scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly of the Chinese algae eater *Gyrinocheilus aymonieri*

Cheng Wang[1,2], Yongrui Lu[1,2], Chengchi Fang[1], Xiaoni Gan[1], Liandong Yang[1]✉ & Shunping He[1]✉

Gyrinocheilus aymonieri **is endemic to Yunnan Province, China. Its mouth has evolved a specialized sucker-like structure as an adaptation to fast-flowing aquatic habitats, representing a remarkable evolutionary example of environmental adaptation. Here, we present a chromosome-level genome assembly of *G. aymonieri* (507 Mb) integrating HiFi, Illumina, and Hi-C sequencing data. The assembly comprises 24 chromosomes with 92.3% anchoring efficiency. Genome annotation revealed 23.02% repetitive elements, 24,078 protein-coding genes, and 55,943 transcript isoforms. The good genomic collinearity also demonstrates the accuracy of the *G. aymonieri* genome assembly at the macroscale. This high-quality genome will provide a crucial resource for understanding cypriniform evolution and different ecological adaptation mechanisms.**

Background & Summary

Gyrinocheilus aymonieri, a freshwater fish native to Southeast Asia, is commonly known as the "Chinese algae eater," "honey sucker," or "sucking loach." This species belongs to the order Cypriniformes, family Gyrinocheilidae, and genus Gyrinocheilus[1]. The family Gyrinocheilidae comprises three small Cypriniform species inhabiting fast-flowing mountain streams. In China, it is endemic to Yunnan Province, specifically distributed in Menghai and Mengla along the lower Lancang River. The species exhibits a specialized sucker-like oral structure for substrate adhesion, facilitating algal scraping from rock surfaces[2,3].

Cypriniformes comprises four suborders: Cyprinoidei, Catostomoidei, Gyrinocheiloidei, and Cobitoidei. Gyrinocheiloidei, the smallest suborder, contains solely the family Gyrinocheilidae. This family, confined to Southeast Asia and China's Yunnan Province, includes three torrent-adapted freshwater species. Despite Cypriniformes' extraordinary species diversity and broad biogeographic distribution, the evolutionary history of this dominant freshwater fish clade remains poorly resolved[4]. Prior studies employing both molecular and morphological datasets have yielded conflicting topologies. Phylogenetic hypotheses based on distinct data types—nuclear genes versus mitochondrial genes—produce inconsistent placements of basal lineages. Some analyses position Catostomidae at the base of Cypriniformes, while others support Gyrinocheilidae as the foundational lineage[5,6]. Crucially, phylogenetic reconstructions using mitochondrial markers and limited nuclear loci have failed to conclusively resolve the Gyrinocheilidae-Catostomidae relationship[7].

The advancement of high-throughput sequencing technologies, particularly single-molecule real-time (SMRT) sequencing, has enabled rapid expansion of fish genomic resources in recent years[8,9]. These genomic datasets provide critical insights into evolutionary trajectories across teleost lineages. Among Cypriniformes suborders, genome assemblies have been established for Cyprinoidei, Catostomoidei, and Cobitoidei, leaving Gyrinocheiloidei as the sole suborder without genomic representation. Here, we present a chromosome-scale reference genome for *Gyrinocheilus aymonieri* generated through an integrative approach combining Illumina short-read sequencing, PacBio HiFi long-read sequencing, and Hi-C chromatin interaction mapping. This resource fills a critical genomic gap, enabling comparative analyses to elucidate cypriniform evolutionary patterns and ecological adaptation mechanisms.

[1]Key Laboratory of Breeding Biotechnology and Sustainable Aquaculture (CAS), State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, 430072, China. [2]These authors contributed equally: Cheng Wang, Yongrui Lu. ✉e-mail: yangliandong1987@163.com; clad@ihb.ac.cn

Library type	Tissue	Raw data (Gb)
WGS Illumina	Muscle	48
PacBio HiFi	Muscle	17
Hi-C	Liver	48
RNA-seq	Spleen	3.8
RNA-seq	Skin	4.7
RNA-seq	Kidney	4.8
RNA-seq	Heart	4.8
RNA-seq	Gill	6.7
RNA-seq	Eye	5.8
RNA-seq	Brain	6.6
RNA-seq	Muscle	5.9

Table 1. Statistics of sequencing data for *G. aymonieri* genome assembly and annotation.

