



OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly and annotation of the moon jellyfish *Aurelia coerulea*

Xinyue Hu^{1,2,3,6}, Yang Guo^{4,6}, Ze Zhang^{1,2,3,5,6} , Liyuan Wang⁴, Duoyuan Chen^{2,3}, Zhenkun Zhuang^{2,3}, Minxiao Wang^{1,4} , Shiping Liu^{1,2,3} & Song Feng⁴

Aurelia coerulea (moon jellyfish), a radially symmetrical metazoan in the phylum Cnidaria, possesses key features pertinent to understanding the evolutionary origins of nervous systems. Here, we employed a combination of long-read sequencing, short-read sequencing, and Hi-C chromatin conformation capture techniques to generate a chromosome-level assembly of the *A. coerulea* genome. The final assembly comprises approximately 554.10 Mb distributed across 21 chromosomes, achieving a scaffold N50 of 24.06 Mb and demonstrating high completeness (protein BUSCO score: 93.0%). Approximately 71.48% of the genome consists of transposable elements. We identified 26,777 protein-coding genes, of which approximately 72.28% have been functionally annotated. This chromosome-level genome provides an essential resource for elucidating early neural evolution and advancing our understanding of cnidarian biology.

Background & Summary

The moon jellyfish (*Aurelia coerulea*) is a globally distributed scyphozoan species of ecological and evolutionary importance, serving as a dominant gelatinous zooplankton in coastal marine ecosystems. The moon jellyfish, characterized by a translucent saucer-shaped body, displays considerable size variation, typically ranging from about 10 centimeters to 30 centimeters in diameter. Highly adaptable, this species thrives across diverse marine environments, ranging from tropical to subarctic waters and from open ocean to brackish estuarine habitats¹. *A. coerulea* reproduces both sexually and asexually. Asexual reproduction occurs either through strobilation budding or podocyst. Strobilation involves sequential developmental stages—from polyp to early strobila, late strobila, and finally ephyra. Budding allows new polyps to form directly from a parent polyp^{2,3}. Podocyst was regarded as the dormant stage produced by polyps when suffering from adverse environmental conditions such as starvation.

A. coerulea plays a vital role as a model organism for studying cnidarian development, symbiosis, and environmental adaptation. And it also has been increasingly recognized as an important player in ecosystem functions and biogeochemical cycles^{4,5}. Adult medusae, characterized by their translucent bell-shaped bodies and trailing tentacles armed with stinging cells (cnidocytes), exert profound impacts on marine food webs through predation on planktonic organisms such as copepods and fish larvae⁶. Population blooms of *A. coerulea*, generally considered to result from climate change and human disturbances to coasts (e.g., eutrophication, overfishing, habitat modification, agriculture, etc), profoundly affect ecosystems by disrupting fisheries, clogging power plant cooling systems, and altering planktonic food webs through increased predation on copepods and other microfauna⁷. Also, characterized by its complex medusoid phase, it serves as a crucial subject for genomic analyses aimed at understanding the evolution of morphological complexity in animals. Comparative studies with closely related species retaining only the simpler polyp stage (e.g., corals and sea anemones) highlight *Aurelia*'s

¹College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China. ²State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, Hangzhou, 310030, China. ³Key Laboratory of Spatial Omics of Zhejiang Province, BGI Research, Hangzhou, 310030, China. ⁴Institute of Oceanology, Chinese Academy of Sciences, Qingdao, 266071, China. ⁵Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, Hangzhou 310024; University of Chinese Academy of Sciences, Hangzhou, China. ⁶These authors contributed equally: Xinyue Hu, Yang Guo, Ze Zhang. e-mail: zhangze21@mails.ucas.ac.cn; fengsong@qdio.ac.cn

Category	Reads(M)	Bases(Gb)	Depth(\times)
WGS	2238.78	223.88	404.04
PacBio	1.77	33.37	60.22
Hi-C	637.82	95.67	172.67
RNA-seq	721.27	72.13	130.17

Table 1. Statistics of sequencing data.

