

A high-quality chromosome-level genome assembly of a deep-sea sea cucumber *Molpadia* sp. (Molpadiidae)

Received: 18 December 2025

Accepted: 22 July 2026

Published online: 03 August 2026

Cite this article as: Wang, J., Xiao, N., Duan, Z. *et al.* A high-quality chromosome-level genome assembly of a deep-sea sea cucumber *Molpadia* sp. (Molpadiidae). *Sci Data* (2026). <https://doi.org/10.1038/s41597-026-07981-y>

Jing Wang, Ning Xiao, Zelin Duan, Zhiyuan Gong & Nansheng Chen

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

A high-quality chromosome-level genome assembly of a deep-sea sea cucumber *Molpadia* sp. (Molpadiidae)

Jing Wang^{1,2,3}, Ning Xiao^{4,5}, Zelin Duan^{1,3,6}, Zhiyuan Gong^{1,3}, Nansheng Chen^{1,2,3,✉}

¹Laboratory of Marine Ecology and Environmental Sciences, Shandong Province Key Laboratory of Marine Biodiversity and Bio-resource Sustainable Utilization, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China.

²Laboratory for Marine Ecology and Environmental Science, Qingdao Marine Science and Technology Center, Qingdao 266237, China.

³Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China.

⁴Laboratory of Marine Organism Taxonomy and Phylogeny, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China.

⁵Marine Biological Museum, Chinese Academy of Sciences, Qingdao 266071, China.

⁶State Key Laboratory of Mariculture Biobreeding and Sustainable Goods, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China

✉e-mail: chenn@qdio.ac.cn

Abstract

Molpadida is one of eight orders in the class Holothuroidea whose phylogenetic positioning has been debated. The absence of a high-quality Molpadida genome has hindered a thorough understanding of their taxonomy and evolution. In this study, we reported the chromosome-level genome assembly of *Molpadia* sp. C3 from South China Sea, representing the first Molpadida species whose chromosome-level genome assembly became available. The *Molpadia* sp. C3 genome assembly has a total length of 2.53 Gb with a contig N50 length of 2.8 Mb and a scaffold N50 length of 107.8 Mb, encompassing 23 chromosomes, highlighting high genome continuity. Benchmarking Universal Single-Copy Orthologs (BUSCO) assessment yielded a completeness estimate of 87.4% for this assembly. We predicted a total of 38,311 protein-coding genes (PCGs), with 32,584 functionally annotated. The availability of this high-quality chromosome-level genome of *Molpadia* sp. C3 contributes to our understanding of the origin and evolution of holothuroids.

Background & Summary

Molpadida Haeckel, 1896 is an order of the class Holothuroidea, which includes eight orders¹. Species in Molpadida, which have been found both in shallow waters and in deep sea^{2,3}, have some special morphological characters, distinct from other sea cucumbers. The body shape of Molpadida species is fusiform, and their hind body is narrowed to form a distinct tail. Although they possess tentacles around the mouths derived from the water vascular systems, they have no true tube feet. Since holothuroids have a suite of homoplastic traits including ossicles, tentacles, and tube feet, and the current study demonstrates that some of these traits have limited utility in systematics, the taxonomic status of Molpadida thus remains unresolved and requires further investigation. Based on the digitiform

to digitate tentacles and plated (imbricate) body walls, Pawson and Fell identified that Molpadida were more related with Apodida, as a subclass Apodacea⁴. Miller *et al* adopted a broad-scale molecular phylogenetic analysis of sequences from mitochondrial (*COI*, 16S rDNA, 12S rDNA) and nuclear (*H3*, 18S rDNA, 28S rDNA) sources, and revealed that Molpadida was clustered with Persiculida and Synallactida, but the support for the clade of Molpadida + Synallactida + Persiculida was low⁵. Based on the phylogenetic analysis of nucleotide sequences of 13 protein coding genes (PCGs) in mitochondrial genomes of sea cucumbers from 6 orders, it was discovered that the order Apodida was placed in a well-supported basal branch, and Molpadida was not recovered as a sister group to Synallactida, instead appeared as a poorly supported subgroup within Synallactida⁶. In addition, bioluminescence is common seen in deep-sea sea cucumbers and other benthic animals⁷. Luminous holothuroids have been found only in the subclass Actinopodae distributed among 5 orders, including Molpadida⁸⁻¹³. Ancestral state reconstruction analysis suggested six evolutionary origins in the five orders, however, the polyphyletic relation of *Molpadia* species in the order Molpadida makes it difficult for an accurate estimation of the evolution of bioluminescence¹³. Meanwhile, research progress in development, nutritional biology, and biological composition of Molpadida species have also been made¹⁴⁻¹⁷. According to the structural elucidation of chilensosides A–D—the glycosides from *Paracaudina chilensis* (Caudinidae, Molpadida), it was demonstrated that their structural similarity to the glycosides isolated from different representatives of the order Dendrochirotida, indicating the closeness of Molpadida to Dendrochirotida³. It is expected that genome analysis may provide better resolution for resolving the taxonomic status of Molpadida.

By now, chromosome-level genomes have been constructed from species of five orders (Dendrochirotida, Synallactida, Holothuriida, Persiculida, and Apodida) in the class Holothuroidea¹⁸⁻²⁵. Molpadida is one of the three orders whose genomes have not been constructed. In this study, the chromosome-level genome of *Molpadia* sp. C3, which is a species of the family Molpadiidae of Molpadida, was constructed using combined PacBio long-read sequencing technology and Hi-C sequencing data-based genome assembly technology. The genome size of *Molpadia* sp. C3 was 2.53 Gb with high continuity, distributed in 23 chromosomes. This availability of this genome of the first species of the order Molpadida will facilitate understanding the taxonomic status of Molpadida, and discovering the details of evolution characters of holothuroids.

Method

Sample collection

The sample *Molpadia* sp. C3 used in this study was collected from the South China Sea (117°49'48"E, 11°58'05"N, 3066 m deep) using the bottom trawling during the cruise by KK2101 cruise (R/V Jiageng) on May 13, 2021 (Fig. 1). Once the sample was brought to the upper deck, it was immediately frozen with liquid nitrogen and then stored at -80°C until for further genomic studies.

