The genome of an apodid holothuroid (*Chiridota heheva*) provides insights into its adaptation to deep-sea reducing environment

Running title: Genome assembly of *Chiridota heheva*

Long Zhang[1,2†], Jian He[1,2†], Peipei Tan[1,2], Zhen Gong[3], Shiyu Qian[4], Yuanyuan Miao[1], Han-Yu Zhang[5], Qi Chen[1,2], Qiqi Zhong[1], Guanzhu Han[3], Jianguo He[1,2,6,7]*, Muhua Wang[1,2]*

Affiliations:

1. State Key Laboratory for Biocontrol, School of Marine Sciences, Sun Yat-sen University, Zhuhai 519000, China.

2. Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519000, China.

3. College of Life Sciences, Nanjing Normal University, Nanjing 210023, China.

4. School of Medicine, Jinan University, Guangzhou 510632, China.

5. Hainan Key Laboratory of Marine Georesource and Prospecting, Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya 572000, China.

6. School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China.

7. Maoming Branch, Guangdong Laboratory for Lingnan Modern Agricultural Science and Technology, Maoming 525435, China.

† These authors contributed equally to this work.

***Correspondence:**

Muhua Wang: wangmuh@mail.sysu.edu.cn

Jianguo He: lsshjg@mail.sysu.edu.cn

Abstract

Cold seeps and hydrothermal vents are deep-sea reducing environments that are characterized by a lack of oxygen, photosynthesis-derived nutrients and a high concentration of reducing chemicals. Apodida is an order of deep-sea echinoderms lacking tube feet and complex respiratory trees, which are commonly found in holothurians. *Chiridota heheva* Pawson & Vance, 2004 (Apodida: Chiridotidae) is one of the few echinoderms that resides in deep-sea reducing environments. Unlike most cold seep and hydrothermal vent-dwelling animals, *C. heheva* does not survive by maintaining an epi- or endosymbiotic relationship with chemosynthetic microorganisms. The species acquires nutrients by extracting organic components from sediment detritus and suspended material. Here, we report a high-quality genome of *C. heheva* as a genomic reference for echinoderm adaptation to reducing environments. *Chiridota heheva* likely colonized its current habitats in the early Miocene. The expansion of the aerolysin-like protein family in *C. heheva* compared with other echinoderms might be involved in the disintegration of microbes during digestion, which in turn facilitates the species' adaptation to cold seep environments. Moreover, several hypoxia-related genes were subject to positive selection in the genome of *C. heheva*, which contributes to their adaptation to hypoxic environments.

Keywords

Reference genome, *Chiridota heheva*, Echinoderms, Cold seep, Aerolysin, Hypoxia

1. Introduction

Echinodermata is a phylum of marine animals comprising 5 extant classes, including Asteroidea (starfish), Ophiuroidea (brittle star), Echinoidea (sea urchin), Crinoidea (feather star), and Holothuroidea (sea cucumber) (Pawson, 2007). Adult echinoderms are characterized by having a body showing pentameral symmetry, a water vascular system with external tube feet (podia), and an endoskeleton consisting of calcareous ossicles (Pechenik, 2015). Echinoderms exhibit a high divergence in morphology, from the star-like architecture in Asteroidea to the worm-like architecture in Holothuroidea (Mooi & David, 2008; Smith et al., 2013).

Compared with other echinoderms, holothurians have a unique body architecture and evolutionary history. The worm-like body of the holothurian preserves the pentameral symmetry structurally along the oral-aboral axis (Li et al., 2020). In addition, holothurians have a soft and stretchable body, in which the ossicles are greatly reduced in size (Pechenik, 2015). Variations in body architecture also exist in Holothuroidea. The order Apodida is a group of holothurians that are found in both shallow-water and deep-sea environments (Pawson & Vance, 2004). Phylogenetic analyses showed that Apodida is sister to other orders of Holothuroidea (Lacey et al., 2005; Miller et al., 2017). Apodid holothurians lack tube feet and complex respiratory trees, making them morphologically distinct from other holothurians (Pechenik, 2015). In contrast to other classes of Echinodermata, which experienced a severe evolutionary bottleneck during the Permian-Triassic mass extinction interval, Holothuroidea

did not experience family-level extinction through the interval. The deposit-feeding lifestyle of holothurians conferred a selective advantage during the primary productivity collapse of the Permian-Triassic mass extinction (Twitchett & Oji, 2005). As the genome of only one shallow-water holothurian (*Apostichopus japonicus*) has been assembled and analyzed (Li et al., 2018; Zhang et al., 2017), it is critical to study the genomes of more holothurians to dissect their evolutionary history and developmental processes.

Cold seeps and hydrothermal vents are deep-sea reducing environments that are characterized by high hydrostatic pressure, low temperature, lack of oxygen and photosynthesis-derived nutrients, and high concentrations of reducing chemicals (Levin, 2005). However, these harsh environments support a variety of macroinvertebrates, including tubeworms, mussels, clams, and gastropods (Vanreusel et al., 2009). Most of these macrobenthos depend on the epi- or endosymbiotic relationships with chemoautotrophic microorganisms for nutrition (Van Dover et al., 2002). Recent genomic analyses have revealed the genetic basis of adaptation in several seep- and vent-dwelling macrobenthos hosting symbiotic bacteria (Li et al., 2019; Sun et al., 2020; Sun et al., 2017; Y. Sun et al., 2021). However, nonsymbiotic animals residing in deep-sea reducing environments are understudied with only one reported genome (Liu et al., 2020).

Hydrocarbon fluid seepage from cold seeps is completely devoid of O_2 and comprises high levels of sulfides. After reacting with sulfides contained in the fluid, any free O_2 is

88 removed from the deep-sea water. There are unique mechanisms of hypoxic adaptation in
89 seep-dwelling animals, as their O_2 consumption rates are similar to those of related shallow-
90 water species (Hourdez & Lallier, 2007). However, the genomic basis of hypoxic adaptation
91 in seep-dwelling animals is still lacking.
92
93 Echinoderms are a rare component of deep-sea chemosynthetic ecosystems (Tunnicliffe,
94 1992). *Chiridota heheva* Pawson & Vance, 2004 (Apodida: Chiridotidae) is one of the few
95 echinoderms that occupies all three types of chemosynthetic ecosystems (hydrothermal vent,
96 cold seep, and whale fall) (Thomas et al., 2020). This suggests that the species is well adapted
97 to deep-sea reducing environments. Unlike most seep- and vent-dwelling species, *C. heheva*
98 does not host chemosynthetic bacteria (Pawson & Vance, 2004). *Chiridota heheva* derives
99 nutrients from a variety of sources, extracting organic components from sediment detritus,
100 suspended material, and wood fragments when available (Carney, 2010; Pawson & Vance,
101 2004). A peltate-digitate tentacle structure allows *C. heheva* to exploit various food sources
102 by switching between deposit and suspension feeding (Thomas et al., 2020). The
103 cosmopolitan distribution and special lifestyle of *C. heheva* make it an ideal model to study
104 adaptation to deep-sea reducing environments in nonsymbiotic animals.
105
106 Here, we assembled and annotated a high-quality genome of *C. heheva* collected from
107 the Haima cold seep in the South China Sea. The evolutionary history of *C. heheva* was
108 investigated by inferring the phylogenetic relationship among echinoderms and the

109 demographic history of *C. heheva* and a shallow-water holothurian (*Apostichopus japonicus*).

110 Additionally, comparative genomic analyses were performed to dissect the genomic basis of

111 adaptation to deep-sea reducing environments in *C. heheva*.

112

2 Methods and Materials

2.1 Sample collection and genome sequencing

The *C. heheva* sample used in this study was collected using manned submersible *Shenhaiyongshi* from the Haima cold seep in the South China Sea (16° 73.228' N, 110° 46.143' E, 1,385 m deep) on August 2, 2019. The *C. heheva* individuals were kept in an enclosed sample chamber placed in the sample basket of the submersible. Once the samples were brought to the upper deck of the mothership, the muscle of the individuals was dissected, cut into small pieces, and immediately stored at -80°C. The samples were then transported to Sun Yat-sen University on dry-ice and stored at -80°C until use.

To construct Nanopore sequencing library, high molecular weight genomic DNA was prepared by the CTAB method. The quality and quantity of the DNA were measured via standard agarose gel electrophoresis and with a Qubit□4.0 Fluorometer (Invitrogen). Sequencing library was constructed and sequenced by Nanopore PromethION platform (Oxford Nanopore Technologies). Additionally, DNA was extracted to construct Illumina sequencing library. The quality and quantity of the DNA were measured via standard agarose gel electrophoresis and with a Qubit□2.0 Fluorometer (Invitrogen). Sequencing library was constructed and sequenced by Illumina Novaseq platform (Illumina).

2.2 Genome assembly

2.2.1 Mitochondrial genome assembly

Low quality (reads with ≥10% unidentified nucleotide and/or ≥50% nucleotides having phred score < 5) and sequencing-adaptor-contaminated Illumina reads were filtered with custom C script. The filtered Illumina reads were then trimmed with Fastp (v0.21.0) (Chen et al., 2018) to obtain high-quality Illumina reads, which were used in the following analyses. Mitochondrial genome of *C. heheva* was assembled using the two-step mode of mitoZ (v2.4) (Meng et al., 2019) with the high-quality Illumina reads. And the assembled genome was annotated using mitoZ (v2.4) with parameter "--clade Echinodermata".

2.2.2 Nuclear genome assembly

The size and heterozygosity of *C. heheva* genome were estimated using high-quality Illumina reads by *k*-mer frequency distribution method. The number of *k*-mers and the peak depth of *k*-mer sizes at 17 was obtained using Jellyfish (v2.3.0) (Marcais & Kingsford, 2011) with the -*C* setting. Genome size was estimated based on the *k*-mer analysis as described previously (Star et al., 2011).

Low quality Nanopore reads were filtered using custom Python script. Two draft genome assemblies were generated using filtered Nanopore reads with Shasta (v0.4.0) (Shafin et al., 2020) and WTDBG2 (v2.5) (Ruan & Li, 2020), respectively. The contigs of the two draft assemblies were subject to error correction using filtered Nanopore reads with Racon (v1.4.16) (Vaser et al., 2017) three times. The corrected contigs were then polished with high-quality Illumina reads with Pilon (v1.23) (Walker et al., 2014) three times. The error-corrected contigs of Shasta assembly and WTDBG2 assembly were assembled into longer

sequences using quickmerge (v0.3) (Chakraborty et al., 2016). The merged contigs were subject to error correction using filtered Nanopore reads with Racon three times, and then using high-quality Illumina reads with Pilon three times. As the heterozygosity of *C. heheva* genome is high, haplotypic duplications in the assembled genome were identified and removed using purge_dups (v1.2.3) (Guan et al., 2020). The completeness and quality of the assembly was evaluated using BUSCO (v4.0.5) (Simao et al., 2015) against the conserved Metazoa dataset (obd10), and SQUAT with high-quality Illumina reads (Yang et al., 2019).

2.3 Genome annotation

2.3.1 Repetitive element annotation

Repetitive elements in the assembly were identified by *de novo* predictions using RepeatMasker (v4.1.0) (https://www.repeatmasker.org/). A *de novo* repeat library for *C. heheva* was built using RepeatModeler (v2.0.1) (Flynn et al., 2020). To identify repetitive elements, sequences from the *C. heheva* assembly were aligned to the *de novo* repeat library using RepeatMasker (v4.1.0). Additionally, repetitive elements in *C. heheva* genome assembly were identified by homology searches against known repeat databases using RepeatMasker (v4.1.0). A repeat landscape of *C. heheva* genome was obtained using an R script that was modified from https://github.com/ValentinaBoP/TransposableElements.