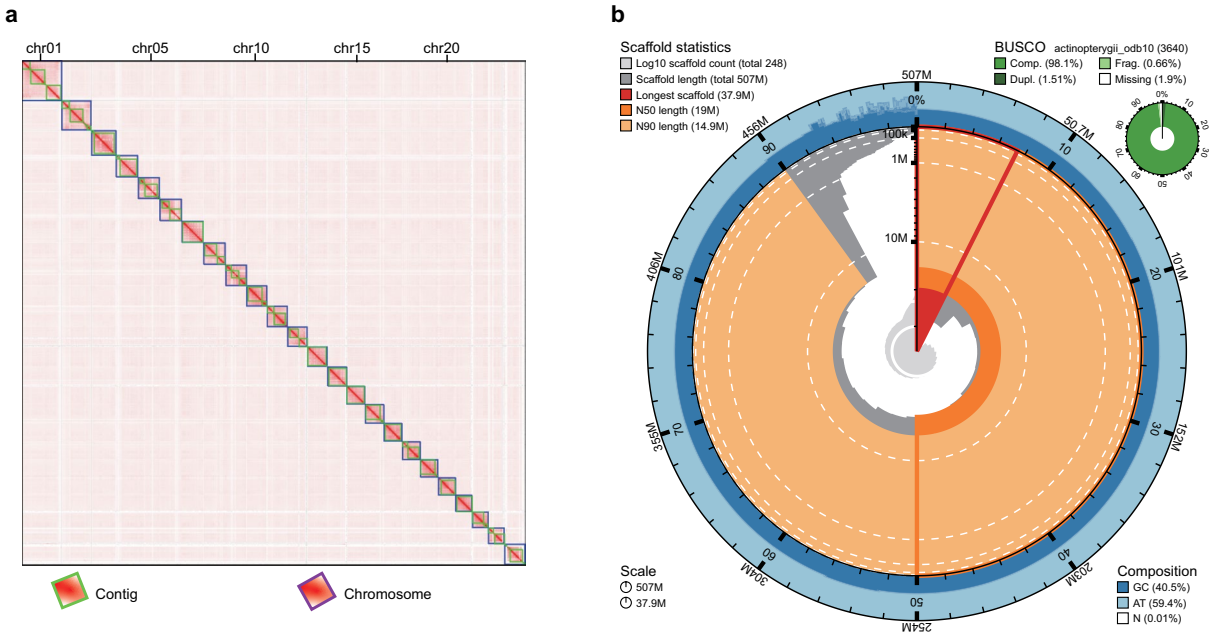

Fig. 1 Characteristics of the *G. aymonieri* genome. (**a**) Hi-C chromatin interaction map of the *G. aymonieri* assembly (2n = 48). Blue boxes represent anchored chromosomes, while green boxes denote assembled contigs. (**b**) Snail plot summary of the assembly statistics and Busco scores for the *G. aymonieri* genome. Dark grey arcs display record lengths in descending size order, with plot radius scaled to the longest assembly record (red). N50 and N90 record lengths are represented by dark and pale orange arcs, respectively. A pale grey spiral indicates cumulative record counts on a logarithmic scale, with white scale lines marking orders of magnitude. The outer ring (blue/pale-blue) shows GC, AT, and N content distributions binned identically to the inner plot. BUSCO scores (top right) were assessed against the actinopterygii core gene set (odb10; n = 3,640).

Methods

Sample collection, preparation and sequencing. An adult fish (~18 cm) was obtained in 2022 from Yunnan Province, China. The sampled fish in this study was permitted by the Ethics Committee of the Institute of Hydrobiology, Chinese Academy of Science. The tissues including muscle, skin, liver, heart, gill, eye, spleen, and kidney were collected and preserved in liquid nitrogen until the extraction of DNA and RNA. The muscles tissues were utilized for DNA sequencing to assembled the genome, while all tissues were utilized for RNA sequencing.

