Genome assembly of a giant isopod *Bathynomus jamesi* provides insights into the body size evolution and adaptation to deep-sea environment

Jianbo Yuan[1,2,#], Xiaojun Zhang[1,2,#], Qi Kou[1,2,#], Yamin Sun[3], Chengzhang Liu[1,2], Shihao Li[1,2], Yang Yu[1,2], Chengsong Zhang[1,2], Songjun Jin[1,2], Jianhai Xiang[1,2,*], Xinzheng Li[1,2,*], Fuhua Li[1,2,*]

Running title: Deep-sea giant isopod genome

[1] CAS and Shandong Province Key Laboratory of Experimental Marine Biology, Department of Marine Organism Taxonomy & Phylogeny, Center for Ocean Mega-Science, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China.

[2] Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China.

[3] Research Center for Functional Genomics and Biochip, Tianjin 300457, China.

[#]These authors contributed equally. [*]Correspondence and requests for materials should be addressed to F.L. (fhli@qdio.ac.cn, Orcid ID: 0000-0001-8693-600X), or to X.L. (lixzh@qdio.ac.cn). or to J.X. (jhxiang@qdio.ac.cn, Orcid ID: 0000-0001-5395-7787).

Abstract

Giant isopods are the most representative group of crustaceans living in the deep sea environment with a huge body size. In order to understand the genetic basis of these large animals to adapt the harsh oligotrophic environment of the deep-sea, the genome of a deep-sea (-898 m) giant isopod *Bathynomus jamesi* was sequenced and its genome characteristics were analyzed. The genome assembly of *B. jamesi* has a total length of 5.89 Gb with a contig N50 of 587.28 Kb, which is among the largest one with high continuity of the sequenced crustacean genomes. The large genome size of *B. jamesi* is mainly attributable to the proliferation of transposable elements, especially for DNA transposons and CR1-type LINEs, which account for more than 84% of the genome. A number of expanded gene families in the genome were enriched in thyroid and insulin hormone signaling pathways, which might have driven the evolution of its huge body size. Transcriptomic analysis showed that several expanded gene families related to glycolysis and vesicular transport were specifically expressed in its digestive organs, revealing the molecular mechanism of nutrient absorption and utilization in oligotrophic environment adaptation. Taken together, the giant isopod genome provides a valuable resource for understanding the body size evolution and adaptation mechanisms of macrobenthos to the deep-sea environment.

Keywords: Giant isopod, Deep-sea, Genome assembly, Body size evolution, Oligotrophic adaptation

1 | INTRODUCTION

Isopods are a large group of crustaceans with more than 10,000 species have been described. So far, Isopoda is one of the limited groups that widely distributed in various environments, as they have been found in all seas at different depths, in fresh and brackish waters, and on land (Fig. 1A) (Hartebrodt, 2020). Therefore, Isopoda is an ideal model for studying migration and speciation, especially for the migration between deep-sea and shallow-water, and between water and land, which are hot research topics attracting world-wide attentions. There are considerable controversies about the hypothesis of the origin of life. Generally, terrestrial organisms should have evolved from marine relatives, whereas some marine organisms (such as mammals) are regarded to have evolved from terrestrial relatives (Foote et al., 2015). Besides, other theories suggest that life may originate from deep-sea hydrothermal vents because their rocky nooks can provide mineral catalysts for some vital biochemical reactions (Herschy et al., 2014; Lindner, Cairns, & Cunningham, 2008). Due to their widespread distribution, comparative genomics of isopods can provide important clues to the migration and evolutionary history of crustaceans.

Notably, isopods are one of the most morphologically diverse groups of crustaceans. Its size ranges from 0.5 cm (dwarf species) to as big as 50 cm for giant isopods (Ono, Tada, & Kose, 2017). Consistent with the Cope-Bergmann's Rule, isopods from deep sea tend to be larger than their relatives in shallower waters (Hunt & Roy, 2006). As the largest extant animals on the planet are aquatic and many of them are deep-sea organisms, the impact of marine habitats and evolutionary adaption

on body size is mysterious (Weber et al., 2020). Besides, body size has always been regarded as one of the most important quantitative traits in evolutionary scrutiny, which is strongly correlated with many physiological and fitness characters (Blanckenhorn, 2000). Thus, isopods provide an excellent model for studying the adaptive evolution of body size. Whereas, even with a great number of species, only two isopods, *Armadillidium vulgare* and *Armadillidium nasatum*, have been sequenced so far, and they are both terrestrial (Becking et al., 2019; Chebbi et al., 2019). Genomics of marine isopods, especially deep-sea species, is far from being understood.

Bathynomids (Crustacea: Isopoda: Cirolanidae) is regarded to be the “supergiant group” of isopods, which is well known for their big size (Brionesfourzan & Lozanoalvarez, 1991; Sankar et al., 2011). Bathynomids inhabit deep-sea benthic environment that are generally found on muddy bottoms at the depth of 170 m to the dark of 2140 m (Cocke, 1986; Sankar et al., 2011). To adapt to the benthic environment, a burrowing behavior has been adaptively applied for bathynomids (Matsui, Moriyama, & Kato, 2011). Besides, in order to adapt to the oligotrophic environment of the deep sea, the full-filled stomach of bathynomids accounts for approximately 2/3 of the whole body, which is conducive to food storage (Fig. 1B). In addition, midgut glands and adipocytes (collectively called "fat body") are distributed throughout the body of bathynomids to store organic reserves (Biesiot, Wang, Perry, & Trigg, 1999). Furthermore, bathynomids are well known for their extremely long hunger strikes (over five years), which should be the longest record to date (Ginn,

Beisel, & Barua, 2014). Larger animals usually have greater absolute energy requirements (Clauss et al., 2003). However, the deep sea conditions are harsh and food resources are limited, which seems to be unsuitable for the survival of giant animals (Martins, Queiroz, Santos, & Bettencourt, 2013; Wang et al., 2019). Therefore, a special mechanism should be developed for these supergiant isopods to adapt to the deep-sea oligotrophic conditions. The giant isopods provide a good model for understanding the mechanism for nutrient storage and utilization.

Deep sea expeditions provide an excellent opportunity for us to learn how animals adapt to the deep-sea environment. During a recent expedition near Hainan Island in the northern South China Sea, a new deep-sea (a depth of 898 m) bathynomid species, *Bathynomus jamesi* sp. nov., was collected and identified (Kou, Chen, Li, He, & Wang, 2017). In this study, a high-quality genome assembly of *B. jamesi* was generated using PacBio sequencing technology. Analysis of the genomic characteristics identified potential factors related to the evolution of the size of the *B. jamesi* genome. Based on the comparisons between the genomes of *B. jamesi* and its terrestrial relatives and other crustaceans, we have identified some expanded gene families related to its body size evolution and deep-sea environment adaptation. This high-quality genome will provide valuable resources for further understanding of the evolutionary history of isopods and their deep-sea environmental adaptation mechanisms.

2| MATERIALS AND METHODS

2.1 | Sampling and sequencing

The specimens of *B. jamesi* were collected by a deep-sea lander at a depth of 898 m near Hainan Island, in the northern South China Sea (17°46.845'N, 110°38.217'E). The specimens were identified as the species *B. jamesi* and kept in 75% ethanol and -80℃ freezer (Kou et al., 2017). The muscle of the legs of *B. jamesi* was collected for DNA extraction and genome sequencing. Total genomic DNA was extracted using TIANamp Marine Animal DNA Kits (Tiangen, Beijing, China), and used for Illumina and PacBio sequencing.

For Illumina sequencing, paired-end libraries with short insert size (350 bp) were constructed according to the instructions of the Illumina library preparation kit (Illumina, San Diego, USA). The constructed libraries were sequenced on an Illumina HiseqX-ten sequencing platform (Illumina, San Diego, USA). The raw sequencing reads were trimmed for quality subsequently using Trimmomatic v.0.35 (http://www.usadellab.org/cms/index.php?page=trimmomatic), and the retained clean reads were used for subsequent analyses.

For PacBio sequencing, genomic DNA was sheared to ~20 Kb, and the short fragments below the size of 10 Kb were filtered out using BluePippin (Sage Science, Beverly, USA). Filtered DNA was then used for the construction of the proprietary SMRTbell library using PacBio DNA Template Preparation Kit. SMRTbell libraries were used for single-molecule real time (SMRT) sequencing using the P6C5 sequencing chemistry (Pacific Biosciences, San Diego, USA), and then sequenced on

the PacBio RSII sequencing platform (Pacific Biosciences, San Diego, USA).

2.2 | RNA extraction and sequencing

In order to perform gene annotation and identification of tissue-specific expression genes, transcriptome sequencing was performed on six tissues of *B. jamesi*, namely gill, hepatopancreas, muscle, stomach, intestine, and nerve. According to the standard manufacturer's protocol, total RNA was isolated and purified from each tissue using TRIzol extraction reagent (Thermo Fisher Scientific, USA). RNA quality was determined by 1% agarose gel electrophoresis, and RNA concentration was assessed using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, USA). Transcriptome libraries were prepared according to the instructions of the TruSeq RNA Library Prep Kit (Illumina, San Diego, USA), and then sequenced on the Illumina HiSeq 2500 platform. The transcriptome reads were mapped to the genome using TopHat v1.2.1 (Trapnell, Pachter, & Lsalzberg, 2009). Then, fragments per kilobase of transcript per million fragments mapped (FPKM) was calculated using Cufflinks v2.2.1 (http://cole-trapnell-lab.github.io/cufflinks/). The differential gene expression analysis was conducted by using edgeR V3.10 (Robinson, McCarthy, & Smyth, 2010).