2.3.2 Protein-coding gene annotation

We applied a combination of homology-based and *de novo* predication methods to build consensus gene models for the *C. heheva* genome assembly. For homology-based gene prediction, protein sequences of *Helobdella robusta*, *Lytechinus variegatus*, *Strongylocentrotus purpuratus*, *Dimorphilus gyrociliatus*, *Apostichopus japonicus* and *Acanthaster planci* were aligned to the *C. heheva* genome assembly using tblastn. The exon-intron structures then were determined according to the alignment results using GenomeThreader (v1.7.0) (Gremme et al., 2005). In addition, *de novo* gene prediction was performed using Augustus (v3.3.2) (Stanke et al., 2006), with the parameters obtained by training the software with protein sequences of *Drosophila melanogaster* and *Parasteatoda tepidariorum.* Two sets of gene models were integrated into a non-redundant consensus gene set using EvidenceModeler (v1.1.1) (Haas et al., 2008). To identify functions of the predicted proteins, we aligned the *C. heheva* protein models against NCBI NR, trEMBL, and SwissProt database using blastp (E-value threshold: 10^{-5}), and against eggNOR database (Huerta-Cepas et al., 2019) using eggNOR-Mapper (Huerta-Cepas et al., 2017). In addition, KEGG annotation of the protein models was performed using GhostKOALA (Kanehisa et al., 2016).

2.4 Phylogenomic analysis

Protein sequences of 15 metazoan species (*A. planci*, *S. purpuratus*, *Lytechinus variegatus*, *A. japonicus*, *Anneissia japonica*, *Saccoglossus kawalevskii*, *Branchiostoma floridae*, *Ciona intestinalis*, *Danio rerio*, *Gallus gallus*, *H. robusta*, *Mus musculus*, *Pelodiscus sinensis*, *Petromyzon marinus*, and *Xenopus laevisproteins*) were downloaded from NCBI.

And protein sequences of *Parastichopus parvimensis* were downloaded from Echinobase (Kudtarkar & Cameron, 2017). OrthoMCL (v2.0.9) (Li et al., 2003) was applied to determine and cluster gene families among these 16 metazoan species and *C. heheva*. Gene clusters with >100 gene copies in one or more species were removed. Single-copy othologs in each gene cluster were aligned using MAFFT (v7.310) (Katoh et al., 2002). And the alignments were trimmed using ClipKit (v1.1.3) (Steenwyk et al., 2020) with "gappy" mode. The phylogenetic tree was reconstructed with the trimmed alignments using a maximum-likelihood method implemented in IQ-TREE2 (v2.1.2) (Minh et al., 2020) with *H. robusta* as outgroup. The best-fit substitution model was selected by using ModelFinder algorithm (Kalyaanamoorthy et al., 2017). Branch supports were assessed using the ultrafast bootstrap (UFBoot) approach with 1,000 replicates (Hoang et al., 2018).

To estimate the divergent time among echinoderms, single-copy orthologs were identified among *A. japonica*, *A. planci*, *A. japonicus*, *P. parvimensis*, *C. heheva*, *L. variegatus* and *S. purpuratus* after running OrthoMCL pipeline as mentioned above. Single-copy orthologs were aligned using MAFFT (v7.310), trimmed using ClipKit (v1.1.3) with 'gappy' mode, and concatenated using PhyloSuite (v1.2.2) (Zhang et al., 2020). Divergent time among 7 echinoderms were estimated using the concatenated alignment with MCMCtree module of the PAML package (v4.9) (Tessmar-Raible & Arendt, 2003). MCMCtree analysis was performed using the maximum-likelihood tree that was reconstructed by IQ-TREE2 as a guide tree and calibrated with the divergent time obtained from TimeTree database (minimum

= 193 million years and soft maximum = 350 million years between *L. variegatus* and *S. purpuratus*).

2.5 Demographic inference of *C. heheva* and *A. japonicus*

Paired-end Illumina reads of *A. japonicus* (Li et al., 2018) were downloaded from NCBI SRA database. The reads of *A. japonicus* were filtered with custom C script and trimmed with fastp (v0.21.0). The Illumina clean reads of *C. heheva* and *A. japonicus* were aligned to the respective reference genome assembly using BWA (v0.7.17) (Li & Durbin, 2009) with "mem" function. Genetic variants were identified using Samtools (v1.9) (Li et al., 2009). Whole genome consensus sequence was generated with the genetic variants using Samtools (v 1.9). PSMC (v0.6.5) (Li & Durbin, 2011) was used to infer the demographic history of *C. heheva* and *A. japonicus* using the whole genome consensus sequences. The substitution mutations rate and generation time of *C. heheva* and *A. japonicus* was set to 1.0×10^{-8} and 3 years according to the previous study of *A. planci* (Hall et al., 2017).

2.6 Homeobox gene analysis

Homeobox genes in *C. heheva* genome were identified by following the procedure described previously (Marletaz et al., 2019). Homeodomain sequences, which were retrieved from HomeoDB database (http://homeodb.zoo.ox.ac.uk) (Zhong et al., 2008), were aligned to *C. heheva* genome assembly using tbalstn. Sequences of the candidate homeobox genes were

238 extracted based on the alignment results. The extracted sequences were aligned against NCBI
239 NR and HomeoDB database to classify the homeobox genes.
240
241 **2.7 Gene family evolution**
242 2.7.1 Gene family expansion and contraction analysis
243 r8s (v1.7) (Sanderson, 2003) was applied to obtain the ultrametric tree of 7 echinoderm
244 species, which is calibrated with the divergent time between *A. planci* and *S. purpuratus* (541
245 mya). CAFÉ (v5) (De Bie et al., 2006) was applied to determine the significance of gene
246 family expansion and contraction among 7 echinoderm species based on the ultrametric tree
247 and the gene clusters determined by OrthoMCL (v2.0.9).
248
249 2.7.2 Evolutionary analysis of *C. heheva* NOD-like receptors (NLRs) and other representative
250 metazoan NLRs
251 We used HMMER (v3.1) to search against the proteome of *C. heheva* with the HMM
252 profile of NACHT domain (PF05729) retrieved from Pfam 34.0 as the query and an *e* cut-off
253 value of 0.01. Proteins identified by the HMM search were retrieved from the proteome and
254 aligned with 964 representative proteins from eukaryotes and prokaryotes (Urbach &
255 Ausubel, 2017), and other representative metazoan NLRs (Yuen et al., 2014) using hmmalign
256 method implemented in HMMER (v3.1) based on the STAND NTPase domain. The
257 alignment was refined by manual editing. The large-scale phylogenetic analysis was
258 performed using an approximate maximum likelihood method implemented in FastTree

(Price et al., 2010). Representative SWACOS and MalT NTPases were used as outgroups (Urbach & Ausubel, 2017). Significant hits clustering with metazoan NLRs were regarded as NLRs, and protein domain organizations were annotated through hmmscan method implemented in HMMER (v3.1).

To explore the evolutionary relationships among *C. heheva* NLRs and other representative metazoan NLRs, we reconstructed the phylogenetic tree of NLRs. The NACHT domains of *C. heheva* NLRs and representative metazoan NLRs were aligned using MAFFT (v7.310), and then refined by manual editing. The representative metazoan NLRs were chosen from literature (Yuen et al., 2014). The phylogenetic tree was reconstructed using a maximum-likelihood method implemented in IQ-TREE2 (v2.1.2). The best-fit substitution model selected by using ModelFinder algorithm. Branch supports were assessed using the UFBoot approach with 1,000 replicates.

2.8 Identification and analysis of positively selected genes

Branch-site models implemented in the codeml module of the PAML package is widely used to identified positively selected genes (PSGs). Thus, we identified PSGs in the *C. heheva* genome within the single-copy orthologs among 7 echinoderm species, based on the branch-site models using GWideCodeML (v1.1) (Macias et al., 2020). *C. heheva* was set as the 'foreground' phylogeny, and the other species were set as the 'background' phylogeny. An alternative branch site model (Model = 2, NSsites = 2, and fix_omega = 0) and a neutral

branch site model (Model = 2, NSsites = 2, fix_omega = 1, and omega = 1) were tested. Genes with Bayesian Empirical Bayes (BEB) sites > 90 % and a corrected *P*-value < 0.1 were identified to have been subject to positive selection.

To investigate *LHPP* gene evolution, sequences of *LHPP* from 8 mammals (*Odobenus rosmarus*, *Orcinus orca*, *Lipotes vexillifer*, *Tursiops truncates*, *Physeter catodon*, *Balaenoptera acutorostrata*, *Mus musculus*, and *Homo sapiens*) and 7 echinoderms (*A. japonica*, *A. planci*, *A. japonicus*, *P. parvimensis*, *C. heheva*, *L. variegatus* and *S. purpuratus*) were aligned using MAFFT (v7.310). To reconstruct the phylogenetic tree, OrthoMCL (v2.0.9) (Li et al., 2003) was applied to determine and cluster gene families among these 15 species. Gene clusters with >100 gene copies in one or more species were removed. Single-copy othologs in each gene cluster were aligned using MAFFT (v7.310) (Katoh et al., 2002). And the alignments were trimmed using ClipKit (v1.1.3) (Steenwyk et al., 2020) with “gappy” mode. The phylogenetic tree was reconstructed with the trimmed alignments using a maximum-likelihood method implemented in IQ-TREE2 (v2.1.2) (Minh et al., 2020). *H. robusta* was used as outgroup. The best-fit substitution model was selected by using ModelFinder algorithm (Kalyaanamoorthy et al., 2017).

3. Results

3.1 Characterization and genome assembly of *C. heheva*

The sequenced sample was collected at a depth of 1,385 meters using manned submersible *Shenhaiyongshi* from the Haima cold seep in the South China Sea (16° 73.228' N, 110° 46.143' E) (**Figure 1**). We sequenced the sample genome on the Nanopore and Illumina sequencing platforms. In total, 42.43 Gb of Nanopore reads and 49.19 Gb of Illumina reads were obtained (**Table S1 and S2**). Species identity of the sequenced individual was first determined according to its morphological characteristics. In addition, we assembled the mitochondrial genome of the individual using Illumina reads. The sequence identity between the published *C. heheva* mitochondrial genome and our assembled genome was 99.74%, which confirmed the species identity of the sequenced individual (S. Sun et al., 2021). Based on the *k*-mer distribution of Illumina reads, the size of the *C. heheva* genome was estimated to be 1.23 Gb with a high heterozygosity of 2% (**Figure S1 and Table S3**). The *C. heheva* genome was assembled into 4,609 contigs, with a total size of 1.107 Gb and contig N50 of 1.22 Mb (**Table 1**). We determined the completeness of the assembled genome by running benchmarking universal single-copy orthologs (BUSCO) and sequencing quality assessment tool (SQUAT) software. BUSCO analysis with metazoan (obd10) gene set showed that the assembled *C. heheva* genome contained 89.6% complete single-copy orthologs (**Table S4**). Additionally, 91.1% of Illumina reads could be aligned to the assembled genome with high confidence in SQUAT analysis (**Table S5**). These results indicate the high integrity of our assembled genome.

