scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Telomere-to-telomere gap-free genome assembly of *Euchiloglanis kishinouyei*

Hao Wang[1,4], Xuefen Yang[1,4], Jisong Liu[2], Haizhou Wang[2], Yongyao Yu[1], Xiaojing Su[1], Yawen Fan[1], Haiping Liu[3] ✉ & Ruibin Yang[1] ✉

Euchiloglanis kishinouyei **is a typical endemic torrent catfish found in the Jinsha River system of the upper Yangtze River in China. It inhabits fast-flowing streams with steep elevation gradients and has evolved unique biological adaptations to thrive in these extreme environments. A high-quality genome provides key insights into the adaptive mechanisms driving its evolution in these harsh conditions. In this study, we successfully assembled the first telomere-to-telomere (T2T) genome of *E. kishinouyei*, marking the first T2T genome assembly of torrent catfish. The genome spans 886.74 Mb, anchored to 27 chromosomes, with over 99% coverage. The quality value (QV) and Benchmarking Universal Single-Copy Ortholog (BUSCO) scores were 46.96 and 98.50%, respectively, reflecting the high quality of the assembly. We identified repetitive elements accounting for 45.59% (404.23 Mb) of the genome and predicted 24,403 protein-coding genes, 94.37% of which were annotated. This high-fidelity genome assembly provides a valuable resource for future research and lays the foundation for exploring the ecological adaptation mechanisms and evolutionary biology of torrent catfish.**

Background & Summary

The upper reaches of the Jinsha River system, located in the Hengduan Mountains of China, are characterized by typical alpine river valley geomorphology, with deep gorges and steep mountain slopes flanking the river[1,2]. The elevation difference between the upstream and downstream sections of the river approaches 1,000 meters (Fig. 1a)[3]. Shaped by the Qinghai-Tibet Plateau uplift, this region exhibits high altitudes, cold and arid climatic conditions, low water temperatures, and significant seasonal variations in flow between the dry and flood seasons[4,5]. Torrential currents carry debris such as twigs and gravel downstream, creating a harsh environment for aquatic life[6,7]. Species of the genus *Euchiloglanis* (order Siluriformes, family Sisoridae) are primarily distributed in the Jinsha River system and its tributaries[5,8]. As torrent catfish endemic to these waters, *Euchiloglanis* has evolved a suite of unique biological adaptations to survive in the extreme torrent environment. Notable adaptations include advanced internal fertilization strategies for reproduction and specialized pectoral fin rays suited for benthic attachment[9]. These characteristics make *Euchiloglanis* an ideal model for studying the mechanisms underlying adaptive evolution in bony fishes inhabiting torrent environments. Among species in this genus, *Euchiloglanis kishinouyei* is one of the most representative[7,10]. However, its natural populations have been severely impacted in recent years by anthropogenic activities, including overfishing, hydropower development, and habitat destruction[11,12]. These threats have caused a sharp decline in population numbers. Consequently, *E. kishinouyei* has been listed as an endangered species in the *Redlist of China's Biodiversity*[13]. This underscores an urgent need to implement effective conservation measures to safeguard its survival and protect the biodiversity and ecological integrity of the Jinsha River system.

Despite the ecological and evolutionary significance of *E.kishinouyei*, the lack of high-quality genomic data presents a major challenge for its research and conservation[14,15]. Genomic resources for torrent catfish remain limited, with existing draft genomes containing gaps, structural breaks, and repetitive regions[16,17]. These limitations hinder investigations into the species' adaptive mechanisms and complicate conservation and management efforts. Recent advancements in long-read sequencing technologies, including PacBio and Oxford Nanopore

[1]College of Fisheries, Huazhong Agricultural University, Wuhan, 430070, China. [2]Yebatan Branch of Huadian Jinshajiang Upstream Hydropower Development Co., Ltd., Ganzi, 627153, China. [3]College of Fisheries, Southwest University, Chongqing, 402460, China. [4]These authors contributed equally: Hao Wang, Xuefen Yang. ✉e-mail: luihappying@163.com; rbyang@mail.hzau.edu.cn