2.3 | Genome size estimation

Genome size of *B. jamesi* was estimated by K-mer analysis, which is widely used for the estimation of genome size and repeat content. Jellyfish was used to calculate K-mer frequencies based on the high-quality reads from the Illumina sequencing data (Marcais & Kingsford, 2011). A K-mer depth distribution was plotted

and the peak depth could be identified. The genome size was estimated as the ratio of the total number of K-mers to the peak depth.

2.4| Genome assembly and quality assessment

The *B. jamesi* genome was *de novo* assembled based on PacBio subreads using FALCON pipeline (https://github.com/PacificBiosciences/FALCON/) with default parameters. The assembled sequences were then polished using Quiver (SMRT Analysis v2.3.0) based on the alignment of PacBio reads to the assembly. Besides, In order to make the genome assembly more accurate, several rounds of iterative error correction were performed using the aforementioned Illumina clean data.

To assess the quality of the genome assembly, Illumina sequencing reads were aligned to the genome using Bowtie2 and the genome coverage was calculated (Langmead & Salzberg, 2012). Besides, the unigenes from the transcriptome data were mapped to the *B. jamesi* genome to assess the completeness of the gene regions. In addition, the sets of Benchmarking universal single-copy orthologs (BUSCO) was used to evaluate the completeness of the genome assembly (http://gitlab.com/ezlab/busco).

2.5| Repetitive sequence annotation

Transposable elements (TEs) in the *B. jamesi* genome were predicted by a combination of *de novo*-based and homology-based approaches. For TE annotation, both RepeatModeler and RepeatMasker were used to perform *de novo* identification (Tarailo-Graovac & Chen, 2009). RepeatMasker was used to identify transposable elements by aligning genome sequences against RepBase (RepBase21.04) and a local

library generated by RepeatModeler with default parameters.

For phylogenetic analysis of TEs, MUSCLE was used for multiple alignments of each cluster of TEs in a fast mode (-maxiters 2 -diags1) (Edgar, 2004). Based on the alignment results, a maximum likelihood (ML) method was used for phylogenetic tree construction with the parameters of "-n 1 -o tl -m 012345". The visualization of the tree was performed on the iTOL (https://itol.embl.de/).

2.6| Protein-coding gene prediction and annotation

Protein-coding genes were predicted through a combination of *de novo* prediction, homology-based prediction and transcriptome-based prediction methods. For *de novo* prediction, the coding regions of the repeat-masked genome were predicted by Augustus v2.5.5 (Stanke, Steinkamp, Waack, & Morgenstern, 2004). For homology-based prediction, protein-coding genes from *Daphnia pulex*, *Eulimnadia texana*, *Litopenaeus vannamei*, *Parhyale hawaiensis*, *Drosophila melanogaster*, *Bombyx mori* and *Anopheles gambiae* were downloaded from NCBI and mapped against the *B. jamesi* genome with Exonerate v2.2.0 (http://www.ebi.ac.uk/~guy/exonerate/). For transcriptome-based prediction, the transcriptome data were aligned to the *B. jamesi* genome using Tophat v2.1.1. Then, Cufflinks v2.2.1 was used to convert the transcripts to gene models (Trapnell et al., 2009). Finally, all gene models predicted by above three methods were integrated into a non-redundant gene set through EvidenceModeler (EVM) (Haas et al., 2008).

Functional annotation of the predicted genes was conducted by blasting against the NR and SwissProt databases using BLASTP program. Protein domains were

annotated by mapping the genome to the InterPro and Pfam databases using InterProScan and HMMER. (Prakash, Jeffryes, Bateman, & Finn, 2017; Zdobnov & Apweiler, 2001). KEGG Automatic Annotation Server (KAAS) was used to annotate the pathways in which the genes might be involved through mapping against the KEGG database (https://www.genome.jp/kaas-bin/kaas_main). The Gene Ontology (GO) classifications of the genes were extracted from the corresponding InterProScan or Pfam results (http://geneontology.org/docs/go-annotations/).

2.7| Gene family analyses

To understand the evolutionary dynamics of the genes, gene family clustering analysis was performed through using OrthoMCL (L. Li, Stoeckert, & Roos, 2003). An all-to-all blast search was conducted on the protein-coding genes of 20 arthropods using BLASTP program, including *B. jamesi*, *A. vulgare*, *Amphibalanus Amphitrite*, *Acyrthosiphon pisum*, *Anopheles gambiae*, *D. pulex*, *E. texana*, *Eurytemora affinis*, *L. vannamei*, *Eriocheir sinensis*, *Procambarus virginalis*, *P. hawaiensis*, *Pediculus humanus*, *Tigriopus californicus*, *Strigamia maritima*, *Ixodes scapularis*, *Tetranychus urticae*, *D. melanogaster, B. mori*, *and Locusta migratoria*.

Expansion and contraction of the gene families among these 20 species were determined. Base on the clustering results calculated by OrthoMCL, gene gain and loss analysis was conducted by CAFE v4.2 [84]. The expansion and contraction of each gene family was examined by comparing cluster size differences between ancestors and each current species. A random birth and death process model was used to identify gene gain and loss along each lineage of the RAxML tree.

2.8| Phylogenetic analysis

According to the results of gene family clustering, 19 single-copy orthologous genes were selected for phylogenetic tree construction. The amino acid sequence alignment was conducted using MUSCLE with the default settings (Edgar, 2004). The conserved alignments of the single-copy genes were concatenated to form the final alignment matrix. Then, the maximum likelihood (ML) method was used for phylogenetic tree construction under the JTT+G+Inv model using RAxML (Stamatakis, 2014). ML phylogeny and branch lengths were obtained by RaxML with 1000 bootstrap replicates. The divergence time estimation was conducted by combining programs of r8s and RAxML (Yang, 1997). Fossil-derived timescales and evolutionary history were obtained from TIMETREE.

3| RESULTS AND DISCUSSION

3.1 | Genome assembly and annotation

To estimate genome size of *B. jamesi*, a total of 235.25 Gb Illumina short reads were generated and utilized for genome survey analysis (Supplementary Table S1). K-mer analysis indicated that the genome size of *B. jamesi* is approximately 5.24 Gb (Supplementary Fig. S1), which was larger than most crustacean genomes reported so far (Supplementary Table S2). The heterozygosity rate of *B. jamesi* was estimated to be 0.69% and the content of repetitive sequences was about 89.7%.

To assemble the genome of *B. jamesi*, 360.80 Gb of PacBio reads with an average length of 13 Kb were generated, covering about 69-fold of the genome

248 (Supplementary Table S1). The PacBio data was assembled into contigs using
249 FALCON, and then polished by raw PacBio and Illumina sequencing data for five
250 rounds. The final assembly was 5.89 Gb in total length with a contig N50 length of
251 587.28 Kb and GC content of 37.28%, showing a higher continuity than the genome
252 of the terrestrial isopod *A. vulgare* (contig N50 of 38.36 Kb) and many other
253 crustacean genomes as well (Table 1)(Chebbi et al., 2019).

254 To assess the quality of genome assembly, the Illumina sequencing data and
255 RNA-seq data were aligned to the *B. jamesi* genome. A total of 99.80% of Illumina
256 reads and 84.23% of RNA-seq reads were mapped on the genome (Supplementary
257 Table S3). BUSCO analysis showed that 94.98% of BUSCOs were covered by the *B.*
258 *jamesi* genome, which was comparable to many recent sequenced crustacean genomes
259 (Table 1, Supplementary Fig. S2, Table S4) (Chebbi et al., 2019; Cui et al., 2021;
260 Yuan et al., 2021; Zhang et al., 2019).

261 A total of 23,221 protein-coding genes were predicated in the *B. jamesi* genome
262 with the average lengths of genes, exons and introns of 936 bp, 223 bp, 3,010 bp,
263 respectively (Table 1). The average exon number per gene was 4.18, which was also
264 similar to that of *A. vulgare* (4.93). The average intron length of genes (3,010 bp) was
265 significantly longer than that of *A. vulgare* (1,872 bp) and many other crustacean
266 genomes with relative smaller genome size (Table 1). It is consistent with the view
267 that genome size is positively correlated with intron size (Wendel et al., 2002). A total
268 of 22,886 predicted genes (98.56%) have been annotated with putative functions
269 through blasting against the databases of NR, Swissprot, interPro and GO

(Supplementary Fig. S3).