3.2 Genome annotation

Repetitive elements represented 624.38 Mb (56.40%) in the *C. heheva* genome assembly (**Table S6**). Long interspersed nuclear elements (LINEs) were the largest class of annotated transposable elements (TEs), making up 9.72% of the genome. DNA transposons, which were the second largest class of TEs, represented 33.59 Mb (3.03%) of the genome. Additionally, the *C. heheva* genome comprised a large proportion (38.39%) of unclassified interspersed repeats. Comparative genomic analysis among *C. heheva* and other echinoderms revealed that the *C. heheva* genome consisted of the largest number of TEs (**Figure 2a and 2b; Table S7**). Repetitive elements constituted 624.38 Mb of the *C. heheva* genome, and they accounted for 253.98 Mb and 218.2 Mb of the genomes of *A. japonicus* and *P. parvimensis*, respectively. The differences in the repeat content were almost consistent with the size differences between the genomes of *C. heheva* and the other two holothurians. This suggests that repeats contributed to the size differences among the genomes of these three holothurians. Notably, the proportion of LINEs in the *C. heheva* genome was substantially higher than that in the genomes of other echinoderms (**Figure 2b**). Kimura distance-based copy divergence analysis identified a recent expansion of LINEs in the *C. heheva* genome (**Figure 2c**). Protein-coding genes were identified in the genome of *C. heheva* through a combination of *ab initio* and homology-based protein prediction approaches. In total, we derived 36,527 gene models in the *C. heheva* genome. The structure of predicted genes in *C. heheva* is slightly different to that of other previously sequenced echinoderm genomes. With longer exon and

intron as well as more exons per gene, genes in *C. heheva* are longer than the ones in *A. japonicus* (**Table S8**).

3.3 Phylogenomic analysis and demographic inference

To investigate the evolutionary history of *C. heheva*, a maximum-likelihood (ML) phylogenetic tree was reconstructed using single-copy orthologs of *C. heheva* and 16 other deuterostomes (**Figure S2**). Consistent with the results of previous analyses, the tree showed that Echinodermata and Hemichordata were sister groups to Chordata. *Chiridota heheva* appeared sister to two other holothurians, which supports the view that Apodida is the sister taxon to the remaining holothuroids (Miller et al., 2017). In addition, divergence times were determined among 7 echinoderms that had whole genome sequences (**Figure 3a**). The divergence time of *A. japonica* and other echinoderms was estimated to be approximately 569 million years (Ma), which is generally consistent with the fossil records (Smith, 1988a; Zamora et al., 2013). *Chiridota heheva* and two other holothurians were estimated to have diverged approximately 429 Ma. As Apodid is the basal taxon in Holothuroidea, these results indicate that holothurians started to diverge in the Early Ordovician (Benton & Twitchett, 2003).

We studied the demographic history of the deep-sea (*C. heheva*) and shallow-water (*A. japonicus*) holothurians by inferring the histories of ancestral population size using the pairwise sequential Markovian coalescent (PSMC) method (**Figure 3b**). *Chiridota heheva*

359 experienced a decline in population size approximately 21 million years ago, which suggests
360 that this species started to colonize the current habitat at the turn of the Miocene. The decline
361 in population size in *A. japonicus* started in the late Miocene (approximately 8 Ma). *Chiridota*
362 *heheva* also experienced a moderate decline in ancestral population size in the early Pliocene.
363
364 **3.4 *Hox/ParaHox* gene clusters**
365 It has been demonstrated that *Hox* genes play a critical role in embryonic development
366 (Pearson et al., 2005). In addition, previous studies proposed that the presence/absence and
367 expression pattern of *Hox* genes might contribute to morphological patterning and embryonic
368 development in echinoderms (Li et al., 2018; Zhang et al., 2017). Therefore, to determine
369 whether *Hox* genes contribute to morphological divergence in Holothuroidea, we identified
370 *Hox* gene clusters and their evolutionary sister complex, the *ParaHox* gene cluster, in the
371 genomes of *C. heheva* and 6 other echinoderms (**Figure 4**). A *Hox* cluster and a *ParaHox*
372 cluster could be identified in the genomes of all 7 species. The gene composition and
373 arrangement of both *Hox* and *ParaHox* clusters were highly consistent between the genomes
374 of *C. heheva* and *A. japonicus*, suggesting that *Hox/ParaHox* genes do not control the
375 development of tube feet and respiratory trees in Apodida. *Hox4* was missing in Echinodeans
376 and holothurians, and *Hox6* was missing in asteroideans and holothurians. These results
377 support the view that the absence of *Hox* genes might have contributed to the morphological
378 divergence of echinoderms.
379

3.5 NLR repertoire in *C. heheva*

NACHT and leucine-rich, repeat-containing proteins (NLRs) are important components of pathogen recognition receptors (PRRs) involved in animal innate immune systems, which can perceive pathogen-associated molecular patterns (PAMPs) of viruses and bacteria (Lange et al., 2011). The bona fide NLRs contain a NACHT (NAIP, CIITA, HET-E, and TP1) domain, which belongs to the signal transduction ATPases with numerous domains (STAND) superfamily, and a series of C-terminal leucine-rich repeats (LRRs) (Ausubel, 2005; Leipe et al., 2004). The Pfam hidden Markov model (HMM) search combined with phylogenetic analysis approach identified only 53 NLRs in *C. heheva* (**Table S9**), compared with a largely expanded set of 203 NLRs in purple sea urchin, a member of the phylum Echinodermata (Hibino et al., 2006). *Chiridota heheva* contained 24 NLRs with one or more N-terminal Death/DED domain, 22 NACHT-only NLRs, 6 NLRs with other domains, including the immunoglobulin V-set domain, which was not identified in sea urchin NLRs, and only one NLR with LRRs (**Table S9**). Taken together, these results indicate that the *C. heheva* NLR repertoire shows different abundances and structural complexities than the sea urchin.

We performed phylogenetic analysis of *C. heheva* NLRs and other representative NLRs of metazoans, including humans, *Amphimedon queenslandica*, *S. purpuratus*, *Acropora digitifera*, *Nematostella vectensis*, *Pinctada fucata*, *Capitella teleta*, mollusks, and arthropods (Yuen et al., 2014). We found that the majority of *C. heheva* NLRs form a monophyletic lineage with sea urchin NLRs (**Figure 5**), supporting the lineage-specific evolution of NLRs

in Echinodermata (Zhang et al., 2010). Given that human IPAF (ice protease-activating factor) and NAIP (neuronal apoptosis inhibitory protein) proteins were reported to have originated before the evolution of vertebrates (Zhang et al., 2010), one *C. heheva* NLR clustering with these two proteins indicates that this NLR may have an ancient independent origin (**Figure 5**).

3.6 Gene family evolution

We performed gene-family analysis based on the phylogenetic tree of 7 echinoderms (**Figure 3a**). Compared with other echinoderms, 66 gene families were expanded, and 25 gene families were contracted in *C. heheva* ($P < 0.05$) (**Tables S10 and S11**). Several significantly expanded gene families are involved in the processes of cell cycle progression, protein folding, and ribosome assembly. As high hydrostatic pressure causes cell cycle delay and affects protein folding (George et al., 2007; Yancey & Siebenaller, 2015), expansion of these families may have contributed to the adaptation of *C. heheva* to cold seep environments.

Aerolysins, which are pore-forming toxins (PFTs), were first characterized as virulence factors in the pathogenic bacterium *Aeromonas hydrophyla* (Abrami et al., 2000; Dal Peraro & van der Goot, 2016). The homologs of aerolysin in eukaryotes (aerolysin-like proteins, ALPs) originated from recurrent horizontal gene transfer (HGT) (Moran et al., 2012). ALPs of different origins possess diverse functions, including immune defense and predation (Galinier et al., 2013; Szczesny et al., 2011; Xiang et al., 2014). The ALPs were significantly

422 expanded in the genome of *C. heheva* (7 copies) compared with other echinoderms (0 or 1
423 copy) ($P < 0.05$) (**Table S10**). To investigate the possible origin and function of *C. heheva*
424 ALPs, we reconstructed the phylogenetic tree of ALPs in echinoderms and diverse species.
425 Interestingly, *C. heheva* ALPs did not cluster with ALPs from other echinoderms, suggesting
426 that ALPs from *C. heheva* and other echinoderms have different origins (**Figure 6**). *Chiridota*
427 *heheva* ALPs form a clade with ALPs from sea anemones (*Nematostella vectensis* and
428 *Ecaiptasia diaphana*). This indicates that ALPs from *C. heheva* and sea anemones might have
429 similar biological functions. It was shown that ALPs secreted by *N. vectensis* are involved in
430 prey digestion (Moran et al., 2012). Therefore, the expansion of the ALP family in *C. heheva*
431 might contribute to the disintegration of microbes during digestion.
432
433 **3.7 Positively selected genes**
434 To better understand the genetic basis of its adaptation to a deep-sea reducing
435 environment, we identified genes undergoing positive selection (PSGs) in *C. heheva*.
436 Compared with 6 other echinoderms, 27 PSGs were identified in the *C. heheva* genome
437 (**Table S12**). Several hypoxia-related genes (*PKM*, *PAN2*, *LHPP*, and *RRP9*) (Benita et al.,
438 2009; Bett et al., 2013; Chen et al., 2021; Luo et al., 2011), were subject to positive selection
439 in *C. heheva*. Cold seeps and hydrothermal vents are characterized by low oxygen
440 concentrations, which are challenging for endemic species (Hourdez & Lallier, 2007).
441 Therefore, the adaptation of *C. heheva* to a deep-sea reducing environment may be attributed
442 to selection against these hypoxia-related genes. Interestingly, the *LHPP* gene is also

443 positively selected in the genomes of cetaceans, which are hypoxia-tolerant mammals (Tian et
444 al., 2017). In addition, comparative genetic analysis showed that cetaceans and *C. heheva*
445 have the same amino acid substitution at position 118 of the LHPP protein (**Figures S3 and**
446 **S4**), which indicates a possible convergent evolution in the *LHPP* during the adaptation of
447 cetaceans and *C. heheva* to hypoxic environments.
448

449 ## 4. Discussion

450 With more than 1,400 extant species, Holothuroidea is one of the largest classes in the
451 phylum Echinodermata (Pawson, 2007). In addition, holothurians are well adapted to diverse
452 marine environments, with habitats ranging from shallow intertidal areas to hadal trenches
453 (Jamieson, 2015; Smirnov et al., 2000). However, due to the lack of body fossils,
454 evolutionary study of Holothuroidea is more difficult than other classes of Echinodermata.
455 The high-quality genome of *C. heheva* presented in this report facilitates the investigation of
456 its evolutionary history. Our phylogenomic analysis revealed that the divergence of
457 echinoderms started in the early Cambrian (~539 Ma), which is consistent with the fossil
458 record. (Bottjer et al., 2006; Smith, 1988b) (**Fig. 3a**). The ancestor of *Chiridota heheva*
459 diverged from the ancestors of two shallow-water holothurians (*A. japonicus* and *P.*
460 *parvimensis*) approximately 429 Ma. As Apodida is the basal taxon in Holothuroidea, these
461 results support the view that holothurians had evolved by the Early Ordovician (Reich, 2010).
462 To better investigate the evolution of holothurians, we inferred the histories of ancestral
463 population sizes of *C. heheva* and *A. japonicus* using PSMC (**Fig. 3b**). *Chiridota heheva*