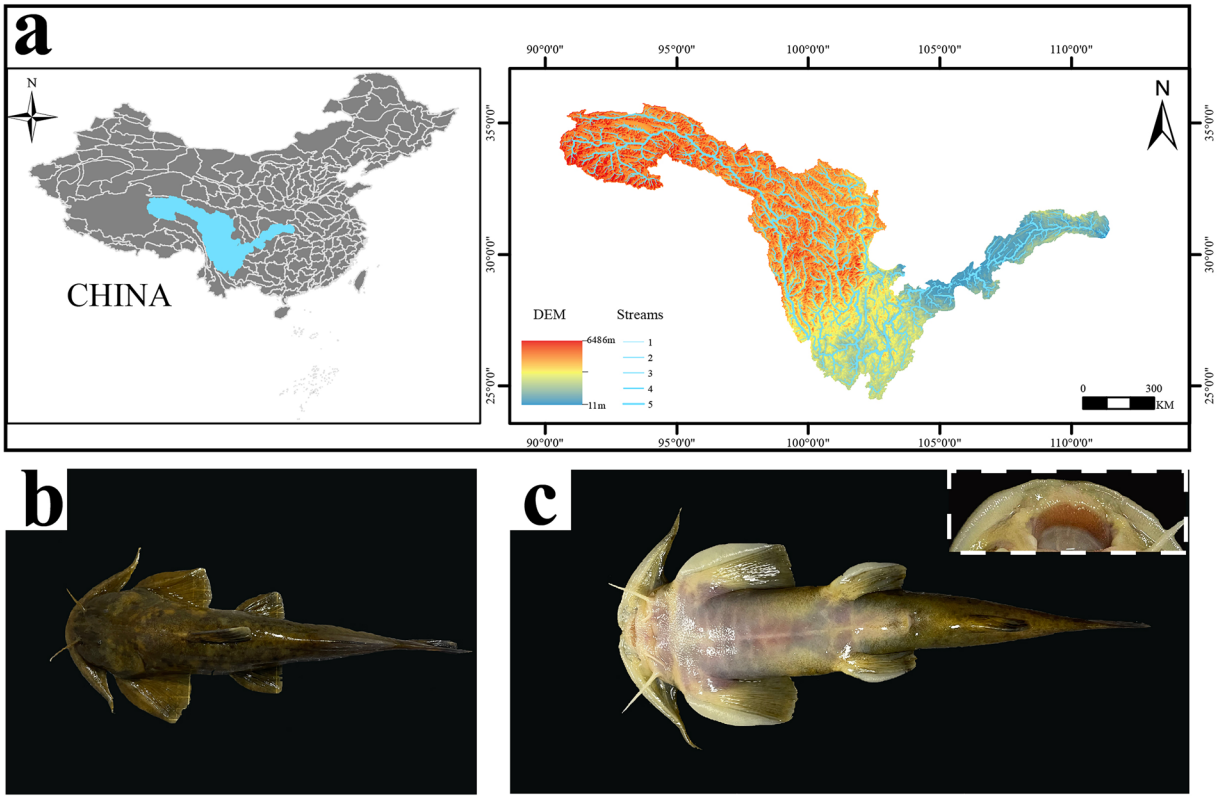

Fig. 1 Habitat analysis and sample presentation of *E. kishinouyei*. (**a**) Altitudinal distribution of *E. kishinouyei* in the Jinsha River Basin; (**b**) Dorsolateral view of *E. kishinouyei*; (**c**) Ventral view showing magnified details, with the inset framed in white highlighting an enlarged view of the tooth band.

Library	Platform	Bases (Gb)	Reads number	Mean read length (bp)	Average coverge (X)
Pacbio HiFi	PACBIO_SMRT	39.31	2,201,937	17,851.01	44.33
ONT	OXFORD_NANOPORE	20.42	509,715	40,061	23.03
Hi-C	DNBSEQ	73.32	464,751,258	150	82.68
NGS	DNBSEQ	60.30	401,986,836	150	68.00
RNA	DNBSEQ	49.14	163,788,236	150	—

Table 1. Statistics of the sequencing data used for genome assembly.

Technologies (ONT), have addressed many limitations of traditional short-read sequencing. These cutting-edge technologies generate highly accurate, continuous, and gap-free telomere-to-telomere (T2T) genomes, resolving issues such as fragmented assemblies and chromosomal discontinuities[18–21]. Additionally, they offer critical insights into complex genomic features like centromere structures and their functions, which are essential for understanding evolutionary adaptations in aquatic species[22].

In this study, we successfully integrated PacBio HiFi, Hi-C, and ONT sequencing data to assemble the T2T genome of *E. kishinouyei*. This represents the first T2T-level genome assembly within the torrent catfish group. The resulting high-quality genome significantly advances genomic resources for species of *Euchiloglanis*, offering key insights into the molecular mechanisms of their adaptations to extreme torrent environments. Furthermore, this genome provides valuable data for the conservation and management of genetic resources in *Euchiloglanis*. It also serves as a critical resource for investigating the adaptive mechanisms of aquatic organisms in fast-flowing riverine habitats. Most importantly, this achievement offers essential genomic data for taxonomic and evolutionary studies in Siluriformes. It lays a solid foundation for comparative genomics research on freshwater fish evolution in extreme environments, contributing to our understanding of how aquatic species adapt to harsh ecological conditions.

Methods

Sample collection and processing. In this study, a mature female individual was collected from Ganzi Tibetan Autonomous Prefecture, Sichuan Province, China, and transported to the laboratory of Fisheries College of Huazhong Agricultural University. The classification of E. *kishinouyei* has been a matter of prolonged debate. The collected specimens are identified based on the taxonomic descriptions in the *Fauna Sinica*, using features

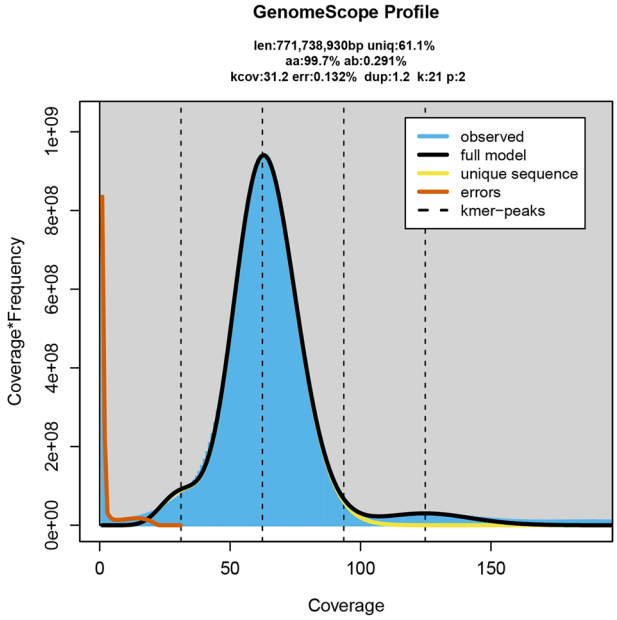

Fig. 2 The k-mer frequency distribution analysis chart for the *E. kishinouyei* genome.

Assembly stage	Total number of sequences	Total length (Mb)	N50 (Mb)	N90 (Mb)	Max sequence (Mb)
HiFiAsm assembly	166 (contig)	890.82	27.26	10.58	50.03
Hi-C assisted	75 (scaffold)	882.90	33.12	23.29	61.37
T2T assembly	75 (scaffold)	886.74	33.25	23.42	61.52

Table 2. Assembly statistics of *E. kishinouyei*.

