



OPEN

DATA DESCRIPTOR

Chromosome-scale genome of the burrowing sea anemone *Paracondylactis sinensis*

Junyuan Li¹, Rongye Tang², Juan Feng¹, Tinghui Xie¹, Sitong Liu¹ & Yang Li²✉

Paracondylactis sinensis is a burrowing sea anemone inhabiting soft sediments along the Chinese coast, representing an ecologically and economically important actinarian species. Despite its unique adaptations to hypoxia and sediment-associated stressors, genomic resources for burrowing sea anemones have been lacking. Here, we report a high-quality, chromosome-level genome assembly of *P. sinensis*. With PacBio HiFi long reads ($39.77 \times$ coverage), Illumina short reads, and Hi-C data, a 210.63 Mb genome with a contig N50 of 8.70 Mb and a scaffold N50 of 9.41 Mb was generated. A total of 93.44% of the assembly was anchored to 19 pseudo-chromosomes. BUSCO analysis indicated 95.91% completeness, confirming high assembly quality. Comprehensive annotation identified 19,420 protein-coding genes, of which 91.35% were functionally annotated. Repetitive elements accounted for 26.43% of the genome, with transposable elements representing 20.47%. This genome provides a crucial reference for understanding the genetic basis of environmental adaptation in *P. sinensis* and supports future efforts in its conservation, aquaculture, and bioactive compound exploration.

Background & Summary

Although the majority of sea anemones are sessile, certain species have evolved a burrowing character by embedding themselves in soft sediments^{1,2}. These burrowing sea anemones play crucial roles in benthic ecosystems. Their movements create microhabitats for small organisms and facilitate biogeochemical cycling and material exchange between water and sediment layers. Unlike other sessile species that are surrounded solely by seawater, burrowing sea anemones inhabit a heterogeneous microenvironment composed of water mixed with sand or mud. To survive under such conditions, they have evolved tolerance to stressors such as hypoxia, pathogens and strongly reducing conditions^{3,4}. Despite their distinct adaptations and ecological significance, genomic resources of burrowing sea anemones are limited, hindering our understanding of their molecular adaptation mechanisms.

Paracondylactis sinensis is a typical burrowing sea anemone and also an economically important species in China. However, due to overfishing and other anthropogenic disturbances, its wild populations have declined⁵. Our previous SSR-based analysis revealed signs of inbreeding in both northern and southern populations, highlighting the urgent need for conservation and artificial breeding efforts⁶. In this study, we present a chromosome-level genome assembly of *P. sinensis*, which provides a valuable resource for investigating its adaptive strategies and supports future studies on population genetics and resource management.

Beyond their ecological roles and economic value, sea anemones are recognized as a natural source of diverse bioactive polypeptides⁷. These include antimicrobial peptides used to defend against external pathogens and venom-derived toxins employed for prey capture—both of which hold promise for pharmaceutical applications such as infection control and wound healing^{8,9}. Compared to sessile species, burrowing sea anemones like *P. sinensis* are exposed to more complex environment with low oxygen contents and distinct microbial communities. These unique conditions may drive the evolution of lineage-specific active compounds with novel biological activities uncharacterized before. The chromosome-level genome generated in this study provides a foundational resource for identifying such compounds and advancing bioprospecting and gene-mining efforts in this ecologically and pharmacologically valuable species.

¹College of Agriculture and Bioengineering, Taizhou Vocational College of Science and Technology, Taizhou, 318020, China. ²Laboratory of Marine Organism Taxonomy and Phylogeny, Qingdao Key Laboratory of Marine Biodiversity and Conservation, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, 266071, China. ✉e-mail: liyang@qdio.ac.cn



Fig. 1 *Paracondylactis sinensis* buried in the sandy substrate of a laboratory culture tank.

Methods

Sample collection and DNA extraction. Samples of the sea anemones *Paracondylactis sinensis* were dug by hand during ebb at the coast of Taizhou, China (121°39.53'E, 28°20.12'N). The sampling procedures complied with relevant regulations of the Institutional Animal Care and Use Committee of the Institute of Oceanology, Chinese Academy of Science. After collection, they were transported to the lab and cultured in an aerated tank with temperature at 20°C and salinity at 33‰ (Fig. 1). One of the *P. sinensis* with active and healthy status was selected and rinsed with sterile seawater for subsequent DNA extraction. The body column of the sample was dissected with sterile scissors and forceps and then flash frozen in liquid nitrogen. DNA for PacBio HiFi sequencing was extracted using the MagAttract® HMW DNA Kit (Qiagen, Germany) according to the manufacturer's instructions. The quantity, integrity and purity of the DNA were measured using a Qubit 3.0 fluorometer, pulsed-field gel electrophoresis (5–80 kV, 17 h) and a NanoDrop® instrument respectively. The value of A260/280 of the extracted DNA was 1.82 and a dominant high-molecular-weight DNA band larger than 48 kb was observed (Fig. 2), indicating its suitability for long-read sequencing. The extracted DNA was sent to Novogene Bioinformatics Technology Co., Ltd. (Tianjin, China) for library construction and sequencing.

Illumina sequencing and genome survey. A paired-end DNA library for genome survey analysis was constructed with an insert size of 350 bp using a Rapid Plus DNA Lib Prep Kit for Illumina (ABclonal, USA). The library was subsequently sequenced on an Illumina NovaSeq X Plus sequencer (2 × 150 bp paired-end reads) (Illumina, USA). A total of 21.92 Gb data was generated. Data cleaning for the Illumina short reads was performed by Trimmomatic 0.36.3¹⁰ with the parameters as follows: “ILLUMINACLIP:TruSeq-3-PE-2.fa:2:30:10 LEADING:15 TRAILING:15 SLIDINGWINDOW:4:15 MINLEN:40”, which was followed by FastUniq 1.1 to remove duplicates¹¹. Finally, 18.48 Gb clean data was obtained (Table 1). GenomeScope 2.0 and Jellyfish 2.2.3¹² were used for genome survey. The k-mer distribution with 17 nt was constructed using clean Illumina short-read data (Fig. 3). According to the formula genome size = $k_{\text{num}}/k_{\text{depth}}$, where k_{num} is the number of k-mers and k_{depth} is the expected depth of k-mers, the genome size was estimated to be 209.71 Mb, with a heterozygosity rate of 0.72% and a repeat rate of 31.47%.