Category	Number
Genome size (bp)	554,095,296
Number of contigs	149
Number of chromosome-scale sequences	21
Number of unplaced scaffolds	104
Contig N50 (bp)	20,249,868
Contig N75 (bp)	16,309,453
Scaffold N50 (bp)	24,061,078
GC content (%)	37.37
Number of genes	26,777
Average gene length (bp)	13,449
Genome BUSCO (BUSCO software)	C:87.0%[S:86.7%,D:0.3%],F:6.6%,M:6.4%,n:954
Protein BUSCO	C:93.0%[S:92.5%,D:0.5%],F:1.8%,M:5.2%,n:954
Transcriptome BUSCO	C:92.6%[S:92.2%,D:0.4%],F:2.4%,M:5.0%,n:954
Gene set Completeness (OMArk)	S:87.86%,D:3.2%[U:3.1%,E:0.09%],M:8.94%

Table 2. Statistics of genome assembly.

significance. Its possession of specialized rhopalial sensory organs and adaptation to a pelagic predatory niche make it an outstanding model for elucidating the origins of nervous and sensory systems³.

Although several genomic studies have been conducted on jellyfish species, including species within the class Scyphozoa, a number of publicly available jellyfish genome assemblies, particularly those generated primarily from short read sequencing, remain fragmented. For instance, early draft scyphozoan assemblies such as *Aurelia aurita* (ARSV1.0; scaffold N50 0.18 Mb)⁸, *Nemopilema nomurai* (draft assembly; scaffold N50 2.71 Mb)⁹, and *Chrysaora quinquecirrha* (draft reference; scaffold N50 733.65 kb)¹⁰ illustrate that many available resources still lack chromosome level continuity. Recent genomic studies have further advanced scyphozoan genomics, including the chromosome level *A. coerulea* assembly reported by Dong *et al.*¹¹, which generated an approximately 566 Mb genome, reported a diploid chromosome number of $2n = 44$, and annotated 32,035 protein coding genes.

However, for many cnidarian taxa, broader comparative analyses can still be limited by the quality and contiguity of currently available assemblies, which affects analyses of structural variation, repetitive elements, and synteny. To complement existing resources and to demonstrate a streamlined, reproducible assembly workflow, we present a chromosome-level genome assembly of *A. coerulea* generated using a combination of PacBio HiFi long-read sequencing, short-read whole-genome sequencing (WGS), and Hi-C chromatin conformation capture (Table 1). The assembled *A. coerulea* genome is 554.10 Mb in size with 37.37% GC content. After Hi-C scaffolding, 98.71% of the assembly was anchored to 21 pseudo-chromosomes (Tables 2, 3, Fig. 1). Transposable elements occupied 71.48% of the genome, of which long terminal repeat (LTR) elements account for 39.82% (Table 4). We predicted 26,777 protein-coding genes in the *A. coerulea* genome, and 72.28% of these genes can be functionally annotated using at least one public database (Table 5). The non-coding RNAs (ncRNAs) including tRNA, rRNA, miRNA, and snRNA, were annotated with a total length of 0.64 Mb (Table 6). In total, assessment of the final genome assembly itself using the metazoan Benchmarking Universal Single-Copy Orthologs (BUSCO) set revealed a completeness of 87.0% (C:87.0% [S:86.7%, D:0.3%], F:6.6%). When assessed on the predicted protein-coding gene set, the completeness reached 93.0% (C:93.0% [S:92.5%, D:0.5%], F:1.8%), indicating high quality of the gene annotation.