Assembly	YLab_EK_001
Total size (Mb)	886.74
Number of Chromosome	27
Number of gap-free chromosomes	27
Number of telomere	52
Chromosome length (Mb)	879.62
Sequences assigned to chromosemes (%)	99.19
BUSCOs (%)	98.54
Repetitive sequence (Mb)	404.23
Repetitive sequence percentage (%)	45.59
Protein coding genes	24,403

Table 3. Statistics of the assembled genome for the *E.kishinouyei*.

such as the tooth band (Fig. 1b,c). Muscle tissue from the experimental fish was selected for whole-genome sequencing, while the transcriptome sequencing samples consisted of heart, brain, spleen, kidney, and skin tissues from the same individual. All samples were collected and immediately frozen in liquid nitrogen until DNA and RNA extraction was performed.

Library preparation and sequencing. For PacBio HiFi sequencing, high molecular weight genomic DNA was extracted using the CTAB method, followed by purification with the Onemore-tech Genomic kit according to the manufacturer's protocol. DNA degradation and contamination were assessed using 1% agarose gels, and purity was measured with a NanoDrop™ One UV-Vis spectrophotometer. DNA concentration was determined by Qubit® 4.0 Fluorometer. DNA was sheared using the Megaruptor 3 system and concentrated with AMPure PB beads. SMRTbell libraries were constructed with the Pacific Biosciences SMRTbell® prep kit 3.0 and size-selected on a PippinHT system (20 Kb). Libraries were sequenced on the PacBio platform, generating 39.31 Gb of raw data. For Oxford Nanopore sequencing, high-quality DNA was isolated from muscle tissue using the SDS method, and library preparation was performed with the SQKLSK110 ligation kit. Libraries were loaded onto R9.4 Spot-On Flow Cells and sequenced with a PromethION sequencer, yielding 20.42 Gb of raw data. For Hi-C sequencing, DNA was extracted from the same individual and processed with crosslinking, chromatin

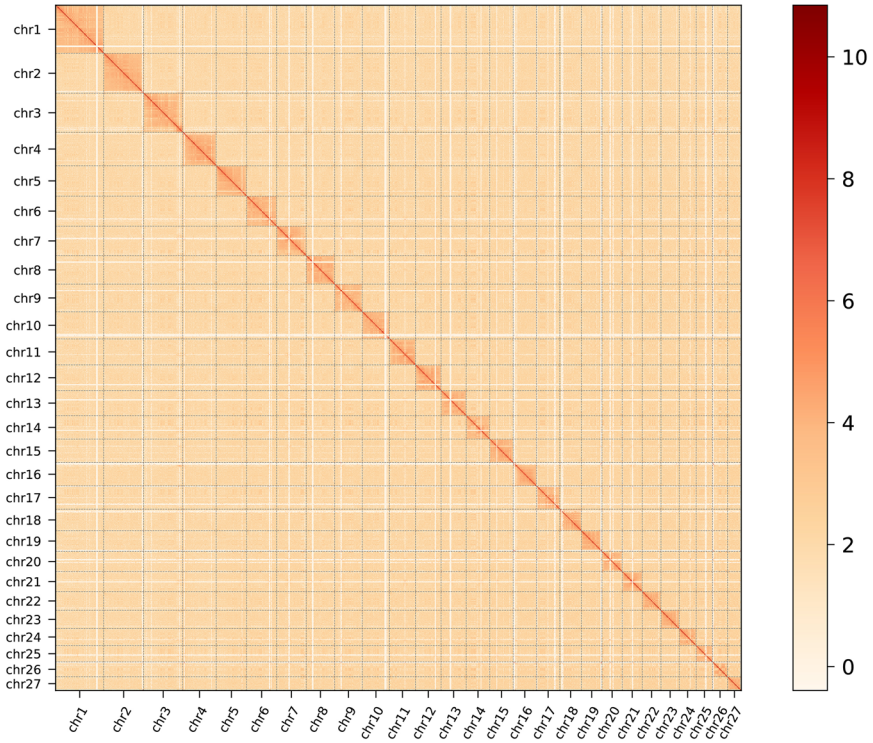

Fig. 3 The heat map of Hi-C contact matrix of the *E. kishinouyei* genome assembly.

Chromosome	Length	Telomere Number
Chr1	61,521,374	2
Chr2	51,418,202	2
Chr3	49,941,094	2
Chr4	43,036,708	2
Chr5	38,877,888	2
Chr6	38,648,619	2
Chr7	37,918,601	2
Chr8	36,554,875	2
Chr9	35,149,686	2
Chr10	35,222,110	2
Chr11	33,245,349	2
Chr12	32,939,503	2
Chr13	32,433,006	2
Chr14	30,270,447	2
Chr15	29,946,560	2
Chr16	30,091,821	2
Chr17	30,039,799	2
Chr18	26,586,328	2
Chr19	26,537,584	2
Chr20	26,035,037	2
Chr21	25,743,129	1
Chr22	24,208,506	2
Chr23	23,421,392	2
Chr24	21,841,209	2
Chr25	21,006,086	1
Chr26	19,255,166	2
Chr27	17,729,107	2

Table 4. Chromosome details and telomere feature sequence counts in the genome.

