scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A near telomere-to-telomere genome assembly of *Pterocryptis cochinchinensis*

Weiting Chen[1,3], Yuqian Ouyang[2,3], Yuqin Fan[1], Feng Huang[1] & Chao Bian[2] ✉

Pterocryptis cochinchinensis **is a small freshwater catfish distributed across the Southeast Asia and southern China. It plays an important role in maintaining inland aquatic biodiversity and serves as a valuable model for phylogeographic research. However, its genomic resources are still limited. Here, we report a near-complete telomere-to-telomere genome assembly of *P. cochinchinensis*, generated by combining MGI short reads (40.6 Gb), PacBio HiFi reads (51.9 Gb), Oxford Nanopore ultra-long reads (15.0 Gb), and Hi-C data (88.8 Gb). The final assembly spans 931.7 Mb, with a contig N50 of 31.4 Mb, and was anchored onto 28 chromosomes. Repetitive sequences account for 46.9% of its genome, and 31,353 protein-coding genes were annotated. BUSCO assessment indicated a genome completeness of 96.7%. This high-quality reference genome will represent a valuable resource for conservation genetics, aquaculture improvement, and comparative genomics within *Pterocryptis*, contributing to biodiversity preservation and the sustainable utilization of freshwater fish resources.**

Background & Summary

Pterocryptis cochinchinensis is a small freshwater catfish, typically inhabiting slow-flowing mountain streams or freshwater caves, and occurring in Thailand, Myanmar, Vietnam, and southern China[1,2]. Formerly assigned to the genus *Silurus*[3], it has since been reclassified to *Pterocryptis*[4]. As a characteristic freshwater species of Southeast Asia, *P. cochinchinensis* holds considerable significance for taxonomic classification, ecological studies, and biodiversity assessment of inland aquatic ecosystems. Moreover, with its relatively wide distribution across southern China, it represents an ideal model for investigating the phylogeography of freshwater fishes in this region[5].

P. cochinchinensis is distinctly different from most common native catfish species. The dorsal surface is dark brown to purplish, the ventral surface is pale gray, and fins are light gray (Fig. 1). It bears two pairs of barbels, with the maxillary barbels extending to the base of the pelvic fins. The species typically inhabits clear, well-oxygenated streams with gravel or sandy substrates, suggesting high sensitivity to environmental disturbance. *P. cochinchinensis* exhibits pronounced photophobic behavior, remaining hidden under rocks or among aquatic vegetation during daylight and becoming active at night. It is carnivorous, feeding on aquatic insects, small fish, and crustaceans. In recent years, wild populations have shown a marked decline, primarily due to their narrow habitat range and high sensitivity to water quality degradation, raising concerns about potential regional endangerment. To date, no effective large-scale aquaculture techniques have been developed, and this species is mainly encountered through small-scale capture in the wild.

However, comprehensive research on this species remains limited, with existing knowledge of its molecular biology, genetics, and ecology still relatively scarce. This lack of information restricts our ability to gain deeper insights into its biological characteristics and evolutionary patterns and hampers efforts in conservation, utilization, and breeding. Previous studies included only a single article reporting the complete mitochondrial genome. Phylogenetic analyses based on complete mitochondrial genomes and cytochrome *b* (Cyt *b*) sequence data further confirmed that *P. cochinchinensis* is more closely related to other members of the genus *Pterocryptis* than to species of *Silurus*, consistent with previous morphological phylogenetic studies[6]. Recent work has produced a genome assembly for *P. cochinchinensis* using Oxford Nanopore long reads and Illumina short reads,

[1]Guangdong Provincial Key Laboratory of Conservation and Precision Utilization of Characteristic Agricultural Resources in Mountainous Area, School of Life Sciences, Jiaying University, Meizhou, 514015, China. [2]Laboratory of Aquatic Genomics, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, 518057, China. [3]These authors contributed equally: Weiting Chen, Yuqian Ouyang. ✉e-mail: bianchao@szu.edu.cn