We constructed and sequenced a SMRT cell sequencing library containing fragments approximately 15–20 kb in size on a PacBio Sequel II platform. We prepared second-generation sequencing libraries using Illumina technology (Illumina, San Diego, USA) with 350 bp inserts. We generated RNA sequencing libraries with an NEBNext® Ultra™ RNA Library Prep Kit (Illumina®) following the manufacturer's recommendations and sequenced RNA on an Illumina NovaSeq 6000 platform (Wuhan Frasergen Technology Co., Ltd.), yielding 150 bp paired-end reads. We prepared Hi-C libraries from liver tissue and conducted sequencing on an Illumina NovaSeq 6000 platform. As a result, 48 Gb of Illumina read data, 17 Gb of PacBio HiFi read data, 43.1 Gb of RNA read data, and 48 Gb of Hi-C read data were obtained (Table 1).

Chromosome ID	Number of Contigs	Chromosome Length (bp)
chr1	3	37,919,197
chr2	5	27,489,376
chr3	3	22,705,696
chr4	2	20,456,773
chr5	4	20,320,894
chr6	3	20,307,039
chr7	1	20,188,782
chr8	2	20,163,935
chr9	3	19,642,735
chr10	3	18,998,020
chr11	2	18,993,287
chr12	3	18,631,157
chr13	2	18,543,815
chr14	1	17,695,479
chr15	1	17,460,932
chr16	2	17,127,520
chr17	2	17,106,523
chr18	2	16,778,736
chr19	2	16,753,024
chr20	1	16,029,282
chr21	1	15,440,243
chr22	2	15,122,221
chr23	2	14,850,106
chr24	4	19,235,996

Table 2. The assembled chromosome statistical results of *G. aymonieri* genome.

Type	Length (bp)	Percentage (%)
SINE	1,163,773	0.23
LINE	6,964,738	1.37
DNA	28,169,041	5.55
LTR	12,250,045	2.42
Unclassified	59,626,637	11.76
Others	8,586,155	1.69
Total	116,742,389	23.02

Table 3. Statistics on transposable elements in *G. aymonieri* genome.

***De novo* assembly of *G. aymonieri* genome.** A chromosome-scale genome assembly of *G. aymonieri* was generated through an integrated approach combining Illumina short-read sequencing, PacBio HiFi long-read sequencing, and Hi-C scaffolding. The assembly pipeline initiated with HiFi long-read processing in hifiasm (v0.18.5)[10] under default parameters, followed by systematic error correction using SMRT Link's gcpp (v11.0) and three rounds of Racon (v1.5.0)[11] polishing. Sequence redundancy was minimized through Purge_haplotigs (v1.1.2)[12] optimization, with residual gaps resolved using TGS-GapCloser[13] (v1.2.1). Chromosomal scaffolding was performed by integrating Hi-C data. Following quality control and adapter trimming with fastp[14] (v0.23.4) software, we aligned the resulting clean paired-end reads to the assembly with BWA-MEM[15] (v0.7.17), processed through the Juicer[16] (v1.6)-3D-DNA[17] (v180922) pipeline, and manually curated using Juicebox Assembly Tools[18] (v2.20.00).

The *G. aymonieri* genome assembly spans 507 Mb with a scaffold N50 of 19 Mb, of which 92.3% was successfully anchored into 24 chromosomes (2n = 48) (Fig. 1a, Table 2), aligning with prior karyotypic analyses[19]. The assembly exhibits a GC content of 40.5% and demonstrates exceptional completeness through BUSCO[20] assessment (actinopterygii_odb10, n = 3,640), with 98.1% of conserved genes identified as complete (96.6% single-copy, 1.5% duplicated) (Fig. 1b). These metrics collectively validate the chromosomal integrity and high contiguity of this chromosome-scale genome resource.