464 experienced a decline in population size approximately 21 Ma. Ocean temperature increased
465 slowly between the late Oligocene and early Miocene (21-27 Ma) after long-term cooling
466 from the end of the Eocene (Zachos et al., 1997; Zachos et al., 2001). Furthermore, species
467 diversity within Echinodermata started to increase in the early Miocene (Kroh, 2007; Oyen &
468 Portell, 2001). These results indicate that *C. heheva* might have colonized the current habitat
469 in the early Miocene when the climate transition improved adaptations in echinoderms. The
470 oceans experienced a decrease in temperature during the late Miocene (7 to 5.4 Ma) (Herbert
471 et al., 2016). A decline in ancestral population size in *A. japonicus* started approximately 7
472 Ma. *Chiridota heheva* also experienced a moderate decline in population size in the early
473 Pliocene. These results suggest that global cooling and environmental changes in the late
474 Miocene were an important driver of demographic changes in both shallow-water and deep-
475 sea holothurians.
476
477 Apodida do not have tube feet or complex respiratory trees, which are commonly found
478 in other holothurians (Barnes, 1982). It was proposed that *Hox* genes might have contributed
479 to the body development of echinoderms (Li et al., 2018). The gene composition and
480 arrangement of the *Hox*/*ParaHox* gene cluster were consistent between *A. japonicus* and *C.*
481 *heheva*, indicating that *Hox* genes were unlikely to have been involved in the morphological
482 divergence between Apodids and other holothurians. There are some inconsistent results
483 regarding the gene composition of *Hox* gene clusters in different echinoderm genomes.
484 Previous studies found that *Hox4* and *Hox6* were missing in the genomes of holothurians (Li

et al., 2018; Zhang et al., 2017), and *Hox4* was missing in the genomes of echinoids (Cameron et al., 2006). Li *et al.* (2018) proposed that *Hox6* was lost before the split of Echinozoa and Asterozoa (Li et al., 2018), while Li *et al.* (2020) suggested that the loss of *Hox4* or *Hox6* was a lineage-specific event (Li et al., 2020). We found that *Hox4* and *Hox 6* were missing in the genomes of both *C. heheva* and *A. japonicus*. In addition, *Hox6* was missing in the genome of *A. planci*, and *Hox4* was missing in the genomes of both *L. variegatus* and *S. purpuratus* (**Fig. 4a**). This suggests that *Hox4* was lost before the split of Echinoidea and Holothuroidea, and *Hox6* was lost in Holothuroidea and Asteroidea. This scenario is not parsimonious, as Holothuroidea is paraphyletic with Asteroidea. As *S. purpuratus Hox6* clusters with *A. planci Hox4* in phylogenetic analysis, Baughman et al. (2014) proposed reclassifying *S. purpuratus Hox6* as *Hox4* (Baughman et al., 2014). Following this argument, *Hox6* was missing in Holothuroidea, Echinoidea, and Asteroidea, and *Hox4* was missing in Holothuroidea (**Fig. 4b**). This supports the view that the loss of *Hox6* occurred before the split of Echinozoa and Asterozoa.

Comparative genomic analysis showed that the ALP gene family was significantly expanded in *C. heheva* compared with other echinoderms. The expansion of the ALP family in *C. heheva* might have contributed to its adaptation to cold seep environments. Cold seeps are areas where methane, hydrogen sulfide, and other hydrocarbons seep or emanate as gas from deep geologic sources (Suess, 2014). Chemosynthetic microbes oxidize the reduced chemicals contained in the fluids to produce energy and fix carbon into organic matter, which

506 supports large benthic communities around the gas source (Levin, 2005). Most seep-dwelling
507 animals survive by hosting chemosynthetic microbes (Petersen & Dubilier, 2009). *Chiridota*
508 *heheva* has a unique feeding habit of acquiring nutrients from sediment detritus, suspended
509 material, and wood fragments when available. The microbial communities of cold seeps are
510 very different from those of other seafloor ecosystems (Ruff et al., 2015). Moreover, some of
511 these microbes have unique cellular structures that might be difficult to disintegrate
512 (Katayama et al., 2020), which impedes nutrient acquisition of *C. heheva* from free-living
513 microbes of cold seeps. As typical pore-forming proteins, aerolysin and related proteins are
514 found in a large variety of species and possess diverse functions (Szczesny et al., 2011). It
515 was proposed that ALPs were derived from recurrent horizontal gene transfer. ALPs of the
516 same origin might have similar functions (Moran et al., 2012). *Chiridota heheva* ALPs and
517 ALPs from other echinoderms are likely to have different origins, as they were clustered with
518 aerolysins from distinct groups of bacteria (**Figure 6**). *Chiridota heheva* ALPs formed a clade
519 with sea anemone ALPs. Furthermore, ALPs from hydra and sea anemones are involved in
520 prey disintegration after predation by lysing cells through pore formation on membranes
521 (Moran et al., 2012; Sher et al., 2008). This suggests that the expansion of the ALP family
522 might have contributed to microbe digestion in *C. heheva*, which in turn facilitated its
523 adaptation to cold seep environments.
524
525 Several genes that are involved in hypoxic responses (*PKM*, *PAN2*, and *LHPP*) and one
526 of the HIF-1 target genes (*PPR9*) were subjected to positive selection in *C. heheva*. The

transcription of the *PKM2* gene is activated by HIF-1. PKM2 promotes transactivation of HIF-1 target genes by directly interacting with the HIF-1α subunit. PKM2 is involved in a feedback loop that reprograms glucose metabolism under hypoxic conditions (Luo et al., 2011). LHPP induces ubiquitin-mediated degradation of PKM2, which results in the inhibition of glycolysis under hypoxia (Chen et al., 2021). Interestingly, the LHPP was also subject to positive selection in cetaceans (Tian et al., 2017). Furthermore, both *C. heheva* and cetaceans have the same amino acid substitution at position 118 of the LHPP protein (**Figs. S3 and S4**). These results suggest that the two interacting genes (*PKM2* and *LHPP*) play a key role in the hypoxic adaptation of these hypoxia-tolerant marine animals.

Acknowledgements

We thank Dr. Kang Ding, and Dr. Zhimin Jian for leading the expedition of TS12-02, the crew of research vessel (R/V) *Tansuoyihao*, the pilot team of the manned submersible *Shenhaiyongshi*, and the onboard diving scientists for their technical support during the cruise. We gratefully acknowledge the National Supercomputing Center in Guangzhou for provision of computational resources. This study was supported by Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. 311021006), National Natural Science Foundation of China (No. 31900309), GuangDong Basic and Applied Basic Research Foundation (No. 2019A1515011644), and National Innovation and Entrepreneurship Training Project for College Student of China (No. 20201126).

548 ## References

549 Abrami, L., Fivaz, M., & van der Goot, F. G. (2000, Apr). Adventures of a pore-forming
550 toxin at the target cell surface. *Trends in Microbiology*, *8*(4), 168-172.
551 https://doi.org/Doi 10.1016/S0966-842x(00)01722-4
552
553 Ausubel, F. M. (2005, Oct). Are innate immune signaling pathways in plants and animals
554 conserved? *Nat Immunol*, *6*(10), 973-979. https://doi.org/10.1038/ni1253
555
556 Barnes, R. D. (1982). *Invertebrate Zoology*. Holt-Sauders International.
557
558 Baughman, K. W., McDougall, C., Cummins, S. F., Hall, M., Degnan, B. M., Satoh, N., &
559 Shoguchi, E. (2014, Dec). Genomic organization of Hox and ParaHox clusters in the
560 echinoderm, Acanthaster planci. *Genesis*, *52*(12), 952-958.
561 https://doi.org/10.1002/dvg.22840
562
563 Benita, Y., Kikuchi, H., Smith, A. D., Zhang, M. Q., Chung, D. C., & Xavier, R. J. (2009,
564 Aug). An integrative genomics approach identifies Hypoxia Inducible Factor-1 (HIF-
565 1)-target genes that form the core response to hypoxia. *Nucleic Acids Res*, *37*(14),
566 4587-4602. https://doi.org/10.1093/nar/gkp425
567
568 Benton, M. J., & Twitchett, R. J. (2003, Jul). How to kill (almost) all life: the end-Permian
569 extinction event. *Trends in Ecology & Evolution*, *18*(7), 358-365. https://doi.org/Doi
570 10.1016/S0169-5347(03)00093-4
571
572 Bett, J. S., Ibrahim, A. F. M., Garg, A. K., Kelly, V., Pedrioli, P., Rocha, S., & Hay, R. T.
573 (2013, Apr 15). The P-body component USP52/PAN2 is a novel regulator of HIF1A
574 mRNA stability. *Biochemical Journal*, *451*, 185-194.
575 https://doi.org/10.1042/Bj20130026
576
577 Bottjer, D. J., Davidson, E. H., Peterson, K. J., & Cameron, R. A. (2006, Nov 10).
578 Paleogenomics of echinoderms. *Science*, *314*(5801), 956-960.
579 https://doi.org/10.1126/science.1132310
580
581 Cameron, R. A., Rowen, L., Nesbitt, R., Bloom, S., Rast, J. P., Berney, K., Arenas-Mena, C.,
582 Martinez, P., Lucas, S., Richardson, P. M., Davidson, E. H., Peterson, K. J., & Hood,
583 L. (2006, Jan 15). Unusual gene order and organization of the sea urchin hox cluster. *J*
584 *Exp Zool B Mol Dev Evol*, *306*(1), 45-58. https://doi.org/10.1002/jez.b.21070
585
586 Carney, R. S. (2010). Stable isotope trophic patterns in echinoderm megafauna in close
587 proximity to and remote from Gulf of Mexico lower slope hydrocarbon seeps. *Deep*
588 *Sea Research Part II: Topical Studies in Oceanography*, *57*(21-23), 1965-1971.
589 https://doi.org/10.1016/j.dsr2.2010.09.027

Chakraborty, M., Baldwin-Brown, J. G., Long, A. D., & Emerson, J. J. (2016, Nov 2). Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res*, *44*(19), e147. https://doi.org/10.1093/nar/gkw654

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018, Sep 1). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884-i890. https://doi.org/10.1093/bioinformatics/bty560

Chen, W. J., Chen, L. H., Wang, J., Wang, Z. T., Wu, C. Y., Sun, K., Ding, B. Y., Liu, N., & Xu, R. X. (2021). LHPP impedes energy metabolism by inducing ubiquitin-mediated degradation of PKM2 in glioblastoma. *Am J Cancer Res*, *11*(4), 1369-1390. https://www.ncbi.nlm.nih.gov/pubmed/33948363
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8085851/pdf/ajcr0011-1369.pdf

Dal Peraro, M., & van der Goot, F. G. (2016, Feb). Pore-forming toxins: ancient, but never really out of fashion. *Nature Reviews Microbiology*, *14*(2), 77-92. https://doi.org/10.1038/nrmicro.2015.3

De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006, May 15). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, *22*(10), 1269-1271. https://doi.org/10.1093/bioinformatics/btl097

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020, Apr 28). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*, *117*(17), 9451-9457. https://doi.org/10.1073/pnas.1921046117

Galinier, R., Portela, J., Mone, Y., Allienne, J. F., Henri, H., Delbecq, S., Mitta, G., Gourbal, B., & Duval, D. (2013, Mar). Biomphalysin, a new beta pore-forming toxin involved in Biomphalaria glabrata immune defense against Schistosoma mansoni. *PLoS Pathog*, *9*(3), e1003216. https://doi.org/10.1371/journal.ppat.1003216