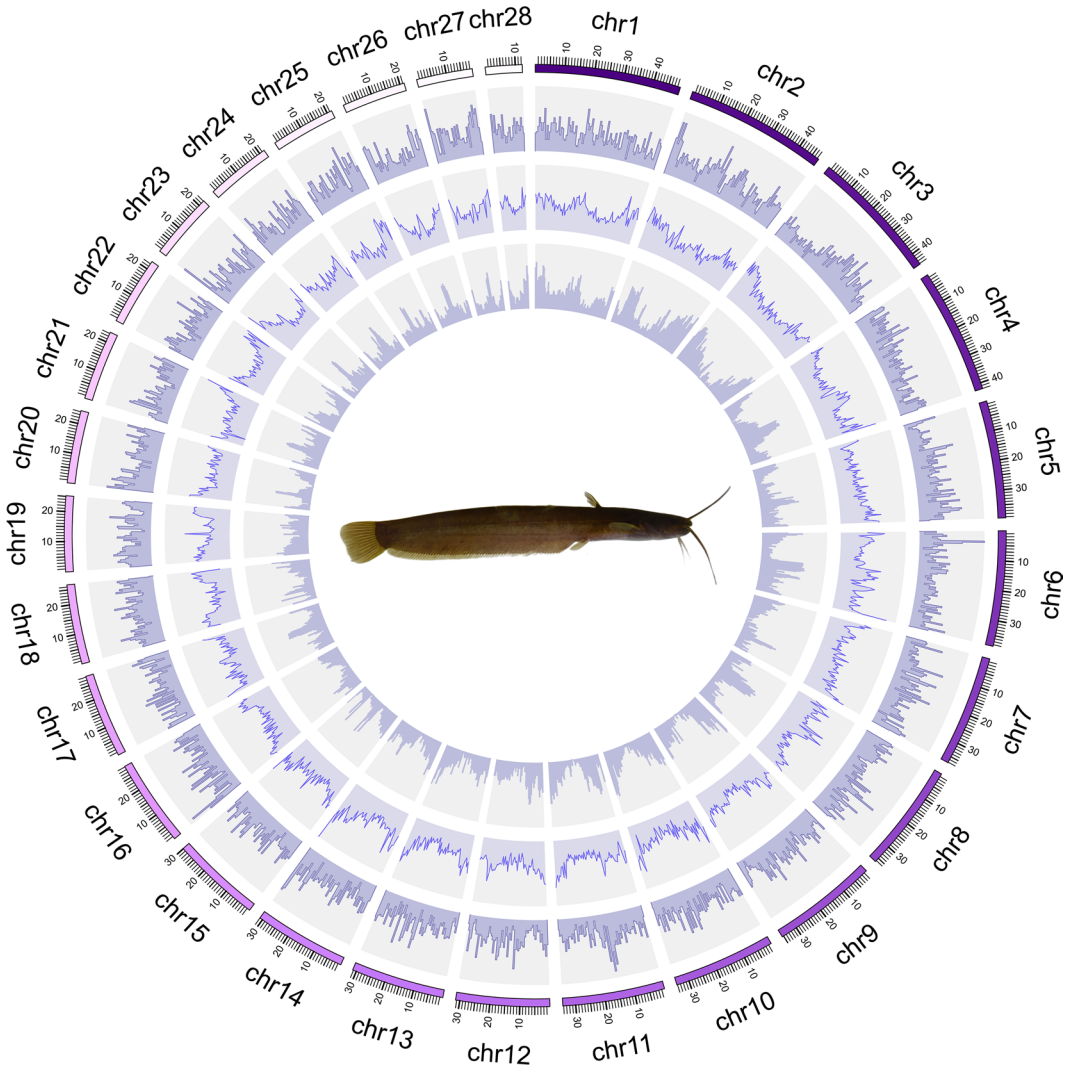

Fig. 1 A Circos plot illustrating the main genomic features. From the outermost to the innermost circles show 28 chromosomes, gene density, GC content, and repetitive sequence density in the *P. cochinchinensis* genome assembly.

alongside chromosome-level assemblies for multiple *Silurus* species, enabling detailed tracing of the evolution of the sex chromosome and the sex-determining gene *amhr2y* across the *Silurus* genus and providing a valuable framework for comparative and evolutionary studies within Siluriformes[7].

In this study, we present a near-complete telomere-to-telomere (T2T) genome assembly of *P. cochinchinensis* that provides a robust foundation for comparative genomics, population genetics, and functional gene discovery. This comprehensive genomic resource also underpins conservation and sustainable management strategies, and opens new avenues for applications in aquaculture, biotechnology, and ecosystem services. Given its ecological sensitivity and potential role as an indicator species for stream health, further integrated studies on the biology, genetics, habitat preferences, and population dynamics of *P. cochinchinensis* are of both theoretical and practical significance.

Materials and Methods

Sample collection, library construction, and genome sequencing. All methods in this study are reported in accordance with ARRIVE guidelines and were performed in accordance with relevant guidelines and regulations under the approval of the Animal Ethics Committee of Jiaying University (Approval No. JYDWLL2026-01).

A female *P. cochinchinensis* individual with 2.5 years old was collected from Shaoguan, Guangdong Province, China. Fresh muscle tissues were dissected and subsequently subjected to a series of genomic analyses, including DNA extraction, library construction, and whole-genome sequencing using multiple sequencing platforms.