3.2 | TEs and genome evolution

According to the Animal Genome Size Database (www.genomesize.com), the C-value of isopods ranges from 1.71 to 8.82 pg, indicating there is a 5.2-fold variation of their genome sizes (Supplementary Table S5). *B. jamesi* has a largest genome (5.89 Gb) among sequenced crustacean genomes (Supplementary Table S2), which is about 3.4-fold larger than that of the *A. vulgare* genome (1.73 Gb). K-mer analysis showed that 89.7% of the *B. jamesi* genome was composed of repetitive sequences, suggesting that repeat proliferation might be the driving force for the genome expansion of *B. jamesi* (Fig. 2A). Based on the RepBase and a local repeat database that generated by RepeatModeler, a total of 5.03 Gb sequences (85.32%) were annotated as repeats, which was significantly higher than other crustaceans ($p < 0.05$, Table 2, Supplementary Table S2). TEs and simple sequence repeats (SSRs) accounted for 84.27% and 0.65% of the *B. jamesi* genome, respectively. Different from *B. jamesi*, the *A. vulgare* genome contained abundant SSRs (18.08%), which is similar to the penaeid shrimp species (19.50% - 23.93%) (Yuan et al., 2021).

Since TEs accounted for 98.77% of the total repeats of *B. jamesi*, we next analyzed TEs in this genome in detail. DNA transposons (35.99%), long interspersed nuclear elements (LINEs, 19.36%) and long terminal repeats (LTRs, 5.95%) were three major classes of TEs in the *B. jamesi* genome (Table 2). The content of LINEs and LTRs in the genome of *B. jamesi* was very similar to its land relative *A. vulgare*. Among them, two typical LINEs (CR1 and Penelope) and two types of LTRs (Pao and

Gypsy) showed apparent proliferation in the genomes of *B. jamesi* and *A. vulgare*. Notably, DNA transposon was the most abundant TE (35.99%) in the *B. jamesi* genome, and its content was significantly higher than that of *A. vulgare* (7.08%, $p < 0.05$). Five types of DNA transposons, including TcMar-Tc1 (6.05%), hAT-hATm (5.77%), Maverick (5.08%), En-Spm (3.28%) and hAT-Tip100 (2.67%), were significantly expanded in the *B. jamesi* genome in comparison with *A. vulgare* (Table 2, $p < 0.05$).

To confirm the time of TE proliferation, we performed a divergence time estimation of TEs. More than 95% of TEs had a divergence rate of <20%, indicating that most TEs in the *B. jamesi* genome are relatively young (Fig. 2B). Thus, a remarkable TE expansion event might have occurred not long ago. The CR1-type LINE was the most abundant TE of both *B. jamesi* and *A. vulgare*, which accounted for 9.13% and 14.46% of the two genomes, respectively (Table 2). However, phylogenetic analysis of the total CR1-type LINEs of the two genomes showed that these TEs proliferated independently in the two isopods, rather than derived from their ancestor (Fig. 2C). In contrast to *B. jamesi*, CR1-type LINEs were relatively more ancient in *A. vulgare* with a divergence rate of >20% (Supplementary Fig. S4). DNA transposon was the most abundant TEs (2.12 Gb) of the *B. jamesi* genome, which were also proliferated in a recent time like that of CR1 (Supplementary Fig. S4). Therefore, the genome expansion of *B. jamesi* driven by proliferation of DNA transposons and LINEs should have occurred in a recent time.

Previous studies suggested that TEs enriched in the promoters of genes play an

important role in regulating gene expressions in response to different stresses (Wicker et al., 2018). Thus, we next analyzed the TEs surrounding genes and calculated their distance to the gene body. Different from previous report that TEs are usually enriched in upstream and downstream of genes immediately (within 2 Kb) (Wicker et al., 2018), TEs in the genome of *B. jamesi* were uniformly distributed surrounding genes (from initiation site to 10Kb), especially for LINEs, LTRs and Maverick of DNA transposons (Supplementary Fig. S5). Exceptionally, TcMar, En-Spm and hAT of DNA transposon and SINEs showed relative enrichment surrounding genes (within 2 Kb). In order to determine which types of TEs should be potentially associated with gene expression, the neighboring TEs of total genes were investigated. It was interesting to find that although many types of TEs (e.g., Maverick, TcMar-Tc1, hAT-hATm, CR1, Penelope and Pao) proliferated significantly in the *B. jamesi* genome, they were less distributed surrounding genes than other genomic regions ($p < 0.05$, Fig. 2D). In contrast, some TEs with lower abundance were significantly enriched in the promoters of genes, including Academ, En-Spm, TcMar-Tigger, hAT-Charlie, RTE-BovB and SINE. Therefore, we suggest the significant proliferation of TEs should perform a more profound impact on the evolution of the whole genome rather than on architecture of protein-coding genes.

3.3 | Comparative genomics

Comparative genomics analysis was performed between *B. jamesi* and 19 other arthropod species, and a total of 16,474 gene families were identified. Among them, 1549 gene families were commonly shared by 20 species, and 364 gene families were

isopod-specific (Fig. 3A, Supplementary Table S6). Besides, 4235 core gene families were shared by four malacostraceans (*B. jamesi*, *A. vulgare*, *E. sinensis* and *L. vannamei*), and 4698 gene families were specific in *B. jamesi* (Fig. 3B).

Based on the 19 orthologous single-copy genes, a phylogenetic tree was constructed (Fig. 3A). As expected, the two isopods (*B. jamesi* and *A. vulgare*) were clustered together and then nested by the other four malacostraceans. Isopods were estimated to be diverged from their ancestor around 376 million years ago (Mya), which is a time of the Late Devonian epoch. The deep-sea isopod (*B. jamesi*) and the terrestrial isopod (*A. vulgare*) were estimated to divergent around 257 Mya, which is consistent with the fossil records of Oniscidea (219.6 – 358.9 Mya) (Lins, Ho, & Lo, 2017). Besides, there is a record showing that the deep-sea isopod *Bathynomus giganteus* has already existed as early as 160 Mya (Shen et al., 2017). Therefore, deep-sea isopods should originate between 160 and 257 Mya.

Based on the phylogenetic tree, the expansion and contraction of gene families were calculated among 20 arthropod species (Fig. 3 A). A total of 226 significantly expanded gene families and 144 contracted families were identified in the *B. jamesi* genome ($p < 0.05$, Supplementary Table S7). The expanded gene families were functional enriched in the gene ontology (GO) terms related to membrane (GO:0016020, membrane; GO:0016021, integral component of membrane), peptidase activity (GO:0008238, exopeptidase activity; GO:0004866, endopeptidase inhibitor activity; GO:0016805, dipeptidase activity; GO:0008235, metalloexopeptidase activity; GO:0070573, metallodipeptidase activity; GO:0008237, metallopeptidase

activity; GO:0004180, carboxypeptidase activity), receptor activity (GO:0004872, receptor activity; GO:0038023, signaling receptor activity; GO:0099600, transmembrane receptor activity; GO:0004888, transmembrane signaling receptor activity; GO:0008066, glutamate receptor activity; GO:0004970, ionotropic glutamate receptor activity; GO:0001653, peptide receptor activity; GO:0004930, G-protein coupled receptor activity), and signal transduction (GO:0007165, signal transduction; GO:0007154, cell communication; GO:0044700, single organism signaling; GO:0050794, regulation of cellular process) (Supplementary Table S8). KEGG analysis significantly linked some of the expanded genes to signal transduction (cAMP signaling pathway, neuroactive ligand-receptor interaction, Cell adhesion molecules (CAMs), and several signaling pathways) and endocrine systems (renin-angiotensin system and thyroid hormone signaling pathway) (Fig. 3C, Supplementary Fig. S6).

3.4 | Gene families related to large body size

B. jamesi is a giant isopod with a body length of > 20 cm, which is significantly larger than its shallow-water and terrestrial relatives (mostly < 1 cm). Comparative genomics approach helps us discover the genetic characteristics associated with the body size evolution of giant isopods.

Comparative genomic analysis of *B. jamesi* and *A. vulgare* showed that the expanded gene families of *B. jamesi* was significantly enriched in the thyroid hormone signaling pathway (p = 2E-06) (Fig. 3C, Supplementary Fig. S6), which is an important pathway in regulating growth, development and metabolism (Mourouzis,

Lavecchia, & Xinaris, 2020). Many gene families related to thyroid hormone synthesis and secretion were significantly expanded and tandem duplicated in the *B. jamesi* genome, including phosphatidylinositol phospholipase C (PLC, 14 members), inositol 1,4,5-triphosphate receptor type 1 (ITPR1, 4 members), low density lipoprotein-related protein 2 (LPR2, 23 members), adenylate cyclase (ADCY, 14 members), serine/threonine-protein kinase mTOR (MTOR, 5 members), tuberous sclerosis 2 (TSC2, 6 members), and mediator of RNA polymerase II transcription subunit (MED, 18 members). Thyroid hormone (TH) signaling is regarded as a key modulator of fundamental biological processes that has been evolutionarily conserved in both vertebrate and invertebrate species. Thyroid peroxidase (TPO), thyroid hormone receptor α (TRα) and β (TRβ), and thyroid receptor-interacting protein 11 (TRIP11) are four key enzymes in TH biosynthesis and signaling transduction. Seven TPO genes, one TRα gene, one TRβ gene, and two TRIP11 genes were identified in the *B. jamesi* genome, indicating the presence of endogenous TH in this deep-sea organism. Whereas, only a single gene encoding TPO and TRIP11 was identified in the *A. vulgare* genome, with the lack of TRα and TRβ.