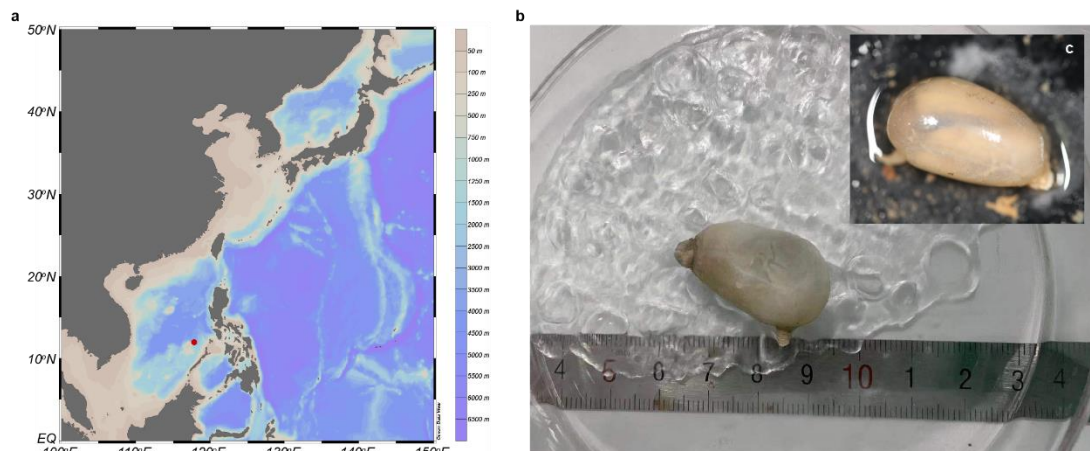


Fig. 1 Illustration of the sampling site and morphology of *Molpadia* sp. C3.

Nucleic acid extraction, library preparation, and sequencing

The animal sample was dissected in liquid nitrogen, and a portion of the muscle was used for DNA and RNA extraction. High-quality genomic DNA was extracted using a modified CTAB method²⁶. To construct DNA short-read library, we used Covaris S220 to shear extracted DNA (1 ug) to 200 bp as the input material for the library construction. The sequencing library was generated using the VAHTS Universal DNA Library Prep Kit for MGI (Vazyme, Nanjing, China) following the manufacturer's instruction. The library was quantified using Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and sequenced on DNBSEQ-T7 platform (BGI, China)²⁷, generating 100.62 Gb (39.77×) of 150 bp paired-end reads (Table 1). To construct the SMRTbell library, we used the SMRTbell Express Template Prep kit 2.0 (Pacific Biosciences). The library was loaded at an on-plate concentration of 120 pM using diffusion loading. SMRT sequencing was performed using a single 8M SMRT Cell on the Sequel II System with Sequel II Sequencing Kit, generating 91.46 Gb of PacBio CCS reads with a coverage depth of 36.15 × (Table 1). A Hi-C library was constructed according to previous studies²⁸, and sequenced on DNBSEQ-T7 platform (BGI, China). The 205.85 Gb raw sequence data were filtered to obtain 192.82 Gb clean data, with Q20=98.47%, which were used for chromosomal level genome assembly (Table 1).

For RNA-seq, the total RNA of muscle tissue was extracted using Trizol reagent (Invitrogen, CA, USA). To construct the library for RNA-seq, we used Covaris S220 to shear extracted RNA (1 ug per sample) to 200 bp. The mRNA was purified from the total RNA using poly(T) oligo attached magnetic beads. Sequencing libraries were generated using the Hieff NGS® Ultima Dual-mode mRNA Library Prep Kit (Yeasten, Shanghai, China) following the manufacturer's recommendations and index codes were added to attribute sequences to the sample. The library quantification was measured using Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). Subsequently, sequencing was performed on the MGI-2000 platform (BGI, China).

Genome size estimation

The short-reads from DNBSEQ-T7 platform were quality filtered by SOAPnuke v2.1.0²⁹. The quality filtered reads were used for genome size estimation. we generated the 19-mer occurrence distribution of sequencing reads from the DNA short-read library using the k -mer method. The occurrences of k -mers were counted by jellyfish 2.1.4³⁰. The characteristics of the genome were estimated using GenomeScope 2.0^{31,32}. The genome size, heterozygosity rate and repeat content of *Molpadia* sp. C3 were 2.29 Gb, 2.45%, and 76.45%, respectively (Fig. 2).

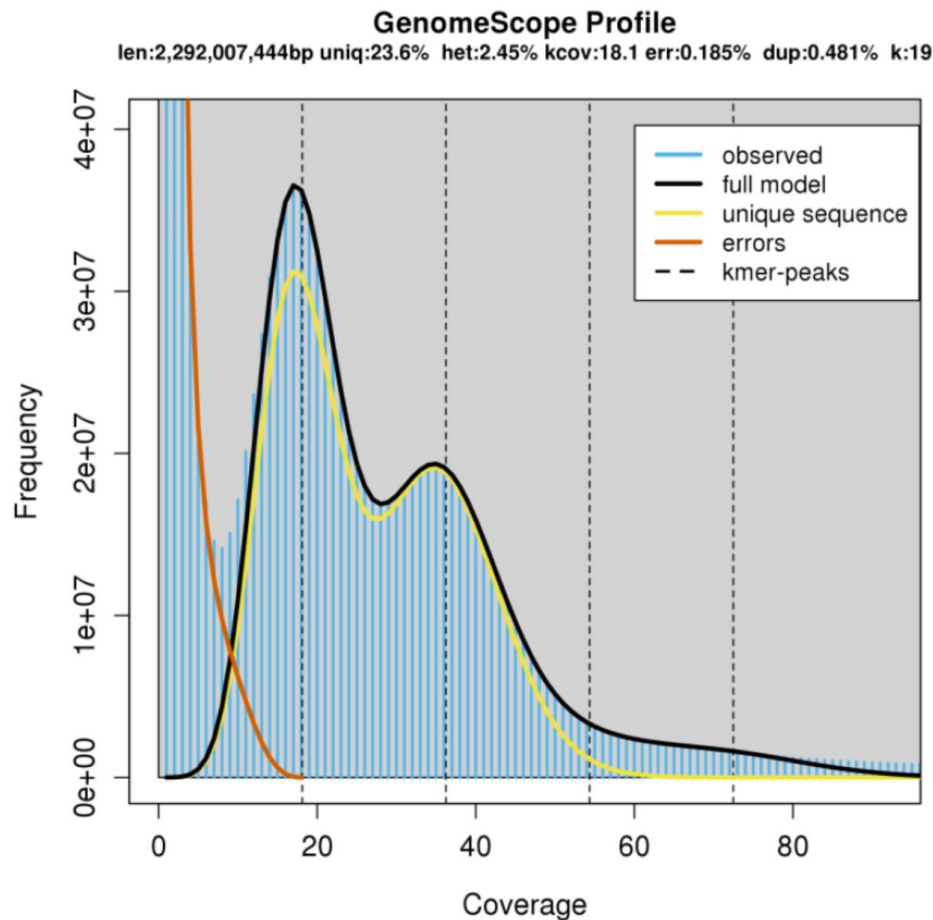


Fig. 2 The k -mer (19-mer) distribution curves for estimation of the genome size, heterozygosity rate, and repeat content.