For short-insert sequencing (~350 bp), the paired-end libraries were constructed, and paired-end sequencing was subsequently performed by the DNBSEQ-T7 sequencing platform (MGI Tech, BGI). Approximately 40.6 Gb of short-read sequencing data was generated. Low-quality bases, duplicated reads, and adapter sequences were filtered out using Fastp v0.21.0 with default parameters[8]. SMRTbell HiFi sequencing libraries were constructed

Parameter	Value
HiFi reads (Gb)	51.9
ONT reads (Gb)	15.0
Hi-C reads (Gb)	88.8
RNA-seq data (Gb)	11.3
Genome size (Mb)	931.7
Contig N50 (Mb)	31.4
BUSCO	96.7%
Protein BUSCO	94.3%
Merqury QV score	43.4
Repeat ratio	46.9%
Gene number	31,353
Average gene length (bp)	13,190.9
Functional gene number	28,079

Table 1. Statistics of the T2T genome assembly and annotation.

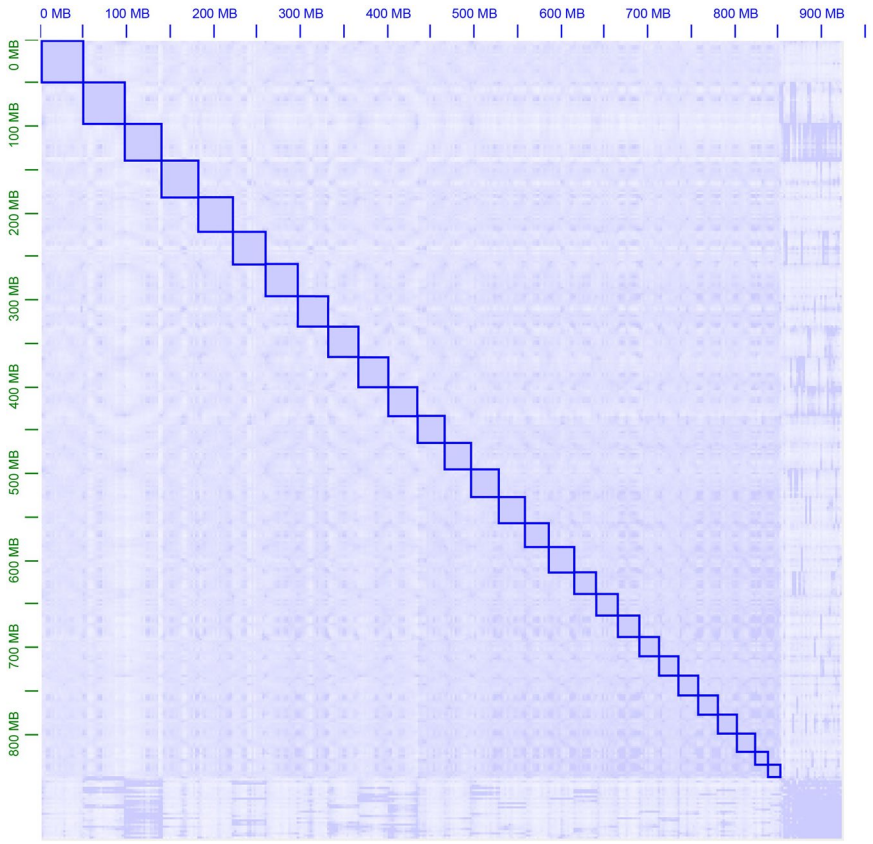

Fig. 2 Genome-wide Hi-C heatmap of the *P. cochinchinensis* genome.

with the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA), and then sequenced on the PacBio Sequel II platform (Pacific Biosciences, Menlo Park, CA, USA). A total of 51.9 Gb of HiFi reads was generated (Table 1). For ONT ultra-long sequencing, the library was constructed using the Ligation Sequencing 1D kit (SQKLSK109, Oxford Nanopore Technologies, Oxford, UK) and subsequently sequenced on a PromethION instrument (Oxford Nanopore Technologies, Oxford, UK), yielding a total of 15.0 Gb of ONT sequencing data (Table 1). Hi-C libraries were prepared using the GrandOmics Hi-C Kit with DpnII digestion, and sequenced on an DNBSEQ-T7 platform (MGI, Shenzhen, China), generating approximately 88.8 Gb of raw reads (Table 1). All resulting reads were subsequently integrated for high-quality genome assembly.