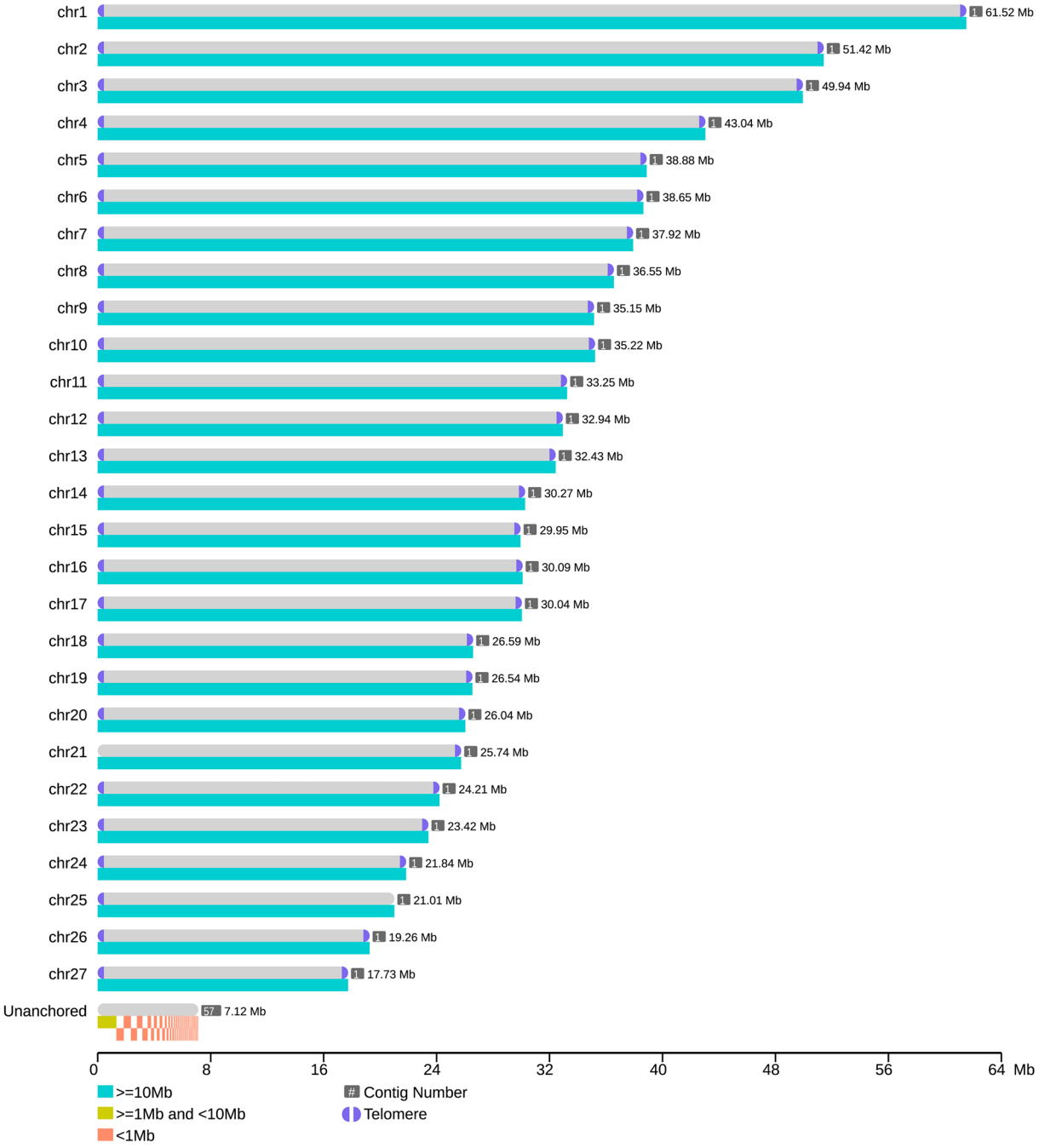

Fig. 4 Contig distribution on chromosomes in the genome.

dissolution, and enzymatic digestion. Crosslinked fragments were ligated, and DNA was purified. Hi-C libraries were prepared, amplified, circulWarized, and sequenced on the DNBSEQ-T7RS platform, generating 73.32 Gb of raw data (Table 1). Quality control and filtering were applied to ensure high-quality clean data, used for downstream analysis.

RNA samples were processed using MGIEasy kits for poly-A enrichment, fragmentation, cDNA synthesis, and circular library preparation, followed by sequencing on the DNBSEQ-T7RS platform. A total of 49.14 Gb of transcriptomic data was generated to assist with genome assembly and annotation (Table 1).

Genome size estimation and heterozygosity analysis. We constructed a small fragment DNA library from DNA extracted from the muscle tissue of *E. kishinouyei*, with a library size of 300–400 bp. The library was then sequenced on the DNBseq platform using paired-end sequencing, generating 60.30 Gb of raw data (Table 1). After filtering with Fastp (v0.23.2)[23], 57.38 Gb of clean data was obtained. Quality assessment of the sequencing data was performed using FastQC (v0.11.3)[24], revealing Q20 parameters at 99.27%, GC content at 41.51%. Randomly sampling 10,000 pairs of reads from the high-quality clean data and aligning them against the NCBI

Type	RepBase TEs		TE Proteins		*De novo*		Combined TEs	
	Length (Mb)	% in Genome	Length (Mb)	% in Genome	Length (Mb)	% in Genome	Length (Mb)	% in Genome
DNA	142.89	16.11	11.48	1.29	83.94	9.47	176.49	19.90
LINE	99.63	11.24	70.66	7.97	50.94	5.75	116.11	13.09
SINE	53.83	6.07	—	0.00	8.56	0.97	59.97	6.76
LTR	34.69	3.91	12.09	1.36	23.75	2.68	47.42	5.35
Satellite	12.77	1.44	—	0.00	0.34	0.04	13.04	1.47
Unknown	1.68	0.19	1.60	0.00	30.62	3.45	32.20	3.63
Total	308.68	34.81	94.17	10.62	197.50	22.27	404.23	45.59

Table 5. Statistics of repetitive sequence annotation result.

