scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Haplotype-resolved chromosomal-level genome assembly of *Chrysaora achlyos* (black sea nettle)

Yueyue Li[1,6], Xuelu Yu[1,6], Bingbing Li[1,6], Yajuan Peng[2,6], Xue Wang[1], Zhengfeng Hao[1,3], Xiao Huang[1], Shirong Lv[1], Yulong Fan[2], Jianbo Jian[4] ✉, Xiaoyu Geng[1] ✉ & Jishun Yang[5] ✉

Chrysaora achlyos **is a distinctive scyphozoan species characterized by its dark pigmentation and large size. In this study, we generated a haplotype-resolved, chromosome-scale genome assembly for *C.achlyos* using PacBio HiFi long reads and Hi-C technology. The resulting assembly comprises two haplotypes with sizes of 246.43 Mb (contig N50 = 7.69 Mb) and 248.65 Mb (contig N50 = 7.23 Mb), Additionally, repetitive sequences were characterized, with 110.52 Mb (~44.85%) in haplotype A and 112.57 Mb (~45.27%) in haplotype B. A total of 20,471 and 20,606 protein-coding genes were predicted in haplotypes A and B, respectively. The sequencing depth, mapping coverage, contig continuity, and BUSCO assessment collectively indicated a high-quality haplotype-resolved genome assembly. This high-quality genome assembly provides a foundation for future studies in comparative evolution, toxicology, ecology, and functional genomics.**

Background & Summary

Cnidaria, as one of the earliest-diverging metazoan lineages, holds a pivotal position in both evolutionary history[1] and marine ecosystems[2–4]. Within this phylum, jellyfish of the class Scyphozoa serve as ideal models for studying cell differentiation, symbiotic relationships, and ecological regulation due to their complex life cycles[5,6] and unique cnidocyte-based predation mechanisms[7,8]. Despite their ecological significance (e.g., impacts on fisheries[9,10] and carbon cycling[11,12]), the lack of genomic resources has severely hindered systematic investigations into their evolutionary adaptations and bioactive compounds (e.g., fluorescent proteins, toxins). To date, only a few draft-level genomes of jellyfish (e.g., *Aurelia aurita*[13], *Nemopilema nomurai*[14], *Cassiopea xamachana*[15]) are available in public databases, with chromosome-level reference genomes being exceptionally scarce[16], significantly limiting comparative phylogenomic analyses and functional gene discovery.

Chrysaora achlyos (commonly known as the black sea nettle), an enigmatic and poorly understood species of the class Scyphozoa (Fig. 1a), was first photographically recorded in 1926 but remained scientifically undescribed until 1997[17]. Representing one of the largest known scyphozoans, *C.achlyos* is morphologically characterized by its distinctive dark pigmentation and impressive size, with bell diameters reaching up to 1 meter. Comparative morphological studies have demonstrated that *C.achlyos* exhibits striking morphological and toxicological parallels with its congener, the Atlantic sea nettle (*Chrysaora quinquecirrha*), including shared nematocyst types (isorhizas and euryteles) and venom characteristics[18]. The venoms extracted from the nematocysts of both oral arms and lappets in *C.achlyos* were demonstrated to induce lethal effects in mice, erythrocyte hemolysis, and hepatocyte cytotoxicity. However, despite recent advances in sequencing technologies, particularly the transformative impact of Hi-C[19] which enables high-continuity genome assembly, *C.achlyos* remains conspicuously lacking genomic resources. While the genome of *C.quinquecirrha* has been successfully assembled and published[20], enabling comparative studies, the complete absence of high-quality genomic data for *C.achlyos* hinders elucidation of the molecular mechanisms underlying its environmental adaptations and venom evolution, despite growing interest in jellyfish toxins for their biomedical and ecological significance[21].

[1]Naval Medical Center of PLA, Naval Medical University, Shanghai, 200052, China. [2]BGI Genomics, Shenzhen, 518083, China. [3]School of Pharmacy, Bengbu Medical University, Bengbu, 233030, China. [4]Guangdong Provincial Key Laboratory of Marine Biotechnology, Shantou University, Shantou, 515063, China. [5]Department of Pharmacy, Changzheng Hospital, Naval Medical University, Shanghai, 200003, China. [6]These authors contributed equally: Yueyue Li, Xuelu Yu, Bingbing Li, Yajuan Peng. ✉e-mail: jbjian@126.com; 15339625597@163.com; jasunyang@foxmail.com