Repetitive element and protein-coding genes annotation in the *G. aymonieri* genome. Repetitive element annotation was conducted through an integrated computational pipeline combining *de novo* repeat library construction with curated database searches. A custom repeat library was first generated using RepeatModeler[21] (v2.0.4) under default parameters. Subsequent integration of this species-specific library with the Dfam database (v3.8) enabled comprehensive transposable element (TE) detection via RepeatMasker[22]

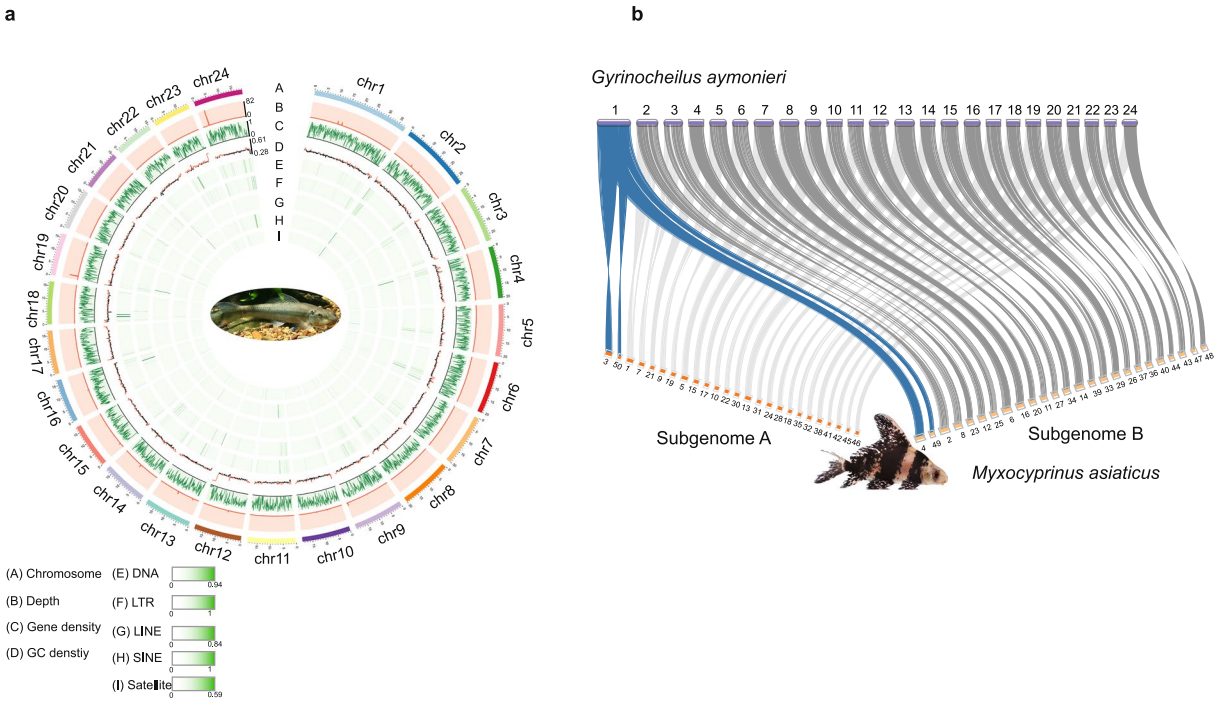

Fig. 2 Genome annotation and synteny of the *G. aymonieri*. (**a**) Genomic landscape of the *G. aymonieri*. The rings, from the outermost to the innermost layer, represent the chromosome of the *G. aymonieri* genome (A), depth (B), gene density (C), GC density (D), DNA transposons (E), LTRs (F), LINEs (G), SINEs (H), and Satellite (I). The analysis was conducted using 100-kb genomic sliding windows. (**b**) The gene synteny map between genomes of *G. aymonieri* and *Myxocyprinus asiaticus*.

		Number	Percent (%)
Total		24,078	
Annotated	EggNog	20,438	84.88
	Swissprot	19,170	79.62
	TrEMBL	21,515	89.36
	NR	21,538	89.45
	InterPro	22,490	93.4
Unannotated		1,418	5.89

Table 4. Summary of functional annotations for predicted genes.

(v4.1.5). The analysis revealed 116.74 Mb of repetitive sequences (23.02% genome coverage), with TE composition dominated by DNA transposons (5.55%), followed by LTR retrotransposons (2.42%), LINEs (1.37%), and SINEs (0.23%) (Table 3).