George, V. T., Brooks, G., & Humphrey, T. C. (2007, Oct). Regulation of cell cycle and stress responses to hydrostatic pressure in fission yeast. *Mol Biol Cell*, *18*(10), 4168-4179. https://doi.org/10.1091/mbc.e06-12-1141

Gremme, G., Brendel, V., Sparks, M. E., & Kurtz, S. (2005, Dec). Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology*, *47*(15), 965-978. https://doi.org/10.1016/j.infsof.2005.09.005

631 Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020, May 1).
632 Identifying and removing haplotypic duplication in primary genome assemblies.
633 *Bioinformatics, 36*(9), 2896-2898. https://doi.org/10.1093/bioinformatics/btaa025
634
635 Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C.
636 R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using
637 EVidenceModeler and the program to assemble spliced alignments. *Genome Biology,*
638 *9*(1). https://doi.org/ARTN R7
639 10.1186/gb-2008-9-1-r7
640
641 Hall, M. R., Kocot, K. M., Baughman, K. W., Fernandez-Valverde, S. L., Gauthier, M. E. A.,
642 Hatleberg, W. L., Krishnan, A., McDougall, C., Motti, C. A., Shoguchi, E., Wang, T.,
643 Xiang, X., Zhao, M., Bose, U., Shinzato, C., Hisata, K., Fujie, M., Kanda, M.,
644 Cummins, S. F., Satoh, N., Degnan, S. M., & Degnan, B. M. (2017, Apr 5). The
645 crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest.
646 *Nature, 544*(7649), 231-234. https://doi.org/10.1038/nature22033
647
648 Herbert, T. D., Lawrence, K. T., Tzanova, A., Peterson, L. C., Caballero-Gill, R., & Kelly, C.
649 S. (2016, Nov). Late Miocene global cooling and the rise of modern ecosystems.
650 *Nature Geoscience, 9*(11), 843-847. https://doi.org/10.1038/Ngeo2813
651
652 Hibino, T., Loza-Coll, M., Messier, C., Majeske, A. J., Cohen, A. H., Terwilliger, D. P.,
653 Buckley, K. M., Brockton, V., Nair, S. V., Berney, K., Fugmann, S. D., Anderson, M.
654 K., Pancer, Z., Cameron, R. A., Smith, L. C., & Rast, J. P. (2006, Dec 1). The immune
655 gene repertoire encoded in the purple sea urchin genome. *Dev Biol, 300*(1), 349-365.
656 https://doi.org/10.1016/j.ydbio.2006.08.065
657
658 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018, Feb 1).
659 UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol, 35*(2),
660 518-522. https://doi.org/10.1093/molbev/msx281
661
662 Hourdez, S., & Lallier, F. H. (2007). Adaptations to hypoxia in hydrothermal-vent and cold-
663 seep invertebrates. *Rev Environ Sci Biotechnol, 6*, 143-159.
664 https://doi.org/https://doi.org/10.1007/s11157-006-9110-3
665
666 Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., &
667 Bork, P. (2017, Aug). Fast Genome-Wide Functional Annotation through Orthology
668 Assignment by eggNOG-Mapper. *Mol Biol Evol, 34*(8), 2115-2122.
669 https://doi.org/10.1093/molbev/msx148
670
671 Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H.,
672 Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019,
673 Jan 8). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated

674 orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res, 47*(D1), D309-D314. https://doi.org/10.1093/nar/gky1085
675
676
677 Jamieson, A. (2015). *The hadal zone: life in the deepest ocean*. Cambridge University Press.
678
679 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017,
680 Jun). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature*
681 *Methods, 14*(6), 587-589. https://doi.org/10.1038/nmeth.4285
682
683 Kanehisa, M., Sato, Y., & Morishima, K. (2016, Feb 22). BlastKOALA and GhostKOALA:
684 KEGG Tools for Functional Characterization of Genome and Metagenome Sequences.
685 *J Mol Biol, 428*(4), 726-731. https://doi.org/10.1016/j.jmb.2015.11.006
686
687 Katayama, T., Nobu, M. K., Kusada, H., Meng, X. Y., Hosogi, N., Uematsu, K., Yoshioka,
688 H., Kamagata, Y., & Tamaki, H. (2020, Dec 14). Isolation of a member of the
689 candidate phylum 'Atribacteria' reveals a unique cell membrane structure. *Nature*
690 *Communications, 11*(1). https://doi.org/ARTN 6381
691 10.1038/s41467-020-20149-5
692
693 Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002, Jul 15). MAFFT: a novel method for
694 rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res,*
695 *30*(14), 3059-3066. https://doi.org/10.1093/nar/gkf436
696
697 Kroh, A. (2007, Sep 14). Climate changes in the Early to Middle Miocene of the Central
698 Paratethys and the origin of its echinoderm fauna. *Palaeogeography*
699 *Palaeoclimatology Palaeoecology, 253*(1-2), 169-207.
700 https://doi.org/10.1016/j.palaeo.2007.03.039
701
702 Kudtarkar, P., & Cameron, R. A. (2017, Jan 1). Echinobase: an expanding resource for
703 echinoderm genomic information. *Database (Oxford), 2017*.
704 https://doi.org/10.1093/database/bax074
705
706 Lacey, K. M. J., McCormack, G. P., Keegan, B. F., & Powell, R. (2005). Phylogenetic
707 relationships within the class holothuroidea, inferred from 18S rRNA gene data.
708 *Marine Biology, 147*(5), 1149-1154. https://doi.org/10.1007/s00227-005-0009-2
709
710 Lange, C., Hemmrich, G., Klostermeier, U. C., Lopez-Quintero, J. A., Miller, D. J., Rahn, T.,
711 Weiss, Y., Bosch, T. C. G., & Rosenstiel, P. (2011, May). Defining the Origins of the
712 NOD-Like Receptor System at the Base of Animal Evolution. *Mol Biol Evol, 28*(5),
713 1687-1702. https://doi.org/10.1093/molbev/msq349
714
715 Leipe, D. D., Koonin, E. V., & Aravind, L. (2004, Oct 8). STAND, a class of P-loop NTPases
716 including animal and plant regulators of programmed cell death: multiple, complex

domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J Mol Biol*, *343*(1), 1-28. https://doi.org/10.1016/j.jmb.2004.08.023

Levin, L. A. (2005). Ecology of cold seep sediments: interactions of fauna with flow, chemistry and microbes. In *Oceanography and marine biology* (pp. 11-56). CRC Press.

Li, H., & Durbin, R. (2009, Jul 15). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., & Durbin, R. (2011, Jul 13). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493-496. https://doi.org/10.1038/nature10231

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Genome Project Data Processing, S. (2009, Aug 15). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

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

Li, Y., Omori, A., Flores, R. L., Satterfield, S., Nguyen, C., Ota, T., Tsurugaya, T., Ikuta, T., Ikeo, K., Kikuchi, M., Leong, J. C. K., Reich, A., Hao, M., Wan, W., Dong, Y., Ren, Y., Zhang, S., Zeng, T., Uesaka, M., Uchida, Y., Li, X., Shibata, T. F., Bino, T., Ogawa, K., Shigenobu, S., Kondo, M., Wang, F., Chen, L., Wessel, G., Saiga, H., Cameron, R. A., Livingston, B., Bradham, C., Wang, W., & Irie, N. (2020, Jul 10). Genomic insights of body plan transitions from bilateral to pentameral symmetry in Echinoderms. *Communications Biology*, *3*(1), 371. https://doi.org/10.1038/s42003-020-1091-1

Li, Y., Tassia, M. G., Waits, D. S., Bogantes, V. E., David, K. T., & Halanych, K. M. (2019, Nov 18). Genomic adaptations to chemosymbiosis in the deep-sea seep-dwelling tubeworm Lamellibrachia luymesi. *BMC Biol*, *17*(1), 91. https://doi.org/10.1186/s12915-019-0713-x

Li, Y. L., Wang, R. J., Xun, X. G., Wang, J., Bao, L. S., Thimmappa, R., Ding, J., Jiang, J. W., Zhang, L. H., Li, T. Q., Lv, J., Mu, C., Hu, X. L., Zhang, L. L., Liu, J., Li, Y. Q., Yao, L. J., Jiao, W. Q., Wang, Y. F., Lian, S. S., Zhao, Z. L., Zhan, Y. Y., Huang, X. T., Liao, H., Wang, J., Sun, H. Z., Mi, X., Xia, Y., Xing, Q., Lu, W., Osbourn, A., Zhou, Z. C., Chang, Y. Q., Bao, Z. M., & Wang, S. (2018, Jun 26). Sea cucumber

760 genome provides insights into saponin biosynthesis and aestivation regulation. *Cell*
761 *Discovery*, *4*. https://doi.org/ARTN 29
762 10.1038/s41421-018-0030-5
763
764 Liu, R., Wang, K., Liu, J., Xu, W., Zhou, Y., Zhu, C., Wu, B., Li, Y., Wang, W., He, S., Feng,
765 C., & Zhang, H. (2020, Jun 1). De Novo Genome Assembly of Limpet Bathyacmaea
766 lactea (Gastropoda: Pectinodontidae): The First Reference Genome of a Deep-Sea
767 Gastropod Endemic to Cold Seeps. *Genome Biology and Evolution*, *12*(6), 905-910.
768 https://doi.org/10.1093/gbe/evaa100
769
770 Luo, W. B., Hu, H. X., Chang, R., Zhong, J., Knabel, M., O'Meally, R., Cole, R. N., Pandey,
771 A., & Semenza, G. L. (2011, May 27). Pyruvate Kinase M2 Is a PHD3-Stimulated
772 Coactivator for Hypoxia-Inducible Factor 1. *Cell*, *145*(5), 732-744.
773 https://doi.org/10.1016/j.cell.2011.03.054
774
775 Macias, L. G., Barrio, E., & Toft, C. (2020, Dec). GWideCodeML: A Python Package for
776 Testing Evolutionary Hypotheses at the Genome-Wide Level. *G3-Genes Genomes*
777 *Genetics*, *10*(12), 4369-4372. https://doi.org/10.1534/g3.120.401874
778
779 Marcais, G., & Kingsford, C. (2011, Mar 15). A fast, lock-free approach for efficient parallel
780 counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764-770.
781 https://doi.org/10.1093/bioinformatics/btr011
782
783 Marletaz, F., Peijnenburg, K. T. C. A., Goto, T., Satoh, N., & Rokhsar, D. S. (2019, Jan 21).
784 A New Spiralian Phylogeny Places the Enigmatic Arrow Worms among Gnathiferans.
785 *Curr Biol*, *29*(2), 312-+. https://doi.org/10.1016/j.cub.2018.11.042
786
787 Meng, G., Li, Y., Yang, C., & Liu, S. (2019, Jun 20). MitoZ: a toolkit for animal
788 mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res*,
789 *47*(11), e63. https://doi.org/10.1093/nar/gkz173
790
791 Miller, A. K., Kerr, A. M., Paulay, G., Reich, M., Wilson, N. G., Carvajal, J. I., & Rouse, G.
792 W. (2017, Jun). Molecular phylogeny of extant Holothuroidea (Echinodermata). *Mol*
793 *Phylogenet Evol*, *111*, 110-131. https://doi.org/10.1016/j.ympev.2017.02.014
794
795 Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler,
796 A., & Lanfear, R. (2020, May 1). IQ-TREE 2: New Models and Efficient Methods for
797 Phylogenetic Inference in the Genomic Era. *Mol Biol Evol*, *37*(5), 1530-1534.
798 https://doi.org/10.1093/molbev/msaa015
799
800 Mooi, R., & David, B. (2008). Radial Symmetry, the Anterior/Posterior Axis, and
801 Echinoderm Hox Genes. *Annual Review of Ecology, Evolution, and Systematics, Vol*
802 *46*, *39*, 43-62. https://doi.org/10.1146/annurev.ecolsys.39.110707.173521

Moran, Y., Fredman, D., Szczesny, P., Grynberg, M., & Technau, U. (2012, Sep). Recurrent horizontal transfer of bacterial toxin genes to eukaryotes. *Mol Biol Evol, 29*(9), 2223-2230. https://doi.org/10.1093/molbev/mss089

Oyen, C. W., & Portell, R. W. (2001, Feb 1). Diversity patterns and biostratigraphy of Cenozoic echinoderms from Florida. *Palaeogeography Palaeoclimatology Palaeoecology, 166*(1-2), 193-218. https://doi.org/Doi 10.1016/S0031-0182(00)00209-1

Pawson, D. L. (2007). Phylum Echinodermata. *Zootaxa, 1668*, 749-764.