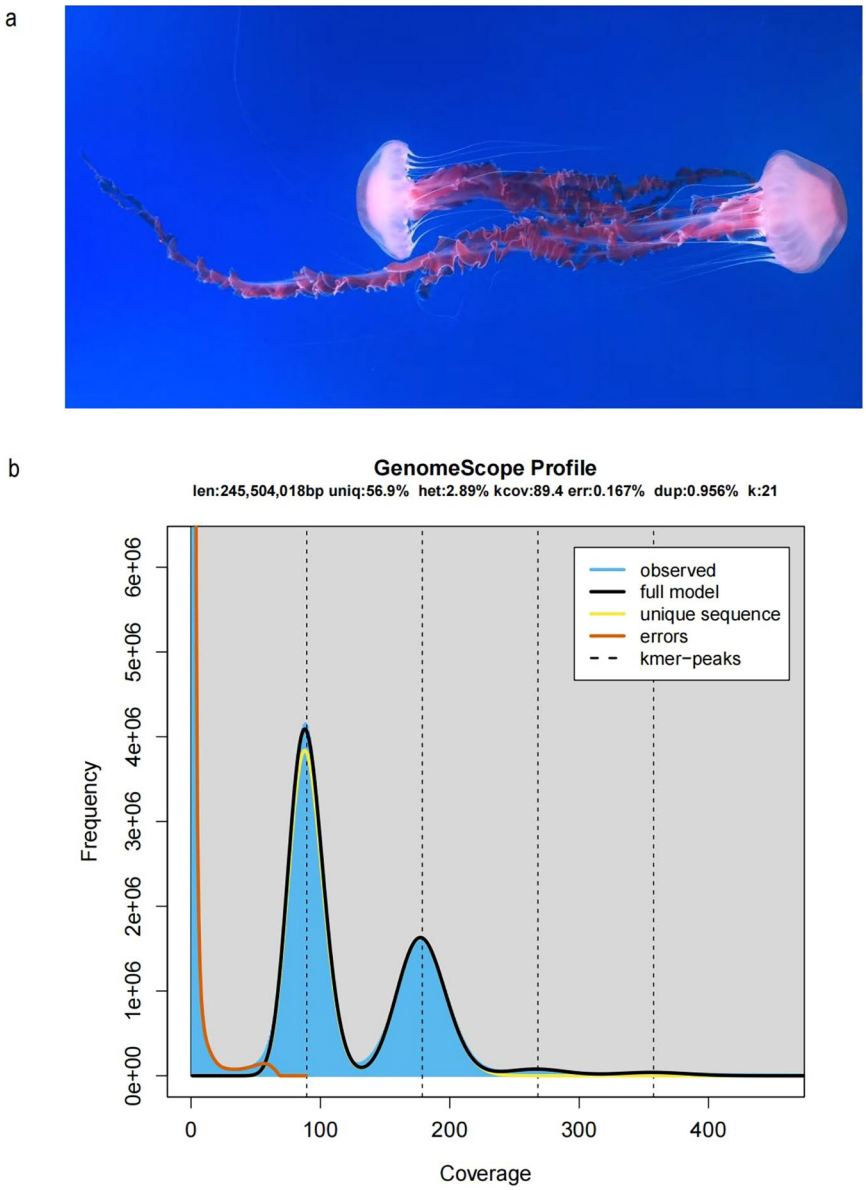

Fig. 1 The characteristics of *Chrysaora achlyos*. (**a**) The picture of the *Chrysaora achlyos*. (**b**) The genomescope plot of the *Chrysaora achlyos* genome survey.

In this study, we present the first haplotype-resolved, chromosome-scale genome assembly of *C.achlyos*, constructed through an integrative approach combining PacBio HiFi sequencing and Hi-C technology. The assembled *C.achlyos* genome had two haplotypes with approximately 246.42 Mb and 248.65 Mb, with 21 chromosomes. Through the BUSCO assessment, the assembly exhibited the completeness value of 94.1% for haplotype A and 93.7% for haplotype B, respectively. The assembled haplotype genomes contained 20,471 and 20,606 predicted genes, respectively, with over 88.3% of them functionally annotated. This haplotype-resolved genome provides an invaluable foundational resource for toxicology, ecological studies, and comparative genomics and evolution.

Methods

Sample collection. One cultured *C.achlyos* medusa individual was collected from the aquarium system at Nanjing Jinniu Lake Wildlife Kingdom (32°28′01″N, 118°57′58″E) on October 28, 2024. The sequenced specimen was approximately three months old at the time of sampling (Fig. 1a). The live individual was promptly transported to the laboratory at BGI-tech (Wuhan) for subsequent sequencing. After starving for 2 days, tissue samples from the bell, oral arms and tentacles were collected. To ensure preservation, all samples were rapidly flash-frozen in liquid nitrogen immediately after collection and stored at −80 °C.

For PacBio long-read sequencing, genomic DNA was extracted using the QIAGEN Blood & Cell Culture DNA Midi Kit (QIAGEN, Germany). The quality of the extracted DNA was assessed through measurements of

	Insert Size	Total length (Gb)	Genome depth	N50 length of reads (bp)
NGS data	300–400 bp	52.70	212.50	150
PacBio HIFI	20 kb	38.97	157.14	20765
Hi-C	300–500 bp	109.35	440.93	150

Table 1. Summary of whole genome sequencing data of *Chrysaora achlyos* genome. Note: Te sequencing depth is calculated based on a genome size of 246 Mb.

DNA concentration, OD260/280 ratio, OD260/230 ratio, and DNA integrity. The fragmented DNA was purified with AMPure PB magnetic beads, and a HiFi library was constructed using the SMRTbell prep kit 3.0 (PacBio, USA). Sequencing was carried out on the PacBio Revio platform (PacBio, USA), yielding 38.97 Gb of Circular Consensus Sequencing (CCS) reads, corresponding to approximately 158× coverage.

The Hi-C library preparation involved fixing three different tissues of *C.achlyos* in 2% formaldehyde[22]. The fixed tissues were homogenized and centrifuged to isolate nuclei. The cross-linked chromatin was digested with MboI, labeled with biotin, and ligated with T4 DNA ligase. After de-crosslinking, the DNA was purified and fragmented into 300–500 bp segments. Fragments containing interaction sites were enriched using streptavidin magnetic beads to construct the Hi-C library. High-throughput sequencing on the DNBSEQ-T7 platform (BGI, China) produced 109.35 Gb of Hi-C reads, equating to approximately 440× coverage.

For transcriptome sequencing, three tissue samples were utilized for RNA extraction using the TRIZOL Reagent (Invitrogen, Carlsbad, CA, USA) in accordance with the manufacturer's protocol. RNA libraries were prepared and sequenced on the DNBSEQ-T7 platform (BGI, China), resulting in 19.88 Gb of RNAseq reads, which were utilized for the genome annotation.

Short-read sequencing and genome survey. The genome size of *C.achlyos* was estimated through short-read sequencing prior to long-read sequencing. The library with a short insert size (300–400 bp) was prepared according to the MGI Library Prep Reagents protocol. A total of 56.80 Gb raw data were generated using the PE 150 sequencing strategy on the DNBSEQ-T7 platform. The adapter and low-quality reads were removed and filtered using Fastp (v0.20.0) pipeline with the parameter (--average_qual 15 -l 150 -w 10)[23]. The high-quality clean short-read data totaling 52.70 Gb were obtained. A 21-mer was utilized to determine the k-mer frequency, and the distribution profile was generated using Jellyfish (v2.2.10)[24]. Genome size and heterozygosity were estimated using GenomeScope 1.0[25]. The estimated genome size is 246.0 Mb with heterozygosity rate of 2.89% (Fig. 1b).