Gene annotation was performed through an integrative approach combining *ab initio* prediction, homology-based inference, and transcriptomic evidence. *Ab initio* gene models were generated using Augustus[23] (v3.4.0) with parameters trained through PASA[24] (v2.5.2)-processed RNA sequencing data. For homology-based annotation, protein sequences from four cypriniform references (*Danio rerio*[25], *Ctenopharyngodon idella*[26], *Megalobrama amblycephala*[27], and *Misgurnus anguillicaudatus*[28]) were aligned to the genome using Exonerate (v2.4.0) and GenomeThreader[29] (v1.7.3) with default splice-site parameters. Transcript-based evidence was derived from StringTie[30] (v2.2.1)-assembled isoforms, which were structurally refined through iterative PASA analysis. EvidenceModeler[31] (v1.1.1) integrated these multi-source predictions using weighted evidence support (weights: transcript = 8, homology = 6, *ab initio* = 1), followed by PASA-mediated incorporation of alternative splicing variants and UTR annotations to produce the final gene set.

For functional annotation of predicted genes, we aligned the predicted genes to the NR (non-redundant protein database), eggnog[32], Swiss-Prot[33], TrEMBL[34], and InterPro[35] databases using BLASTP[36]. In total, we successfully predicted 24,078 protein-coding genes, and 55,943 transcript isoforms within the genome. Further, 22,660 genes, which accounts for 94.1% of the total number of predicted genes, were successfully assigned with at least one functional annotation database (Fig. 2a, Table 4).

Genome synteny between *G. aymonieri* and *Myxocyprinus asiaticus*. Based on annotated protein-coding sequences and gene structure annotations, we conducted chromosomal collinearity analysis

between *G. aymonieri* (n = 24) and *Myxocyprinus asiaticus* (n = 50) using JCVI[37] software (V1.4.17)[37]. In addition to the teleost-specific whole-genome duplication (WGD) occurring approximately 350 million years ago (Ma)[38–41], *M. asiaticus* (a catostomid fish) experienced an ancestral Cat-4R WGD event[42]. Synteny analysis revealed that the *M. asiaticus* genome could be partitioned into subgenome A and subgenome B. Each chromosome in *G. aymonieri* corresponds to two homologous chromosomes in *M. asiaticus*, reflecting the Cat-4R duplication. Notably, chromosomal fission events were identified between *G. aymonieri* chromosome 1 and *M. asiaticus* subgenome A chromosomes 3/50, as well as subgenome B chromosomes 4/49 (Fig. 2b). In conclusion, the good collinearity between the two genomes reveals the high quality and completeness of the genome we assembled.

Data Records

The raw data of PacBio HiFi long reads, Hi-C data and RNA-seq data for the eight tissues from *G. aymonieri* were submitted to the National Center for Biotechnology Information (NCBI) SRA (Sequence Read Archive) database with accession number SRP589447[43] under BioProject accession number PRJNA1269254. The assembled genome of *G. aymonieri* was deposited at GenBank under accession JBPGVV000000000[44]. Furthermore, files of the assembled genome, gene annotation were deposited in Figshare database[45].

Technical Validation

The BUSCO was used to evaluate the quality of the genome assembly. We assessed assembly completeness using BUSCO[20] (v4.1.2) with the reference actinopterygii gene set (n = 3,640). The final genome assembly showed a BUSCO completeness of 98.1%, consisting of 3,516 (96.6%) single-copy BUSCOs, 55 (1.5%) duplicated BUSCOs, 24 (0.7%) fragmented BUSCOs, and 45 (1.2%) missing BUSCOs. Comparison of BUSCO results with *Gymnocypris eckloni*[46] (87.5%) and *Beaufortia pingi*[47] (97.0%) revealed the high genome assembly quality of *G. aymonieri*.

Data availability

The raw sequencing dataset has been deposited to NCBI (PRJNA1269254), the assembled genome of *G. aymonieri* was deposited at GenBank (JBPGVV000000000), and the assembled genome and gene annotation has been deposited to figshare database (https://doi.org/10.6084/m9.figshare.29176637).