Genome assembly

For preliminary assembly, SMRTbell libraries were sequenced on a PacBio Sequel II system and consensus reads (HiFi reads) were generated using ccs v4.0.0 (<https://github.com/pacificbiosciences/unanimity>) with the parameter '-minPasses 3'. To further validate and improve the assemblies, we generated 91.46 Gb (36.15×) PacBio HiFi reads (Table 1). These long (~15 kb) and highly accurate (>99%) HiFi reads were assembled using hifiasm v0.16.1 with the parameter '-l 3'³³. For anchored contigs, the clean reads generated from the Hi-C library and were mapped to the *Molpadia* sp. C3 preliminary assembly using Juicer v1.5.6 and 3D-DNA v180922. The Hi-C heatmap was visualized using Juicebox 1.11.08³⁴. The genome landscape illustrating the

length, repeat element density, gene density and GC content was created by circos-0.69-9³⁵. The assembled genome of *Molpadia* sp. C3 consisted of 23 chromosomes (Fig. 3; Table 2), with a contig anchoring rate of 93.69%. The assembled *Molpadia* sp. C3 genome has a total length of 2.53 Gb with a contig N50 length of 2.78 Mb and a scaffold N50 length of 107.84 Mb (Table 3).

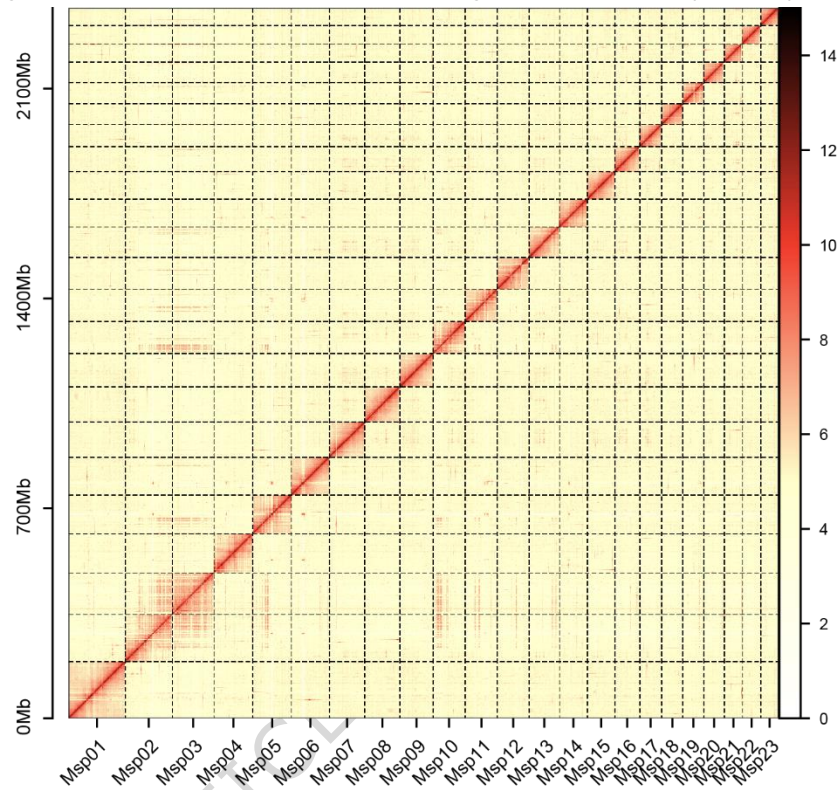


Fig. 3 Hi-C analysis of *Molpadia* sp. C3 genome contigs. Chromosomes are arranged in the size order from left to right and from bottom to top. The color bar illuminates the logarithm of the contact density from red (16) to white (0) in the plot.

Table 1. Summary of sequencing data for *Molpadia* sp. C3 assembly.

Technology	Total data (Gb)	Reads number	Depth (×)
PacBio HiFi	91.46	6,225,556	36.15
DNBSEQ-T7	100.62	670,769,794	39.77
Hi-C	192.82	1,285,478,358	76.21
RNA-seq	7.22	24,078,529	2.85

Table 2. The length and contig number of chromosomes in *Molpadia* sp. C3 genome.

Chromosome	Length (bp)
Msp01	188,645,477
Msp02	157,228,960
Msp03	138,887,095
Msp04	130,120,612
Msp05	128,696,844

Msp06	126,228,028
Msp07	118,587,980
Msp08	116,917,259
Msp09	110,371,485
Msp10	107,839,791
Msp11	107,284,356
Msp12	105,757,129
Msp13	101,784,798
Msp14	92,716,339
Msp15	91,523,847
Msp16	82,940,393
Msp17	74,118,355
Msp18	70,476,329
Msp19	69,615,592
Msp20	67,966,936
Msp21	61,192,432
Msp22	61,052,007
Msp23	58,344,515

Table 3. Assembly statistics of *Molpadia* sp. C3 genome assembly

Sample ID	Contig length	Scaffold length	Contig number	Scaffold number
Total length (bp)	2,526,520,357	2,527,727,857	2,477	32
Max length (bp)	15,788,927	188,645,477	-	-
Number $\geq$ 10000 bp			2,443	32
Number $\geq$ 20000 bp	-	-	2,377	32
N50	2,778,330	107,839,791	287	10
N60	2,232,148	105,757,129	388	12
N70	1,784,782	91,523,847	515	15
N80	1,290,398	72,759,105	681	18
N90	744,559	61,192,432	931	22
GC content (%)	39.55	39.53	-	-
Anchored rate (%)		93.69%		

Genome annotation

Repeats annotation

Repeat sequences of the *Molpadia* sp. C3 genome were identified by homology alignment and *de novo* prediction. For homology-based analysis, known transposable elements (TEs) within the *Molpadia* sp. C3 genome were identified according to the Repbase TE library^{36,37} employing RepeatMasker (v4.0.7)³⁸ and RepeatProteinMask (v4.0.7). For *de novo* prediction, we constructed a *de novo* repetitive elements library of the *Molpadia* sp. C3 genome using RepeatModeler (v2.0.2) and

LTR_FINDER³⁹, and then the library was supplied to RepeatMasker to identify the repeat contents³⁸. We also identified tandem repeats using the Tandem Repeat Finder (TRF, v4.10.0) package⁴⁰. The majority (63.70%) of the *Molpadia* sp. C3 genome was estimated to be repetitive elements (Table 4). The top 3 most frequent categories of repetitive elements in the *Molpadia* sp. C3 genome were Class I/LTR (27.29%), Class I/LINE (21.24%) and Class I/SINE (9.02%). The distribution of repetitive elements was highly uneven, with regional content of the repetitive elements in the *Molpadia* sp. C3 genome, with peaks occurring at regions with low gene density (Fig. 4). In addition, the LINE transposons were more abundant in the chromosomes Msp02 and Msp03 (Fig. 4).

