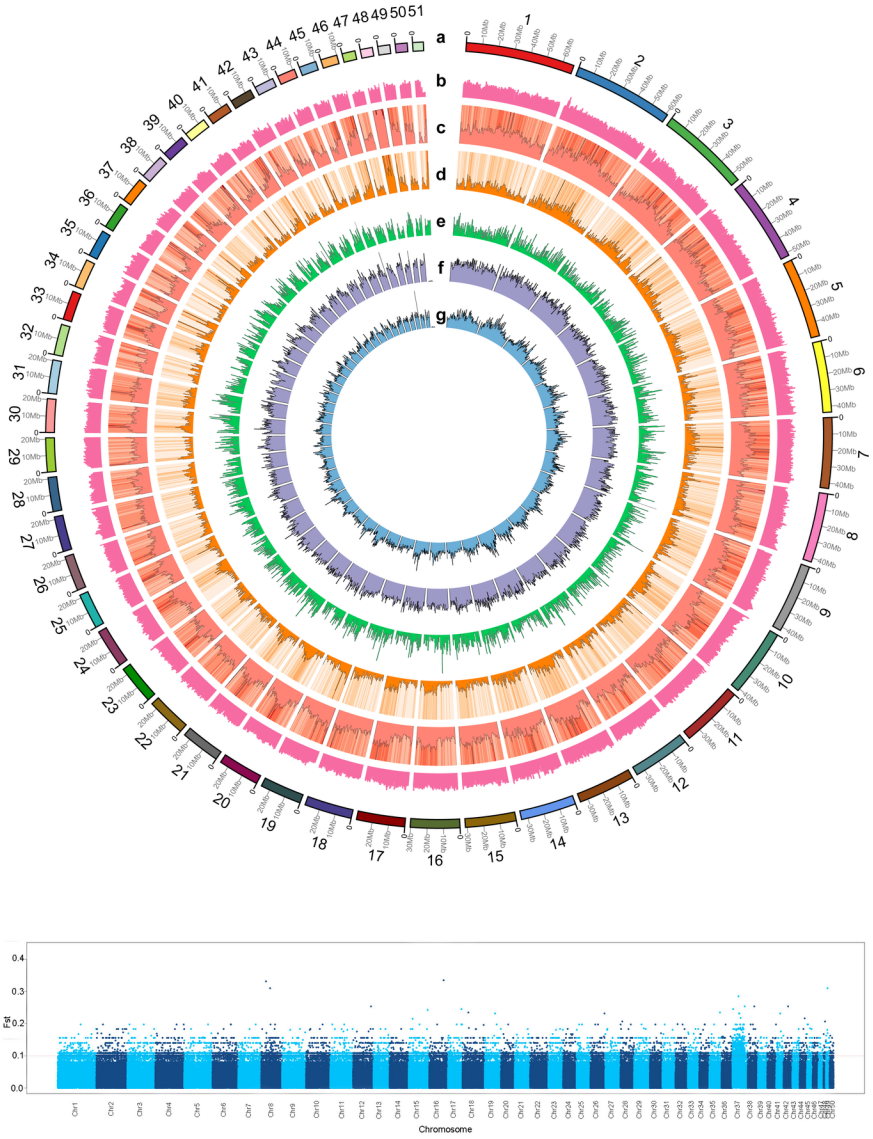
Figure 7. Phylogenetic analysis based on 968 single-copy orthologous genes.