PacBio HiFi sequencing and Hi-C sequencing. For HiFi-read sequencing, a 15 kb SMRTbell library was constructed with a molecular weight of genomic DNA larger than 15 kb, following the instructions of the SMRTbell Express Template Prep Kit 3.0 (PacBio, USA). Fragment size distribution of the library detected by Agilent 5200 (Agilent, USA) revealed a dominant insertion peak at approximately 16 kb, with the majority of the fragments distributed between 10 and 40 kb (Fig. 4). The library was subsequently sequenced on a PacBio Revio system with a Revio SPRQ polymerase kit (PacBio, USA). A total of 8.34 Gb high-quality HiFi reads (39.77 × genome coverage) were finally produced (Table 1). The length distribution of the HiFi reads showed a major peak at approximately 11 kb (Fig. 5), with a mean read length of 11,344 bp and an N50 at the length of 13,891 bp.

For a haploid assembly, the sample used for Hi-C scaffolding was the same individual as one used for PacBio HiFi sequencing and Illumina sequencing. Dissected body column tissue was cross-linked (4% formaldehyde, 5 min, RT), quenched (0.20 M glycine, 1 min, RT), chilled (> 15 min, ice), lysed (10 mM NaCl, 0.20% IGEPAL CA-630, 10 mM Tris-HCl, 1 × protease inhibitors), digested with DpnII, biotin-labeled, and end-repaired¹³. Library with insert size 350 bp was prepared with the Rapid Plus DNA Library Prep Kit for Illumina (ABclonal, USA) and sequenced on an Illumina NovaSeq X Plus sequencer (2 × 150 bp paired-end reads) (Illumina, USA). After quality control performed by Trimmomatic 0.36.3¹⁰ and FastUniq 1.1¹¹, a total of 15.92 Gb Hi-C reads with 88.12 × genome coverage was generated (Table 1).

Chromosome-level genome assembly. The genome was preliminarily assembled using hifiasm 0.20.0¹⁴ with the generated PacBio HiFi reads under the parameters “-l 3 -s 0.5”. Sequencing contamination was screened by FCS-GX 0.5.5¹⁵, and no sequences were detected to originate from foreign organisms. The contig-level genome assembly was 210.63 Mb in length, closely matching the estimated genome size (209.71 Mb) by genome survey analysis. The contig number of this initial assembly was 181, with contig N50 as long as 8.70 Mb (Table 2). The

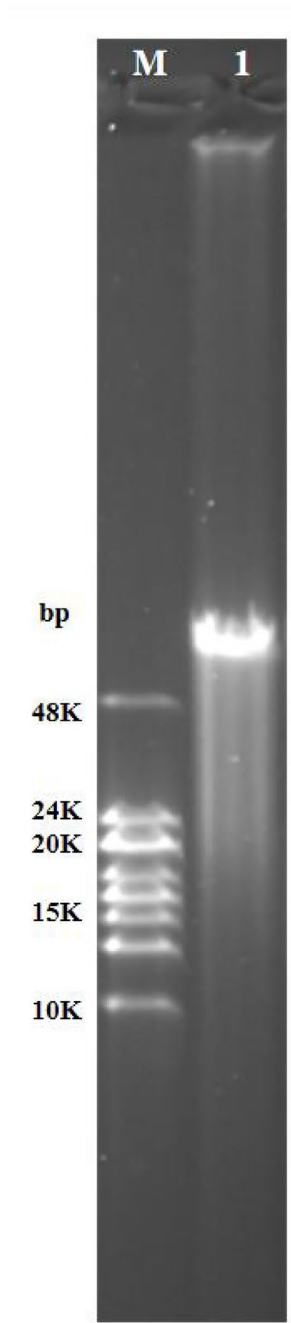


Fig. 2 Pulsed-field gel electrophoresis (PFGE) assessment of the extracted high-molecular-weight genomic DNA. Lane M: DNA size marker (10–48 kb); lane 1: the extracted DNA.

	Illumina	PacBio	Hi-C
Clean data (Gb)	18.48	8.34	15.92
Sequencing Coverage (×)	88.12	39.77	75.91
GC content (%)	39.46	39.02	38.93
Q20 (%)	99.78	\	99.98
Q30 (%)	97.93	\	96.61

Table 1. Sequencing data used for the assembly and annotation of *Paracondylactis sinensis* genome.

GC content of the assembly was 39.02%, which was close to that of the model sea anemone species *Nematostella vectensis* (40.72%; GCA_932526225¹⁶), suggesting a comparable base composition to published actiniarian genome.

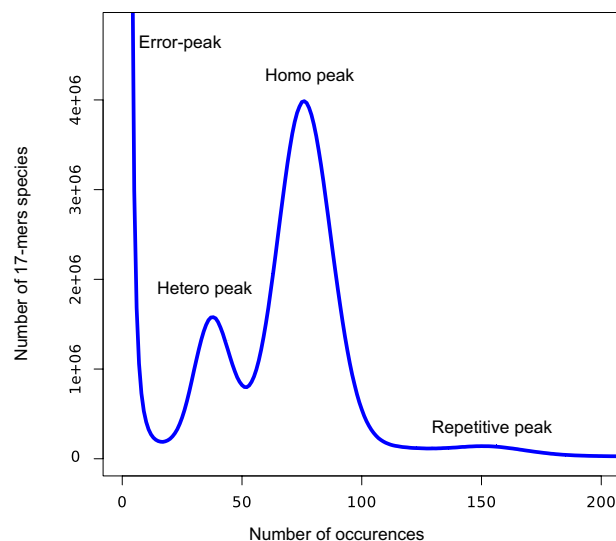


Fig. 3 17-mer frequency distribution in the *Paracondylactis sinensis* genome.