Long-read sequencing and Hi-C sequencing. With the estimated genome size of approximately 246.0 Mb, the PacBio sequencing were employed for genome assembly. A PacBio library (insert size ~20 kb) was constructed for *C.achlyos* following the SMRTbell Express Template Prep Kit 2.0 protocol (Pacific Biosciences, USA). A total number of 1,931,158 highly accurate long reads were sequenced using the PacBio Revio SMRT cell with HiFi model. A total of 38.97 Gb HiFi long reads, with an N50 length of 20,765 bp, were generated for the genome assembly (Table 1). The Hi-C data can be useful for anchoring the contigs to chromosomes. In this study, approximately 1.5 g of muscle tissue from *C.achlyos* was fixed in 1% formaldehyde at room temperature for a duration of 10 to 30 minutes. The cross-linked DNA of *C.achlyos* was digested using MboI (New England Biolabs, Ipswich, USA). The resulting cohesive ends were labeled with biotin through incubation with biotin-14-dATP and Klenow fragment. Following the completion of proximity ligation, crosslinking reversal, and DNA purification, the Hi-C products were subsequently enriched and utilized for library construction. Then, a total of 116.90 Gb of Hi-C raw data were sequenced with DNBSEQ-T7 platform. The raw reads were processed using Fastp v0.20.0 pipeline with the parameter (--average_qual 15 -l 150 -w 10)[23]. A total of 109.35 Gb of clean data, with a Q20 value of 98.11%, was utilized to anchor the contigs to chromosomes (Table 1).

Genome assembly. PacBio HiFi reads and Hi-C short reads were combined in Hifiasm (v0.24.0)[26], employing the Hi-C Integrated Assembly mode. Haplotype-phased contigs were constructed by co-assembling with phasing parameters (--h1/--h2), resulting in two haplotype assemblies, designated as haplotype A and haplotype B. To anchor contigs onto chromosomes, Hi-C clean data were mapped to the assembled contigs using BWA (v0.7.12)[27]. Erroneous mappings (MAPQ = 0) and duplicates were subsequently filtered using the Juicer pipeline (v1.5)[28] to generate the interaction matrix. Approximately 243 Mb reads (67%) were then utilized to anchor the contigs into chromosomes via the 3D-DNA pipeline (v180,922)[29], which was also employed to remove selected short contigs using default parameters.

Manual corrections were applied using Juicebox (v2.17.00)[28] to rectify misinsertions and optimize contig orientations, thereby ensuring the overall chromosome structure. The final assembly is 495.07 Mb (Table 2). The statistics of genome after Hi-C anchoring assembly (Table 3). Among them, 488.60 Mb representing approximately 98.69% of the total genome size, were anchored to 42 chromosomes (Figs. 2, 3).

To compare the genome consistency between the haplotype A and the haplotype B, we used MuMmer (v4.0.0)[30] to identify similar regions with parameters "--maxmatch -c 65 -b 2000 -l 200" at the genome level (Fig. 4).

Genome annotation. The assembled *C.achlyos* genome was annotated by identifying repetitive elements and protein-coding genes. For repeat identification, the homolog and ab initio annotation were integrated to

	Contig		Scaffold	
	Size(bp)	Number	Size(bp)	Number
N10	13,350,082	2	15,720,866	2
N20	11,868,334	4	14,342,470	4
N30	10,170,320	7	14,033,379	5
N40	9,393,571	9	13,065,937	7
N50	7,686,517	12	11,675,418	9
N60	6,192,464	16	11,343,892	11
N70	4,580,992	20	10,447,002	14
N80	3,441,839	26	9,201,437	16
N90	2,364,992	35	8,575,064	19
Max Length	14,388,921	—	17,903,875	—
Total Length	246,390,916	142	246,425,416	73
GC Ratio	0.37	—	0.37	—
N10	13,170,273	2	15,792,483	2
N20	11,219,524	4	14,584,477	4
N30	9,028,433	7	14,060,771	5
N40	7,764,343	10	12,782,586	7
N50	7,228,620	13	11,729,090	9
N60	6,078,872	17	11,559,873	12
N70	4,321,756	22	10,964,054	14
N80	3,492,275	28	9,335,182	16
N90	2,247,300	37	8,493,434	19
Max Length	13,430,372	—	16,375,998	—
Total Length	248,612,092	116	248,649,092	42
GC Ratio	0.37	—	0.37	—

Table 2. (1) The assembly statistics of the Chrysaora achlyos hapA genome. (2) The assembly statistics of the *Chrysaora achlyos* hapB genome.

hapA Chromosome ID	hapA Size (bp)	hapB Chromosome ID	hapB Size (bp)
aChr01	17,903,875	bChr01	16,375,998
aChr02	15,720,866	bChr02	15,792,483
aChr03	14,342,470	bChr03	14,614,023
aChr04	14,388,921	bChr04	14,584,477
aChr05	14,033,379	bChr05	14,060,771
aChr06	13,065,937	bChr06	13,170,273
aChr07	13,350,082	bChr07	12,782,586
aChr08	11,544,160	bChr08	11,729,090
aChr09	11,675,418	bChr09	11,574,143
aChr10	10,447,002	bChr10	11,670,947
aChr11	11,343,892	bChr11	11,299,100
aChr12	11,968,862	bChr12	11,559,873
aChr13	10,381,174	bChr13	11,749,783
aChr14	11,039,525	bChr14	10,964,054
aChr15	10,502,991	bChr15	10,850,196
aChr16	9,201,437	bChr16	9,259,037
aChr17	9,192,191	bChr17	9,335,182
aChr18	8,964,271	bChr18	9,028,433
aChr19	8,374,520	bChr19	8,335,984
aChr20	8,575,064	bChr20	8,493,434
aChr21	7,686,517	bChr21	7,663,928

Table 3. The statistics of genome after Hi-C anchoring assembly.

annotate the repeat sequences. The homologous sequences of *C.achlyos* genome were detected and classified using RepeatProteinMask (v4.0.7)[31] and RepeatMasker (v4.0.7)[32] based on the Repbase library[33]. For ab initio annotation, the de novo repetitive element database was constructed using RepeatModeler (v1.0.4)[34] and LTRharvest[35].