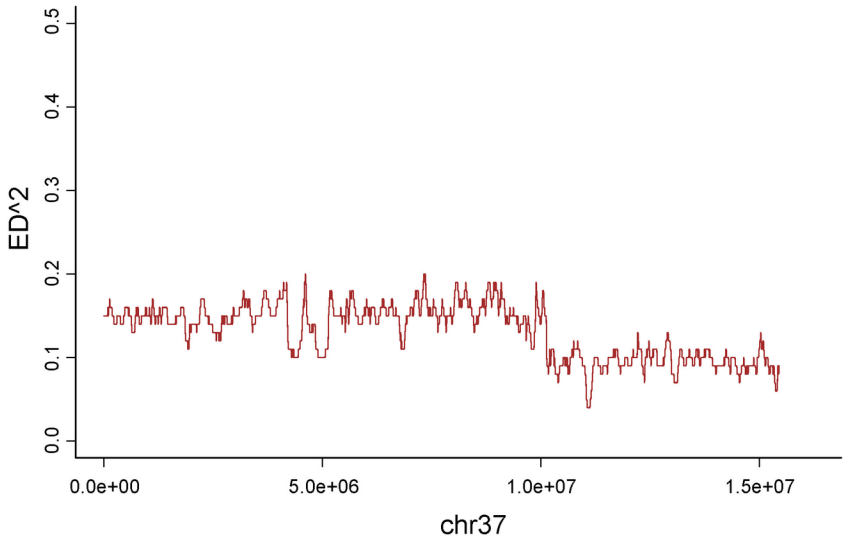
Figure 8. Demographic history of *C. japonica* .

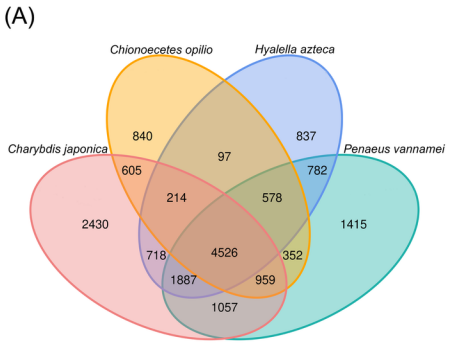
Figure 9. Relative evolution rate analysis of different species.

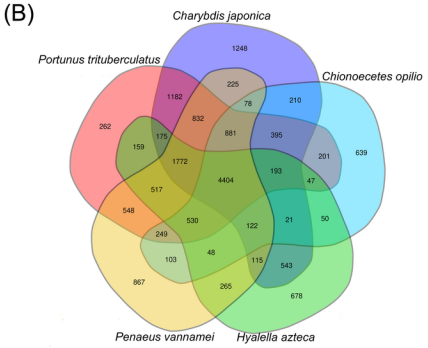
Figure 10. Overall structure (left) and schematic (right) of an SCF ubiquitin ligase. Note: Ub: Ubiquitin; UBA: Ubiquitin activating enzyme; UBC: Ubiquitin conjugating enzyme; RBX1: E3 ubiquitin-protein ligase.

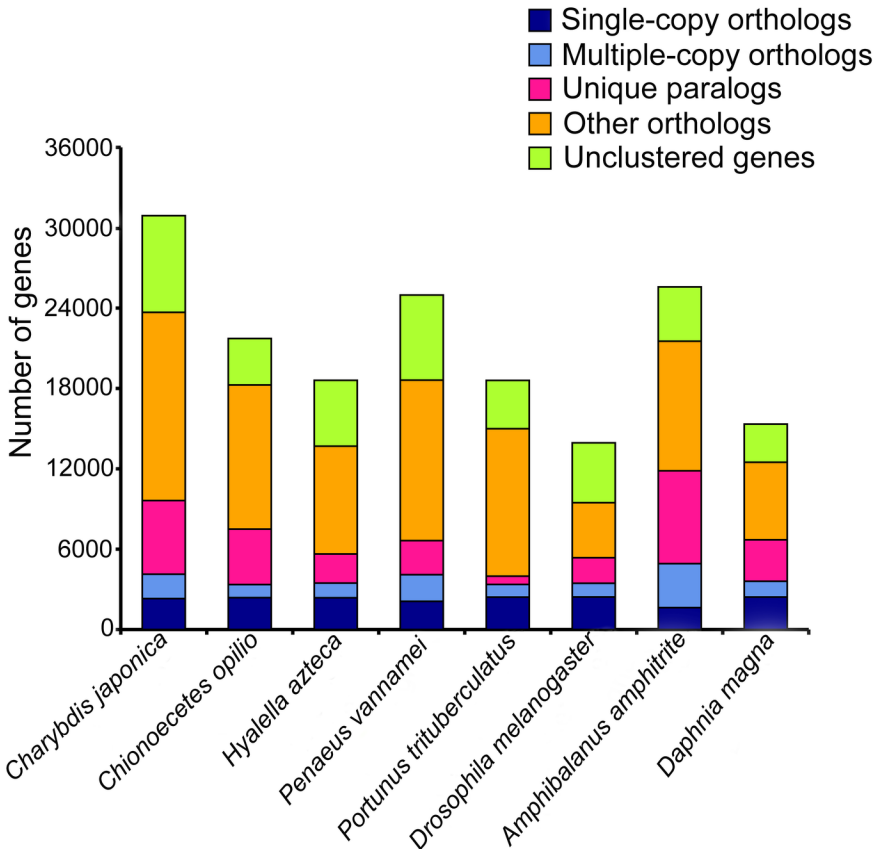
Figure 11. Schematic of hemocyanin binding to oxygen (Markl, 2013).