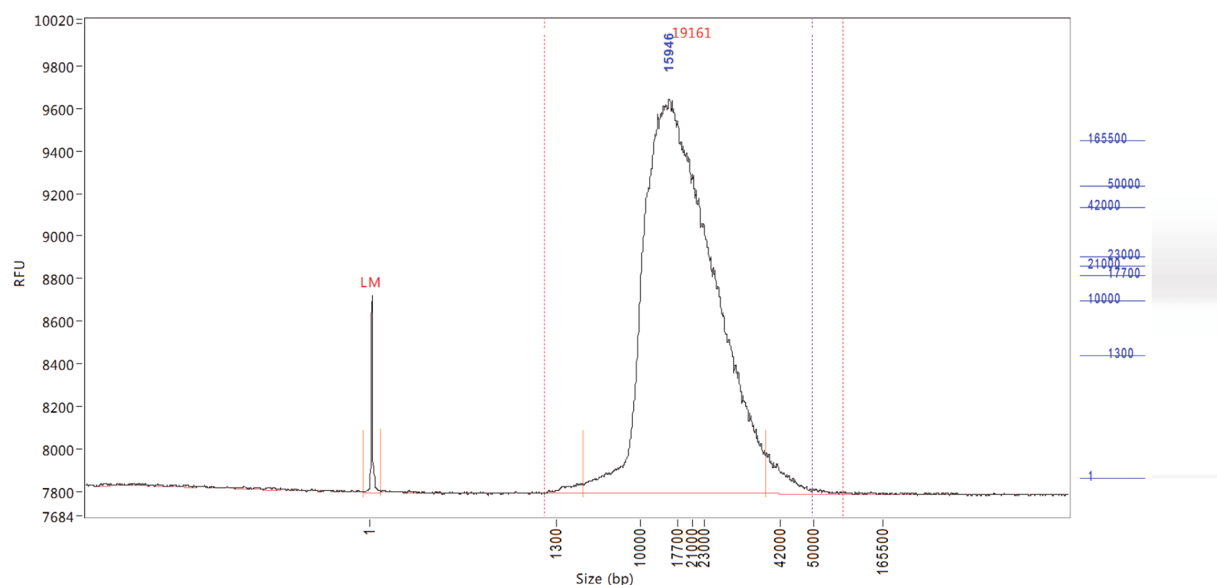


Fig. 4 Quality check of the SMRTbell library by Agilent 5200.

Hi-C reads were used to scaffold the draft assembly to the chromosome level. Paired-end Hi-C reads were aligned to the contig-level genome using Chromap 0.2.3¹⁷ with the “--preset hic” and “--remove-pcr-duplicates” options enabled. The resulting SAM alignments were sorted and converted to BAM format using SAMtools 1.16¹⁸. Contig clustering and scaffolding were performed with YaHS 1.2.2¹⁹ under default parameters. The scaffolded assembly and Hi-C contact maps were further processed using the pre and post modules of Juicer Tools 1.22.01²⁰ to generate hic files and finalize chromosome-scale scaffolds through manual review. Finally a total of 196.80 Mb (93.44%) of the genome was anchored to 19 pseudo-chromosomes (Fig. 6, Table 3), with the sizes ranging from 5.40 Mb to 25.01 Mb. The scaffold N50 of the chromosome-level genome was 9.41 Mb (Table 2). The total length of the unanchored scaffolds was 13.83 Mb, comprising 147 sequences, of which 86 scaffolds (58.50%) were distributed between 30–60 kb (Fig. 7). These unplaced scaffolds lacked sufficient Hi-C linkage evidence for confident anchoring with contact map largely blank (Fig. 8), which may be attributable to repeat-associated multi-mapping and/or conflicting placements across pseudochromosomes that precluded reliable assignment to a unique chromosomal locus.

Annotation of the repetitive sequences. Repetitive sequences were annotated using a combination of homology-based and *de novo* approaches. A species-specific repeat library was built from the assembled genome using RepeatModeler 2.0²¹ with the -LTRStruct option enabled. Homology-based annotation was performed with RepeatMasker 4.1²² against the Actiniaria repeat database using the parameters of -e rmbblast, -nolow, -s,

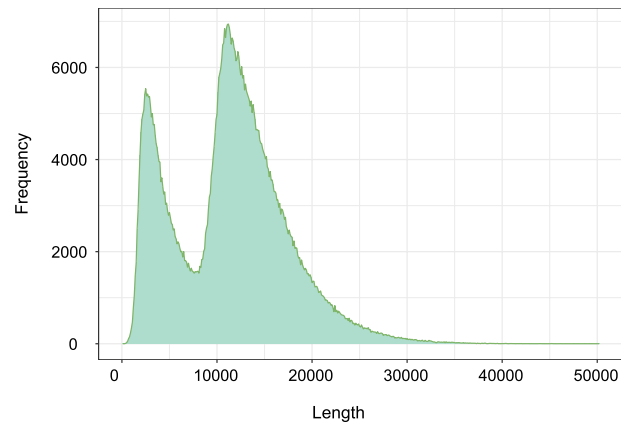


Fig. 5 Length distribution diagram of HiFi-reads.