In addition to the thyroid hormone signaling, the insulin signaling is also important for growth and development. A set of common genes involved in the insulin signaling pathway were identified to be tightly associated with the body size of mammals (Bouwman et al., 2018). In the *B. jamesi* genome, the insulin signaling pathway was also under significant enrichment of expanded gene families (p = 0.0078). Insulin growth factor 2 (IGF2) is an essential peptide hormone of the insulin

signaling pathway (R. W. Li & Sperling, 2001), and a single IGF2-like gene was identified in both *B. jamesi* and *A. vulgare*. Whereas, some other genes related to the insulin signaling were significantly expanded in the *B. jamesi* genome in contrast to *A. vulgare*, including insulin-like growth factor-binding protein (IGFBP) and insulin enhancer protein (ISL). IGFs are normally bound to IGFBPs in great affinities that higher than IGF receptors (IR), and function as modulators of IGF availability and activity (Hwa, Oh, & Rosenfeld, 1999). A total of fourteen IGFBP genes were identified in the *B. jamesi* genome, which was significantly more than that of *A. vulgare* (four copies). In contrast, the number of genes encoding IR was similar in the two isopod genomes (seven and five members in *B. jamesi* and *A. vulgare* respectively). ISL is a LIM-homeodomain transcription factor that involved in insulin secretion and metabolic, and also mediates glycolysis (Guo et al., 2021). Seven ISL genes were tandemly located in the genome of *B. jamesi*, while only one ISL gene was identified in the *A. vulgare* genome. These results indicated that the key genes of the growth-related hormone signaling have been significantly replicated and expanded in the *B. jamesi* genome, which might be associated with its large body size.

3.5 | Gene families related to deep-sea adaptation

To adapt to the deep-sea oligotrophic environments, the mechanisms of food storage and utilization of giant isopods should under strong selective pressure. In accordance, giant isopods have developed a huge stomach to store food and have an extraordinary long hunger strike (> 5 years) (Fig. 1B) (Ginn et al., 2014).

In order to identify potential genes related the nutrient storage, absorption and

424 utilization, RNA-seq sequencing and analysis were performed on various tissues of *B.*
425 *jamesi*. A total of 901 genes were identified to be specifically highly expressed in
426 digestive organs, including stomach and intestine. Functional enrichment analysis of
427 these differentially expressed genes indicated that they were enriched in the pathways
428 of mismatch repair, insulin signaling and resistance, endocytosis, glycolysis, and so
429 on (Fig. 4A). Glycolysis is an important metabolic process in which glucose is broken
430 down to produce energy. The genes involved in the glycolysis pathway were mostly
431 highly expressed in stomach, intestine and muscle (Fig. 4B). Among them,
432 phosphoglucomutase-2 (PGM2) is a transferase that plays an important role in
433 carbohydrate metabolism of both glycogenolysis and glyconeogenesis (Morava, 2014).
434 Seven genes encoding PGM2 were identified in the *B. jamesi* genome, whereas only
435 one PGM2 gene was identified in the *A. vulgare* genome. Besides, these genes were
436 tandem duplicated on scaffold281 and scaffold7261, and mostly high-expressed in
437 stomach and intestine. Similar results were also identified in the genes encoding
438 acetyl-CoA synthetase (ACSS1_2) and alcohol dehydrogenase (ADH), both of which
439 participate in TCA cycle for ATP production. A total of five *ACSS1_2* and 21 *ADH*
440 were identified in the *B. jamesi* genome, which were significantly more than that of *A.*
441 *vulgare* (one *ACSS1_2* and seven *ADH*), and these genes were also highly expressed
442 in the stomach and intestine. Therefore, the glycolysis of *B. jamesi* might be
443 strengthened, which should support sufficient energy for the activity of this species.
444 Besides energy production, the molecule transportation is also important for the
445 absorption and utilization of food. Vesicular transport is an important procedure of

446 transporting macromolecules through the membrane, which has been identified to be
447 under strong natural selection in the deep-sea crustaceans (Yuan et al., 2020).
448 Endocytosis is an essential process of the vesicular transport mechanisms, which
449 actively transporting molecules into the cell by engulfing it with its membrane. The
450 pathway of endocytosis were significantly enriched by differentially expressed genes
451 (p = 0.0018), and a large number of them were specifically expressed in the stomach
452 and intestine (Fig. 4C). Besides, some expanded gene families were identified to be
453 involved in vesicular transport, and annexin B9 (AnxB9) was a representative one
454 among them. AnxB9 is a functional protein that involved in the formation of
455 multivesicular bodies and regulation of protein trafficking, and even stabilizing the
456 endomembrane system during stress (Monika Tjota et al., 2011). A total of 53 genes
457 encoding AnxB9 were identified in the *B. jamesi* genome, which were significantly
458 more than that of *A. vulgare* (eight genes) and other crustaceans (seven genes on
459 average). These AnxB9 genes were mostly tandem duplicated in the *B. jamesi* genome
460 (Fig. 5), and some of them were highly expressed in the stomach, intestine and muscle.
461 Therefore, the gene family expansion and their specific expression in digestive organs
462 play an important role in the energy supply of giant isopod, and help these organisms
463 adapt to the oligotrophic conditions of the deep-sea environments.
464
465 **4| CONCLUSIONS**
466 A genome of a deep-sea giant isopod *B. jamesi* was successfully assembled,
467 which is the first high-quality genome of aquatic isopods. Comparative genomic

analyses provided new insights into the evolution of genome size and body size of animals, and the adaptation mechanisms of the deep-sea extreme environments. The isopod genomes will shed lights on the migration and evolution history of the crustaceans inhabiting deep-sea, shallow water and land. Furthermore, the genomic resources also provide tools for broader studies on the ecology, evolution, and conservation of isopods.

Acknowledgements

We acknowledge financial support from the Natural Science Foundation of China (42176105, 31830100, 31972782, and 41876167), the National Key Research & Development Program of China (2018YFD0900404 and 2018YFD0900103), and the China Agriculture Research system-48 (CARS-48). We acknowledge the support from High Performance Computing Center, Institute of Oceanology, CAS.

References:

Becking, T., Chebbi, M. A., Giraud, I., Moumen, B., Laverre, T., Caubet, Y., . . . Cordaux, R. (2019). Sex chromosomes control vertical transmission of feminizing Wolbachia symbionts in an isopod. *Plos Biology, 17*(10). doi:ARTN e300043810.1371/journal.pbio.3000438

Biesiot, P. M., Wang, S. Y., Perry, H. M., & Trigg, C. (1999). Organic reserves in the midgut gland and fat body of the giant deep-sea isopod Bathynomus giganteus. *Journal of Crustacean Biology, 19*(3), 450-458. doi:Doi 10.2307/1549253

Blanckenhorn, W. U. (2000). The evolution of body size: What keeps organisms small? *Quarterly Review of Biology, 75*(4), 385-407. doi:Doi 10.1086/393620

Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Ponce, C. H., Sargolzaei, M., Schenkel, F. S., . . . Hayes, B. J. (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet, 50*(3), 362-367. doi:10.1038/s41588-018-0056-5

Brionesfourzan, P., & Lozanoalvarez, E. (1991). Aspects of the Biology of the Giant Isopod Bathynomus-Giganteus Edwards,A.Milne, 1879 (Flabellifera, Cirolanidae), Off the Yucatan Peninsula. *Journal of Crustacean Biology, 11*(3), 375-385. doi:Doi 10.2307/1548464

Chebbi, M. A., Becking, T., Moumen, B., Giraud, I., Gilbert, C., Peccoud, J., & Cordaux, R. (2019). The Genome of Armadillidium vulgare (Crustacea, Isopoda) Provides Insights into Sex Chromosome Evolution in the Context of Cytoplasmic Sex Determination. *Molecular Biology and Evolution, 36*(4), 727-741. doi:10.1093/molbev/msz010

Clauss, M., Frey, R., Kiefer, B., Lechner-Doll, M., Loehlein, W., Polster, C., . . . Streich, W. J. (2003). The maximum attainable body size of herbivorous mammals: morphophysiological constraints on foregut, and adaptations of hindgut fermenters. *Oecologia, 136*(1), 14-27. doi:10.1007/s00442-003-1254-z

Cocke, B. T. (1986). Deep-sea isopods in aquaria. *Tropical Fish Hobbyist, 35*, 48-52.

Cui, Z. X., Liu, Y., Yuan, J. B., Zhang, X. J., Ventura, T., Ma, K. Y., . . . Chu, K. H. (2021). The Chinese mitten crab genome provides insights into adaptive plasticity and developmental regulation. *Nature Communications, 12*(1). doi:Artn 2395 10.1038/S41467-021-22604-3

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res, 32*(5), 1792-1797. doi:10.1093/nar/gkh340

Foote, A. D., Liu, Y., Thomas, G. W., Vinar, T., Alfoldi, J., Deng, J., . . . Gibbs, R. A. (2015). Convergent evolution of the genomes of marine mammals. *Nature Genetics, 47*(3), 272-275. doi:10.1038/ng.3198

Ginn, F., Beisel, U., & Barua, M. (2014). Flourishing with awkward creatures: Togetherness, vulnerability, killing. *Environmental Humanities, 4*(1), 113-123.