Across core chromosome scale metrics, our assembly is broadly consistent with the previously published resource, supporting the robustness and reproducibility of chromosome level *A. coerulea* genome references. At the same time, differences in contig level contiguity and predicted gene number between the two assemblies are likely attributable to technical factors such as sequencing data characteristics, assembly and scaffolding decisions, and annotation pipelines rather than biological differences. We also note that minor differences in pseudo chromosome number among assemblies, such as 21 in our resource versus 22 reported previously, can arise from alternative merge and split decisions during Hi-C based scaffolding at ambiguous boundaries, and should be interpreted as assembly supported representations in the absence of cytogenetic validation. Taken together, the two assemblies provide complementary value. One offers stronger contig-level continuity, while the other is supported by deeper sequencing coverage. These high-quality genomic data of the *A. coerulea* represent a valuable resource for further studies on cnidarian phylogeny, genome evolution, and ecological adaptation.

Sequences	Length (bp)	Percentage (%)
Chr1	49,417,701	8.92%
Chr2	35,538,025	6.41%
Chr3	34,177,997	6.17%
Chr4	32,475,843	5.86%
Chr5	31,721,956	5.73%
Chr6	31,029,773	5.60%
Chr7	26,040,675	4.70%
Chr8	25,802,007	4.66%
Chr9	24,061,078	4.34%
Chr10	23,630,219	4.26%
Chr11	23,611,818	4.26%
Chr12	23,311,773	4.21%
Chr13	22,695,286	4.10%
Chr14	22,514,845	4.06%
Chr15	21,875,200	3.95%
Chr16	21,417,956	3.87%
Chr17	20,971,611	3.78%
Chr18	19,491,548	3.52%
Chr19	19,341,539	3.49%
Chr20	19,297,652	3.48%
Chr21	18,499,408	3.34%
Total length	546,923,910	98.71%
Unplaced scaffolds	7,171,386	1.29%

Table 3. Statistics of Hi-C scaffolding.

Methods

Sampling and sequencing. Specimens of *A. coerulea* were collected in the Jiaozhou Bay, China (36.07°N, 120.16°E) and were cultivated in the Institute of Oceanology, Chinese Academy of Sciences. Samples were preserved in RNAlater immediately after sampling (Thermo Fisher Scientific) and stored at -80°C prior to nucleic acid extraction. Total genomic DNA was isolated from the tissue of a single individual using the QIAamp DNA Mini Kit (Qiagen) for subsequent whole-genome sequencing (WGS). For WGS, DNA fragmentation was performed using a Covaris E220 sonicator, followed by size selection of approximately 200 bp fragments with AMPure XP beads (Beckman). The selected fragments underwent eight cycles of PCR amplification and were sequenced on the DNBSEQ platform (BGI) in a paired-end 100 bp configuration (Table 1). High-molecular-weight (HMW) DNA for PacBio HiFi sequencing was obtained from the same individual used for WGS. The extraction followed a CTAB-based method optimized for HMW DNA preparation, including tissue grinding in liquid nitrogen and the final purification with AMPure XP beads (Beckman Coulter) to retain long fragments. The size distribution of DNA was verified using a Femto Pulse system (Agilent). Long-read sequencing was conducted on the PacBio Sequel II system. DNA quality was assessed using Qubit fluorometer (Thermo Fisher Scientific) and pulsed-field gel electrophoresis (BioRad). Subsequently, DNA was sheared to ~ 15 kb using g-TUBE (Covaris), followed by end-repair, and size-selected using BluePippin (Sage Science). One SMRT cell was sequenced in circular consensus sequencing (CCS) mode (Table 1). For Hi-C library construction, *A. coerulea* tissue was freeze-dried, powdered, resuspended in nuclei isolation buffer, and incubated in 0.5% SDS for 10 min at 62°C . Nuclei were then collected by centrifugation, and nuclear DNA was digested using the restriction enzyme *Mbo*I (NEB). Digested fragments were subsequently end-filled, biotinylated, and ligated using T4 DNA ligase (NEB). Post-purification, DNA was sheared, biotinylated fragments were captured using Dynabeads MyOne Streptavidin T1 (Invitrogen), and the captured DNA was amplified and sequenced on the NovaSeq. 6000 platform (Illumina) in a paired-end 150 bp layout (Table 1). To facilitate genome annotation, RNA sequencing (RNA-seq) was performed. Total RNA extracted with TRIzol (Invitrogen) was reverse-transcribed into complementary DNA (cDNA) using HiscriptII (Vazyme). The resulting cDNA fragments were subjected to terminal repair, A-tailing, and connecting adaptors. Strand-specific library was purified with AMPure XP beads (Beckman Coulter) and sequenced on the DNBSEQ platform, generating 7.21 Gb of 100 bp paired-end data.