Survey		
K-mer	17	
Depth	76	
N _{k-mer}	16,42,90,87,914	
Genome size (Mb)	209.71	
Heterozygous rate (%)	0.72	
Repeat rate (%)	31.47	
Assembly		
Assembly size (bp)	21,06,27,143	
Number of scaffolds	166	
Scaffold N50 (bp)	94,08,176	
Scaffold N90 (bp)	68,43,887	
Number of contigs	181	
Contig N50 (bp)	86,98,258	
Contig N90 (bp)	24,84,273	
GC(%)	39.02	
BUSCO (% of total BUSCO)		
Complete BUSCOs (C)	915 (95.91%)	932 (97.69%)
Complete and single-copy BUSCOs (S)	915 (95.91%)	926 (97.06%)
Complete and duplicated BUSCOs (D)	0	6 (0.63%)
Fragmented BUSCOs (F)	21 (2.20%)	6 (0.63%)
Missing BUSCOs (M)	18 (1.89%)	16 (1.68%)
CEGMA		
Highly conserved CEGs (% of all highly conserved CEGs)	239 (96.37)	
Mapping rate		
Illumina short reads mapping rate	99.70%	
PacBio long-reads mapping rate	99.93%	

Table 2. Statistics of the genome survey, genome assembly and quality assessments.

-a, and -gff. *De novo* repeat annotation was then conducted by masking the genome with the species-specific library generated by RepeatModeler 2.0. Results from both homology-based and *de novo* runs were combined using the combineRMFiles.pl utility supplied with RepeatMasker 4.1 to produce a comprehensive repeat annotation. Finally, 26.43% of the genome at the length of 55.68 Mb was annotated as repeats. Among the repetitive sequences, 43.11 Mb were transposable elements (TEs), accounting for 20.47% of the genome (Table 4). The proportion of repeats in the genome of *P. sinensis* was similar to that of the model actinarian species *Exaiptasia diaphana* (24.24%; GCA_024322265²³) predicted in our previous study²⁴.

Gene prediction and functional annotation. Gene structure prediction was performed by integrating evidence from transcriptome data, protein homology, and *ab initio* prediction. RNA-Seq data from different tissues of *P. sinensis* were downloaded (NCBI SRA accessions: SRR28341359, SRR28327126, SRR28327127, and SRR28327128)⁶ and aligned to the soft-masked genome using HISAT2 2.2.1²⁵ with default parameters. The alignments were sorted and indexed using SAMtools 1.16¹⁸. Gene prediction was then conducted with BRAKER3

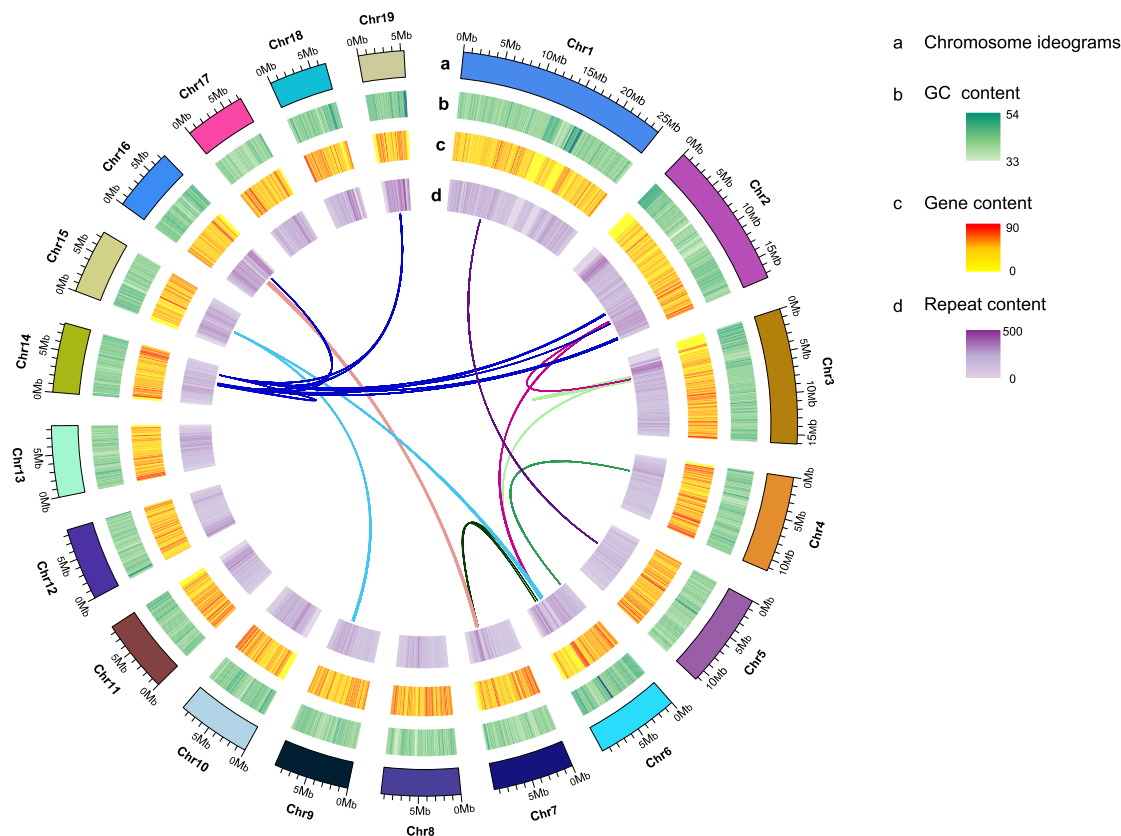


Fig. 6 Circos plot of 19 pseudochromosomal linkage groups of *Paracondylactis sinensis* genome showing the marker distribution in 1 Mb sliding windows (from the outer to inner circle: GC content, gene density and repeat density; lines in the middle depict intrachromosomal synteny).