Guo, T., Bai, Y. H., Cheng, X. J., Han, H. B., Du, H., Hu, Y., . . . Ji, J. F. (2021). Insulin gene enhancer protein 1 mediates glycolysis and tumorigenesis of gastric cancer through regulating glucose transporter 4. *Cancer Communications, 41*(3), 258-272. doi:10.1002/cac2.12141

Haas, B., Salzberg, S., Zhu, W., Pertea, M., Allen, J., Orvis, J., . . . Wortman, J. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology, 9*(1), R7.

Hartebrodt, L. (2020). The Biology, Ecology, and Societal Importance of Marine Isopods. *Encyclopedia of the World's Biomes*, 567-572.

Herschy, B., Whicher, A., Camprubi, E., Watson, C., Dartnell, L., Ward, J., . . . Lane, N. (2014). An Origin-of-Life Reactor to Simulate Alkaline Hydrothermal Vents. *Journal of Molecular Evolution, 79*(5-6), 213-227. doi:10.1007/s00239-014-9658-4

Hunt, G., & Roy, K. (2006). Climate change, body size evolution, and Cope's Rule in deep-sea ostracodes. *Proc Natl Acad Sci U S A, 103*(5), 1347-1352. doi:10.1073/pnas.0510550103

Hwa, V., Oh, Y., & Rosenfeld, R. G. (1999). The insulin-like growth factor-binding

protein (IGFBP) superfamily. *Endocrine Reviews, 20*(6), 761-787. doi:Doi 10.1210/Er.20.6.761

Kou, Q., Chen, J., Li, X., He, L., & Wang, Y. (2017). New species of the giant deep-sea isopod genus Bathynomus (Crustacea, Isopoda, Cirolanidae) from Hainan Island, South China Sea. *Integr Zool, 12*(4), 283-291. doi:10.1111/1749-4877.12256

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

Li, L., Stoeckert, C. J., Jr., & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res, 13*(9), 2178-2189. doi:10.1101/gr.1224503

Li, R. W., & Sperling, A. K. (2001). IGF2 Locus. *Brenner's Encyclopedia of Genetics (Second Edition), 2*, 12-14.

Lindner, A., Cairns, S. D., & Cunningham, C. W. (2008). From Offshore to Onshore: Multiple Origins of Shallow-Water Corals from Deep-Sea Ancestors. *Plos One, 3*(6). doi:ARTN e2429 10.1371/journal.pone.0002429

Lins, L. S. F., Ho, S. Y. W., & Lo, N. (2017). An evolutionary timescale for terrestrial isopods and a lack of molecular support for the monophyly of Oniscidea (Crustacea: Isopoda). *Organisms Diversity & Evolution, 17*(4), 813-820. doi:10.1007/s13127-017-0346-2

Marcais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics, 27*(6), 764-770. doi:10.1093/bioinformatics/btr011

Martins, E., Queiroz, A., Santos, R. S., & Bettencourt, R. (2013). Finding immune gene expression differences induced by marine bacterial pathogens in the Deep-sea hydrothermal vent mussel Bathymodiolus azoricus. *Biogeosciences, 10*(11), 7279-7291. doi:10.5194/bg-10-7279-2013

Matsui, T., Moriyama, T., & Kato, R. (2011). Burrow Plasticity in the Deep-Sea Isopod Bathynomus doederleini (Crustacea: Isopoda: Cirolanidae). *Zoological Science, 28*(12), 863-868. doi:10.2108/zsj.28.863

Monika Tjota, Seung-Kyu Lee, Juan Wu, Janice A. Williams, Mansi R. Khanna, & Thomas, G. H. (2011). Annexin B9 binds to βH-spectrin and is required for multivesicular body function in Drosophila *Journal of cell science, 124*(17), 2914-2926.

Morava, E. (2014). Galactose supplementation in phosphoglucomutase-1 deficiency; review and outlook for a novel treatable CDG. *Molecular Genetics and Metabolism, 112*(4), 275-279. doi:10.1016/j.ymgme.2014.06.002

Mourouzis, I., Lavecchia, A. M., & Xinaris, C. (2020). Thyroid Hormone Signalling: From the Dawn of Life to the Bedside. *Journal of Molecular Evolution, 88*(1), 88-103. doi:10.1007/s00239-019-09908-1

Ono, A., Tada, T., & Kose, H. (2017). Research on Giant Isopod Concerning the Importance of Biodiversity and its Publicity by using ICT. *2017 31st Ieee International Conference on Advanced Information Networking and Applications Workshops (Ieee Waina 2017)*, 449-454.

585 doi:10.1109/Waina.2017.91
586 Prakash, A., Jeffryes, M., Bateman, A., & Finn, R. D. (2017). The HMMER Web
587 Server for Protein Sequence Similarity Search. *Curr Protoc Bioinformatics, 60*,
588 3 15 11-13 15 23. doi:10.1002/cpbi.40
589 Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor
590 package for differential expression analysis of digital gene expression data.
591 *Bioinformatics, 26*(1), 139-140. doi:10.1093/bioinformatics/btp616
592 Sankar, R., Rajkumar, M., Sun, J., Gopalakrishnan, A., Vasanthan, T. M., Ananthan,
593 G., & Trilles, J. P. (2011). First record of three giant marine Bathynomids
594 (Crustacea, Isopoda, Cirolanidae) from India. *Acta Oceanologica Sinica, 30*(1),
595 113-117. doi:10.1007/s13131-011-0097-4
596 Shen, Y. J., Kou, Q., Zhong, Z. X., Li, X. Z., He, L. S., He, S. P., & Gan, X. N. (2017).
597 The first complete mitogenome of the South China deep-sea giant isopod
598 Bathynomus sp (Crustacea: Isopoda: Cirolanidae) allows insights into the
599 early mitogenomic evolution of isopods. *Ecology and Evolution, 7*(6),
600 1869-1881. doi:10.1002/ece3.2737
601 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and
602 post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313.
603 doi:10.1093/bioinformatics/btu033
604 Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: a
605 web server for gene finding in eukaryotes. *Nucleic Acids Research, 32*(Web
606 Server issue), W309-312. doi:10.1093/nar/gkh379
607 Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive
608 elements in genomic sequences. *Curr Protoc Bioinformatics, Chapter 4*, Unit
609 4 10. doi:10.1002/0471250953.bi0410s25
610 Trapnell, C., Pachter, L., & Lsalzberg, S. (2009). TopHat: discovering splice junctions
611 with RNA-Seq. *Bioinformatics, 25*(9), 1105.
612 Wang, K., Shen, Y., Yang, Y., Gan, X., Liu, G., Hu, K., . . . He, S. (2019). Morphology
613 and genome of a snailfish from the Mariana Trench provide insights into
614 deep-sea adaptation. *Nature Ecology & Evolution, 3*(5), 823-833.
615 doi:10.1038/s41559-019-0864-8
616 Weber, J. A., Park, S. G., Luria, V., Jeon, S., Kim, H. M., Jeon, Y., . . . Church, G. M.
617 (2020). The whale shark genome reveals how genomic and physiological
618 properties scale with body size. *Proc Natl Acad Sci U S A, 117*(34),
619 20662-20671. doi:10.1073/pnas.1922576117
620 Wendel, J. F., Cronn, R. C., Alvarez, I., Liu, B., Small, R. L., & Senchina, D. S.
621 (2002). Intron size and genome size in plants. *Molecular Biology and*
622 *Evolution, 19*(12), 2346-2352. doi:DOI
623 10.1093/oxfordjournals.molbev.a004062
624 Wicker, T., Gundlach, H., Spannagl, M., Uauy, C., Borrill, P., Ramirez-Gonzalez, R.
625 H., . . . Sequencing, I. W. G. (2018). Impact of transposable elements on
626 genome structure and evolution in bread wheat. *Genome Biology, 19*. doi:Artn
627 103 10.1186/S13059-018-1479-0
628 Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum

likelihood. *Comput Appl Biosci, 13*(5), 555-556.
Yuan, J. B., Zhang, X. J., Gao, Y., Zhang, X. X., Liu, C. Z., Xiang, J. H., & Li, F. H. (2020). Adaptation and molecular evidence for convergence in decapod crustaceans from deep-sea hydrothermal vent environments. *Molecular Ecology, 29*(20), 3954-3969. doi:10.1111/mec.15610
Yuan, J. B., Zhang, X. J., Wang, M., Sun, Y. M., Liu, C. Z., Li, S. H., . . . Li, F. H. (2021). Simple sequence repeats drive genome plasticity and promote adaptive evolution in penaeid shrimp. *Communications Biology, 4*(1). doi:ARTN 186 10.1038/s42003-021-01716-y
Zdobnov, E. M., & Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics, 17*(9), 847-848.
Zhang, X. J., Yuan, J. B., Sun, Y. M., Li, S. H., Gao, Y., Yu, Y., . . . Xiang, J. H. (2019). Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nature Communications, 10*. doi:ARTN 356 10.1038/s41467-018-08197-4

Data Accessibility

All PacBio long-read sequencing data are available in the NCBI SRA database under accession ID of SRR16962112-SRR16962114. The genome assembly is available in the NCBI under Bioproject ID PRJNA776076. The genome assembly, predicted genes, repeats, and all raw sequencing data of genome and transcriptome are also available on the database at the link (username: sph; password: sph@8786@326): http://210.72.156.40/download/sphaeromadae/.