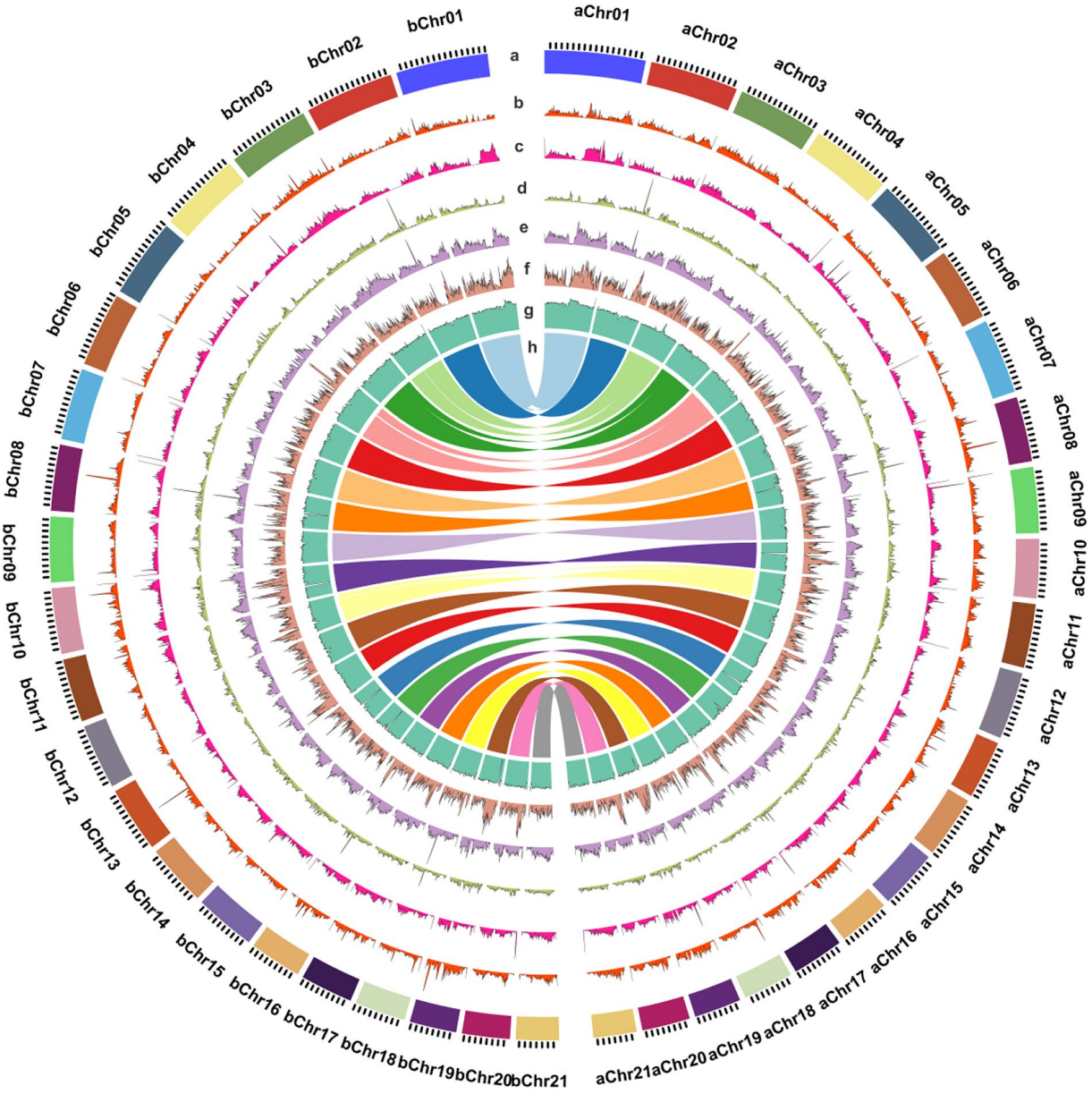

Fig. 2 The circos plot of the *Chrysaora achlyos* genome. (**a**) the length of chromosomes in the size of 1 Mb; (**b**) DNA class of repeats in 100 kb sliding windows; (**c**) LINE of repeats in 100 kb sliding windows; (**d**) LTR of repeats in 100 kb sliding windows; (**e**) DNA transposon of repeats in 100 kb sliding windows; (**f**) gene density in 100 kb sliding windows; (**g**) GC content in non-overlapping 100 kb windows; **(h)** the paths at the center of the figures represent collinear blocks between chromosomes.

The Tandem repeats and long terminal repeats were identified using Tandem Repeat Finder (v4.10.0)[36] and LTR finder (v1.0.7)[37], respectively. Finally, repetitive sequences were characterized, with 110.52 Mb (~44.85%) in haplotype A and 112.57 Mb (~45.27%) in haplotype B (Table 4).

The annotation of protein-coding genes was performed using three strategies including RNA-Seq assisted annotation, *de novo* prediction and homology-based prediction. The transcriptome data were sequenced from three different tissues of *C.achlyos*. The sequenced transcriptome data were mapped onto the two distinct haplotypes genomes of *C.achlyos* using Hisat2 (v2.1.0)[38] with parameters "--sensitive --no-discordant --no-mixed -I 1 -X 500 --max-intronlen 500000". Then, the mapped reads of three samples were assembled using Stringtie (v1.3.5)[39] with parameters (-j 1 -c 1 -f 0.01 -g 50 -s 5). The coding sequence was predicted using TransDecoder (v5.5.0) (https://github.com/TransDecoder/TransDecoder) with default parameters. For *de novo* prediction, the gene models of the *C.achlyos* genome were predicted using Augustus (v3.2.1)[40] with default parameters. For homology-based prediction, protein sequences of seven representative species, namely *Acropora millepora* (GCF_013753865.1), *Acropora muricata* (GCF_036669905.1), *Caenorhabditis elegans* (GCF_000002985.6), *Exaiptasia diaphana* (GCF_001417965.1), *Hydra vulgaris* (GCF_038396675.1), *Nematostella vectensis* (GCF_932526225.1) and *Rhopilema esculentum* (GCF_013076305.1) were retrieved from NCBI (National