Genome assembling and Hi-C scaffolding. A genome survey was conducted with WGS data using Jellyfish (v2.2.6)¹² at K-mer 17, resulting in an estimated genome size of 555.84 Mb and heterozygosity of 2.08% for *A. coerulea*. The genome was assembled from PacBio HiFi long-read data utilizing hifiasm (v0.16.1) with parameters: $-k\ 45 -r\ 2 -a\ 2 -m\ 2,000,000 -p\ 20,000 -l\ 0$ ¹³. Following assembly, PacBio reads were realigned to the draft assembly using minimap2 (v2.14)¹⁴, and duplicated regions were identified and removed using Purge_Dups (v1.2.3) with default parameters. Potential contaminant contigs were detected by Kraken2¹⁵, and those classified as Bacteria were removed. The decontaminated contig-level assembly was assessed using BUSCO (v5.2.2)¹⁶ and Compleasm (v0.2.6)¹⁷ with metazoan odb10 (Table 2). The quality control of Hi-C data was performed using

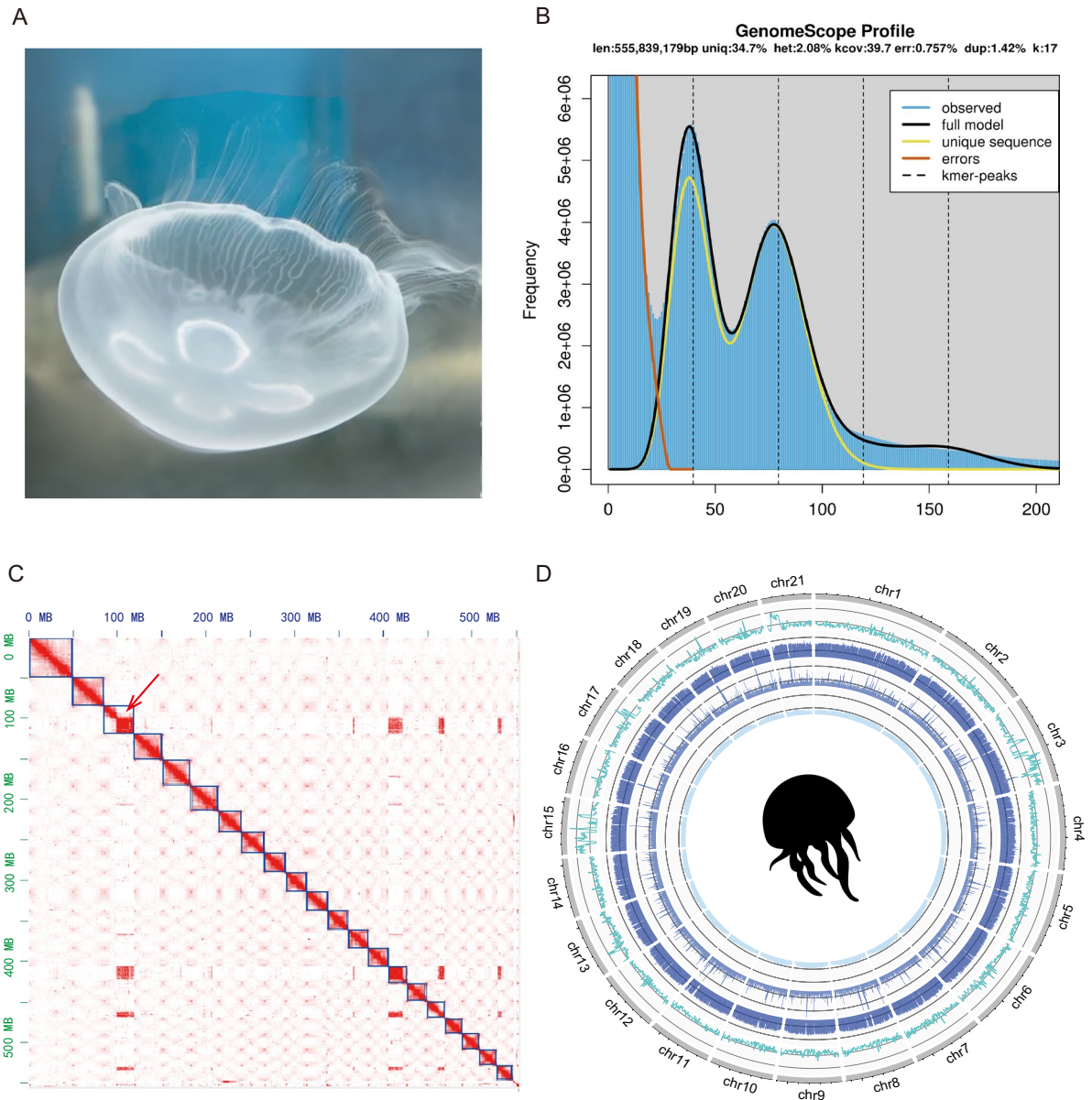


Fig. 1 Genomic characteristics of *A. coerulea*. **(A)** Image of adult *A. coerulea*. **(B)** Genome size estimation using GenomeScope version 1.0 with 17-mer. **(C)** Genome-wide all-by-all Hi-C matrix. The red arrow on Chromosome 3 indicates a potential assembly discontinuity. Users should exercise caution when performing fine-scale downstream analyses in this specific genomic interval. **(D)** Circos view of the assembled chromosomes showing marker distributions at 2-Mb sliding windows from outer to inner circle: gene density, GC content, tandem repeat density, transposable element density.