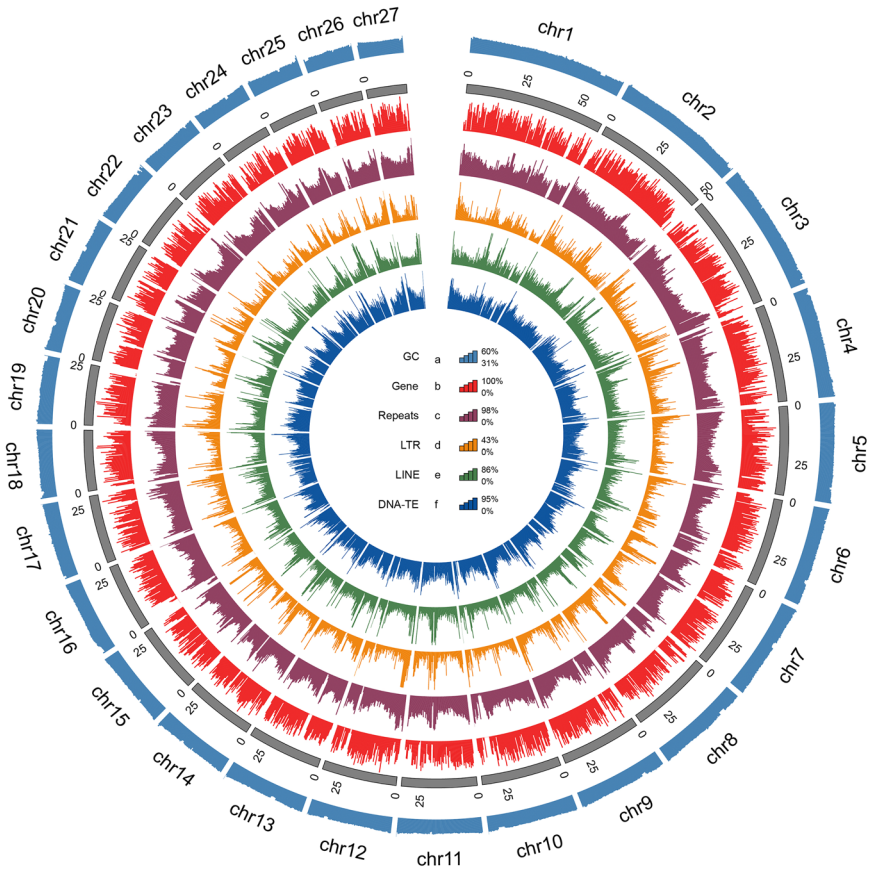

Fig. 5 Circos plot illustrating the genome of the *E. kishinouyei* genome. The plot includes the following components, arranged from outside to inside: (**a**) GC content, (**b**) gene density, (**c**) Repeats content, (**d**) LTR content, (**e**) LINE content and (**f**) DNA-TE content.

database using Blastn (v2.11.0+)[25], no significant exogenous contamination was detected. We obtained the K-mer frequency distribution using Jellyfish (v2.3.0)[26], and then estimated the genome heterozygosity, repeat content, and genome size from the sequencing reads using GenomeScope (v2.0)[27] based on the K-mer frequency distribution data. The analysis results indicated that, with an assumed ploidy of 2, the estimated haploid genome size is 771.74 Mb, with a heterozygosity rate of 0.291% and a repeat content of 38.88% (Fig. 2).

Draft genome assembly. The draft genome assembly of *E. kishinouyei* was performed using PacBio HiFi data combined with ONT ultra-long reads and Hi-C reads. The assembly was carried out with HiFiasm (v0.19.6)[28]. Redundancy removal was accomplished using purge_haplotigs (v1.0.4)[29], and data quality control was performed with Fastp (v0.20.0). The resulting assembly consisted of 166 contigs, with a total genome length of 890.82 Mb (Table 2). The longest contig was 50.03 Mb, the average contig length was 5.37 Mb, and the N50 length was 27.26 Mb (Table 2). The GC content was 41.52%.

Hi-C-Assisted chromosome-level genome assembly. Hi-C-assisted scaffolding utilizes cis-interactions to split, anchor, order, orient, and merge contigs or scaffolds, resulting in chromosome-level

Gene set		Protein coding gene number	Average gene length (bp)	Average CDS length (bp)	Average exon per gene	Average exon length (bp)	Average intron length (bp)
De novo	Genscan	27,166	20,046	1,541	7.75	198.86	2,742
	AUGUSTUS	28,431	12,288	1,370	6.97	196.62	1,830
Homolog	*P. anteanalis*	39,441	13,959	1,438	8.06	178.36	1,773
	maculatum	23,185	13,981	1,411	7.96	177.27	1,806
	D. rerio	59,049	24,375	2,027	11.20	181.07	2,192
	B. yarrelli	17,351	20,762	1,714	9.33	183.73	2,287
	T. fulvidraco	60,775	24,734	2,151	12.17	176.80	2,022
	H. wyckioides	48,462	26,980	2,210	12.61	175.23	2,134
Liftoff	*P. anteanalis*	23,571	15,875	1,541	8.80	175.11	1,837
	G. maculatum	20,763	12,679	1,341	7.50	178.88	1,745
Trans ORF	RNAseq	16,427	27,261	1,894	11.77	457.39	2,032
BUSCO		3,652	13,843	1,622	10.20	158.93	1,328
MAKER		23,012	23,023	1,526	10.11	423.12	2,058
HiFAP		24,403	18,913	1,612	9.42	299.24	1,912

Table 6. Statistics of gene prediction in the genome.

Total gene number	Total transcript number	Mean length of transcripts (bp)	Median length of transcripts (bp)	Mean length of exon (bp)	Median length of exon (bp)
24,403	42,924	23,152.70	12,030	176.06	125

Table 7. Statistics of gene sets with isoforms in the genome.