Total RNA was extracted from muscle tissue of *P. cochinchinensis* using TRIzol Reagent (Invitrogen, Waltham, MA, USA). RNA concentration and purity were assessed with a NanoDrop 2000 spectrophotometer (Invitrogen, USA), and RNA integrity was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies,

Chr ID	Chr Length (bp)	Gap number	Telomere number	GC (%)	Centromere start	Centromere end
Chr1	48,944,905	0	2	38.7	13,571,522	13,739,542
Chr2	47,593,778	0	2	39.0	35,780,539	35,942,907
Chr3	44,243,564	0	2	38.9	7,130,821	7,319,996
Chr4	42,061,174	0	2	38.8	26,479,697	26,866,086
Chr5	39,365,477	0	2	38.6	36,923,148	37,034,095
Chr6	38,563,929	0	2	38.9	17,387,551	17,496,483
Chr7	37,186,343	1	2	38.7	434,933	886,934
Chr8	36,451,459	0	2	39.0	348,620	456,698
Chr9	35,506,938	0	2	38.7	1	210,973
Chr10	34,187,571	0	2	38.9	32,507,973	32,835,408
Chr11	34,267,867	0	2	38.8	2,990,272	3,166,949
Chr12	31,427,384	0	2	39.1	6,686,242	7,232,397
Chr13	31,297,520	0	2	38.8	28,121,295	28,250,123
Chr14	31,020,200	0	2	39.2	201,714	330,483
Chr15	30,086,553	0	2	38.8	482,674	691,501
Chr16	29,179,150	0	2	38.8	27,176,469	27,302,460
Chr17	28,270,690	0	2	38.9	4,324,197	4,608,079
Chr18	26,708,902	0	2	39.0	17,955,900	18,088,625
Chr19	25,228,651	1	2	39.1	3,448,407	3,568,860
Chr20	24,397,128	0	2	38.9	22,674,791	22,815,622
Chr21	23,458,928	0	2	38.8	22,21,141	2368,025
Chr22	22,936,506	0	2	39.1	331,681	513,016
Chr23	22,160,249	0	2	39.0	11,189,874	11,358,141
Chr24	21,814,211	0	2	39.0	12,630,630	12,771,046
Chr25	21,658,472	0	2	39.2	7,311,528	7,520,124
Chr26	21,471,095	0	2	39.0	2,184,795	2,294,763
Chr27	18,816,046	0	2	39.1	10,787,383	10,976,346
Chr28	12,347,523	0	2	39.1	1,889,202	2,021,044

Table 2. Summary of assembled chromosomes in the *P. cochinchinensis* genome.

Santa Clara, CA, USA). Only RNA samples with an RNA integrity number (RIN) > 7.0 were used for subsequent library preparation. cDNA libraries were generated using DNA nanoball (DNB) technology for the DNBSEQ platform (MGI Tech, BGI-Shenzhen, China) and sequenced in paired-end mode on the MGISEQ-2000 platform (MGI Tech) with a read length of 150 bp.

Genome size prediction, genome assembly and evaluation. A k-mer-based approach was applied to investigate the genomic characteristics of *P. cochinchinensis*, utilizing the short-insert sequencing reads[9]. The genome size of *P. cochinchinensis* was estimated via the 17-mer depth frequency distribution analysis using the formula: genome size = K_{num}/K_{depth}, where the K_{num} denotes the count of 17-mer number derived from sequencing reads, and the K_{depth} represents the peak depth of k-mers at the highest frequency. Through k-mer analysis, we estimated that its genome size is approximately 863.9 Mb, with an estimated heterozygosity of 0.5% and repeat ratio of 31.3%.

The initial assembly was performed by using NextDenovo v2.5.2[10] with default parameters. Hi-C sequencing reads were aligned to the draft contig assembly using Bowtie2 v2.4.5[11], and valid interaction pairs were extracted with HiC-Pro v2.8.1[12]. Contig anchoring and orientation were performed using Juicer v1.5[13] and 3D-DNA v170123[14]. Juicebox v1.11.08[15] was used to visualize the Hi-C contact matrices, correct errors in the assembly, and remove redundant contigs. Gap filling was performed using LR_GapCloser v1.1[16] and TGS-GapCloser v1.0.1[17]. Telomeric and centromeric regions were identified using the TeloExplorer and CentroMiner modules implemented in the quarTeT software[18]. The final genome assembly was 931.7 Mb in size, with a contig N50 of 31.4 Mb (Table 1), and was anchored onto 28 pseudochromosomes (Figs. 1, 2). We ultimately identified a complete set of 56 telomeres, with telomeric sequences present at both ends of each chromosome, and centromeres detected on all 28 chromosomes (Table 2).

The completeness and accuracy of the final assembly were evaluated using Benchmarking Universal Single Copy Ortholog (BUSCO v5.4.5)[19] with the actinopterygii_odb10 database and Merqury[20], respectively, yielding 96.7% complete BUSCO genes and a Merqury QV score of 43.4 (Table 1), corresponding to a per-base accuracy of approximately 0.9999543. These results indicate the high quality and completeness of our genome assembly.