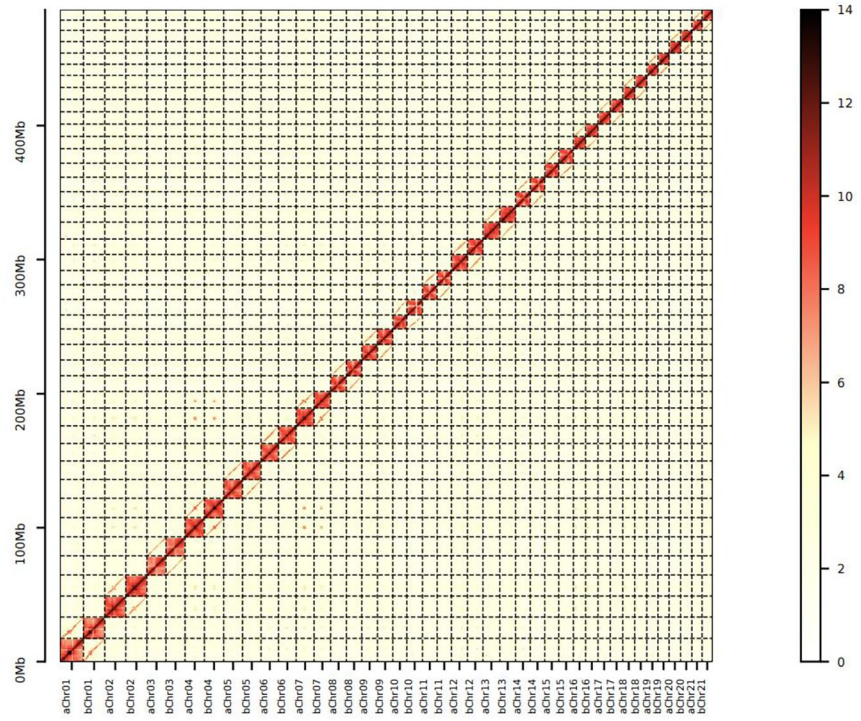

Fig. 3 Hi-C intra-chromosomal interaction map in *Chrysaora achlyos* with 400 Kb.

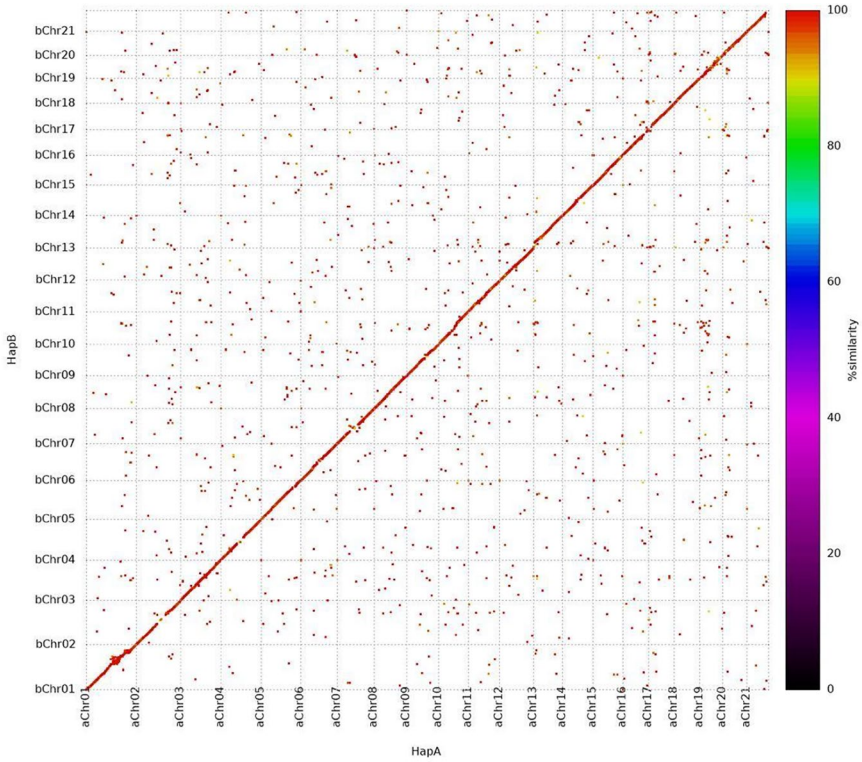

Fig. 4 The sequence comparison between hapA and hapB genomes.

Center for Biotechnology Information). We predicted gene structures using GeMoMa (v1.9)[41], and Evidence Modeler (v1.1.1)[42,43] was further used to merge the results of RNA-Seq-assisted annotation, de novo prediction, and homology-based prediction for the final gene set. Finally, a total of 20,471 and 20,606 predicted genes were identified in haplotype A and haplotype B of the *C.achlyos* genome, respectively (Table 5).

	RepBase TEs		TE Proteins		De novo		Combined TEs	
	Length (bp)	%in Genome	Length (bp)	% in Genome	Length (bp)	% in Genome	Length (bp)	% in Genome
DNA	3,602,536	1.46	415,538	0.17	16,851,130	6.84	19,175,846	7.78
LINE	2,991,042	1.21	1,779,925	0.72	16,586,574	6.73	18,217,187	7.39
SINE	996,153	0.40	0	0.00	3,039,647	1.23	3,751,123	1.52
LTR	4,172,123	1.69	4,778,234	1.94	59,946,850	24.33	61,246,840	24.85
Other	2,836	0.00	0	0.00	0	0.00	2,836	0.00
Unknown	0	0.00	0	0.00	19,009,183	7.71	19,009,183	7.71
Total	11,203,469	4.55	6,973,430	2.83	107,509,530	43.63	110,522,107	44.85
DNA	3,706,090	1.49	509,380	0.20	18,941,988	7.62	21,443,432	8.62
LINE	2,924,885	1.18	1,823,314	0.73	16,227,320	6.53	17,876,809	7.19
SINE	795,182	0.32	0	0.00	3,933,227	1.58	4,424,562	1.78
LTR	4,280,287	1.72	5,126,781	2.06	61,978,827	24.93	63,336,863	25.47
Other	1,327	0.00	0	0.00	0	0.00	1,327	0.00
Unknown	0	0.00	0	0.00	17,350,976	6.98	17,350,976	6.98
Total	11,136,870	4.48	7,459,218	3.00	109,751,444	44.14	112,567,560	45.27

Table 4. (1) The summary of interspersed repeat contents in the hapA genome assembly. Note: This statistical table does not contain Tandem Repeats, some elements may partly include another element domain.Combined: the non-redundant consensus of all repeat prediction/classifcation methods employed. LINE, long interspersed nuclear elements; SINE, short interspersed nuclear elements; LTR, long terminal repeat. (2) The summary of interspersed repeat contents in the hapB genome assembly. Note: This statistical table does not contain Tandem Repeats, some elements may partly include another element domain. Combined: the non-redundant consensus of all repeat prediction/classifcation methods employed. LINE, long interspersed nuclear elements; SINE, short interspersed nuclear elements; LTR, long terminal repeat.