Code availability

No specific code was used in this study. The data analyses adhered to the manuals and protocols offered by the creators of the corresponding bioinformatics tools, the parameter settings of which were outlined in the methods section.

Received: 26 June 2025; Accepted: 26 September 2025;
Published online: 18 November 2025

References

1. 伍献文, 罗., 林人端. 双孔鱼科(Gyrinocheilidae)鱼类的演统发育和分类位置. 动物分类学报, 307–311 (1979).
2. Benjamin, M. The oral sucker of Gyrinocheilus aymonieri (Teleostei: Cypriniformes). *Journal of Zoology* **1**, 211–254, https://doi.org/10.1111/j.1096-3642.1986.tb00638.x (1986).
3. Apitanakul, P. & Wetchakul, W. Some biological aspects of siamese algae eater, Gyrinocheilus aymonieri (Tirant, 1884) in the Yom river, Sukhothai province. (2002).
4. Dai, W. *et al.* Phylogenomic Perspective on the Relationships and Evolutionary History of the Major Otocephalan Lineages. *Scientific Reports* **8**, 205, https://doi.org/10.1038/s41598-017-18432-5 (2018).
5. Liu, S.-Q. *et al.* Phylogenetic relationships of the Cobitoidea (Teleostei: Cypriniformes) inferred from mitochondrial and nuclear genes with analyses of gene evolution. *Gene* **508**, 60–72, https://doi.org/10.1016/j.gene.2012.07.040 (2012).
6. He, S. *et al*. Phylogenetic position of the enigmatic genus Psilorhynchus (Ostariophysi: Cypriniformes): Evidence from the mitochondrial genome. *Molecular Phylogenetics and Evolution* **47**, 419–425, https://doi.org/10.1016/j.ympev.2007.10.012 (2008).
7. Near, T. J. & Thacker, C. E. Phylogenetic classification of living and fossil ray-finned fishes (Actinopterygii). *Bulletin of the Peabody Museum of Natural History* **65**, 3–302 (2024).
8. Nakano, K. *et al.* Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. *Human cell* **30**, 149–161 (2017).
9. Yue, G. & Wang, L. Current status of genome sequencing and its applications in aquaculture. *Aquaculture* **468**, 337–347 (2017).
10. 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, https://doi.org/10.1038/s41592-020-01056-5 (2021).
11. Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* **27**, 737–746, https://doi.org/10.1101/gr.214270.116 (2017).
12. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460, https://doi.org/10.1186/s12859-018-2485-7 (2018).
13. 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**. https://doi.org/10.1093/gigascience/giaa094 (2020).
14. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* **2**, e107, https://doi.org/10.1002/imt2.107 (2023).
15. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).
16. 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).
17. Dudchenko, O. *et al*. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *science* (2017).
18. 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).
19. Rainboth, W. J., Buth, D. G. & Joswiak, G. R. Electrophoretic and Karyological Characters of the Gyrinocheilid Fish, Gyrinocheilus aymonieri. *Biochemical Systematics and Ecology* **14**, 531–537, https://doi.org/10.1016/0305-1978(86)90014-1 (1986).