Table 4. Summary statistics of repetitive sequences in *Molpadia* sp. C3 genome.

Type	Length(bp)	Percentage (%)
Class I/LTR/Copia	11,895,229	0.47
Class I/LTR/Gypsy	126,383,004	5.00
Class I/LTR/Other	551,636,756	21.82
Class I/SINE	227,911,161	9.02
Class I/LINE	536,948,286	21.24
Class I/Other	4,764	0.00
Class II/EnSpm	36,717,150	1.45
Class II/Harbinger	8,547,275	0.34
Class II/hAT	93,016,623	3.68
Class II/Helitron	21,949,446	0.87
Class II/Mariner	126,614	0.01
Class II/MuDR	10,049,580	0.40
Class II/P	4,939,648	0.20
Class II/Other	94,505,917	3.74
Other	181,137,886	7.17
Unknown	3,886,113	0.15
Total	1,610,027,158	63.70

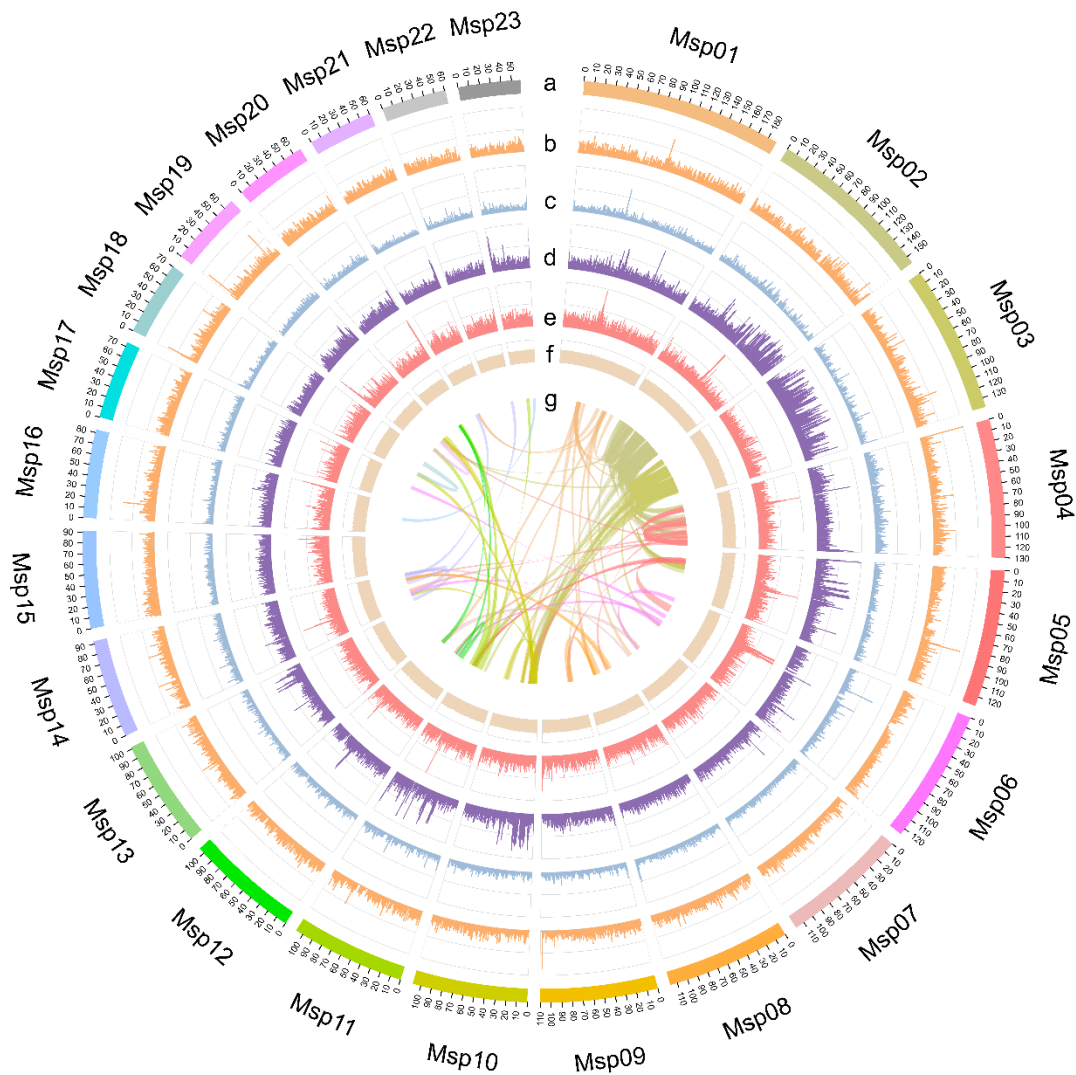


Fig. 4 The genomic landscape of *Molpadia* sp. C3 (from outer to inner circles): a. the 23 chromosomes; b. gene density; c. DNA transposon density; d. LINE transposon density; e. LTR transposon density; f. GC density; and g. a colinear relationship across the genome, drawn in 1 Mb non-overlapping windows.