HiC-Pro v3.2¹⁸ (Table 1), after which contigs were scaffolded using 3D-DNA¹⁹. Assembled chromosomes were visualized and adjusted in Juicebox (v2.17)²⁰, and 98.71% of the contigs were anchored to 21 chromosomes (Table 3, Fig. 1C). The final assembly is 554.10 Mb with a scaffold N50 length of 24.06 Mb (Table 2, Fig. 1D).

Repeat annotation. Repetitive sequences were comprehensively annotated using a combination of de novo, homology-based, and structural searches. Tandem repeats were identified with Tandem Repeats Finder (v4.09.1) using a MaxPeriod of 2000²¹. For transposable elements (TEs), a de novo repeat library was first constructed by integrating outputs from LTR_Finder (v1.0.6, parameter “-C”)²² and RepeatModeler (v1.0.8, default parameters)²³, the latter of which incorporates RECON and RepeatScout. This custom library was then combined with the curated Repbase (v21.01)²⁴ database to create a non-redundant master library. The entire genome was screened against this master library using RepeatMasker (v4.0.6)²⁵ with parameters “-nolow -norna -no_is”. Finally, annotations from all sources were merged, and overlaps were resolved to generate a consolidated set of repetitive elements. The final statistics are summarized in Table 4.

Type	Length(bp)	Percentage of genome(%)
DNA	2,556,700	0.44%
LINE	4,147,021	0.72%
SINE	441,123	0.08%
LTR	167,115,005	29.05%
Unclassified TEs	154,793,227	26.92%
Unknown	82,096,937	14.27%
Total	411,150,013	71.48%

Table 4. Statistics of transposable elements (TE) annotation. Note: Percentages were calculated based on the total physical length of scaffolded chromosomes (575,209,453 bp), including gaps between anchored contigs.