Pawson, D. L., & Vance, D. J. (2004). Chirodota heheva, new species, from western Atlantic deep-sea cold seeps and anthropogenic habits (Echinodermata: Holothuroidea: Apodida). *Zootaxa, 534*, 1-12.

Pearson, J. C., Lemons, D., & McGinnis, W. (2005, Dec). Modulating Hox gene functions during animal body patterning. *Nat Rev Genet, 6*(12), 893-904. https://doi.org/10.1038/nrg1726

Pechenik, J. A. (2015). *Biology of the Invertebrates*. McGraw-Hill.

Petersen, J. M., & Dubilier, N. (2009, Oct). Methanotrophic symbioses in marine invertebrates. *Environ Microbiol Rep, 1*(5), 319-335. https://doi.org/10.1111/j.1758-2229.2009.00081.x

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010, Mar 10). FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One, 5*(3), e9490. https://doi.org/10.1371/journal.pone.0009490

Reich, M. (2010, Dec). The oldest synallactid sea cucumber (Echinodermata: Holothuroidea: Aspidochirotida). *Palaeontologische Zeitschrift, 84*(4), 541-546. https://doi.org/10.1007/s12542-010-0067-8

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

Ruff, S. E., Biddle, J. F., Teske, A. P., Knittel, K., Boetius, A., & Ramette, A. (2015, Mar 31). Global dispersion and local diversification of the methane seep microbiome. *Proc Natl Acad Sci U S A, 112*(13), 4015-4020. https://doi.org/10.1073/pnas.1421865112

Sanderson, M. J. (2003, Jan 22). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics, 19*(2), 301-302. https://doi.org/10.1093/bioinformatics/19.2.301

Shafin, K., Pesout, T., Lorig-Roach, R., Haukness, M., Olsen, H. E., Bosworth, C., Armstrong, J., Tigyi, K., Maurer, N., Koren, S., Sedlazeck, F. J., Marschall, T., Mayes, S., Costa, V., Zook, J. M., Liu, K. J., Kilburn, D., Sorensen, M., Munson, K. M., Vollger, M. R., Monlong, J., Garrison, E., Eichler, E. E., Salama, S., Haussler, D., Green, R. E., Akeson, M., Phillippy, A., Miga, K. H., Carnevali, P., Jain, M., & Paten, B. (2020, Sep). Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nat Biotechnol, 38*(9), 1044-1053. https://doi.org/10.1038/s41587-020-0503-6

Sher, D., Fishman, Y., Melamed-Book, N., Zhang, M., & Zlotkin, E. (2008, Jan). Osmotically driven prey disintegration in the gastrovascular cavity of the green hydra by a pore-forming protein. *FASEB J, 22*(1), 207-214. https://doi.org/10.1096/fj.07-9133com

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015, Oct 1). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31*(19), 3210-3212. https://doi.org/10.1093/bioinformatics/btv351

Smirnov, A. V., Gebruk, A. V., Galkin, S. V., & Shank, T. (2000). New species of holothurian (Echinodermata: Holothuroidea) from hydrothermal vent habitats. *J. Mar. Biol. Ass. U.K., 80*(2), 321-328. https://doi.org/10.1017/s0025315499001897

Smith, A. B. (1988a). Fossil evidence for the relationships of extant echinoderm classes and their time of divergence. In C. R. C. Paul & A. B. Smith (Eds.), *Echinoderm phylogeny and evolutionary biology* (pp. 85-97). Clarendon Press.

Smith, A. B. (1988b). Fossil evidence for the relationships of extant echinoderm classes and their times of divergence. In C. R. C. Paul & A. B. Smith (Eds.), *Echinoderm phylogeny and evolutionary biology* (pp. 85-97). Clarendon Press.

Smith, A. B., Zamora, S., & Alvaro, J. J. (2013, Jan). The oldest echinoderm faunas from Gondwana show that echinoderm body plan diversification was rapid. *Nature Communications, 4*. https://doi.org/ARTN 1385
10.1038/ncomms2391

Stanke, M., Schoffmann, O., Morgenstern, B., & Waack, S. (2006, Feb 9). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics, 7*. https://doi.org/Artn 62
10.1186/1471-2105-7-62

Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrom, M., Gregers, T. F., Rounge, T. B., Paulsen, J., Solbakken, M. H., Sharma, A., Wetten, O. F., Lanzen, A., Winer, R., Knight, J., Vogel, J. H., Aken, B., Andersen, O., Lagesen, K., Tooming-Klunderud, A., Edvardsen, R. B., Tina, K. G., Espelund, M., Nepal, C., Previti, C., Karlsen, B. O., Moum, T., Skage, M., Berg, P. R., Gjoen, T., Kuhl, H., Thorsen, J., Malde, K., Reinhardt, R., Du, L., Johansen, S. D., Searle, S., Lien, S., Nilsen, F., Jonassen, I., Omholt, S. W., Stenseth, N. C., & Jakobsen, K. S. (2011, Sep 8). The genome sequence of Atlantic cod reveals a unique immune system. *Nature*, *477*(7363), 207-210. https://doi.org/10.1038/nature10342

Steenwyk, J. L., Buida, T. J., 3rd, Li, Y., Shen, X. X., & Rokas, A. (2020, Dec). ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol*, *18*(12), e3001007. https://doi.org/10.1371/journal.pbio.3001007

Suess, E. (2014, Oct). Marine cold seeps and their manifestations: geological control, biogeochemical criteria and environmental conditions. *International Journal of Earth Sciences*, *103*(7), 1889-1916. https://doi.org/10.1007/s00531-014-1010-0

Sun, J., Chen, C., Miyamoto, N., Li, R., Sigwart, J. D., Xu, T., Sun, Y., Wong, W. C., Ip, J. C. H., Zhang, W., Lan, Y., Bissessur, D., Watsuji, T. O., Watanabe, H. K., Takaki, Y., Ikeo, K., Fujii, N., Yoshitake, K., Qiu, J. W., Takai, K., & Qian, P. Y. (2020, Apr 8). The Scaly-foot Snail genome and implications for the origins of biomineralised armour. *Nature Communications*, *11*(1), 1657. https://doi.org/10.1038/s41467-020-15522-3

Sun, J., Zhang, Y., Xu, T., Zhang, Y., Mu, H. W., Zhang, Y. J., Lan, Y., Fields, C. J., Hui, J. H. L., Zhang, W. P., Li, R. S., Nong, W. Y., Cheung, F. K. M., Qiu, J. W., & Qian, P. Y. (2017, May). Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nature Ecology & Evolution*, *1*(5). https://doi.org/ARTN 0121
10.1038/s41559-017-0121

Sun, S., Sha, Z., & Xiao, N. (2021, Apr 22). The first two complete mitogenomes of the order Apodida from deep-sea chemoautotrophic environments: New insights into the gene rearrangement, origin and evolution of the deep-sea sea cucumbers. *Comp Biochem Physiol Part D Genomics Proteomics*, *39*, 100839. https://doi.org/10.1016/j.cbd.2021.100839

Sun, Y., Sun, J., Yang, Y., Lan, Y., Ip, J. C., Wong, W. C., Kwan, Y. H., Zhang, Y., Han, Z., Qiu, J. W., & Qian, P. Y. (2021, Jul 13). Genomic signatures supporting the symbiosis and formation of chitinous tube in the deep-sea tubeworm Paraescarpia echinospica. *Mol Biol Evol*. https://doi.org/10.1093/molbev/msab203

930 Szczesny, P., Iacovache, I., Muszewska, A., Ginalski, K., van der Goot, F. G., & Grynberg,
931 M. (2011). Extending the aerolysin family: from bacteria to vertebrates. *PLoS One, 6*(6), e20349. https://doi.org/10.1371/journal.pone.0020349
932
933
934 Tessmar-Raible, K., & Arendt, D. (2003, Aug). Emerging systems: between vertebrates and
935 arthropods, the Lophotrochozoa. *Current Opinion in Genetics & Development, 13*(4),
936 331-340. https://doi.org/10.1016/S0959-437x(03)00086-8
937
938 Thomas, E. A., Liu, R. Y., Amon, D., Copley, J. T., Glover, A. G., Helyar, S. J., Olu, K.,
939 Wiklund, H., Zhang, H. B., & Sigwart, J. D. (2020, Nov 20). Chiridota heheva-the
940 cosmopolitan holothurian. *Marine Biodiversity, 50*(6). https://doi.org/ARTN 110
941 10.1007/s12526-020-01128-x
942
943 Tian, R., Yin, D., Liu, Y., Seim, I., Xu, S., & Yang, G. (2017). Adaptive Evolution of Energy
944 Metabolism-Related Genes in Hypoxia-Tolerant Mammals. *Front Genet, 8*, 205.
945 https://doi.org/10.3389/fgene.2017.00205
946
947 Tunnicliffe, V. (1992). The nature and origin of the modern hydrothermal vent fauna. *Palaios, 7*(4), 338-350. https://doi.org/http://doi.org/10.2307/3514820
948
949
950 Twitchett, R. J., & Oji, T. (2005). Early Triassic recovery of echinoderms. *C. R. Palevol, 4*(6-
951 7), 531-542. https://doi.org/10.1016/j.crpv.2005.02.006
952
953 Urbach, J. M., & Ausubel, F. M. (2017, Jan 31). The NBS-LRR architectures of plant R-
954 proteins and metazoan NLRs evolved in independent events. *Proc Natl Acad Sci U S A, 114*(5), 1063-1068. https://doi.org/10.1073/pnas.1619730114
955
956
957 Van Dover, C. L., German, C. R., Speer, K. G., Parson, L. M., & Vrijenhoek, R. C. (2002,
958 Feb 15). Evolution and biogeography of deep-sea vent and seep invertebrates. *Science, 295*(5558), 1253-1257. https://doi.org/10.1126/science.1067361
959
960
961 Vanreusel, A., Andersen, A. C., Boetius, A., Connelly, D., Cunha, M. R., Decker, C., Hilario,
962 A., K.A., K., Maignien, L., Olu, K., Pachiadaki, M., Ritt, B., Rodrigus, C., Sarrazin,
963 J., Tyler, P., Van Gaever, S., & Vanneste, H. (2009). Biodiversity of cold seep
964 ecosystems along the European margins. *Oceanography, 22*(1), 110-127.
965
966 Vaser, R., Sovic, I., Nagarajan, N., & Sikic, M. (2017, May). Fast and accurate de novo
967 genome assembly from long uncorrected reads. *Genome Res, 27*(5), 737-746.
968 https://doi.org/10.1101/gr.214270.116
969
970 Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A.,
971 Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: an integrated tool

for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, *9*(11), e112963. https://doi.org/10.1371/journal.pone.0112963