Protein coding gene annotation

Protein coding gene (PCG) structures were predicted using three basic strategies: de novo, homology-based, and transcriptome sequencing-based prediction. For homology-based gene prediction, gene sets from five echinoderms *Acanthaster planci*⁴¹, *Apostichopus japonicus*¹⁹, *Chiridota heheva*²², *Hansenothuria* sp. M1226 (unpublished), and *Strongylocentrotus purpuratus*⁴² were used. GeMoMa (v1.8) and Genewise (v2.4.1) were used to align the protein sequences of *Molpadia* sp. C3 to the five other species for homology-based gene prediction. For RNA-seq-aided gene prediction, we first mapped short RNA reads to the reference genome using HISAT2 (v2.2.1), and then assembled transcripts using StringTie (v2.2.1) with the default parameters. The transcripts were used to predict

Open Reading Frames (ORFs) using TransDecoder (v5.5.0). Based on the genome sequence, we used Augustus (v3.4.0)⁴³⁻⁴⁵ to perform *de novo* gene prediction. Finally, EvidenceModeler (EVM, v1.1.1) was applied for combining predicted results from three strategies above. A total number of 38,311 PCGs were annotated in the *Molpadia* sp. C3 genome (Table 5). The average gene length, average coding sequence (CDS) length, average exon length, and average intron length were 26,693 bp, 1,667 bp, 317 bp, and 5,879 bp, respectively.

Table 5. Statistics of PCG structure prediction in *Molpadia* sp. C3 genome.

Type	Gene Set	Number	Average transcript length(bp)	Average CDS length(bp)	Average exons per gene	Average exon length(bp)	Average intron length(bp)
De novo	Augustus	43,961	21,451.83	1,736.78	5.22	333.06	4,677.85
	<i>Acanthaster planci</i>	8,458	75,217.38	1,698.90	9.85	172.41	8,303.43
	<i>Apostichopus japonicus</i>	23,712	33,860.95	1,258.81	5.00	251.88	8,155.26
	<i>Chiridota heheva</i>	11,827	18,341.00	1,121.18	3.80	294.80	6,143.00
	<i>Hansenothuria</i> sp. M1226	13,111	22,600.84	2,056.35	8.48	242.41	2,745.53
Homolog	<i>Strongylocentrotus purpuratus</i>	10,776	48,641.56	1,670.86	8.26	202.23	6,467.87
RNAseq	transcripts	19,173	32,485.11	925.69	4.91	531.73	7,641.78
Final	-	38,311	26,693.41	1,666.66	5.26	317.22	5,879.68

Gene functions were assigned according to the best match by aligning the protein sequences to the Swiss-Prot⁴⁶ using Blastp (v2.2.26, with a threshold of E-value $\leq 1e^{-5}$). The motifs and domains were annotated using InterProScan (v5.59-91.0)⁴⁷ by searching against publicly available databases Pfam⁴⁸. The Gene Ontology (GO) IDs for each gene were assigned according to the corresponding InterPro entry. We predicted the proteins function by transferring annotations from the closest BLAST hit (E-value $<10^{-5}$) in the Swiss-Prot⁴⁶ database and DIAMOND (v2.1.9) hit (E-value $<10^{-5}$) / BLAST hit (E-value $<10^{-5}$) in the National Center for Biotechnology Information (NCBI) Non-Redundant (NR). We also mapped gene set to a KEGG pathway and identified the best match for each gene. Of these PCGs, 32,584 (85.05%) had functional annotations assigned using comparisons to public databases, including NCBI nr, Swiss-Prot⁴⁶, KEGG⁴⁹, KOG, TrEMBL⁴⁶, InterPro and Gene Ontology⁵⁰ (Table 6), supporting the high quality of the genome.

Non coding RNA gene annotation

We used tRNAscan-SE (v1.3.1) algorithms⁵¹ with default parameters to identify the genes associated with tRNA. For rRNA identification, we selected the invertebrate_rRNA database from BLASTN as references, and used Blast (v2.2.26) to search for rRNA. MiRNAs and snRNAs were identified by

cmsearch (v1.1.4) against the Rfam (v14.1) database⁴⁸ with default parameters. All the ncRNA annotation were summarized in Table 7.

Table 6. Statistics of gene function annotation.

Type	Number	Percentage (%)
Total	38,311	100.00
NR	30,832	80.48
Swissport	17,586	45.90
KEGG	20,525	53.57
KOG	16,870	44.03
TrEMBL	30,533	79.70
InterPro	25,751	67.22
GO	15,685	40.94
Annotated	32,584	85.05
Unannotated	5,727	14.95

Table 7. Statistics of ncRNA annotation in *Molpadia* sp. C3 genome

Type		Copy	Average length (bp)	Total length(bp)	% of genome
miRNA	miRNA	14	81.71428571	1144	0.000045
tRNA	tRNA	4827	73.81499896	356305	0.014096
rRNA	rRNA	1137	108.826737	123736	0.004895
	18S	135	248.6222222	33564	0.001328
	28S	25	53.28	1332	0.000053
	5.8S	0	0	0	0
	5S	977	90.93142272	88840	0.003515
snRNA	snRNA	2091	138.278814	289141	0.011439
	CD-box	31	123.0967742	3816	0.000151
	HACA-box	2	265	530	0.000021
	splicing	2058	138.3843537	284795	0.011267

Genome quality assessment

To examine the assembly integrity, the assembled genome and predicted PCGs were subjected to BUSCO (v5.1.2)⁵² using the metazoan_db10 database to evaluate the completeness of the genome,

respectively. We found that 87.4%-87.6% of the core genes used by BUSCO were identified as full length in the *Molpadia* sp. C3 genome assembly (Table 8). We further tested the completeness of the predicted PCGs using BUSCO, which identified 85.12% of the core genes as full length.

Table 8. Completeness assessment of *Molpadia* sp. C3 by BUSCO

Type	Contig		Scaffold		Gene set	
	Number of genes	Percentage	Number of genes	Percentage	Number of genes	Percentage
Complete BUSCOs (C)	836	87.6%	834	87.4%	812	85.12%
Complete and single-copy BUSCOs (S)	734	76.9%	809	84.8%	748	78.41%
Complete and duplicated BUSCOs (D)	102	10.7%	25	2.6%	64	6.71%
Fragmented BUSCOs (F)	36	3.8%	32	3.4%	81	8.49%
Missing BUSCOs (M)	82	8.6%	88	9.2%	61	6.39%
Total BUSCO groups	954		954		954	