	Gene set	Number	Average gene length (bp)	Average CDS length (bp)	Average exon per gene	Average exon length (bp)	Average intron length (bp)
De novo	Augustus	22409	5931.2	1734	6.8	254.94	723.45
	Snap	20116	10985.63	2003.37	11.23	178.39	878.01
Homolog	*Acropora millepora*	7149	7842.32	1473.82	7.17	205.67	1032.84
	Acropora muricata	6772	8493.05	1506.16	7.76	194.01	1033.06
	Caenorhabditis elegans	706	5866.94	899.9	4.49	200.55	1424.34
	Exaiptasia diaphana	5474	8195.25	1372.67	7.27	188.88	1088.58
	Hydra vulgaris	5388	9316.93	1519.01	8.24	184.32	1076.9
	Nematostella vectensis	5645	9565.86	1534.82	8.74	175.7	1038.23
	Rhopilema esculentum	12316	9498.64	1736.48	8.33	208.41	1058.64
RNA-seq	transcript	8052	10166.43	1384.9	8.47	163.57	1176.05
Final	—	20471	6767.33	1740.48	7.2	241.6	810.28
De novo	Augustus	22582	5940.39	1733.66	6.8	254.93	725.23
	Snap	30593	6725.19	1378.86	7.45	185	828.48
Homolog	*Acropora millepora*	7167	7764.83	1481.32	7.09	208.92	1031.7
	Acropora muricata	6800	8387.08	1509.82	7.63	197.81	1036.88
	Caenorhabditis elegans	696	6008.18	893.96	4.48	199.55	1469.65
	Exaiptasia diaphana	5477	8067.69	1373.9	7.18	191.42	1083.58
	Hydra vulgaris	5453	9169.6	1523.75	8.08	188.57	1079.82
	Nematostella vectensis	5606	9446.04	1515.75	8.62	175.87	1040.91
	Rhopilema esculentum	12331	9520.07	1745.76	8.31	210.09	1063.56
RNA-seq	transcript	7881	10251.05	1400.19	8.53	164.22	1176.01
Final	—	20606	6752.79	1738.69	7.18	242.08	811.04

Table 5. (1) The statistics of gene structure prediction in *Chrysaora achlyos* hapA genome. (2) The statistics of gene structure prediction in *Chrysaora achlyos* hapB genome.

Functional annotation. The predicted gene sets of both haplotypes were functionally annotated using diamond (v0.8.23) with an E-value threshold of 1E-5 based on the five databases including NR (NCBI nonredundant protein), TrEMB (http://www.uniprot.org), KOG, KEGG (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg/), Swiss-Prot (http://www.gpmaw.com/html/swiss-prot.html). The protein motifs and domains were identified using the InterProScan with InterPro 93.0[44]. The GO Ontology (GO) was classified from

		Total	NR	Swissprot	KEGG	KOG	Trembl	Interpro	GO	Overall
hapA	Number	20,471	17,736	11,574	14,340	11,114	17,618	15,891	9,696	18,179
	Percentage(%)	100	86.64	56.54	70.05	54.29	86.06	77.63	47.36	88.80
hapB	Number	20,606	17,829	11,518	14,494	11,162	17,662	15,790	9,654	18,201
	Percentage(%)	100	86.52	55.89	70.34	54.17	85.71	76.62	46.85	88.32

Table 6. The functional annotations for predicted genes in *Chrysaora achlyos* hapA and *Chrysaora achlyos* hapB.

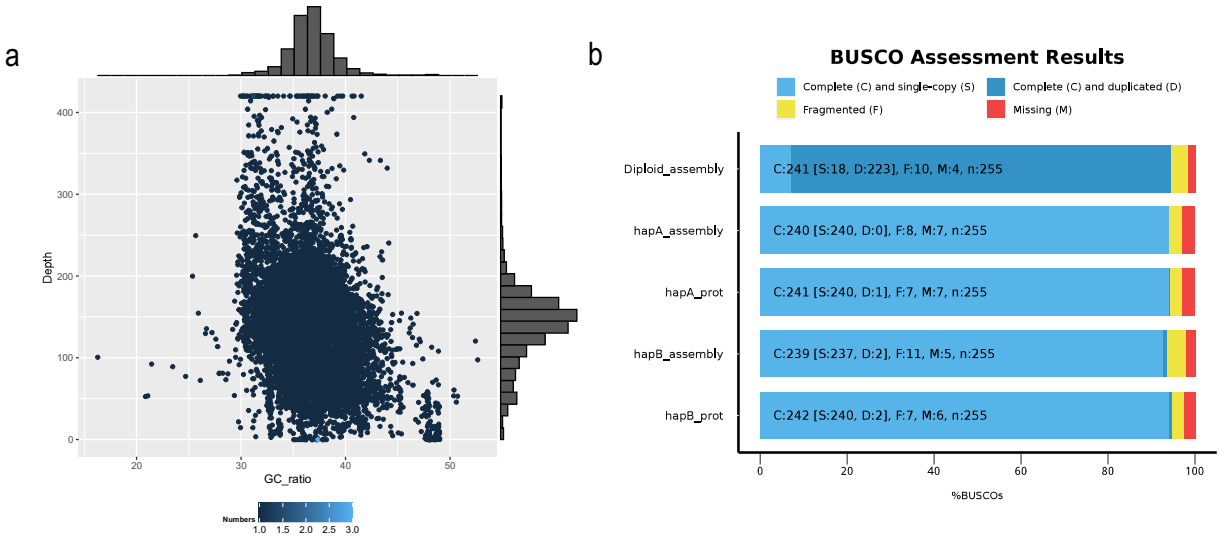

Fig. 5 Genome evaluation of *C.achlyos*. (**a**) GC content and depth distribution of HiFi reads in *C. achlyos* genome. (**b**) The BUSCO evaluation results of the assembly and annotation results.

the results of InterProScan[45]. Finally, a total of 20,471 and 20,606 protein-coding genes were annotated from hapA and hapB, respectively. 97.91% of the 40,220 proteins were annotated by at least one database (Table 6).