Number	Length (bp)	Scaffold number	Percentage of whole genome (%)
1	2,50,14,327	Chr 1	11.88
2	1,75,78,697	Chr 2	8.35
3	1,60,70,250	Chr 3	7.63
4	1,14,02,762	Chr 4	5.41
5	1,13,32,616	Chr 5	5.38
6	99,10,978	Chr 6	4.71
7	95,87,682	Chr 7	4.55
8	94,08,176	Chr 8	4.47
9	89,67,751	Chr 9	4.26
10	88,02,227	Chr 10	4.18
11	86,98,258	Chr 11	4.13
12	86,26,659	Chr 12	4.10
13	84,61,090	Chr 13	4.02
14	82,49,100	Chr 14	3.92
15	76,77,218	Chr 15	3.64
16	75,30,589	Chr 16	3.58
17	72,33,425	Chr 17	3.43
18	68,43,887	Chr 18	3.25
19	54,03,863	Chr 19	2.57
Total	19,67,99,555		93.44

Table 3. Length of 19 pseudomolecules in the assembly of *Paracondylactis sinensis* genome.

3.0.3²⁶, which integrates GeneMark-ETP 1.00, AUGUSTUS 3.4.0, and extrinsic evidence. RNA-Seq alignments were incorporated via the --bam option, while protein homology information was provided using the --prot_seq option. Protein sequences from seven closely related sea anemone species—*Actinostola* sp.²⁷, *Alvinactis* sp.²⁸,

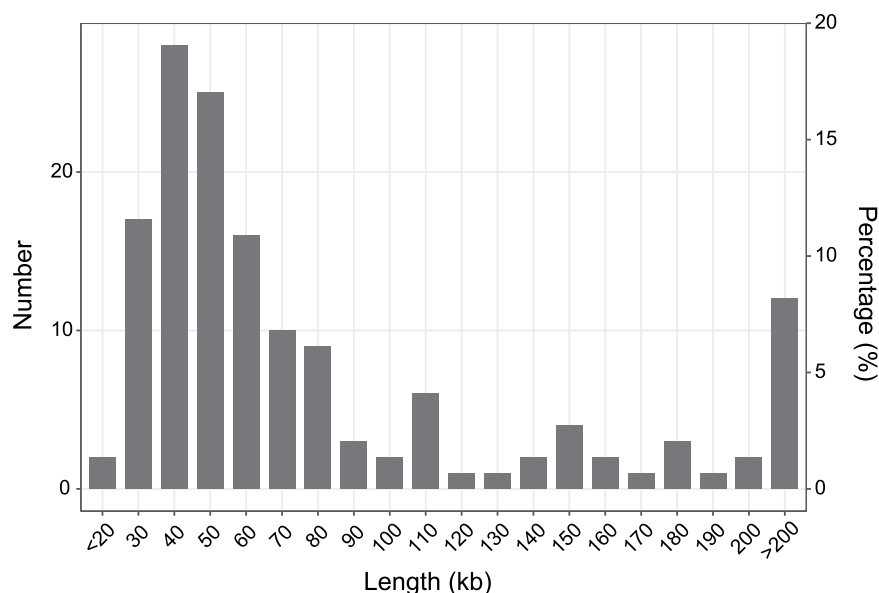


Fig. 7 Length distribution of the unanchored scaffolds.

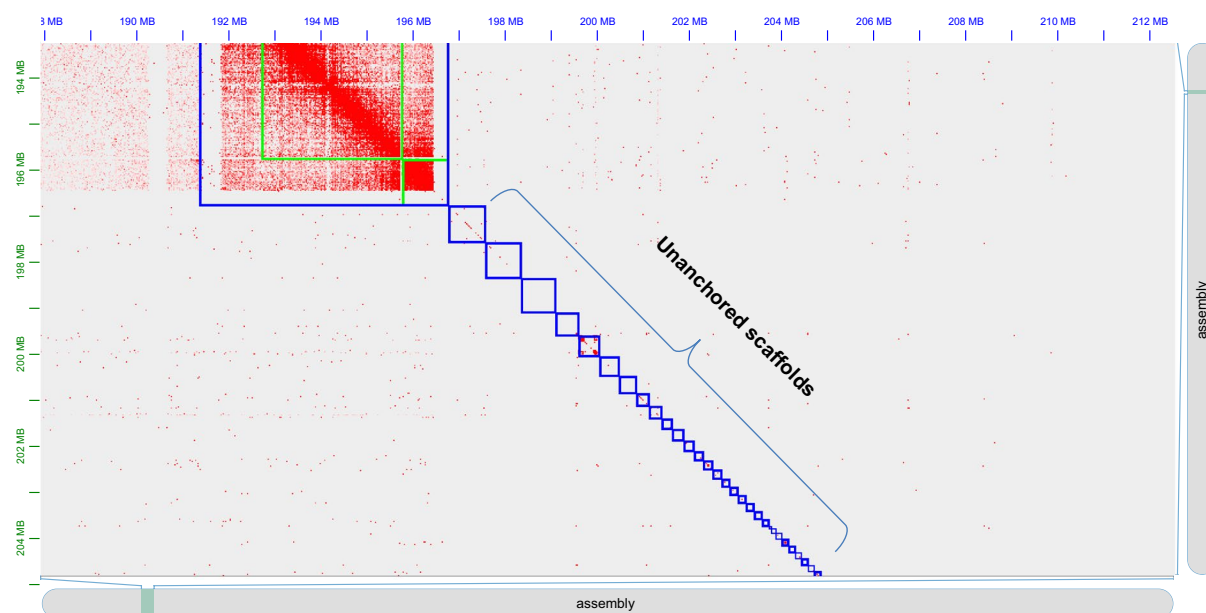


Fig. 8 Unanchored scaffolds showed in the Hi-C contact map.