Repeat element annotation and gene annotation. For repeat annotation, A *de novo* repeat library was constructed using RepeatModeler v2.0.41[21] and LTR-FINDER v1.0.6[22] both with default parameters. RepeatMasker v4.1.7[23] was applied against both the Repbase TE v21.01[24] and the *de novo* repeat library

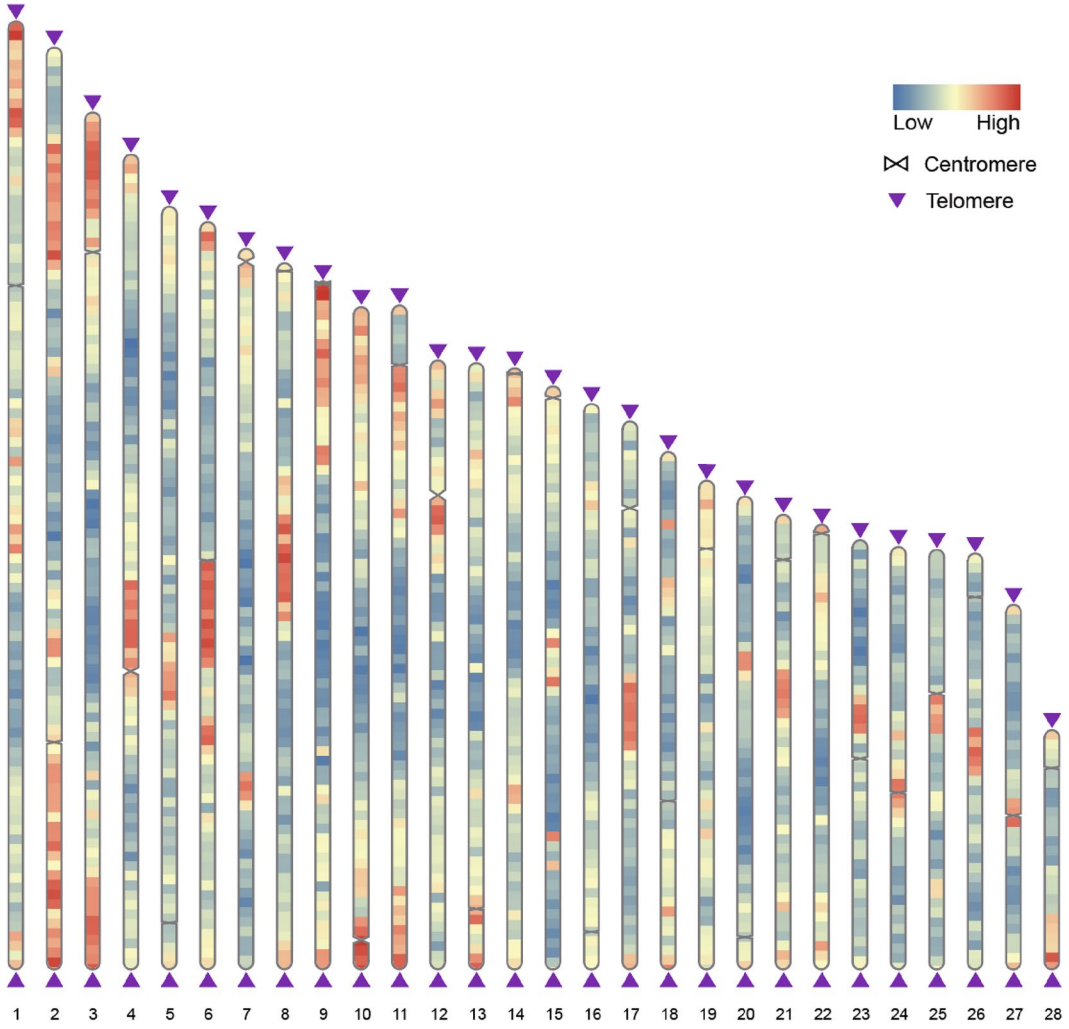

Fig. 3 Genome distribution across whole *P. cochinchinensis* genome.

to identify and mask repetitive sequences. Additionally, we used RepeatProteinMask v4.1.7[23] implemented in RepeatMasker to detect the TE-relevant proteins. We also predicted tandem repeats using Tandem Repeat Finder v4.09[25].