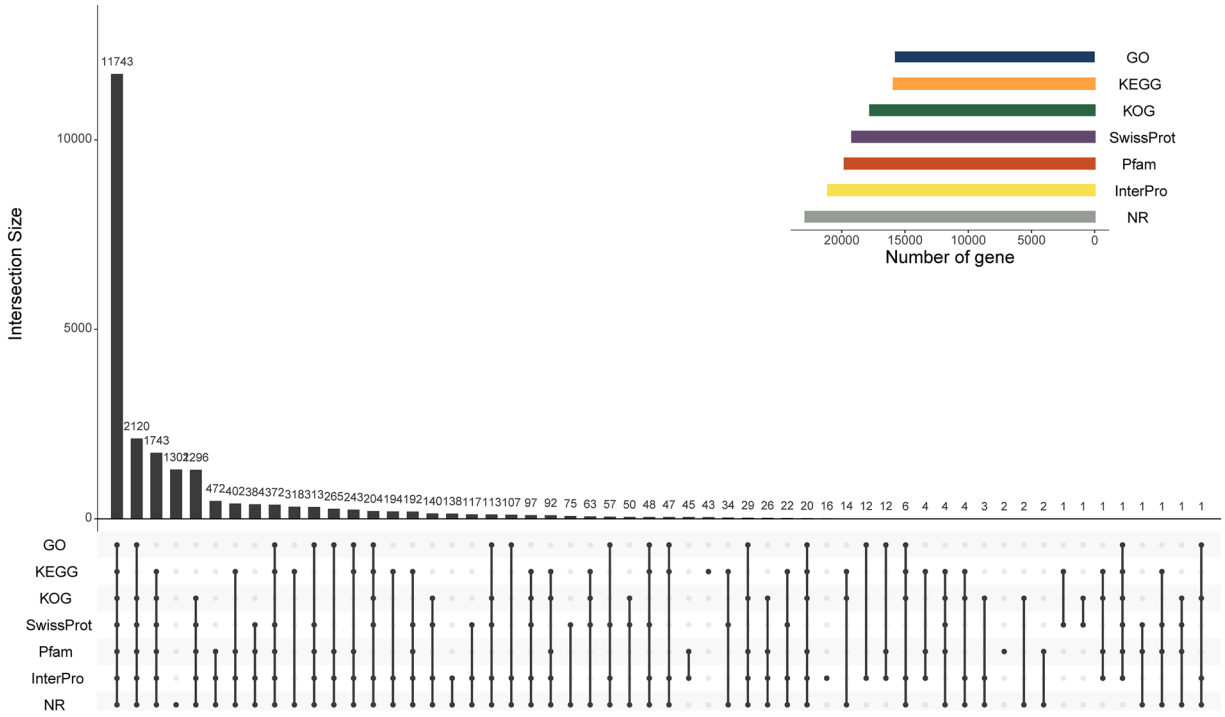

Fig. 6 Statistics of functional annotation result.

genome assembly. The Hi-C sequencing data obtained were preprocessed using HiCUP (v0.7.2)[30]. The assembled *E. kishinouyei* genome was then scaffolded using Juicer (v1.6)[31], which constructed interaction maps to correct any contig assembly errors. Subsequently, the assembled *E. kishinouyei* genome was used as a reference to evaluate the effectiveness of the Hi-C data. The final set of effective paired-end reads (after deduplication) accounted for 30.24% of the total reads. Filtered alignment results were utilized to adjust and anchor the genome, ultimately anchoring the entire genome sequence to 27 chromosomes (Tables 2, 3), with an anchoring rate of 99.19%. After mounting, a chromosome Hi-C interaction map was constructed (Fig. 3), and visualization along with error correction were performed using JuiceBox (v1.11.08)[32].

Type		Copy	Average length (bp)	Total length (bp)	Percentage of genome (%)
miRNA	—	1,376	88	121,536	0.0137
tRNA	—	21,848	77	1,677,355	0.1892
rRNA	rRNA	612	552	337,646	0.0381
	18S	179	1,566	280,297	0.0316
	28S	0	0	0	0.0000
	5.8S	173	154	26,642	0.0030
	5S	260	118	30,707	0.0035
snRNA	snRNA	4,379	147	643,153	0.0725
	CD-box	569	184	104,931	0.0118
	HACA-box	69	150	10,350	0.0012
	splicing	3,735	141	526,783	0.0594
	scaRNA	6	182	1,089	0.0001

Table 8. Statistics of non-coding RNA annotation.

Type	Mapping rate (%)	Average sequencing depth	Coverage (%)	Coverage at least 10X (%)	QV
Short reads	99.59	64.66	99.05	95.51	41.31
Long reads	99.95	44.10	99.51	98.14	46.96

Table 9. Sequence accuracy and consistency evaluation.

Telomere identification and gap closure. Ultra-long reads from ONT sequencing were aligned to the terminal 100 bp of the chromosomes using minimap2 (v2-2.24)[33], with artifact-containing reads removed. Alignment statistics were analyzed, and the read with the median extendable length was designated as the reference (ref), and the remaining reads were treated as queries. Ref and query reads were reassembled to obtain consensus sequences. These consensus sequences were then aligned to the ends of the chromosomes using blastn (2.11.0+), and sequences with coverage ≥90% were used to replace the telomere sequences on the chromosomes based on their alignment positions. Gaps between contigs were filled using TGS-GapCloser (v1.2.0)[34], utilizing the coverage relationship between the ONT ultra-long reads and the already assembled contigs to extend the contigs. Pilon (v1.23)[35] was then used to correct the extended and gap-filled genome with short-read sequencing data, resulting in the final telomere-to-telomere genome assembly of *E. kishinouyei* (Table 3). By searching the entire genome for the telomeric repeat motif (TTAGGG), 52 telomeric sequences were identified at the ends of the 27 chromosomes (Table 4). The positions of the identified telomeres on the chromosomes and their distribution across the contigs were annotated and displayed (Fig. 4).