Category	Number	Percentage(%)
Total	26,777	/
Swissprot	9,687	36.18%
KEGG	6,502	24.28%
TrEMBL	16,832	62.86%
Interpro	10,785	40.28%
GO	6,232	23.27%
Overall	19,355	72.28%

Table 5. Statistics of gene functional annotation.

Type	Number	Length (bp)	Percentage (%)	
miRNAs	2	148	0.0001%	
tRNAs	1643	244,825	0.0044%	
rRNAs	rRNA	360	374,239	0.0675%
	18S	59	105,610	0.0191%
	28S	64	240,035	0.0433%
	5S	58	19,778	0.0036%
	5.8S	179	8,816	0.0015%
snRNA	10	21,609	0.0039%	

Table 6. Statistics of ncRNA annotation.

Gene prediction and non-coding RNA annotation. Protein-coding genes were annotated through an integrated pipeline combining *ab initio* prediction, homology-based approaches, and transcriptomic evidence. Augustus (v3.1)²⁶ was utilized for *ab initio* gene prediction. Homology-based annotation involved aligning gene sets from four closely related jellyfish species (*A. aurita atlantic*⁸, *Rhopilema esculentum*²⁷, *Chrysaora quinquecirrha*²⁸, and *Morbakka virulenta*⁸) to the *A. coerulea* genome using Blast (v2.2.26)²⁹. Alignment hits were linked to candidate gene regions using GenBlastA³⁰, and gene models were subsequently predicted using GeneWise (v2.2.0)³¹ based on candidate gene sequences and their flanking 2-kb genomic regions. Additionally, RNA-seq data were mapped to the genome assembly using HISAT (v2.1.0)³². Transcriptomic evidence for annotation was subsequently generated using StringTie (v1.3.4)³², and potential coding regions were identified with TransDecoder (v5.7.1) (github.com/TransDecoder/TransDecoder). Results from the three prediction methods were integrated using EvidenceModeler (EVM) (v1.1.1)³³ with parameters set to --segmentSize 500000 --overlapSize 200000. The integration weights assigned were: AUGUSTUS 1, GeneWise 3, and transdecoder 10. All annotated protein-coding genes were functionally annotated by searching against the public databases, including Swiss-Prot (v201709), KEGG (v87.0), InterPro (v55.0), and TrEMBL (v201709) (Table 5). Completeness of the predicted gene set was assessed using BUSCO (v5.2.2)¹⁶ (Table 2), and the gene annotation quality was assessed using OMArk (v0.3.0)³⁴.

Non-coding RNAs (ncRNAs), including transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs), and microRNAs (miRNAs), were predicted in the genome assembly. tRNAs were identified using tRNAscan-SE (v1.3.1)³⁵ with default parameters. Invertebrate rRNA sequences were aligned against the genome assembly using barrnap (v0.9) (<https://github.com/tseemann/barrnap>) software with its database of eukaryotes. For miRNA and snRNA annotation, candidate regions were first identified by aligning the genome assembly against the Rfam database (v14.1)³⁶ using BLAST software²⁹ (e-value = 1). Subsequently, snRNAs and miRNAs were annotated using INFERNAL (v1.1.1)³⁷ with default parameters (Table 6).

Data Records

All genomic and transcriptomic datasets generated for *A. coerulea* are hosted across the NCBI and Figshare repositories. The final chromosome-level genome assembly is accessible via the NCBI Assembly database under accession JBPOCX000000000³⁸. Supporting this assembly, all raw sequencing reads are archived in the NCBI Sequence Read Archive (SRA) under the study accession SRP596850³⁹ (associated with BioProject PRJNA1284069), which encompasses data from WGS (SRX29512917), PacBio HiFi (SRX29505274), Hi-C (SRX29497767), and four biological replicates of RNA-seq (SRX29505275–SRX29505278). Additional curated genomic resources are provided through a dedicated Figshare⁴⁰ repository. This repository includes the genome assembly sequences in FASTA format (AcoGenome.fasta) and the standardized structural gene models in GFF format (AcoGenome.gff).