Xiang, Y., Yan, C., Guo, X., Zhou, K., Li, S., Gao, Q., Wang, X., Zhao, F., Liu, J., Lee, W. H., & Zhang, Y. (2014, May 6). Host-derived, pore-forming toxin-like protein and trefoil factor complex protects the host against microbial infection. *Proc Natl Acad Sci U S A*, *111*(18), 6702-6707. https://doi.org/10.1073/pnas.1321317111

Yancey, P. H., & Siebenaller, J. F. (2015, Jun). Co-evolution of proteins and solutions: protein adaptation versus cytoprotective micromolecules and their roles in marine organisms. *J Exp Biol*, *218*(12), 1880-1896. https://doi.org/10.1242/jeb.114355

Yang, L. A., Chang, Y. J., Chen, S. H., Lin, C. Y., & Ho, J. M. (2019, Apr 18). SQUAT: a Sequencing Quality Assessment Tool for data quality assessments of genome assemblies. *Bmc Genomics*, *19*(Suppl 9), 238. https://doi.org/10.1186/s12864-019-5445-3

Yuen, B., Bayes, J. M., & Degnan, S. M. (2014, Jan). The Characterization of Sponge NLRs Provides Insight into the Origin and Evolution of This Innate Immune Gene Family in Animals. *Mol Biol Evol*, *31*(1), 106-120. https://doi.org/10.1093/molbev/mst174

Zachos, J., Flower, B., & Paul, H. (1997). Orbitally paced climate oscillations across the Oligocene/Miocene boundary. *Nature*, *388*, 567-570. https://doi.org/https://doi.org/10.1038/41528

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001, Apr 27). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, *292*(5517), 686-693. https://doi.org/10.1126/science.1059412

Zamora, S., Lefebvre, B., Álvaro, J. J., Clausen, S., Elicki, O., Fatka, O., Jell, P., Kouchinski, A., Lin, J.-P., Nardin, E., Parsley, R., Rozhnov, S., Sprinkle, J., Sumrall, C. D., Vizcaïno, D., & Smith, A. B. (2013). Global Cambrian echinoderm diversity and palaeobiogeography. In D. A. T. Harper & T. Servais (Eds.), *Early Palaeozoic biogeography and palaeogeography* (Vol. 38, pp. 151-164). Geological Society.

Zhang, D., Gao, F., Jakovlic, I., Zou, H., Zhang, J., Li, W. X., & Wang, G. T. (2020, Jan). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources*, *20*(1), 348-355. https://doi.org/10.1111/1755-0998.13096

Zhang, Q., Zmasek, C. M., & Godzik, A. (2010, May). Domain architecture evolution of pattern-recognition receptors. *Immunogenetics*, *62*(5), 263-272. https://doi.org/10.1007/s00251-010-0428-1

Zhang, X. J., Sun, L. N., Yuan, J. B., Sun, Y. M., Gao, Y., Zhang, L. B., Li, S. H., Dai, H., Hamel, J. F., Liu, C. Z., Yu, Y., Liu, S. L., Lin, W. C., Guo, K. M., Jin, S. J., Xu, P., Storey, K. B., Huan, P., Zhang, T., Zhou, Y., Zhang, J. Q., Lin, C. G., Li, X. N., Xing, L. L., Huo, D., Sun, M. Z., Wang, L., Mercier, A., Li, F. H., Yang, H. S., & Xiang, J. H. (2017, Oct). The sea cucumber genome provides insights into morphological evolution and visceral regeneration. *Plos Biology*, *15*(10). https://doi.org/ARTN e2003790 10.1371/journal.pbio.2003790

Zhong, Y. F., Butts, T., & Holland, P. W. (2008, Sep-Oct). HomeoDB: a database of homeobox gene diversity. *Evol Dev*, *10*(5), 516-518. https://doi.org/10.1111/j.1525-142X.2008.00266.x

Data Accessibility

Raw reads and genome assembly are accessible in NCBI under BioProject number PRJNA752986. Assembled genome sequences are accessible under Whole Genome Shotgun project number JAIGNY000000000. Raw reads and genome assembly are also available at the CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with accession number CNP0002134. The genome assembly and related annotation files are available at Figshare (https://doi.org/10.6084/m9.figshare.15302229).

Author Contributions

M.W and J.G.H. conceived of the project and designed research; J.H. collected the sample; P.T., L.Z, Y.M., Q.C., Q.Z., L.Z. assembled and annotated the genome; L.Z., Z.G., J.H., M.W., S.Q., Y.W. performed the evolutionary analyses; M.W., G.H. wrote the paper the manuscript with the contribution from all authors.

1045 **Table 1 Genome assembly statistics of *C. heheva* and *A. japonicus***

1046

	C. heheva	*A. japonicus* (Li et al., 2018)	*A. japonicus* (Zhang et al., 2017)
Estimated genome size (Gb)	~1.23	~1.0	~1.0
Assembled genome size (bp)	1,106,937,276	952,279,490	804,993,085
Number of contigs	4,609	21,303	7,058
Contig N50 (bp)	1,221,604	45,411	190,269
Scaffold N50 (bp)	-	195,518	486,650

1047

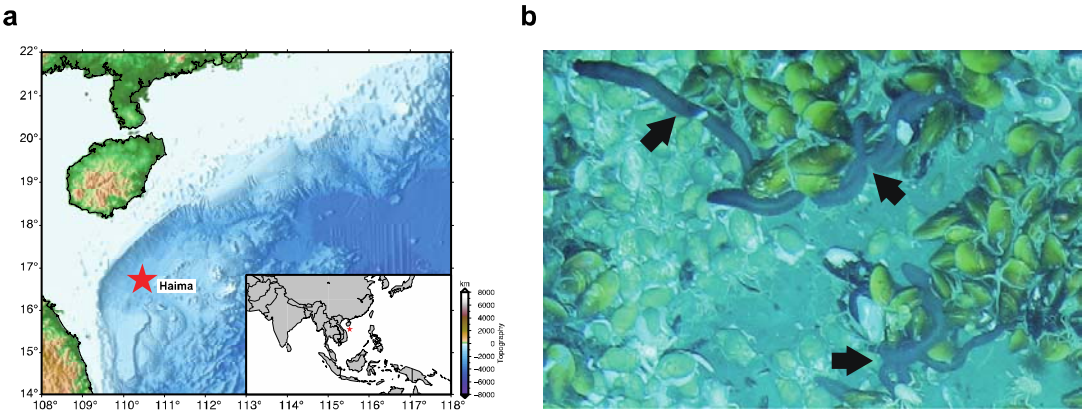


Figure 1. Collection of *C. heheva*. (a) Map showing the sampling site at the Haima cold seep of northern South China Sea (16° 73.228’ N, 110° 46.143’ E). (b) *C. heheva* at the sampling site (depth: 1,385 m), where they cohabit with deep-sea mussels. *C. heheva* individuals are indicated by black arrows.

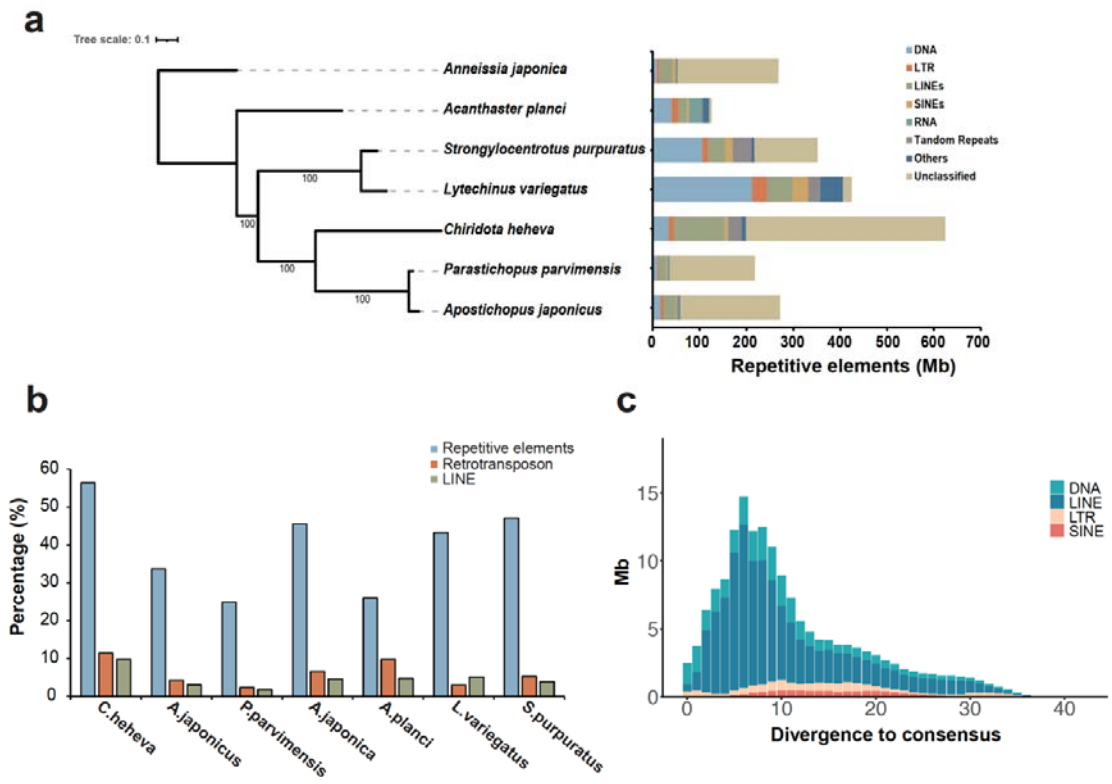


Figure 2. Landscape of transposable elements in echinoderm genomes. (a) Comparison of the occurrence and composition of repetitive elements in the genomes of 7 echinoderms. (b) Comparison of the proportion of repetitive elements, retrotransposon, and long interspersed nuclear elements (LINEs) in the genomes of 7 echinoderms. The proportions of repetitive elements and LINEs are higher in the genome of *C. heheva* than that in other echinoderms. (c) Transposable element accumulation profile in *C. heheva* genome. A recent burst of LINEs was observed in *C. heheva*.