For gene annotation, transcriptome-based prediction and homology-based prediction were integrated to achieve a comprehensive annotation of gene structures. For transcriptome-based prediction, the pooled RNA-seq reads were mapped to the reference genome using HISAT2 v.2.2.1[26], followed by transcript assembly using Cufflinks v2.2.1[27]. For homology-based prediction, homologous protein sequences from four species were downloaded from the NCBI database, including *Danio rerio* (Zebrafish; GCF_000002035.6)[28], *Oryzias latipes* (Medaka; GCF_002234675.1)[29], *Kryptopterus vitreolus* (Glass catfish; GCA_044706155.1)[30] and *Hemibagrus wyckioides* (GCF_019097595.1)[31]. These homologous proteins were aligned to the assembly of *P. cochinchinensis* using tBLASTn[32] with an e-value $\leq 10^{-5}$, respectively. The resulting alignments were refined and used for gene structure prediction with GeneWise[33] v2.4.1. All gene models predicted by the above methods were integrated into a non-redundant gene set using Evidence Modeler (EVM v1.1.1)[34]. Functional annotation of the predicted genes was performed using four public databases: TrEMBL[35], Swiss-Prot[35], InterPro[36], and KEGG[37]. The annotation of protein-coding genes in the *P. cochinchinensis* genome resulted in a total of 31,353 genes. These genes were annotated in functional databases, with an average mRNA length of 13,190.9 bp, containing an average of 8.7 exons per gene and an average exon length of 186.5 bp. Notably, 89.6% (28,079) of the predicted genes were mapped to at least one functional database from above indicated four databases, highlighting the reliability of the gene annotation for *P. cochinchinensis* (Table 1, Fig. 3).

A comprehensive annotation of non-coding RNAs (ncRNAs) was performed using multiple approaches. Transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) were predicted using tRNAscan-SE v2.0.12[38] and RNAmmer v1.2[39], respectively. Furthermore, other conserved ncRNA elements, including miRNAs, snRNAs, and snoRNAs, were annotated by searching against the Rfam database v14.10[40] using the cmscan program implemented in Infernal v1.1.4[41]. Finally, 4,649 ncRNAs were annotated, including 343 miRNAs, 1,448 tRNAs, 1,737 rRNAs and 1,121 snRNAs (Table 3).

Type		Copy	Average length(bp)	Total length(bp)	% of genome
miRNA		343	100.4	34,454	0.003698
tRNA		1,448	75.5	109,255	0.011727
rRNA	rRNA	1,737	100.8	175,168	0.018801
	18S	204	202.0	41,198	0.004422
	28S	179	252.9	45,261	0.004858
	5S	1,344	65.2	87,594	0.009402
snRNA	snRNA	1,121	147.6	165,473	0.017761
	CD-box	99	136.7	13,529	0.001452
	HACA-box	37	162.4	6,010	0.000645
	splicing	976	148.8	145,267	0.015592

Table 3. Non-coding RNA annotation of the genome of *P. cochinchinensis*.

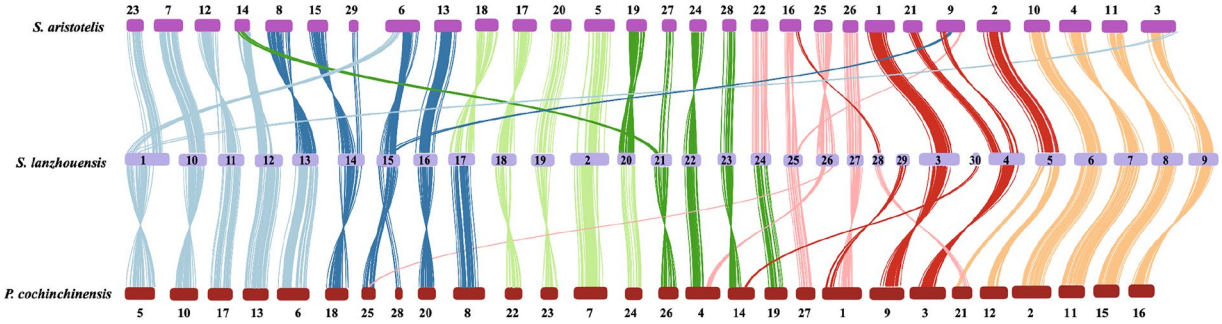

Fig. 4 Genome synteny relationship of *S. aristotelis* (top), *S. lanzhouensis* (middle) and *P. cochinchinensis* (bottom).