Repetitive sequence annotation. The assembled genome sequence of *E. kishinouyei* was subjected to repeat sequence prediction and annotation. TRF (v4.09)[36] was used to predict tandem repeats de novo based on the characteristics of repetitive sequences. RepeatMasker (open-4.0.9) and its internal RepeatProteinMask were employed to perform homology-based prediction using the RepBase library (http://www.girinst.org/repbase). Additionally, a de novo repeat library was constructed using RepeatModeler (open-1.0.11)[37] and LTR_FINDER_parallel (v1.0.7)[38], followed by RepeatMasker for repeat prediction from this custom library. The results from these three methods were integrated after removing redundancies. In the assembled genome of *E. kishinouyei*, 45.59% (404.23 Mb) of the sequence was annotated as repetitive sequences. Among these, DNA transposons accounted for 19.90% of the genome size, long terminal repeats (LTRs) made up 5.35%, long interspersed nuclear elements (LINEs) constituted 13.09%, and short interspersed nuclear elements (SINEs) accounted for 6.76% (Table 5 and Fig. 5).

Protein-coding gene prediction and annotation. We performed gene structure and functional annotation on the assembled *E. kishinouyei* genome. Prior to annotation, we masked the repetitive sequences identified in the previous steps. De novo gene prediction was conducted using Genscan[39] and AUGUSTUS (v3.3.2)[40]. Homology-based prediction was performed using miniprot (v0.11-r234)[41] and Liftoff (v1.6.3)[42]. Protein sequences from *Pareuchiloglanis anteanalis*, *Glyptosternon maculatum*, *Danio rerio*, *Bagarius yarrelli*, *Hemibagrus wyckioides* and *Tachysurus fulvidraco* were aligned to our assembled genome[16,43–46]. These protein sequences were sourced from the National Center for Biotechnology Information (NCBI) database. For transcriptome-based prediction, we sequenced the transcriptomes of heart, brain, spleen, kidney, and skin tissues of *E. kishinouyei*. The transcriptome data were aligned to the assembled genome using Stringtie (v1.3.5)[47] for transcriptome prediction, followed by coding region prediction using TransDecoder (v5.5.0). BUSCO was used with the AUGUSTUS self-training optimization mode and a lineage-specific set of conserved orthologous genes for prediction. Finally, the results were integrated using Maker2 (v2.31.10)[48] and HIFAP, leading to the successful prediction of 24,403 protein-coding genes. These genes had an average length of 18,913 bp (Table 6), an average CDS length of 1,612 bp, and an average of 9.42 exons. We obtained a total of 42,924 transcripts for the predicted protein-coding genes, with an average transcript length of 23,152.7 bp (Table 7).

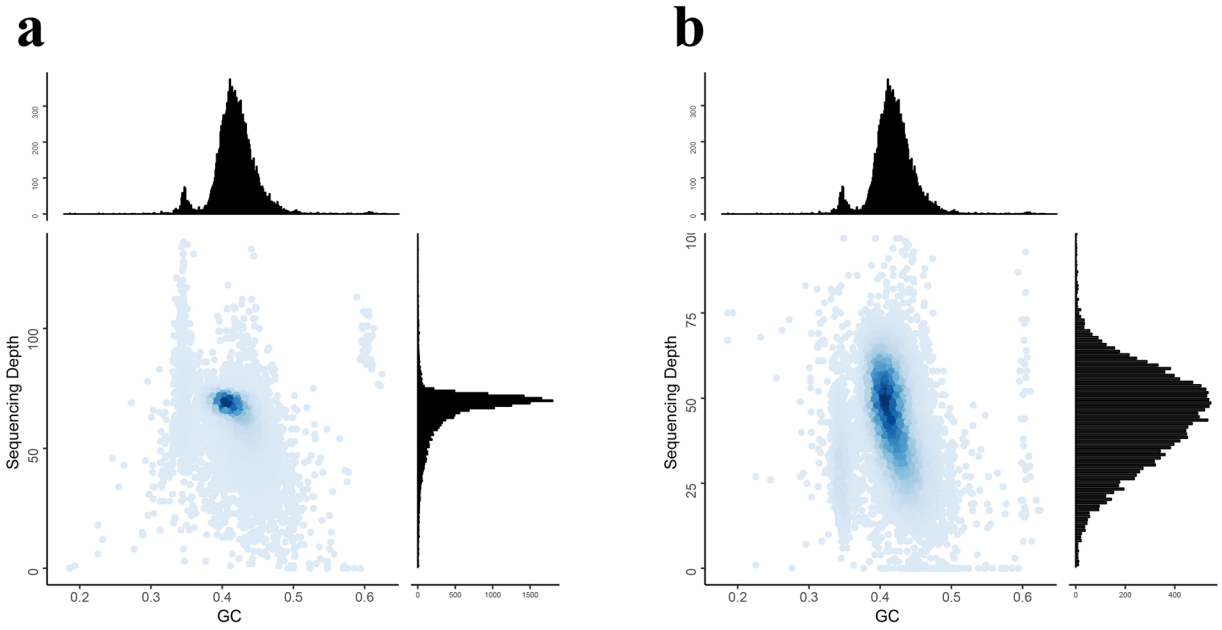

Fig. 7 GC content and depth distribution correlation assessment. (**a**) Short reads data analysis; (**b**) Long reads data analysis.

Software	BUSCO		Compleasm	
Type	Proteins	Percentage (%)	Proteins	Percentage (%)
Complete BUSCOs	3,527	96.90	3,528	96.92
Complete Single-Copy BUSCOs	3,464	95.16	3,019	82.94
Complete Duplicated BUSCOs	63	1.73	509	13.98
Fragmented BUSCOs	24	0.66	25	0.69
Missing BUSCOs	89	2.45	87	2.37
Total BUSCO groups searched	3,640	100.00	3,640	100.00
Database	actinopterygii_odb10		actinopterygii_odb10	

Table 10. BUSCO assessment result.

Functional and metabolic pathway annotations were performed using the Non-redundant Protein Sequences Database (NR), SwissProt, TrEMBL, Clusters of Orthologous Groups (KOG), Transcription Factors Database (TF), InterPro, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Pfam databases. Out of the predicted genes, 23,028 received at least one functional annotation, accounting for approximately 94.37% of all predicted genes. Additionally, 40,654 mRNAs were functionally annotated, representing about 94.71% of the total number of transcripts (Fig. 6).