Gene family and phylogenomic analyses

Gene family analysis was conducted using OrthoFinder (v2.5.5)⁵³ to examine the orthologous genes shared by *Molpadia* sp. C3 and other related species including *A. japonicus*¹⁹, *Stichopus monotuberculatus*²⁰, *C. heheva*²², *Holothuria scabra*²³, *S. purpuratus*⁴³, *Amphiura filiformis*⁵⁴, *Asterias amurensis*⁵⁵, *Glyptometra* sp. CNS01629⁵⁶, and *Branchiostoma floridae*⁵⁷. Based on these 10 species, a total number of 1,338 single-copy orthologous genes were obtained for phylogenetic analysis. First, multiple alignments of protein sequences were conducted by MUSCLE (version 3.8.31)⁵⁸; then, the CDS alignment was constructed based on the protein alignments. The nucleotides at position 2 (phase 1) of each codon were extracted to construct the phylogenetic tree using RAxML (version 8.2.13)⁵⁹ with the model "GTRGAMMA.", with *B. floridae* as the outgroup. The divergence times were estimated MCMCTree in PAML (v4.9j)⁶⁰ and we calibrated the model using the divergence times between *B. floridae* – *A. japonicus* (590 Ma–896 Ma), *A. filiformis* – *S. purpuratus* (461 Ma–600 Ma), *S. purpuratus*

– *H. scabra* (371 Ma–464 Ma), and *H. scabra*-*A. japonicus* (158 Ma–308 Ma) obtained from the TimeTree Website. According to the phylogenetic tree, *Molpadia* sp. C3 and the other sea cucumbers clustered in a single clade, and *Molpadia* sp. C3 diverged from *A. japonicus* and *S. monotuberculatus* approximately 139 million years ago (mya) (Fig. 5).

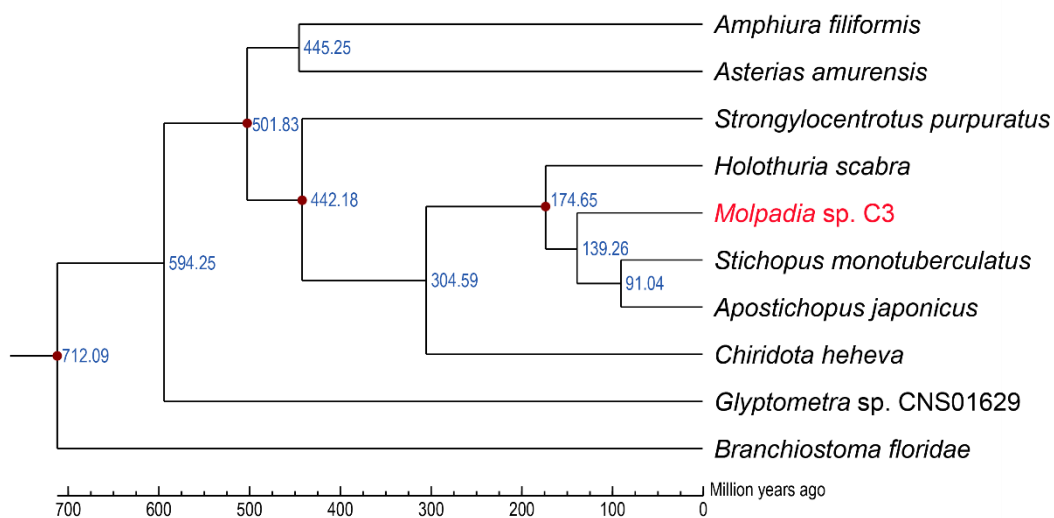


Fig. 5 Phylogenetic analysis between *Molpadia* sp. C3 and other related species.

Data Records

The data that support the findings of this study are openly available in public repositories National Genomics Data Center (NGDC)⁶¹, NCBI and figshare. All raw sequencing data of the short-read, PacBio, Hi-C, and RNA-seq sequencing data for *Molpadia* sp. C3 were deposited at NGDC under the BioProject accession number PRJCA051471. The raw short-reads, PacBio, Hi-C, and RNA-seq sequencing data are available in the Genome Sequence Archive (GSA)⁶² with the accession numbers CRR2351251⁶³, CRR2351252-CRR2351255⁶⁴⁻⁶⁷, CRR2351256-CRR2351259⁶⁸⁻⁷¹, and CRR2351260⁷², respectively. The final genome assembly data and annotation files of repeats annotation and PCG annotation are available in NCBI⁷³ and figshare⁷⁴, respectively.

Technical Validation

To evaluate the assembly completeness and integrity, we implemented two complementary validation approaches. First, the completeness of *Molpadia* sp. C3 genome assembly was evaluated using the BUSCO⁵² (in the metazoa_odb10 database), and the completeness was 87.4% (84.8% single-copied genes and 2.6% duplicated genes) (Table 8). The Hi-C heatmap revealed a well-structured interaction pattern in and around the chromosome inversion regions (Fig. 3).

Data Availability

The data that support the findings of this study are openly available in public repositories NGDC, NCBI and figshare. All raw sequencing data of the short-read, PacBio, Hi-C, and RNA-seq sequencing data

for *Molpadia* sp. C3 were deposited at NGDC under the BioProject accession number PRJCA051471. The raw short-reads, PacBio, Hi-C, and RNA-seq sequencing data are available in the Genome Sequence Archive (GSA)⁶² with the accession numbers CRR2351251⁶³, CRR2351252-CRR2351255⁶⁴⁻⁶⁷, CRR2351256-CRR2351259⁶⁸⁻⁷¹, and CRR2351260⁷² (<https://ngdc.cncb.ac.cn/gsa/browse/CRA033897>), respectively. The final genome assembly data and annotation files of repeats annotation and PCG annotation are available in NCBI (<https://identifiers.org/ncbi/insdc:JCADBF000000000.1>) and figshare (<https://doi.org/10.6084/m9.figshare.30899549>), respectively.

Code Availability

No custom scripts were used in this study. All software and pipelines were executed in accordance with the manuals and programs of the published bioinformatic tools. All programs used in this study are publicly available, with version and parameters explicitly described in the Methods section.