Chromosomal synteny analysis. To investigate the chromosomal differences among *P. cochinchinensis, Silurus lanzhouensis* and *Silurus aristotelis*, we performed a comparative chromosomal synteny analysis. The data of *S. lanzhouensis* and *S. aristotelis* were downloaded from NCBI database (*S. lanzhouensis*: GCA_039877035.1; *S. aristotelis*: GCA_946808225.1). The assemblies of the three species were used as input datasets for interspecific synteny analysis with MUMmer v4.0.0 beta1[42]. Homologous alignments between the two genomes were conducted using the Nucmer module. Alignment blocks with a sequence identity greater than 90% and a length exceeding 2 kb were retained for downstream analysis. Finally, chromosomal synteny relationships and structural variations were ultimately visualized through RectChr (https://github.com/BGI-shenzhen/RectChr).

The extensive chromosomal synteny was observed among three species, with the connecting lines covering most of the chromosomes. Their genomic structures have retained a certain degree of conservation. The crossing of chromosomal connection lines and the corresponding fragments in the diagram indicate some structural variations (Fig. 4).

Data Records

The final assembled genome and detailed annotation and protein files are openly available in the Figshare repository (https://doi.org/10.6084/m9.figshare.30244429)[43].

The genomic sequencing data for *P. cochinchinensis* has been archived in the National Genomics Data Center (NGDC) database, accessible via the BioProject accession PRJCA047698[44], and also deposited in to NCBI with accession number: GCA_054855095.1[45]. The raw sequencing reads are available in the NCBI Sequence Read Archive (SRA) under the accession numbers[46]: SRR37059788 (Hi-C), SRR37059789 (HiFi), SRR37059790 (ONT) and SRR37092960 (RNA).

Technical Validation

We generated a near-T2T assembly for *P. cochinchinensis*. It contains 28 chromosomes that comprised a total length of 860.7 Mb with a contig N50 of 31.4 Mb. Genome completeness was evaluated with BUSCO v5.4.5 in genome mode using MetaEuk as the gene predictor and the Actinopterygii_odb10 lineage dataset. The analysis revealed that 96.7% (3,523) of the BUSCO genes were complete, comprising 95.1% (3,463) single-copy and 1.6% (60) duplicated genes, with 0.4% (16) fragmented and 2.9% (101) missing. When assessed at the protein level, the BUSCO v5.4.5 analysis using the same dataset demonstrated a completeness of 94.3% (3,430), with 92.2% (3,355) single-copy and 2.1% (75) duplicated, while 0.9% (34) were fragmented and 4.8% (176) were absent. These research findings fully demonstrate the high completeness and excellent quality of the *P. cochinchinensis* genome assembly and gene annotation. The high quality and reliability of the constructed genome assembly were further validated by a Merqury QV score of 43.4.

Data availability

The raw sequencing data are available in the NCBI SRA under accession numbers[46]: SRR37059788 (Hi-C), SRR37059789 (HiFi), SRR37059790 (ONT) and SRR37092960 (RNA). The genome assembly is available at NCBI with accession GCA_054855095.1[45] and the annotations are available at Figshare (https://doi.org/10.6084/m9.figshare.30244429)[43].

Code availability

All genome assembly and gene annotation were conducted strictly following the standard guidelines and protocols of the respective bioinformatics tools. No custom code was developed for this study.

Received: 13 October 2025; Accepted: 25 March 2026;
Published online: 31 March 2026