*Exaiptasia diaphana*²³, *Nematostella vectensis*²⁹, *Scolanthus callimorphus*²⁹, *Actinia tenebrosa*³⁰, and *Actinoscyphia liui*²⁴—were downloaded from NCBI and Figshare databases, which were used to guide evidence-based gene structure prediction. To further improve the structural annotation by incorporating untranslated regions (UTRs), PASA 2.5.3³¹ was employed for transcript-based refinement. Transcripts assembled by trinity 2.15.1³² were aligned to the genome using minimap2 2.30³³, and integrated into the annotation framework in two rounds. The first round generated transcript alignments and preliminary structural models, while the second round updated the BRAKER3 gene models by adding UTR annotations and refining exon-intron boundaries based on transcript evidence. The representative transcript for each gene was then selected using AGAT 1.4.1³⁴. This integrative workflow produced a total of 19,420 genes (Table 5), with a gene length of 123,281,089 bp, exon length of 43,511,086 bp, CDS length of 30,009,737 bp, and intron length of 79,770,003 bp. After blasted in public databases (NR, SwissProt, eggNOG, KEGG, Pfam and GO), a total of 17,740 genes (91.35%) were functionally annotated (Table 6). The total number of predicted genes in *P. sinensis* (19,420) is comparable to that reported for other sea anemones, such as *Nematostella vectensis* with 19,231 predicted genes under accession number GCA_932526225¹⁶) and *Anthopleura sola* (GCA_023349425^{35,36}) with 19,899 genes estimated in our previous study²⁹, although this information was not reported in the original genome publication³³.

Category	Count	bp covered	Percentage (%)
DNA transposons	4700	3427586	1.627
Maverick	234	465533	0.221
Crypton	1730	387795	0.184
Crypton-A	524	338800	0.161
hAT-Ac	414	171073	0.081
hAT-Tip100	123	148854	0.071
TcMar-Tc1	113	133544	0.063
CMC-Chapaev	26	101876	0.048
Sola-3	41	83869	0.040
Kolobok-T2	54	68717	0.033
Merlin	8	67868	0.032
PIF-Harbinger	71	64255	0.031
hAT-hAT5	239	39889	0.019
P	50	36838	0.017
TcMar-ISRm11	56	20247	0.010
Academ-1	4	15592	0.007
CMC-EnSpm	11	13745	0.007
Academ-H	19	8419	0.004
Helitron	715	1131084	0.537
Unclassified	268	129588	0.062
LTR	15311	5114151	2.428
Ngaro	8094	1716681	0.815
Unknown	2755	1311486	0.623
Gypsy	129	309022	0.147
Pao	496	266029	0.126
ERVK	45	49418	0.023
ERV1	26	1732	0.001
Unclassified	3766	1459783	0.693
LINE	2893	1532878	0.728
L2	1035	652716	0.310
Penelope	1214	417362	0.198
L1	241	256984	0.122
CRE	158	83930	0.040
Rex-Babar	72	35492	0.017
CR1	65	24260	0.012
L1-Tx1	18	16409	0.008
Dong-R4	19	10820	0.005
RTE-BovB	35	8122	0.004
Unclassified	36	26783	0.013
SINE	442	78230	0.037
MIR	301	57613	0.027
tRNA-Deu	51	11171	0.005
B2	90	9446	0.004
Unknown	179511	32960837	15.649
Low complexity	16948	1162779	0.552
Simple repeat	105464	8898344	4.225
Satellite	37	37095	0.018
rRNA	1005	2465081	1.170
Total	146800	22716144	26.434

Table 4. Repeat composition of *Paracondylactis sinensis* genome.

Data Records

The sequencing data generated in this study have been deposited in the National Center for Biotechnology Information (NCBI) under BioProject number PRJNA1334318. Specifically, the Illumina, PacBio HiFi, and Hi-C datasets are available in the NCBI Sequence Read Archive under accession numbers SRR35627867, SRR35627868, and SRR35627866³⁷, respectively. The chromosome-level assembly of *Paracondylactis sinensis* genome was deposited in the NCBI genome database under accession number GCA_054491775³⁸ and

Methods	Software	Gene number
<i>Ab initio</i> (RNA/protein hints-guided)	AUGUSTUS	17,740
<i>Ab initio</i> (unsupervised gene prediction)	GeneMark.hmm3	838
RNA-Seq-based ORF detection	GeneMarkS-T	842
Integration	BRAKER3	19,420

Table 5. Gene prediction through BRAKER 3 integrating.

Database	Annotated number	Percentage (%)
GO	7,797	40.15
KEGG	8,433	43.42
Pfam	4,299	22.14
SwissProt	12,317	63.42
eggNOG	12,816	65.99
NR	17,729	91.29
All annotated	17,740	91.35

Table 6. Functional annotation of the predicted genes.

the corresponding BioSample ID is SAMN51854516. The genome annotation results are accessible from the Figshare database (<https://doi.org/10.6084/m9.figshare.30209509>)³⁹.

Technical Validation

The completeness of the chromosome-level genome was evaluated by benchmarking universal single-copy orthologue (BUSCO 5.4.7) analysis⁴⁰, Core Eukaryotic Genes Mapping Approach (CEGMA 2.5)⁴¹ and a k-mer based approach implemented in Merqury 1.3⁴². BUSCO analysis with the metazoan (odb10) gene set revealed that the genome of *P. sinensis* and its predicted genes covered 95.91% and 97.69% of the “core” metazoan genes, respectively (Table 2). CEGMA analysis indicated the genome assembly covered 96.37% of highly conserved proteins in eukaryotic genomes. K-mer based evaluation was performed using a 21-mer database constructed from Illumina short reads. The result showed a high consensus accuracy (QV = 47.45, i.e., an error rate of 1.80×10^{-5}), but a relatively lower completeness of 87.82% for the assembly. The spectra-cn plot indicated that most missing k-mers (the “read-only” part) were concentrated around the heterozygous peak (Fig. 9a), potentially representing heterozygous regions in the haplotype assembly. To confirm this, the completeness of a diploid representation which included both haplotype-resolved assemblies generated by hifiasm 0.20.0¹⁴ was evaluated. The k-mer completeness of the diploid increased to 99.31% and the “read-only” k-mers were not aggregated around primary or heterozygous peaks (Fig. 9b). This result indicated that the lower completeness of the primary assembly mainly reflects unrepresented allele-specific k-mers rather than widespread assembly errors.