Technical Validation

Sequencing was performed on a PacBio library, which yielded long reads with an N50 of 17.7 kb, providing approximately 60× coverage. The assembled genome size was 554.10 Mb, consistent with the 555 Mb estimate from Jellyfish. Assembly accuracy was high, evidenced by a Merquy (v1.3)⁴¹ quality value of 61.35. The assembly comprised 149 contigs exhibiting an N50 of 20.2 Mb and an N75 of 16.3 Mb. Hi-C data processing yielded a valid read rate of 19.36%, and subsequent scaffolding successfully anchored 98.71% of contigs onto 21 chromosomes. Whole-genome sequencing (WGS) reads were aligned to the final genome assembly using BWA MEM (v0.7.17) (github.com/lh3/bwa), and the mapping rate calculated by samtools (v1.9)⁴² (flagstat, excluding secondary mappings) was 98.15%. Furthermore, PacBio HiFi reads aligned with minimap2 (v2.17)¹⁴ (“-ax asm20”) achieved a 92.37% mapping rate. BUSCO analysis (v5.2.2)¹⁶ in protein identified 904 out of 954 BUSCOs (887 complete), corresponding to a completeness of 93.0%. Additionally, Compleasm (v0.2.6)¹⁷ estimated the genome completeness at 92.56%. Gene annotation quality, assessed using OMArk (v0.3.0)³⁴, showed a completeness of 91.06%.

Data availability

All sequencing data, including WGS, PacBio, Hi-C, RNA-seq, as well as the assembly (JBPOCX000000000)³⁸ have been deposited at the NCBI (National Centre for Biotechnology Information) repository under project PRJNA1284069, SRP596850³⁹. The genome assembly and annotations of *A. coerulea* are also available at Figshare⁴⁰.

Code availability

The parameters of the software were default. No specific script was used in this work.

Received: 7 July 2025; Accepted: 19 March 2026;

Published online: 10 April 2026

References

- Ge, M. *et al.* Comparative proteomic analysis of *Aurelia coerulea* for its locomotion system molecular structure-function inference. *J. Proteomics* **209**, 103509 (2019).
- Liu, Q. *et al.* Copper-induced oxidative stress inhibits asexual reproduction of *Aurelia coerulea* polyps. *Ecotoxicol. Environ. Saf.* **285**, 117112 (2024).
- Gold, D. A. *et al.* The genome of the jellyfish *Aurelia* and the evolution of animal complexity. *Nat. Ecol. Evol.* **3**, 96–104 (2019).
- Ohdera, A. H. *et al.* The microbiome of a Pacific moon jellyfish *Aurelia coerulea*. *PLoS One* **19**, e0298002 (2024).
- Sun, T. *et al.* Physiological and transcriptomic responses of *Aurelia coerulea* polyps to acidified seawater conditions. *Mar. Environ. Res.* **196**, 106441 (2024).
- Lewis Ames, C. Medusa: A Review of an Ancient Cnidarian Body Form. *Results Probl. Cell Differ.* **65**, 105–136 (2018).
- Li, X. *et al.* The medusa of *Aurelia coerulea* is similar to its polyp in molecular composition and different from the medusa of *Stomolophus meleagris* in toxicity. *Toxicol. Off. J. Int. Soc. Toxicology* **210**, 89–99 (2022).
- Khalturin, K. *et al.* Medusozoan genomes inform the evolution of the jellyfish body plan. *Nat. Ecol. Evol.* **3**, 811–822 (2019).
- Kim, H.-M. *et al.* The genome of the giant Nomura’s jellyfish sheds light on the early evolution of active predation. *BMC Biol.* **17**, 28 (2019).
- Xia, W. *et al.* Frontiers | High-Quality Genome Assembly of *Chrysaora quinquecirrha* Provides Insights Into the Adaptive Evolution of Jellyfish. <https://doi.org/10.3389/fgene.2020.00535>.
- Dong, Z. *et al.* Genomic and single-cell analyses reveal genetic signatures of swimming pattern and diapause strategy in jellyfish. *Nat. Commun.* **15**, 5936 (2024).
- Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinforma. Oxf. Engl.* **27**, 764–770 (2011).
- 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).
- Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinforma. Oxf. Engl.* **34**, 3094–3100 (2018).
- Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
- Seppy, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol. Biol. Clifton NJ* **1962**, 227–245 (2019).
- compleasm: a faster and more accurate reimplement of BUSCO - PubMed. <https://pubmed.ncbi.nlm.nih.gov/37758247/>.
- Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
- Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* **3**, 95–98 (2016).
- Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* **3**, 99–101 (2016).
- Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
- Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–268 (2007).
- Smit, A. F. & Hubley, R. Repeatmodeler. <https://github.com/Dfam-consortium/RepeatModeler> (2008).
- Smit, A. F. Repeat-Masker Open-3.0. <http://www.repeatmasker.org> (2004).

25. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
26. Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–439 (2006).
27. Li, Y. *et al.* Chromosome-level reference genome of the jellyfish *Rhopilema esculentum*. *GigaScience* **9**, gaa036 (2020).
28. Xia, W.-X. *et al.* High-continuity genome assembly of the jellyfish *Chrysaora quinquecirrha*. *Zool. Res.* **42**, 130–134 (2021).
29. BLAST: improvements for better sequence analysis - PubMed. <https://pubmed.ncbi.nlm.nih.gov/16845079/>.
30. She, R., Chu, J. S.-C., Wang, K., Pei, J. & Chen, N. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res.* **19**, 143–149 (2009).
31. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).
32. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
33. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
34. Quality assessment of gene repertoire annotations with OMArk - PubMed. <https://pubmed.ncbi.nlm.nih.gov/38383603/>.
35. 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).
36. Kalvari, I. *et al.* Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–D200 (2021).
37. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinforma. Oxf. Engl.* **29**, 2933–2935 (2013).
38. NCBI GenBank. https://identifiers.org/ncbi/insdc.gca:GCA_051233565.1 (2025).
39. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRP596850> (2025).
40. Hu, X. *et al.* Chromosome-level genome assembly and annotation of the moon jellyfish *Aurelia coerulea*. 566285649 Bytes figshare <https://doi.org/10.6084/M9.FIGSHARE.29465030.V2> (2025).
41. 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 (2020).
42. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).

Acknowledgements

This work was supported by the National Key Research and Development Program of China (No. 2023YFC3108200), the Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (No. 2022QNLM030004), the National Natural Science Foundation of China (No. 42176136 and No. 42211540390), the “Future Partner Network” Project of the Chinese Academy of Sciences (No. 058GJHZ2022097FN), the Youth Innovation promotion Association Project, Chinese Academy of Sciences, the Huiquan Scholar, Institute of Oceanology, Chinese Academy of Sciences to Song Feng, the Zhejiang Science and Technology Department (No. 2024C03004), and the Hangzhou Leading Innovation Team Project (No. TD2023003).

Author contributions

Y.G. and Z.Zhang. conceived the idea. Y.G. and S.F. collected the sample and performed the experiments. X.H., Y.G., and Z.Zhang. performed the analysis. X.H., Z.Zhang., and L.W. wrote the manuscript. Y.G., D.C., and Z.Zhuang. contributed to the revision of the manuscript. M.W., S.L., and S.F. supervised the study. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Ze Zhang or S.F.

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 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

© The Author(s) 2026