References

1. Kobayakawa, M. Systematic revision of the catfish genus *Silurus*, with description of a new species from Thailand and Burma. *Jpn. J. Ichthyol.* **36**, 155–186, https://doi.org/10.11369/jji1950.36.155 (1989).
2. Kottelat, M. *Fishes of Laos* (WHT Publications, 2001).
3. Haig, J. Studies on the classification of the catfishes of the Oriental and Palaearctic family Siluridae. *Rec. Indian Mus.* **48**, 59–116 (1951).
4. Bornbusch, A. H. Redescription and reclassification of the silurid catfish *Apodoglanis furnessi* Fowler (Siluriformes: Siluridae), with diagnoses of three intrafamilial silurid subgroups. *Copeia* **1991**, 1070–1084, https://doi.org/10.2307/1446109 (1991).
5. Wan, C. *Population variation and affinity geography of Silurus cochinchinensis in South China* (Master's thesis, South China Normal University, 2011).
6. Xu, J., Han, C. & Huang, J. R. The complete mitogenome of the sheatfish *Pterocryptis cochinchinensis* (Siluriformes: Siluridae) and phylogenetic implications. *Mitochondrial DNA A* **27**, 2785–2786, https://doi.org/10.3109/19401736.2015.1053066 (2016).
7. Zheng, S. *et al*. The origin, evolution, and translocation of sex chromosomes in *Silurus* catfish mediated by transposons. *BMC Biol.* **23**, 54, https://doi.org/10.1186/s12915-025-02160-8 (2025).
8. Chen, S. fastp 1.0: An ultra-fast all-round tool for FASTQ data quality control and preprocessing. *iMeta* e70078 https://doi.org/10.1002/imt2.70078 (2025).
9. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770, https://doi.org/10.1093/bioinformatics/btr011 (2011).
10. Hu, J. *et al*. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol.* **25**, 107, https://doi.org/10.1186/s13059-024-03252-4 (2024).
11. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359, https://doi.org/10.1038/nmeth.1923 (2012).
12. Servant, N. *et al*. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259, https://doi.org/10.1186/s13059-015-0831-x (2015).
13. Durand, N. C. *et al*. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98, https://doi.org/10.1016/j.cels.2016.07.002 (2016).
14. Dudchenko, O. *et al*. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95, https://doi.org/10.1126/science.aal3327 (2017).
15. Durand, N. C. *et al*. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101, https://doi.org/10.1016/j.cels.2015.07.012 (2016).
16. Xu, G. C. *et al*. LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience* **8**, giy157, https://doi.org/10.1093/gigascience/giy157 (2019).
17. Xu, M. *et al*. TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *Gigascience* **9**, giaa094, https://doi.org/10.1093/gigascience/giaa094 (2020).
18. Lin, Y. *et al*. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Hortic. Res.* **10**, uhad127, https://doi.org/10.1093/hr/uhad127 (2023).
19. 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).
20. 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).
21. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457, https://doi.org/10.1073/pnas.1921046117 (2020).
22. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268, https://doi.org/10.1093/nar/gkm286 (2007).
23. Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **Chapter 4**, Unit 4.10, https://doi.org/10.1002/0471250953.bi0410s05 (2004).
24. Jurka, J. *et al*. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467, https://doi.org/10.1159/000084979 (2005).
25. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580, https://doi.org/10.1093/nar/27.2.573 (1999).
26. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360, https://doi.org/10.1038/nmeth.3317 (2015).
27. Trapnell, C. *et al*. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578, https://doi.org/10.1038/nprot.2012.016 (2012).
28. Howe, K. *et al*. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503, https://doi.org/10.1038/nature12111 (2013).
29. Kasahara, M. *et al*. The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447**, 714–719, https://doi.org/10.1038/nature05846 (2007).
30. Bian, C. *et al*. A telomere-to-telomere chromosome-scale genome assembly of glass catfish (*Kryptopterus vitreolus*). *Sci. Data* **12**, 483, https://doi.org/10.1038/s41597-025-04841-z (2025).
31. Shao, F. *et al*. Chromosome-level genome assembly of the Asian Red-tail catfish (*Hemibagrus wyckioides*). *Front. Genet.* **12**, 747684, https://doi.org/10.3389/fgene.2021.747684 (2021).
32. Mount, D. W. Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc.* **2007**, pdb.top17, https://doi.org/10.1101/pdb.top17 (2007).

33. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995, https://doi.org/10.1101/gr.1865504 (2004).
34. 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).
35. Boeckmann, B. *et al*. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370, https://doi.org/10.1093/nar/gkg095 (2003).
36. Hunter, S. *et al*. InterPro: the integrative protein signature database. *Nucleic Acids Res.* **37**, D211–D215, https://doi.org/10.1093/nar/gkn785 (2009).
37. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361, https://doi.org/10.1093/nar/gkw1092 (2017).
38. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096, https://doi.org/10.1093/nar/gkab688 (2021).
39. Lagesen, K. *et al*. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108, https://doi.org/10.1093/nar/gkm160 (2007).
40. Griffiths-Jones, S. *et al*. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124, https://doi.org/10.1093/nar/gki081 (2005).
41. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935, https://doi.org/10.1093/bioinformatics/btt509 (2013).
42. 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).
43. Bian, C. A near telomere-to-telomere genome assembly of *Pterocryptis cochinchinensis*. *figshare* https://doi.org/10.6084/m9.figshare.30244429 (2025).
44. Bian, C. *Pterocryptis cochinchinensis* genome assembly. *National Genomics Data Center* https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA047698 (2025).
45. *GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_054855095.1 (2025).
46. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP672553 (2025).

Acknowledgements

This study was supported by the Guangdong Province Key Construction Discipline Research Ability Improvement Project (2021ZDJS072), National Key Research and Development Program of China (2023YFE0205100), Startup Fund for the Talents of Jiaying University (2022RC38), Natural Science Foundation of Jiaying University (2023KJY19) and Guangdong Basic and Applied Basic Research Foundation (no.2026A1515011346).

Author contributions

W.C. and C.B. conceived this project; Y.F. and F.H. collected samples; C.B. and Y.O. analyzed the data and wrote the manuscript. W.C., C.B. and Y.O. revised the manuscript. All authors have read and approved the final manuscript for publication.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.B.