Author Contributions

F.L., X.L., J.X., J.Y., Q.K., and X.Z. initiated, managed, and drove the barnacle genome sequencing project. Q.K., X.Z. and J.Y. collected the animal material. J.Y., X.Z. and Y.S. prepared DNA sequencing and analysis. J.Y., Y.S., and C.L. performed genome assembly, gene annotation, genome structure analyses, and phylogenetic analyses. X.Z. and J.Y. conducted transcriptome sequencing and analysis. S.L., Y.Y., C.Z., and S.J conducted the genetics analysis. C.L., J.Y. and Y.S. submitted the genome data. J.Y., X.Z. and Q.K. wrote the manuscript and additional supplementary files. F. L., J.X., X.L., S.L., Y.Y. and S.J revised the manuscript. All authors read and approved the final manuscript.

Tables

Table 1. Summary of genome assembly and characteristics of *B. jamesi* and other three crustaceans.

Species	*B. jamesi*	*A. vulgare*	*L. vannamei*	*E. sinensis*
Genome size (bp)	5,892,409,081	1,725,108,002	1,618,026,442	1,562,256,418
Number of Contigs	22,827	52,740	50,304	12,722
Contig N50 (bp)	587,279	38,359	57,650	26,045
Contig N90 (bp)	108,712	18,318	14,641	2,670
Genome GC percent%	37.28%	29.15%	35.68%	46.39%
BUSCOs coverage (%)	94.80%	91.38%	94.00%	91.20%
Repeat percentage (%)	85.32%	69.54%	49.39%	45.30%
Gene number	23,221	19,051	25,572	28,033
Gene average length (bp)	936	1259	1,546	1,078
Exon number per gene	4.18	4.93	5.94	3.26
Exon average length (bp)	223	181	260	330
Intron average length (bp)	3,010	1,872	1,484	1,602

712 **Table 2. Comparison of the repeats among four crustaceans.**

Repeats	*B. jamesi*	*A. vulgare*	*L. vannamei*	*E. sinensis*
Total length	5.90 Gb	1.73 Gb	1.66 Gb	1.56 Gb
Repeats	85.32%	69.54%	49.39%	35.57%
DNA	35.99%	7.08%	9.33%	2.30%
DNA/En-Spm	3.28%	0.00%	6.39%	0.82%
DNA/Maverick	5.08%	0.63%	0.80%	0.10%
DNA/Merlin	0.37%	0.28%	0.00%	0.01%
DNA/TcMar-Mariner	0.87%	0.21%	0.06%	0.00%
DNA/TcMar-Tc1	6.05%	1.23%	0.03%	0.02%
DNA/hAT-Ac	1.41%	2.18%	0.00%	0.11%
DNA/hAT-Charlie	1.04%	0.11%	1.00%	0.09%
DNA/hAT-hATm	5.77%	0.81%	0.00%	0.00%
DNA/hAT-Tip100	2.67%	0.36%	0.00%	0.00%
LINE	19.36%	20.24%	2.82%	9.72%
LINE/CR1	9.13%	14.46%	0.25%	4.06%
LINE/Jockey	1.06%	0.63%	0.06%	0.05%
LINE/L2	1.80%	0.62%	0.35%	0.36%
LINE/Penelope	3.61%	1.26%	0.45%	0.04%
LINE/RTE-BovB	0.62%	3.00%	0.77%	0.91%
SINE	1.00%	0.00%	0.06%	0.29%
LTR	5.95%	5.89%	0.62%	1.79%
LTR/ERV1	0.24%	0.00%	0.02%	0.01%
LTR/Pao	2.48%	2.32%	0.00%	0.19%
LTR/Gypsy	2.76%	3.22%	0.22%	1.28%
Unknown	21.97%	14.87%	3.42%	10.39%
Satellite	0.31%	0.00%	0.10%	0.00%
Simple repeat	0.65%	18.08%	23.93%	6.90%
Low complexity	0.01%	3.57%	9.49%	2.04%

713
714
715
716
717
718
719
720
721
722
723
724
725
726

Figures

Fig. 1. The distributions and phenotypes of isopods. (A) The distributions of various isopods from the land to deep-sea environments. **(B)** The morphology of the giant isopod, *B. jamesi*.

Fig. 2. The evolution of transposable elements (TEs) and genome size. (A) The relationship between the genome size and repeat content. The repeat contents and genome sizes of the sequenced crustacean genomes were summarized in the supplementary Table S2. The TE content and the genome size was positively correlated with the Pearson correlation r = 0.68 and *p*-value = 0.00275. **(B)** Kimura distance-based copy divergence analyses of TEs in the two isopod genomes, *B. jamesi* and *A. vulgare*. The graphs represent genome coverage for each TE superfamily in the different genomes analyzed. Clustering was performed according to their Kimura distances (K-value from 0 to 50). **(C)** Phylogenetic tree of the CR1 LINEs from *B. jamesi* (yellow) and *A. vulgare* (dark gray). **(D)** Enrichment analyses of TE families within gene promoters. The closest TE was calculated for each gene, and the content of the closest TEs were calculated and compared with that of the whole genome.

Fig. 3. Comparative genomes analyses of *B. jamesi* and its relatives. (A) Phylogenetic tree and divergence times of *B. jamesi* and other arthropods. The number of significantly expanded (+, green) and contracted (−, red) gene families is designated on each branch. **(B)** Number of gene families shared among four Malacostraca species shown as a Venn diagram. **(C)** KEGG enrichment analysis of the expanded gene families of *B. jamesi*. The enrichment analysis was performed by using the toolkit from Omicshare (https://www.omicshare.com/). The enriched KEGG terms was referred to the supplementary Fig. S6.

Fig. 4. The differential gene expressions in six tissues of *B. jamesi*. (A) KEGG enrichment analysis of the highly expressed genes in stomach and intestine. The top 20 significantly enriched KEGG terms were displayed in the plot. **(B)** Expression level of the genes involved in the glycolysis of *B. jamesi*. **(C)** Expression level of the genes involved in the endocytosis of *B. jamesi*.

Fig. 5. Phylogenetic tree of the genes encoding AnxB9. The AnxB9 genes from various crustaceans were used for the tree construction, which labeled in various colors. A cluster of AnxB9 genes was specific expanded in the *B. jamesi* genome (gray background), and these genes were tandem duplicated in the genome. The circles with different colors indicate the genes located on different scaffolds.

Supplementary materials

This PDF file contains supplementary tables S1-S8 and supplementary figures S1–S6.