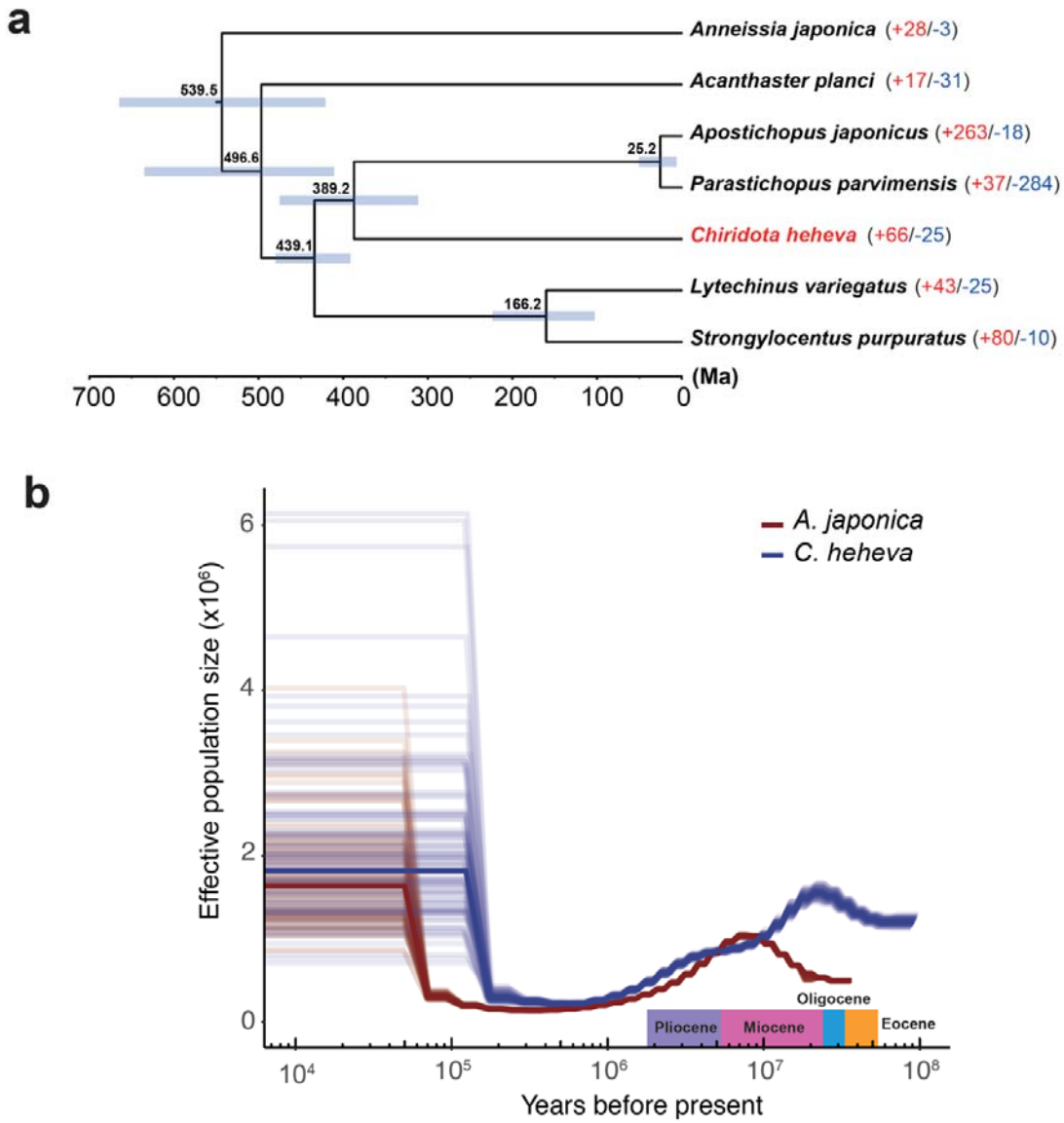


Figure 3. Evolutionary history of *C. heheva*. (a) A species tree of 7 echinoderm species. In total, 988 single-copy orthologs were used to reconstruct the phylogenetic tree. The divergence time between species pairs was listed above each node, and 95% confidence internal of the estimated divergence time was denoted as blue bar. The numbers of protein families that were significantly expanded (red) of contracted (blue) ($P < 0.05$) in each species are denoted beside the species names. (b) Demographic history of *C. heheva* (blue) and *A. japonicus* (red). The changes of ancestral population size of *C. heheva* and *A. japonicus* were inferred using the PSMC method. Time in history was estimated by assuming a generation time of 3 years and a mutation rate of 1.0×10^{-8}.

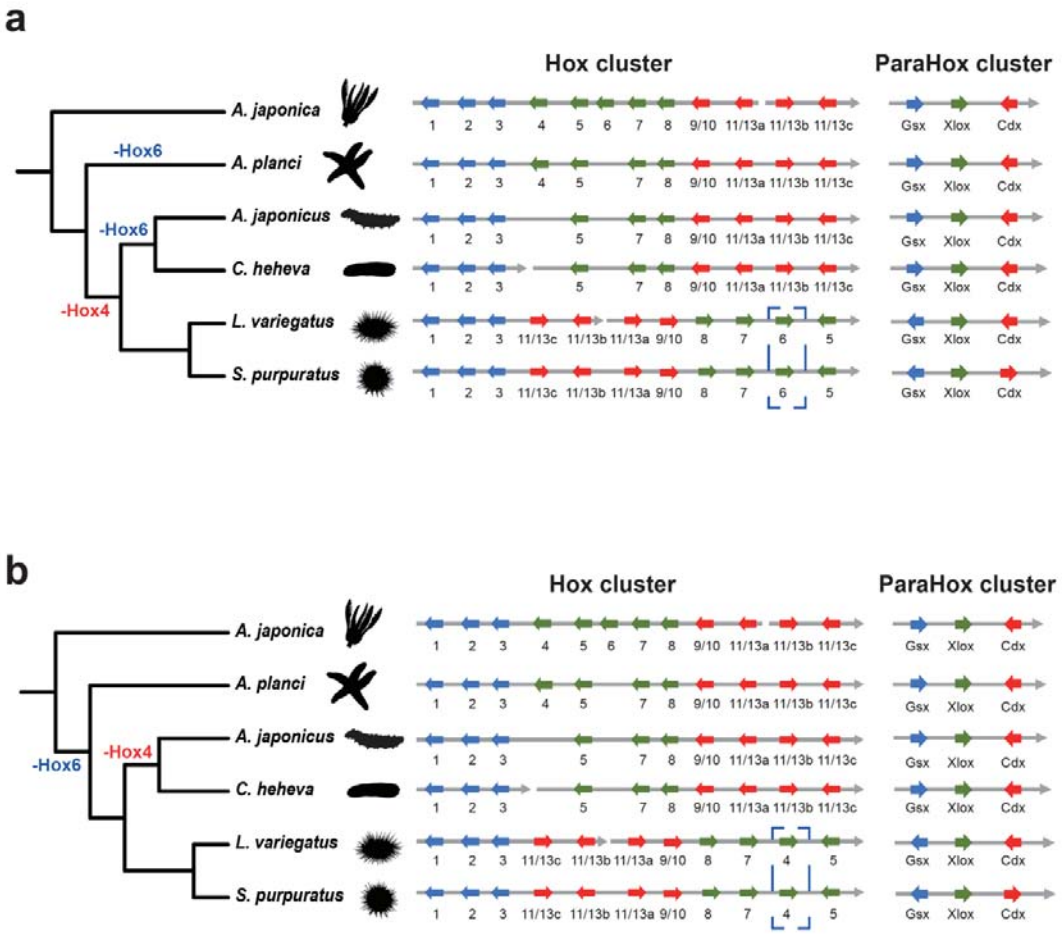
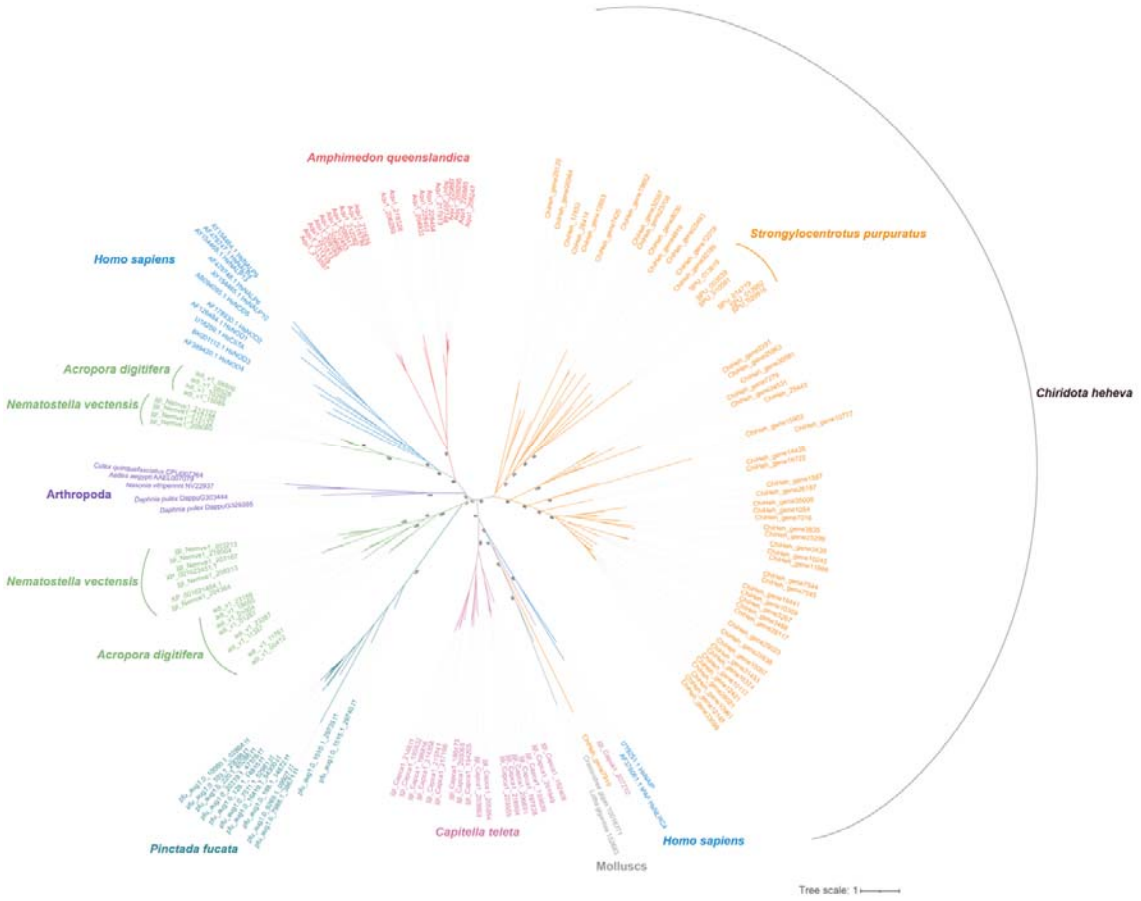


Figure 4. Genomic organization of *Hox* and *ParaHox* gene clusters in echinoderms. *Hox* and *ParaHox* genes are indicated by arrows. The gene composition and orientation of *Hox* and *ParaHox* clusters are consistent between two holothurians (*C. heheva*, *A. japonicus*). (a) The organization of *Hox* genes in echinoderms if *Hox4* is lost in echinoids (*L. variegatus*, *S. purpuratus*). (b) The organization of *Hox* genes in echinoderms if *Hox6* is lost in echinoids.

1084



1085

1086 **Figure 5. Evolutionary relationships among *C.heheva* NLRs and other representative**
1087 **metazoan NLRs.** The unrooted phylogenetic tree was reconstructed based on the NACHT
1088 domain sequences using a maximum likelihood method. The values near the nodes are
1089 ultrafast bootstrap (UFBoot) values. NLRs from different types of species are highlighted by
1090 branches of different colors. The species name is shown near the corresponding lineage.

1091

Figure 6. Evolutionary relationship with aerolysin-like proteins (ALPs) from *C. heheva* and other species. The unrooted phylogenetic tree was reconstructed using a maximum likelihood method. The values near the nodes are ultrafast bootstrap (UFBoot) values. ALPs from different types of species are highlighted by branches of different colors. The species name is shown near the corresponding lineage. ALPs from *C. heheva* do not cluster with ALPs from other echinoderms (*A. japonicus*, *P. parvimensis*), but with the ones from sea anemones (*N. vectensis*, *E. diaphana*).