The trimmed Illumina short reads and PacBio long reads were mapped against the assembled genome using BWA 0.7.19⁴³. The mapping rates were 99.70% and 99.93%, respectively (Table 2), also indicating high accuracy of the assembly.

Data availability

The raw sequencing data are available in the NCBI Sequence Read Archive (SRA) database (<https://identifiers.org/ncbi/insdc.sra:SRP627585>), the genome assembly is available in NCBI GenBank (https://identifiers.org/ncbi/insdc.gca:GCA_054491775.1), and the genome and annotation files are available via Figshare (<https://doi.org/10.6084/m9.figshare.30209509.v1>).

Code availability

Genome survey:

- (1) Trimmomatic: parameter: ILLUMINACLIP:TruSeq. 3-PE-2.fa:2:30:10 LEADING:15 TRAILING:15 SLIDINGWINDOW:4:15 MINLEN:40.
- (2) FastUniq: parameter: -t q -c 1.
- (3) Jellyfish:
count: parameter: -G 2 -m 17 -C.
histo: all parameters were set as default.
stats: all parameters were set as default.
- (4) GenomeScope: parameter: -p 2 -k 17 -m 10000.

Contamination screen:

- (1) FCS-GX: parameter: --tax-id 1 --div None --split-fasta true --action-report true.

Genome assembly:

- (1) Hifiasm: parameter: -l 3.
- (2) Chromap: parameter: --preset hic --remove-pcr-duplicates.
- (3) Yaws: all parameters were set as default.

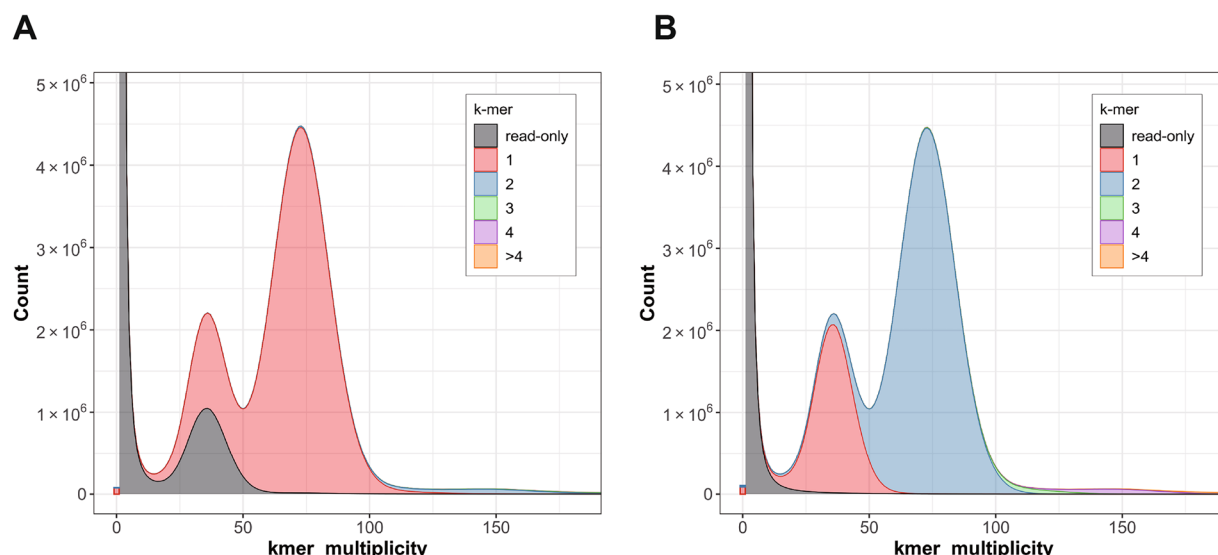


Fig. 9 Merqury copy number spectrum (spectra-cn) plots for haplotype (A) and diploid (B) assemblies of a *Paracondylactis sinensis* genome. Colors represent the different copy numbers of a certain k-mer found in the assembly. “read-only” means the k-mer found only in the Illumina reads.

- (4) Juicer: all parameters were set as default.
 (5) Samtools: all parameters were set as default.

Genome annotation:

- (1) RepeatModeler: parameter: BuildDatabase -name, -LTRStruct.
 (2) RepeatMasker: parameter: -e rmbblast -gff -s -a -nolow.
 (3) Hisat2: all parameters were set as default.
 (4) Trinity: parameter: --genome_guided_bam --genome_guided_max_intron 20000.
 (5) BRAKER3: all parameters were set as default.
 (6) PASA: parameter:
 Launch_PASA_pipeline.pl -C -R --ALIGNERS minimap2.
 Launch_PASA_pipeline.pl -A.
 (7) AGAT: all parameters were set as default.

Completeness evaluation

- (1) BUSCO (for genome assembly): parameter: -m genome -l metazoa_odb10.
 (2) BUSCO (for translated protein sequence): parameter: -m prot -l metazoa_odb10.
 (3) CEGMA: all parameters were set as default.
 (4) Merqury: parameter: meryl count k = 21.

Read mapping:

- (1) BWA:
 short reads: all parameters were set as default.
 PacBio HiFi reads: parameter: -x pacbio.
 (2) Picard: all parameters were set as default.
 (3) Samtools: all parameters were set as default.