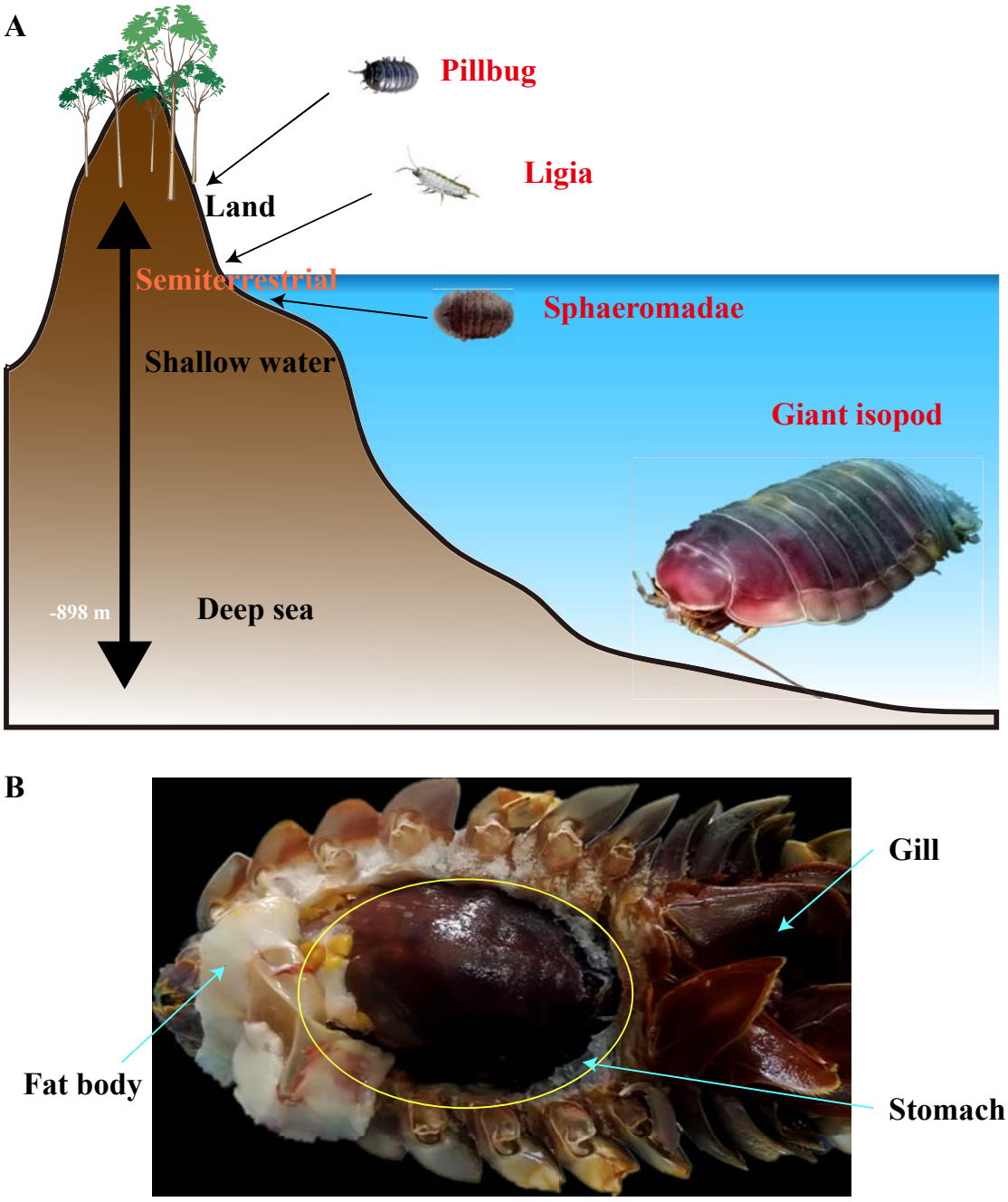

Fig. 1. The distributions and phenotypes of isopods. (A) The distributions of various isopods from the land to deep-sea environments. **(B)** The morphology of the giant isopod, *B. jamesi*.

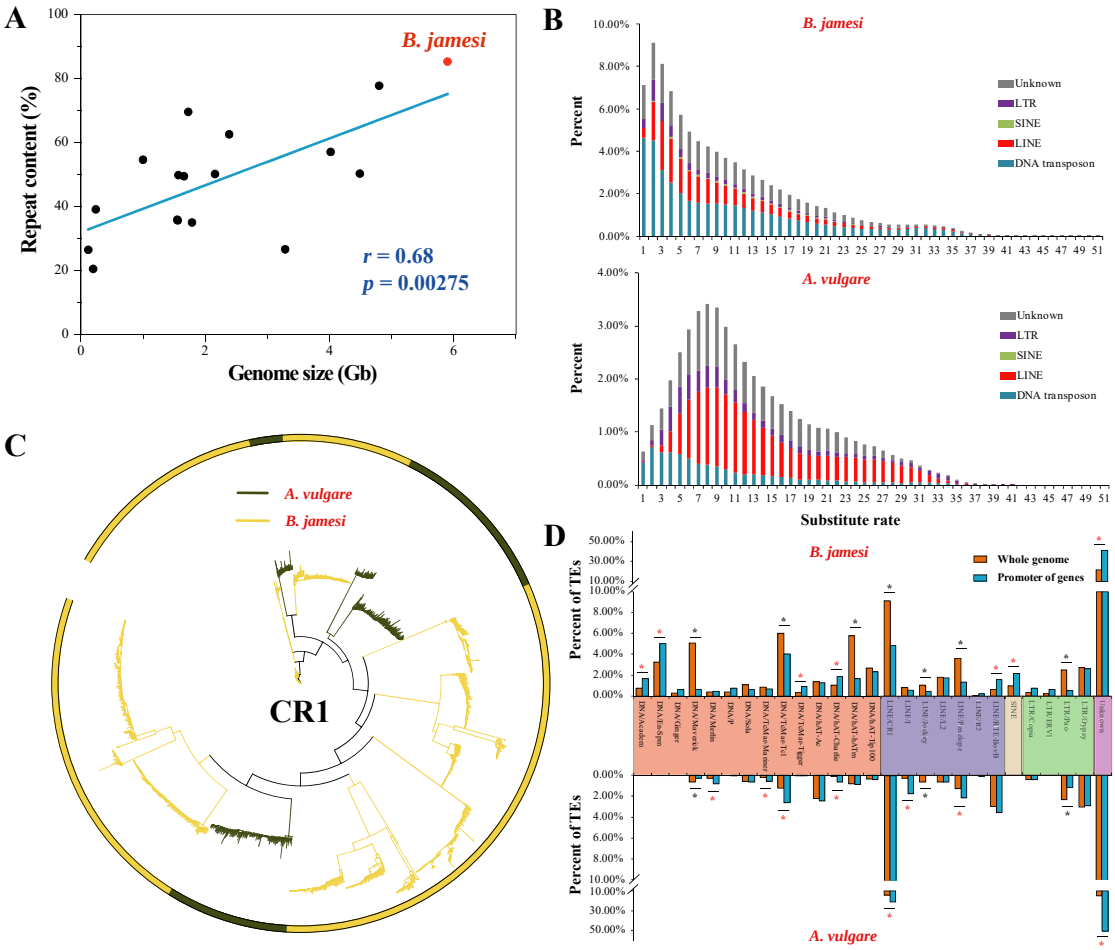

Fig. 2. The evolution of transposable elements (TEs) and genome size. (A) The relationship between the genome size and repeat content. The repeat contents and genome sizes of the sequenced crustacean genomes were summarized in the supplementary Table S2. The TE content and the genome size was positively correlated with the Pearson correlation r = 0.68 and *p*-value = 0.00275. **(B)** Kimura distance-based copy divergence analyses of TEs in the two isopod genomes, *B. jamesi* and *A. vulgare*. The graphs represent genome coverage for each TE superfamily in the different genomes analyzed. Clustering was performed according to their Kimura distances (K-value from 0 to 50). **(C)** Phylogenetic tree of the CR1 LINEs from *B. jamesi* (yellow) and *A. vulgare* (dark gray). **(D)** Enrichment analyses of TE families within gene promoters. The closest TE was calculated for each gene, and the content of the closest TEs were calculated and compared with that of the whole genome.

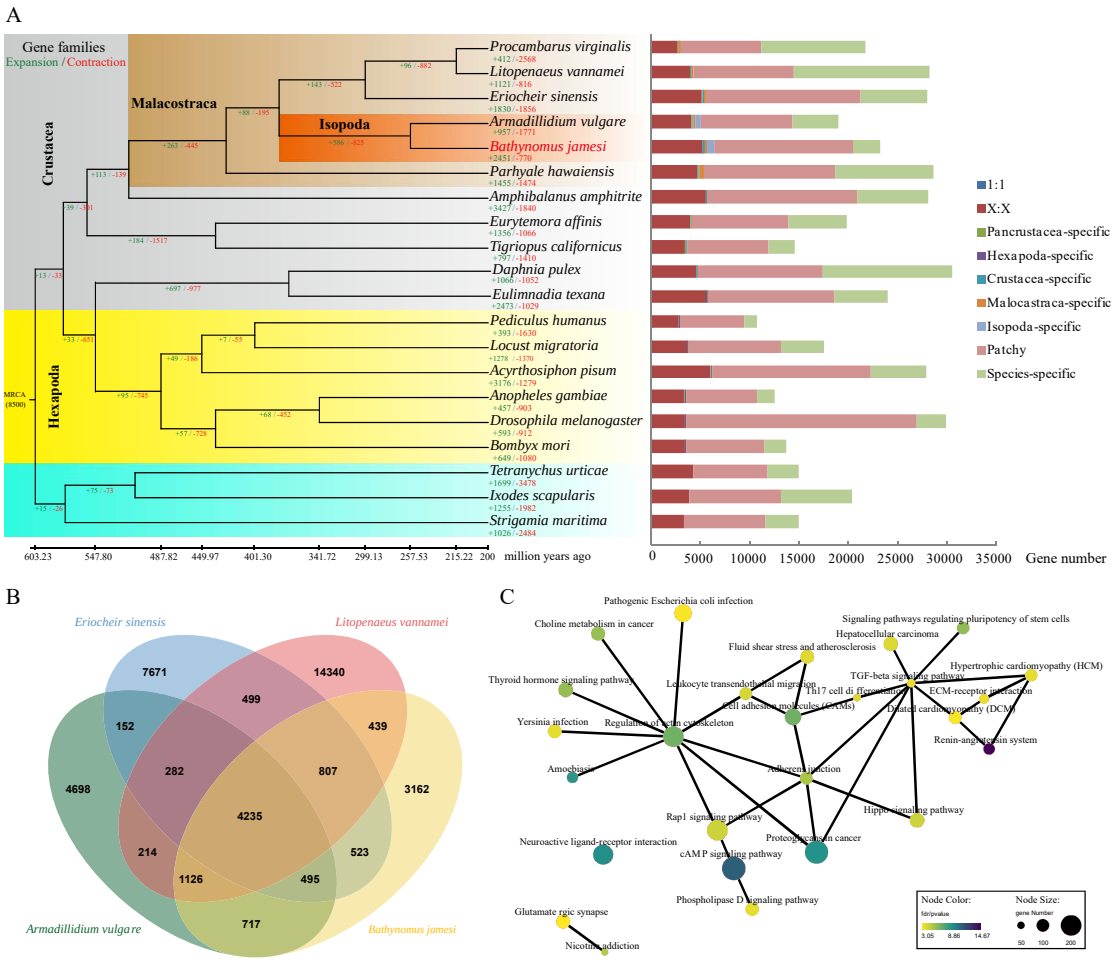

Fig. 3. Comparative genomes analyses of *B. jamesi* and its relatives. (A) Phylogenetic tree and divergence times of *B. jamesi* and other arthropods. The number of significantly expanded (+, green) and contracted (−, red) gene families is designated on each branch. **(B)** Number of gene families shared among four Malacostraca species shown as a Venn diagram. **(C)** KEGG enrichment analysis of the expanded gene families of *B. jamesi*. The enrichment analysis was performed by using the toolkit from Omicshare (https://www.omicshare.com/). The enriched KEGG terms was referred to the supplementary Fig. S6.