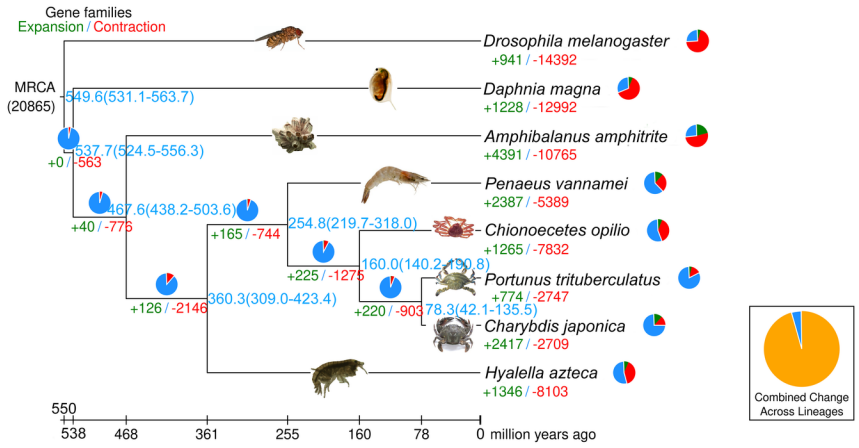


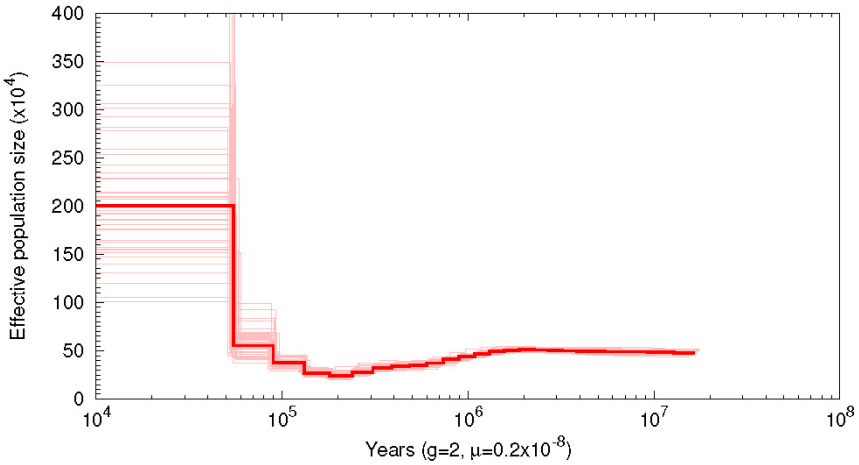


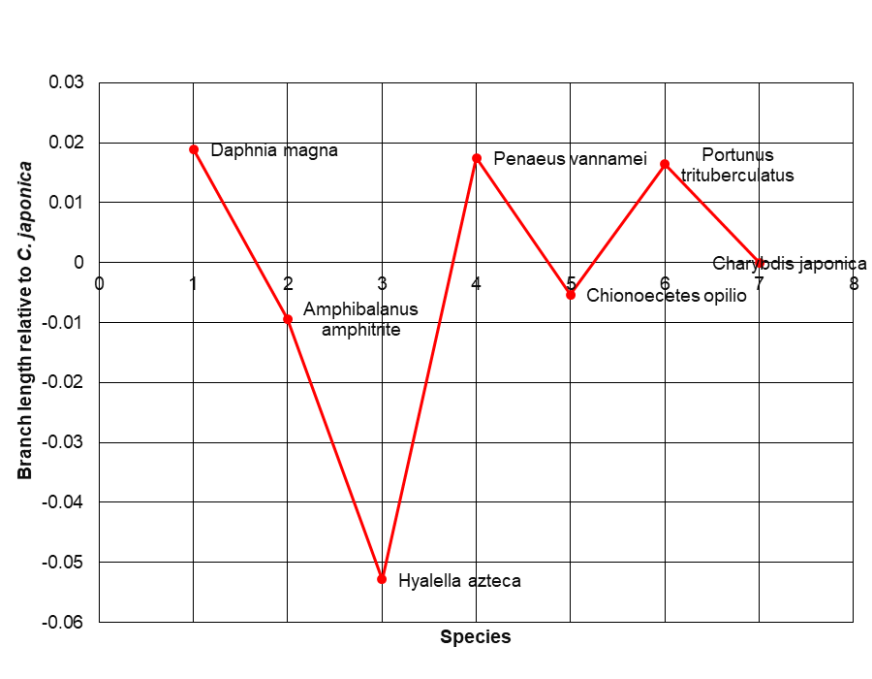


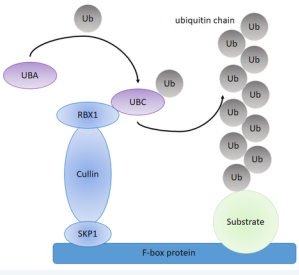


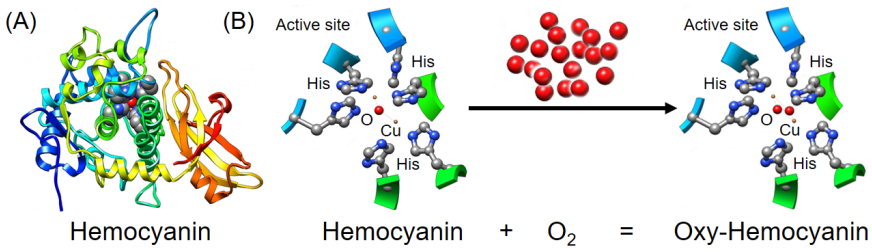












Hosted file

Table 1.docx available at https://authorea.com/users/658087/articles/713131-charybdis-japonica-genome-provides-insights-into-desiccation-adaptation-and-sex-determining-region

Hosted file

Table 2.docx available at https://authorea.com/users/658087/articles/713131-charybdis-japonica-genome-provides-insights-into-desiccation-adaptation-and-sex-determining-region

Hosted file

Table 3.docx available at https://authorea.com/users/658087/articles/713131-charybdis-japonica-genome-provides-insights-into-desiccation-adaptation-and-sex-determining-region

Hosted file

Table 4.docx available at https://authorea.com/users/658087/articles/713131-charybdis-japonica-genome-provides-insights-into-desiccation-adaptation-and-sex-determining-region

Hosted file

Table 5.docx available at https://authorea.com/users/658087/articles/713131-charybdis-japonica-genome-provides-insights-into-desiccation-adaptation-and-sex-determining-region

Hosted file

Table 6.docx available at https://authorea.com/users/658087/articles/713131-charybdis-japonica-genome-provides-insights-into-desiccation-adaptation-and-sex-determining-region

Hosted file

Table 7.docx available at https://authorea.com/users/658087/articles/713131-charybdis-japonica-genome-provides-insights-into-desiccation-adaptation-and-sex-determining-region

Hosted file

Table 8.docx available at https://authorea.com/users/658087/articles/713131-charybdis-japonica-genome-provides-insights-into-desiccation-adaptation-and-sex-determining-region

Hosted file

Table 9.docx available at https://authorea.com/users/658087/articles/713131-charybdis-japonica-genome-provides-insights-into-desiccation-adaptation-and-sex-determining-region

Hosted file

Table 10.docx available at https://authorea.com/users/658087/articles/713131-charybdis-japonica-genome-provides-insights-into-desiccation-adaptation-and-sex-determining-region