Received: 8 October 2025; Accepted: 5 February 2026;

Published online: 14 February 2026

References

- Mangum, D. C. Burrowing behavior of the sea anemone *Phyllactis*. *Biol. Bull.* **138**, 316–325 (1970).
- Ivanova, N. Y. Classification and evolution of the burrowing sea anemones (Anthozoa: Actiniaria: Athenaria): a review of the past and current views. *Zoosyst. Rossica* **29**, 213–237 (2020).
- Wang, Y. *et al.* Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. *Appl. Environ. Microbiol.* **78**, 8264–8271 (2012).
- Sassaman, C. & Mangum, C. P. Adaptations to environmental oxygen levels in infaunal and epifaunal sea anemones. *Biol. Bull.* **143**, 657–678 (1972).
- Li, J. *et al.* The complete mitochondrial genome of an economic sea anemone (*Paracondylactis sinensis*) in the East China Sea. *Mitochondrial DNA Part B. Resources* **8**, 977–980 (2023).
- Li, J. *et al.* Development of microsatellite markers and evaluation of the genetic diversity of the edible sea anemone *Paracondylactis sinensis* (Cnidaria, Anthozoa) in China. *Biodivers. Data J.* **12**, e134363 (2024).
- Ramírez-Carreto, S., Miranda-Zaragoza, B., Simões, N., González-Muñoz, R. & Rodríguez-Almazán, C. Marine Bioprospecting: enzymes and stress proteins from the sea anemones *Anthopleura dowii* and *Lebrunia neglecta*. *Mar. Drugs* **22**, 12 (2024).
- Mizuno, M., Ito, Y. & Morgan, B. P. Exploiting the nephrotoxic effects of venom from the sea anemone, *Phyllodiscus semoni*, to create a hemolytic uremic syndrome model in the rat. *Mar. Drugs* **10**, 1582–1604 (2012).

9. Graftskaia, E. N. *et al.* Discovery of novel antimicrobial peptides: A transcriptomic study of the sea anemone *Cnidopus japonicus*. *J. Bioinf. Comput. Biol.* **16**, 1840006 (2018).
10. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
11. Xu, H. *et al.* FastUniq: A fast *de novo* duplicates removal tool for paired short reads. *PLOS ONE* **7**, e52249 (2012).
12. Vurtture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, **14** (2017).
13. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
14. 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).
15. Astashyn, A. *et al.* Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biology* **25**, 60 (2024).
16. NCBI GenBank https://identifiers.org/ncbi/insdc.gca:GCA_932526225 (2022).
17. Zhang, H., Song, L. & Wang, X. Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat. Commun.* **12**, 6566 (2021).
18. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
19. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).
20. Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
21. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA*. **117**, 9451–9457 (2020).
22. Chen, N. Using Repeatmasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.* **5**, 4.10.11–14.10.14 (2004).
23. Shum, C. W. *et al.* Genome of the sea anemone *Exaiptasia pallida* and transcriptome profiles during tentacle regeneration. *Front. Cell Dev. Biol.* **10**, 900321 (2022).
24. Li, J. *et al.* Chromosome-level genome assembly of a deep-sea Venus flytrap sea anemone sheds light upon adaptations to an extremely oligotrophic environment. *Mol. Ecol.* **33**, e17504 (2024).
25. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
26. Gabriel, L. *et al.* BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **34**, 769–777 (2024).
27. Liu, C. *et al.* Whole genome sequencing of a novel sea anemone (*Actinostola* sp.) from a deep-sea hydrothermal vent. *Sci. Data* **11**, 102 (2024).
28. Zhou, Y. *et al.* Genetic adaptations of sea anemone to hydrothermal environment. *Sci. Adv.* **9**, eadh0474 (2023).
29. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
30. Surm, J. M., Stewart, Z. K. & Papanicolaou, A. The draft genome of *Actinia tenebrosa* reveals insights into toxin evolution. *Ecol. Evol.* **9**, 11314–11328 (2019).
31. Haas, B. J. *et al.* Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic. Acids Res.* **31**, 5654–5666 (2003).
32. Haas, B. J. *et al.* *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
33. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
34. Dainat, J. *et al.* AGAT: Another GFF analysis toolkit to handle annotations in any GTF, GFF format. *Zenodo* (2020).
35. NCBI GenBank https://identifiers.org/ncbi/insdc.gca:GCA_023349425 (2022).
36. Li, J. Y. Annotation files of five sea anemones (*Actinocyphia liui*, *Diadumene lineata*, *Anthopleura sola*, *Telmatactis stephensoni* and *Actinia tenebrosa*). *figshare* <https://doi.org/10.6084/m9.figshare.25966645.v4> (2025).
37. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRP627585> (2026).
38. NCBI GenBank https://identifiers.org/ncbi/insdc.gca:GCA_054491775.1 (2026).
39. Li, J. Y. The genome and annotation file of sea anemone *Paracondylactis sinensis*. *figshare* <https://doi.org/10.6084/m9.figshare.30209509.v1> (2026).
40. Simão, F., Waterhouse, R. M., Panagiotis, I., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 3210–3212 (2015).
41. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
42. Rhie, A. *et al.* Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
43. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

Acknowledgements

This work was supported by Shandong Provincial Natural Science Foundation (No. ZR2025MS656), the Taizhou Science and Technology Plan Project (No. 25nyb03), and National Natural Science Foundation of China (No. 32570619). Thanks are given to the Oceanographic Data Center, IOCAS, for providing computing power.

Author contributions

Junyuan Li: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization and Writing. Rongye Tang: Conceptualization, Investigation, Project administration, Resources, Supervision, and Validation. Juan Feng: Investigation, Resources and Validation. Tinghui Xie: Resources and Validation. Sitong Liu: Software. Yang Li: Conceptualization, Validation, Funding acquisition, Project administration, Resources, Supervision, Writing–review & editing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026