References

1. WoRMS Editorial Board (2025). World Register of Marine Species. Available from <https://www.marinespecies.org> at VLIZ. Accessed 2025-10-29. doi:10.14284/170.
2. Thandar, A. S. New species and a new record of sea cucumbers from deep waters of the South African temperate region (Echinodermata: Holothuroidea). *Zootaxa* **2013**, 30–42 (2009).
3. Silchenko, A. S., Avilov, S. A., Andrijaschenko, P. V., Popov, R. S., Chingizova, E. A., Grebnev, B. B., Rasin, A. B. & Kalinin, V. I. The isolation, structure elucidation and bioactivity study of chilensosides A, A1, B, C, and D, holostane triterpene Di-, Tri- and tetrasulfated pentaosides from the sea cucumber *Paracaudina chilensis* (Caudinidae, Molpadida). *Molecules* **27**, 7655 (2022).
4. Pawson, D. L. & Fell, H. A revised classification of the dendrochirote holothurians. *Breviora* **214**, 1-7 (1965).
5. Miller, A. K., Kerr, A. M., Paulay, G., Reich, M., Wilson, N. G., Carvajal, J. I. & Rouse, G. W. Molecular phylogeny of extant Holothuroidea (Echinodermata). *Mol. Phylogenet. Evol.* **111**, 110-131 (2017).
6. Sun, S. E., Ni, G., Wang, W., Sha, Z., Xiao, N. & Zeng, X. Highly rearranged gene orders and rapid evolutionary rates in the mitochondrial genomes of Apodida (Echinodermata: Holothuroidea). *Front. Mar. Sci.* **10** (2023).
7. Martini, S., Kuhn, L., Mallefet, J. & Haddock, S. H. D. Distribution and quantification of bioluminescence as an ecological trait in the deep sea benthos. *Sci. Rep.* **9**, 14654 (2019).
8. Herring, P. J. New observations on the bioluminescence of echinoderms. *J. Zool.* **172**, 401-418 (1974).
9. Herring, P. J. The Spectral Characteristics of Luminous Marine Organisms. *Proc. R. Soc. B.* **220**, 183-217 (1983).
10. Widder, E. A., Latz, M. I. & Case, J. F. Marine bioluminescence spectra measured with an optical multichannel detection system. *Biol. Bull.* **165**, 791-810 (1983).
11. Robison, B. H. Bioluminescence in the benthopelagic holothurian *Eynipniastes eximia*. *J. Mar. Biol. Assoc. U. K.* **72**, 463-472 (1992).

12. Johnsen, S., Frank, T. M., Haddock, S. H. D., Widder, E. A. & Messing, C. G. Light and vision in the deep-sea benthos: I. Bioluminescence at 500–1000 m depth in the Bahamian Islands. *J. Exp. Biol.* **215**, 3335-3343 (2012).
13. Bessho-Uehara, M., Mallefet, J. & Haddock, S. in *The World of Sea Cucumbers* (ed Jean-François Hamel Annie Mercier, Andrew D. Suhrbier, Christopher M. Pearce) Ch. 22, 361-375 (Academic Press, 2024).
14. Birk, M. H., Blicher, M. E. & Garm, A. Deep-sea starfish from the Arctic have well-developed eyes in the dark. *Proc. R. Soc. B.* **285**, 20172743 (2018).
15. Kalinin, V. I., Malyutin, A. N. & Stonik, V. A. Caudinoside A - A new triterpene glycoside from the holothurian *Paracaudina ransonetii*. *Chem. Nat. Compd.* **22**, 355-356 (1986).
16. McEuen, F. S. & Chia, F.-S. Larval development of a molpadiid holothuroid, *Molpadia intermedia* (Ludwig, 1894) (Echinodermata). *Can. J. Zool.* **63**, 2553-2559 (1985).
17. Amaro, T., Bianchelli, S., Billett, D. S. M., Cunha, M. R., Pusceddu, A. & Danovaro, R. The trophic biology of the holothurian *Molpadia musculus*: implications for organic matter cycling and ecosystem functioning in a deep submarine canyon. *Biogeosciences* **7**, 2419-2432 (2010).
18. Chen, T., Ren, C., Wong, N.-K., Yan, A., Sun, C., Fan, D., Luo, P., Jiang, X., Zhang, L., Ruan, Y., Li, J., Wu, X., Huo, D., Huang, J., Li, X., Wu, F., E, Z., Cheng, C., Zhang, X., Wang, Y. & Hu, C. The *Holothuria leucospilota* genome elucidates sacrificial organ expulsion and bioadhesive trap enriched with amyloid-patterned proteins. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2213512120 (2023).
19. Sun, L., Jiang, C., Su, F., Cui, W. & Yang, H. Chromosome-level genome assembly of the sea cucumber *Apostichopus japonicus*. *Sci. Data* **10**, 454 (2023).
20. Zhong, S., Ma, X., Jiang, Y., Liu, X., Zeng, M., Zhao, L., Huang, L., Huang, G., Zhao, Y., Qiao, Y. & Chen, X. The draft genome of the tropical sea cucumber *Stichopus monotuberculatus* (Echinodermata, Stichopodidae) reveals critical genes in fucosylated chondroitin sulfates biosynthetic pathway. *Front. Genet.* **14** (2023).
21. Zhang, L., He, J., Tan, P., Gong, Z., Qian, S., Miao, Y., Zhang, H., Tu, G., Chen, Q., Zhong, Q., Han, G., He, J. & Wang, M. The genome of an apodid holothuroid (*Chiridota heheva*) provides insights into its adaptation to a deep-sea reducing environment. *Commun Biol* **5**, 224 (2022).
22. Pu, Y., Zhou, Y., Liu, J. & Zhang, H. A high-quality chromosomal genome assembly of the sea cucumber *Chiridota heheva* and its hydrothermal adaptation. *GigaScience* **13** (2024).
23. Zhong, S., Liu, X., Ma, X., Chen, X., Jiang, Y., Zeng, M., Zhao, L., Huang, L., Huang, G., Zhao, Y., Liu, H. & Qiao, Y. Chromosomal-level genome assembly and annotation of the tropical sea cucumber *Holothuria scabra*. *Sci. Data* **11**, 474 (2024).
24. Jiang, C., Wu, Q., Su, F., Cui, W., Chen, T. & Sun, L. Chromosome-level genome assembly of the sea cucumber, *Colochirus anceps*. *Sci. Data* **12**, 1643 (2025).
25. Lau, N.-S., Miswan, N., Rosli, N., Gong, S., Irfan, I., Ganasan, M., Woo, S. P. & Jaya-Ram, A. A chromosomal-level genome assembly of the tropical sea cucumber *Stichopus fusiformiossa*. *Sci. Data* **12**, 973 (2025).
26. Porebski, S., Bailey, L. G. & Baum, B. R. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* **15**, 8-15 (1997).
27. Zhu, X., Li, S., Liu, L., Li, S., Luo, Y., Lv, C., Wang, B., Cheng, C. H. K., Chen, H. & Yang, X.