Data Records

The sequencing data of *C.achlyos* have been deposited in the NCBI SRA database under BioProject accession PRJNA1303233[46], including DNA short-read survey data (SRA: SRR34916834), DNA short-read Hi-C data (SRA: SRR34916836), DNA long-read PacBio HiFi data (SRA: SRR34916835), and RNA short-read data (SRR34916837-SRR34916839). This Whole Genome Shotgun project has been deposited at GenBank under the accession JBVTIV000000000[47] and JBVTIW000000000[48]. The files of genome assembly and gene structure annotation have been deposited at Figshare[49].

Technical Validation

Evaluation of the *C.achlyos* genome quality and gene set. To evaluate the quality of the *C.achlyos* genome, the short-reads were mapped to the haplotype A genome with a mapping rate of 98.67% and the haplotype B genome with a mapping rate of 98.74% using BWA (v0.7.12)[27]. Additionally, the HiFi long-reads were also mapped to the *C.achlyos* genome using Minimap2[50], the haplotype A genome with the mapping rate of 99.81% and the haplotype B genome with the mapping rate of 99.95%. The analysis of sequencing depth and GC content based on HiFi reads revealed that the GC values were tightly clustered around 40% with a pronounced peak, suggesting minimal contamination (Fig. 5a).

To ensure the purity of the *C. achlyos* genome assembly, we performed contamination screening using fcs-gx (v0.5.5)[51] via the Galaxy platform (https://usegalaxy.org/) with parameters strictly aligned to the National Center for Biotechnology Information (NCBI) genome submission requirements. The results revealed low levels of potential contaminants in both haplotypes (haplotype A and haplotype B). The cleaned haplotype A and haplotype B genomes meet NCBI's purity threshold (<1% contaminant content), ensuring reliability for subsequent analyses. All identified contaminant sequences were systematically removed from the final assemblies.

The Quality Value (QV) of the haplotype A assembly estimated using Merqury (v1.3)[52] is 66.80, with a k-mer completeness of 70.65. For the haplotype B assembly, the QV is 70.89, and the k-mer completeness is 71.08 (Table 7). The quality and completeness of the genome assembly were rigorously assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO, v5.4.3)[53]. The BUSCO were used based on the eukaryota_odb10 dataset. For haplotype A and haplotype B, the BUSCO evaluation of genome were 94.1% and 93.7%, respectively (Table 8). The BUSCO evaluation of the gene set demonstrated that 94.5% of the BUSCOs were classified as Complete in haplotype A, with 94.1% being single-copy and complete BUSCOs, while 0.4% were identified as

Description	hapA	hapB
Consensus quality (QV)	66.80	70.89
K-mer completeness	70.65	71.08

Table 7. The QV of *Chrysaora achlyos* hapA and *Chrysaora achlyos* hapB.

Type	hapA		hapB	
	Genome	Percentage (%)	Genome	Percentage (%)
Complete BUSCOs (C)	240	94.1	239	93.7
Complete and single-copy BUSCOs (S)	240	94.1	237	92.9
Complete and duplicated BUSCOs (D)	0	0.0	2	0.8
Fragmented BUSCOs (F)	8	3.1	11	4.3
Missing BUSCOs (M)	7	2.8	5	2.0
Total BUSCO groups searched	255	—	255	—

Table 8. Genome evaluation of *Chrysaora achlyos*.

Description	hapA		hapB	
Type	Proteins	Percentage(%)	Proteins	Percentage(%)
Complete BUSCOs (C)	241	94.5	242	94.9
Complete and single-copy BUSCOs (S)	240	94.1	240	94.1
Complete and duplicated BUSCOs (D)	1	0.4	2	0.8
Fragmented BUSCOs (F)	7	2.7	7	2.7
Missing BUSCOs (M)	7	2.8	6	2.4
Total BUSCO groups searched	255	—	255	—

Table 9. The BUSCOs evaluation of gene set in *Chrysaora achlyos* hapA and *Chrysaora achlyos* hapB.

duplicated and complete BUSCOs. For haplotype B, the BUSCO assessment revealed comparable completeness: 94.9% of BUSCOs were classified as Complete (94.1% single-copy and complete, 0.8% duplicated and complete (Table 9). Both haplotypes exhibited a high score of completeness with over 93% (Fig. 5b).

These results indicated that the assembly covered most of the genetic regions, further confirming the high quality of the assembled genomes for both haplotypes.

Data availability

The raw sequencing data from the deep mutational scanning experiments have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA1303233 (https://identifiers.org/ncbi/bioproject:PRJNA1303233). The files of genome assembly and gene structure annotation have been deposited at Figshare (https://doi.org/10.6084/m9.figshare.30020194).

Code availability

No specific software or code was developed for this study. The experiments and bioinformatics were performed in accordance with the protocols detailed in the Methods section.

Received: 4 September 2025; Accepted: 27 March 2026;
Published online: 09 April 2026

References

1. Weissbourd, B. *et al*. A genetically tractable jellyfish model for systems and evolutionary neuroscience. *Cell* **184**, 5854–5868. e5820 (2021).
2. Tinta, T., Zhao, Z., Bayer, B. & Herndl, G. J. Jellyfish detritus supports niche partitioning and metabolic interactions among pelagic marine bacteria. *Microbiome* **11**, 156 (2023).
3. Irvine, T., Costello, J. H., Gemmell, B. J., Sutherland, K. R. & Colin, S. P. Impactful feeding ecology of a predatory guild of oceanic jellies. *bioRxiv*, 2024.2008. 2019.608650 (2024).
4. Tinta, T., Klun, K. & Herndl, G. J. The importance of jellyfish–microbe interactions for biogeochemical cycles in the ocean. *Limnology and Oceanography* **66**, 2011–2032 (2021).
5. Peng, S. *et al*. Microbiota regulates life-cycle transition and nematocyte dynamics in jellyfish. *Iscience* **26** (2023).
6. López-Martínez, J. *et al*. Multiple reproduction forms in the polyps of the cannonball jellyfish Stomolophus sp. 2: Probable life-cycle reversal. *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology* **339**, 239–252 (2023).
7. Karabulut, A. *et al*. The architecture and operating mechanism of a cnidarian stinging organelle. *Nature Communications* **13**, 3494 (2022).
8. Park, S. *et al*. The nematocyst's sting is driven by the tubule moving front. *Journal of The Royal Society Interface* **14**, 20160917 (2017).
9. Bosch-Belmar, M. *et al*. Jellyfish impacts on marine aquaculture and fisheries. *Reviews in Fisheries Science & Aquaculture* **29**, 242–259 (2020).