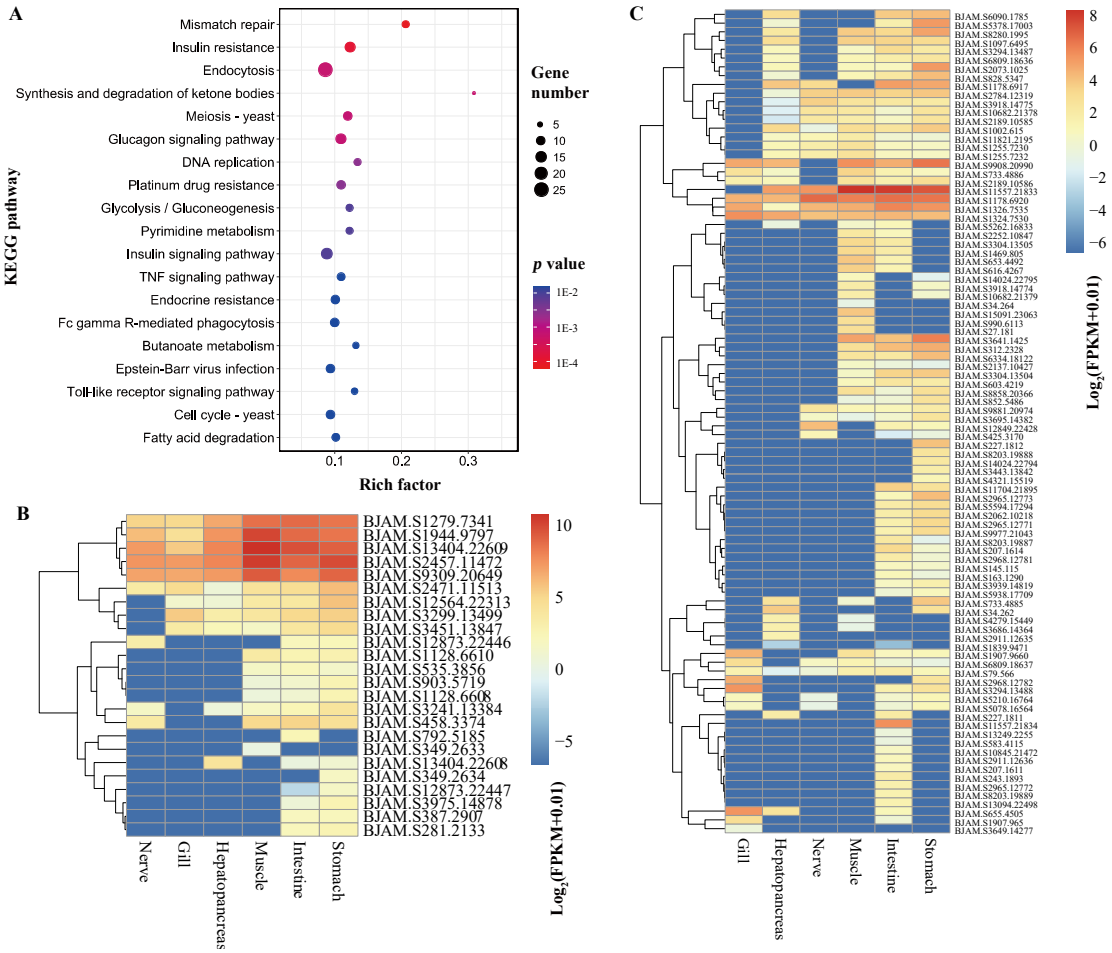

Fig. 4. The differential gene expressions in six tissues of *B. jamesi*. (A) KEGG enrichment analysis of the highly expressed genes in stomach and intestine. The top 20 significantly enriched KEGG terms were displayed in the plot. **(B)** Expression level of the genes involved in the glycolysis of *B. jamesi*. **(C)** Expression level of the genes involved in the endocytosis of *B. jamesi*.

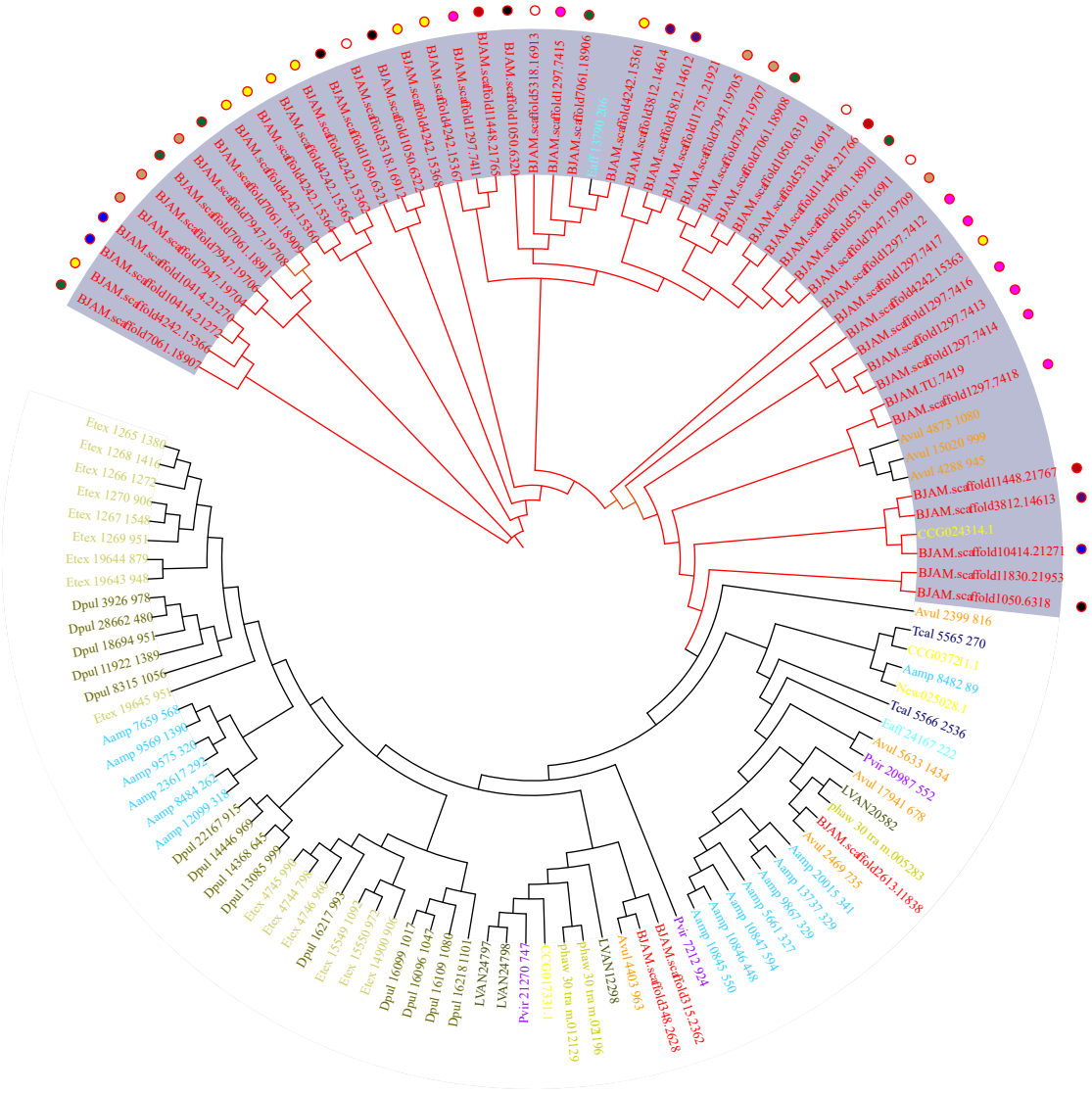

Fig. 5. Phylogenetic tree of the genes encoding AnxB9. The AnxB9 genes from various crustaceans were used for the tree construction, which labeled in various colors. A cluster of AnxB9 genes was specific expanded in the *B. jamesi* genome (gray background), and these genes were tandem duplicated in the genome. The circles with different colors indicate the genes located on different scaffolds.