- Genome sequencing and analysis of *Thraustochytridae* sp. SZU445 provides novel insights into the polyunsaturated fatty acid biosynthesis pathway. *Mar. Drugs* **18**, 118 (2020).
28. Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S. & Aiden, E. L. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
 29. Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., Li, Z., Zhang, X., Wang, J., Yang, H., Fang, L. & Chen, Q. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**, 1-6 (2018).
 30. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764-770 (2011).
 31. Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J. & Schatz, M. C. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202-2204 (2017).
 32. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
 33. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170-175 (2021).
 34. Robinson, J. T., Turner, D., Durand, N. C., Thorvaldsdóttir, H., Mesirov, J. P. & Aiden, E. L. Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst.* **6**, 256-258 (2018).
 35. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. & Marra, M. A. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639-1645 (2009).
 36. Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16**, 418-420 (2000).
 37. Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O. & Walichiewicz, J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462-467 (2005).
 38. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc.* **25**, 4.10.11-14.10.14 (2009).
 39. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265-268 (2007).
 40. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573-580 (1999).
 41. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001949145.1/
 42. <https://www.echinobase.org/echinobase/static-echinobase/ftpDatafiles.jsp>
 43. Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S. & Morgenstern, B. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435-439 (2006).
 44. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465-W467 (2005).

45. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309-W312 (2004).
46. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. & Schneider, M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365-370 (2003).
47. Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R. & Hunter, S. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).
48. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J. & Punta, M. Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222-D230 (2014).
49. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27-30 (2000).
50. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25-29 (2000).
51. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955-964 (1997).
52. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
53. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 1–14 (2015).
54. Parey, E., Ortega-Martinez, O., Delroisse, J., Piovani, L., Czarkwiani, A., Dylus, D., Arya, S., Dupont, S., Thorndyke, M., Larsson, T., Johannesson, K., Buckley, K. M., Martinez, P., Oliveri, P. & Marlétaz F. The brittle star genome illuminates the genetic basis of animal appendage regeneration. *Nat Ecol Evol.* **8**, 1505–1521 (2024).
55. Wang, Y., Wang, Y., Yang, Y. Ni, G., Li, Y. & Chen, M. Chromosome-level genome assembly of the northern Pacific seastar *Asterias amurensis*. *Sci. Data.* **10**, 767 (2023).
56. Wang, J., Sun, S., Mei, Z. Sha, Z. & Chen N. A high-quality chromosome-level genome assembly of a feather star *Glyptometra* sp. from a deep seamount. *Sci. Data* **13**, 598 (2026).
57. NCBI GenBank https://identifiers.org/ncbi/insdc.gca:GCF_000003815.2 (2020).
58. Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
59. Stamatakis A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* **22**, 2688–90 (2006).
60. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
61. Members, C.-N. & Partners. Database Resources of the National Genomics Data Center, China

National Center for Bioinformation in 2025. *Nucleic Acids Res.* **53**, D30-D44 (2025).

62. Zhang , S., Chen , X., Jin , E., Wang , A., Chen , T., Zhang , X., Zhu , J., Dong , L., Sun , Y., Yu , C., Zhou , Y., Fan , Z., Chen , H., Zhai , S., Sun , Y., Chen , Q., Xiao , J., Song , S., Zhang , Z., Bao , Y., Wang , Y. & Zhao, W. The GSA family in 2025: A broadened sharing platform for multi-omics and multimodal data. *Genomics, Proteomics Bioinf.* **23** (2025).
63.

63.	<i>NGDC</i>	<i>Genome</i>	<i>Sequence</i>	<i>Archive</i>
https://ngdc.cncb.ac.cn/gsa/browse/CRA033897/CRR2351251 (2025).				
64.

64.	<i>NGDC</i>	<i>Genome</i>	<i>Sequence</i>	<i>Archive</i>
https://ngdc.cncb.ac.cn/gsa/browse/CRA033897/CRR2351252 (2025).				
65.

65.	<i>NGDC</i>	<i>Genome</i>	<i>Sequence</i>	<i>Archive</i>
https://ngdc.cncb.ac.cn/gsa/browse/CRA033897/CRR2351253 (2025).				
66.

66.	<i>NGDC</i>	<i>Genome</i>	<i>Sequence</i>	<i>Archive</i>
https://ngdc.cncb.ac.cn/gsa/browse/CRA033897/CRR2351254 (2025).				
67.

67.	<i>NGDC</i>	<i>Genome</i>	<i>Sequence</i>	<i>Archive</i>
https://ngdc.cncb.ac.cn/gsa/browse/CRA033897/CRR2351255 (2025).				
68.

68.	<i>NGDC</i>	<i>Genome</i>	<i>Sequence</i>	<i>Archive</i>
https://ngdc.cncb.ac.cn/gsa/browse/CRA033897/CRR2351256 (2025).				
69.

69.	<i>NGDC</i>	<i>Genome</i>	<i>Sequence</i>	<i>Archive</i>
https://ngdc.cncb.ac.cn/gsa/browse/CRA033897/CRR2351257 (2025).				
70.

70.	<i>NGDC</i>	<i>Genome</i>	<i>Sequence</i>	<i>Archive</i>
https://ngdc.cncb.ac.cn/gsa/browse/CRA033897/CRR2351258 (2025).				
71.

71.	<i>NGDC</i>	<i>Genome</i>	<i>Sequence</i>	<i>Archive</i>
https://ngdc.cncb.ac.cn/gsa/browse/CRA033897/CRR2351259 (2025).				
72.

72.	<i>NGDC</i>	<i>Genome</i>	<i>Sequence</i>	<i>Archive</i>
https://ngdc.cncb.ac.cn/gsa/browse/CRA033897/CRR2351260 (2025).				
73. NCBI GenBank <https://identifiers.org/ncbi/insdc:JCADBF000000000.1> (2026).
74. Wang, J. & Chen, N. Chromosome-level genome assembly of *Molpadia* sp. C3. <https://doi.org/10.6084/m9.figshare.30899549>. (2025).

Acknowledgements

We acknowledge the support of the Research Vessel Jiageng of Xiamen University and Prof. Junlong Zhang from Institute of Oceanology, Chinese Academy of Sciences collected the animal sample *Molpadia* sp. C3 for us.

Authors' contributions

N.C. conceived of and supervised the execution of this research project. N.X. identified the sample and carried out taxonomical identification. J.W. and Z.D processed the sample. J.W. performed analysis and wrote the manuscript. Z.G. and N.C. revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDB42000000), the Taishan Scholar Project Special Fund (to Nansheng Chen), the project ZR2022QD066 supported by Shandong Provincial Natural Science Foundation (to Jing Wang), and the National Natural Science Foundation of China (No. 42376135 to Ning Xiao).

Competing interests

The authors declare no competing interests.

ARTICLE IN PRESS