20. Simao, 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).
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. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* Chapter 4, Unit 4 10, https://doi.org/10.1002/0471250953.bi0410s25 (2009).
23. Hoff, K. J. & Stanke, M. Predicting Genes in Single Genomes with AUGUSTUS. *Curr Protoc Bioinformatics* **65**, e57, https://doi.org/10.1002/cpbi.57 (2019).
24. Haas, B. J. *et al*. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654–5666, https://doi.org/10.1093/nar/gkg770 (2003).
25. 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).
26. Wu, C. S. *et al*. Chromosome-level genome assembly of grass carp (Ctenopharyngodon idella) provides insights into its genome evolution. *BMC Genomics* **23**, 271, https://doi.org/10.1186/s12864-022-08503-x (2022).
27. Liu, H. *et al*. A chromosome-level assembly of blunt snout bream (Megalobrama amblycephala) reveals an expansion of olfactory receptor genes in freshwater fish. *Mol Biol Evol*. https://doi.org/10.1093/molbev/msab152 (2021).
28. Sun, B. *et al*. The chromosome-level genome and key genes associated with mud-dwelling behavior and adaptations of hypoxia and noxious environments in loach (Misgurnus anguillicaudatus). *BMC Biology* **21**, 18, https://doi.org/10.1186/s12915-023-01517-1 (2023).
29. Gremme, G., Brendel, V., Sparks, M. E. & Kurtz, S. Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology* **47**, 965–978, https://doi.org/10.1016/j.infsof.2005.09.005 (2005).
30. Pertea, M. *et al*. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295, https://doi.org/10.1038/nbt.3122 (2015).
31. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* **9**, R7, https://doi.org/10.1186/gb-2008-9-1-r7 (2008).
32. Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol*. https://doi.org/10.1093/molbev/msab293 (2021).
33. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* **28**, 45–48, https://doi.org/10.1093/nar/28.1.45 (2000).
34. Kriventseva, E. V., Servant, F. & Apweiler, R. Improvements to CluSTr: the database of SWISS-PROT+TrEMBL protein clusters. *Nucleic Acids Res* **31**, 388–389, https://doi.org/10.1093/nar/gkg035 (2003).
35. Blum, M. *et al*. InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res* **53**, D444–D456, https://doi.org/10.1093/nar/gkae1082 (2024).
36. Camacho, C. *et al*. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421, https://doi.org/10.1186/1471-2105-10-421 (2009).
37. Tang, H. *et al*. JCVI: A versatile toolkit for comparative genomics analysis. *Imeta* **3**, e211, https://doi.org/10.1002/imt2.211 (2024).
38. Harris, P. M. & Mayden, R. L. Phylogenetic relationships of major clades of Catostomidae (Teleostei: Cypriniformes) as inferred from mitochondrial SSU and LSU rDNA sequences. *Mol Phylogenet Evol* **20**, 225–237, https://doi.org/10.1006/mpev.2001.0980 (2001).
39. Tan, M. & Armbruster, J. W. Phylogenetic classification of extant genera of fishes of the order Cypriniformes (Teleostei: Ostariophysi). *Zootaxa* **4476**, 6–39, https://doi.org/10.11646/zootaxa.4476.1.4 (2018).
40. Christoffels, A. *et al*. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol Biol Evol* **21**, 1146–1151, https://doi.org/10.1093/molbev/msh114 (2004).
41. Vandepoele, K., De Vos, W., Taylor, J. S., Meyer, A. & Van de Peer, Y. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc Natl Acad Sci USA* **101**, 1638–1643, https://doi.org/10.1073/pnas.0307968100 (2004).
42. Krabbenhoft, T. J. *et al*. Chromosome-Level Genome Assembly of Chinese Sucker (Myxocyprinus asiaticus) Reveals Strongly Conserved Synteny Following a Catostomid-Specific Whole-Genome Duplication. *Genome Biology and Evolution* **13**. https://doi.org/10.1093/gbe/evab190 (2021).
43. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP589447 (2025).
44. Cheng, W. *et al*. NCBI GenBank. *The genome assembly of Gyrinocheilus aymonieri* https://identifiers.org/ncbi/insdc:JBPGVV000000000 (2025).
45. Genome and gene annotation files of *Gyrinocheilus aymonieri*, *figshare*, https://doi.org/10.6084/m9.figshare.29176637.
46. Wang, F. *et al*. Chromosome-level assembly of Gymnocypris eckloni genome. *Sci Data* **9**, 464, https://doi.org/10.1038/s41597-022-01595-w (2022).
47. Shen, Q. *et al*. Chromosome-level genome assembly of the butterfly hillstream loach Beaufortia pingi. *Scientific Data* **11**, 1260, https://doi.org/10.1038/s41597-024-04144-9 (2024).

Acknowledgements

This research was supported by the grant from the National Key Research and Development Program of China (2022YFF0608200), the National Natural Science Foundation of China (Grant Nos. 91731301, 32170438) and the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB31000000).

Author contributions

S.H. and L.Y. conceived this study. C.W., Y.L., C.F. and X.G. collected the samples and performed the experiments; C.W. and Y.L. performed the research and analyzed the data. C.W. drafted the manuscript. All authors have 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 L.Y. or S.H.