10. Báez, J. C. *et al*. Effects of environmental conditions and jellyfish blooms on small pelagic fish and fisheries from the Western Mediterranean Sea. *Estuarine, Coastal and Shelf Science* **264**, 107699 (2022).
11. Sweetman, A. K. & Chapman, A. First assessment of flux rates of jellyfish carcasses (jelly-falls) to the benthos reveals the importance of gelatinous material for biological C-cycling in jellyfish-dominated ecosystems. *Frontiers in Marine Science* **2**, 47 (2015).
12. Sweetman, A. K. *et al*. Jellyfish decomposition at the seafloor rapidly alters biogeochemical cycling and carbon flow through benthic food-webs. *Limnology and Oceanography* **61**, 1449–1461 (2016).
13. Gold, D. A. *et al*. The genome of the jellyfish Aurelia and the evolution of animal complexity. *Nature ecology & evolution* **3**, 96–104 (2019).
14. Kim, H.-M. *et al*. The genome of the giant Nomura's jellyfish sheds light on the early evolution of active predation. *BMC biology* **17**, 28 (2019).
15. Ohdera, A. *et al*. Box, stalked, and upside-down? Draft genomes from diverse jellyfish (Cnidaria, Acraspeda) lineages: Alatina alata (Cubozoa), Calvadosia cruxmelitensis (Staurozoa), and Cassiopea xamachana (Scyphozoa). *GigaScience* **8**, giz069 (2019).
16. Li, Y. *et al*. Chromosome-level reference genome of the jellyfish Rhopilema esculentum. *GigaScience* **9**, giaa036 (2020).
17. Martin, J. W., Gershwin, L.-A., Burnett, J. W., Cargo, D. G. & Bloom, D. A. Chrysaora achlyos, a remarkable new species of scyphozoan from the Eastern Pacific. *The Biological Bulletin* **193**, 8–13 (1997).
18. Radwan, F. F., Gershwin, L.-A. & Burnett, J. W. Toxinological studies on the nematocyst venom of Chrysaora achlyos. *Toxicon* **38**, 1581–1591 (2000).
19. Pal, K., Forcato, M. & Ferrari, F. Hi-C analysis: from data generation to integration. *Biophysical reviews* **11**, 67–78 (2019).
20. Xia, W.-X. *et al*. High-continuity genome assembly of the jellyfish Chrysaora quinquecirrha. *Zoological Research* **42**, 130 (2021).
21. Nisa, S. A. *et al*. Jellyfish venom proteins and their pharmacological potentials: A review. *International Journal of Biological Macromolecules* **176**, 424–436 (2021).
22. van Berkum, N. L. *et al*. Hi-C: a method to study the three-dimensional architecture of genomes. *J Vis Exp*, https://doi.org/10.3791/1869 (2010).
23. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
24. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
25. Vurture, G. W. *et al*. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
26. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature methods* **18**, 170–175 (2021).
27. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754–1760 (2009).
28. Durand, N. C. *et al*. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell systems* **3**, 95–98 (2016).
29. Dudchenko, O. *et al*. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
30. Marçais, G. *et al*. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol* **14**, e1005944, https://doi.org/10.1371/journal.pcbi.1005944 (2018).
31. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile Dna* **6**, 11 (2015).
32. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
33. Jurka, J. *et al*. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research* **110**, 462–467 (2005).
34. Chen, N. Using Repeat Masker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics* **5**, 4.10. 11–14.10. 14 (2004).
35. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC bioinformatics* **9**, 18 (2008).
36. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* **27**, 573–580 (1999).
37. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic acids research* **35**, W265–W268 (2007).
38. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* **37**, 907–915 (2019).
39. Kovaka, S. *et al*. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome biology* **20**, 278 (2019).
40. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
41. Keilwagen, J., Hartung, F. & Grau, J. GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. *Gene prediction: Methods and protocols*, 161–177 (2019).
42. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**, R7, https://doi.org/10.1186/gb-2008-9-1-r7 (2008).
43. Korf, I. Gene finding in novel genomes. *BMC bioinformatics* **5**, 59 (2004).
44. Mulder, N. & Apweiler, R. InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol Biol* **396**, 59–70, https://doi.org/10.1007/978-1-59745-515-2_5 (2007).
45. Ashburner, M. *et al*. Gene ontology: tool for the unification of biology. *Nature genetics* **25**, 25–29 (2000).
46. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP607485 (2025).
47. Li, Y.-Y. whole genome shotgun sequencing project. *GenBank* https://identifiers.org/ncbi/insdc:JBVTIV000000000 (2026).
48. Li, Y.-Y. whole genome shotgun sequencing project. *GenBank* https://identifiers.org/insdc:JBVTIW000000000 (2026).
49. Li, Y.-Y. Haplotype-resolved chromosomal-level genome assembly of Chrysaora achlyos. *figshare. Dataset.* https://doi.org/10.6084/m9.figshare.30020194 (2025).
50. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574, https://doi.org/10.1093/bioinformatics/btab705 (2021).
51. Astashyn, A. *et al*. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol* **25**, 60, https://doi.org/10.1186/s13059-024-03198-7 (2024).
52. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**, 245, https://doi.org/10.1186/s13059-020-02134-9 (2020).
53. 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, https://doi.org/10.1093/bioinformatics/btv351 (2015).

Acknowledgements

This work was financially supported by the Shanghai Natural Science Foundation General Project (25ZR1401400), and the Military Project (CHJ24C025, CB2023A03).

Author contributions

J.S.Y., J.B.J. and X.Y.G. designed and conceived the study. X.L.Y. and X.Y.G. collected and prepared the samples. Y.J.P., Y.L.F. and J.B.J. performed bioinformatics analysis. Y.Y.L., X.L.Y., B.B.L. and Y.J.P. wrote the manuscript with significant contributions from X.W., Z.F.H., X.H. and S.R.L. J.S.Y. provided the financial support. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.J., X.G. or J.Y.

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.