Non-coding RNA prediction and annotation. Non-coding RNAs (ncRNAs) were predicted and annotated using methods specific to their characteristics. To identify tRNA sequences within the genome, we used tRNAscan-SE (v1.3.1)[49] based on the distinctive structural features of tRNAs. Given the high conservation of rRNAs, we downloaded rRNA sequences from closely related species from the NCBI database as reference sequences and then used BLASTN (v2.11.0+) to align these sequences against our genome to identify rRNAs. Using the covariance models of the Rfam family, we employed INFERNAL, included in the Rfam (v14.8)[50] package, to annotate miRNA and snRNA sequences within the genome. Through these methods, a total of 1,376 miRNAs, 21,818 tRNAs, 612 rRNAs, and 4,379 snRNAs were annotated (Table 8).

Data Records

All raw data for the assemblies in this study, including short reads for genome survey, PacBio HiFi, Hi-C, and ONT long reads, as well as RNA-seq data for gene annotation, have been deposited in the NCBI database under BioProject accession number PRJNA1196141[51]. The short-read sequencing data are available under accession number SRR31839237, PacBio data under SRR31967650, Hi-C data under SRR31789112, ONT data under SRR31793075, and RNA-seq data under the following accession numbers: SRR31967654, SRR31967653, SRR31967652, SRR31967651, and SRR31754171[51]. The genome assembly is available for public access at the NCBI GenBank under the accession number GCA_048932625.1[52]. Genome annotations, predicted coding sequences and protein sequences can be accessed through the Figshare[53].

Technical Validation

Genome assembly an annotation evaluation. To evaluate the quality and accuracy of the assembled genome, we aligned the short reads against the assembled genome using BWA (0.7.12-r1039)[54]. The alignment covered 99.59% of the genome, with an average sequencing depth of 64.66×. Regions with at least 10× coverage accounted for 95.51% of the genome. Additionally, long reads with an average depth of 44.10× were mapped to the genome using minimap2 (2.24-r1122) and achieved 99.51% coverage, with 98.14% of the genome covered by at least 10× (Table 9). These results ensure the high quality and accuracy of the genome assembly. Consistency was assessed using Merquery (1.3)[55], yielding quality values of 41.31% for short reads and 46.96% for long reads, further emphasizing the high quality of the assembly (Table 9). In addition, we constructed a GC content and depth distribution correlation analysis plot to visually present and evaluate the sequencing uniformity. (Fig .7). Finally, we assessed the quality of genome annotation based on predicted proteins using BUSCO (v5.7.1)[56] and Compleasm (v0.2.6)[57]. The analyses identified 96.90% and 96.92% of single-copy genes, respectively (Table 10), highlighting the high completeness and accuracy of our genome annotation.

Code availability

No special codes or scripts were used in this work, and data processing was carried out based on the protocols and manuals of the corresponding bioinformatics software.

Received: 5 February 2025; Accepted: 24 April 2025;
Published online: 07 May 2025

References

1. Zhan, J. *et al*. Mass movements along a rapidly uplifting river valley: an example from the upper Jinsha River, southeast margin of the Tibetan Plateau. *Environ Earth Sci* **77**, 634 (2018).
2. Ma, X., Kang, J., Chen, W., Zhou, C. & He, S. Biogeographic history and high-elevation adaptations inferred from the mitochondrial genome of Glyptosternoid fishes (Sisoridae, Siluriformes) from the southeastern Tibetan Plateau. *BMC Evol Biol* **15**, 233 (2015).
3. Sun, X., Chen, J., Li, Y. & Rene, N. N. Landslide Susceptibility Mapping along a Rapidly Uplifting River Valley of the Upper Jinsha River, Southeastern Tibetan Plateau, China. *Remote Sens.* **14**, 1730 (2022).
4. He, S., Cao, W. & Chen, Y. The uplift of Qinghai-Xizang (Tibet) Plateau and the vicariance speciation of glyptosternoid fishes (Siluriformes: Sisoridae). *Sci China Life Sci.* **44**, 644–651 (2001).
5. Li, Y., Ludwig, A. & Peng, Z. Geographical differentiation of the *Euchiloglanis* fish complex (Teleostei: Siluriformes) in the Hengduan Mountain Region, China: Phylogeographic evidence of altered drainage patterns. *Ecol Evol* **7**, 928–940 (2017).
6. Regan, C. T. Reports on a Collection of Batrachia, Reptiles and Fish from nepal and the Western Himalayas–Fishes. *Rec. Zool. Surv. India* 0157–0158 (1907).
7. Zhou, W., Li, X. & Thomson, A. W. Two new species of the Glyptosternine catfish genus *Euchiloglanis* (Teleostei: Sisoridae) from southwest China with redescriptions of *E. davidi* and *E. kishinouyei*. *Zootaxa* **2871**, 1–18 (2011).
8. Guo, X. *Molecular Phylogeny of Sisoridae and Species Validity of Euchiloglanis in China*. (Southwest China Normal University Chongqing, 2003).
9. Huang, J., Du, J., Wang, C. & Zhao, G. Reproduction environment, sexual system and reproduction behavior of *Euchiloglanis kishinouyei* Kimur. *Sw. China J. Agric. Sci.* **16**, 119–121 (2003).
10. Chu, X. Systematics and evolutionary pedigree of the glyptosternoid fishes (family Sisoridae). *Acta Zootax. Sin* **4**, 72–82 (1979).
11. Lin, P. *et al*. Ecological drivers shaping mainstem and tributary fish communities in the upper Jinsha River, southeastern Qinghai-Tibet Plateau. *Ecological Processes* **14**, 8 (2025).
12. Yan, T. *et al*. Fish Community Structure and Biomass Particle-Size Spectrum in the Upper Reaches of the Jinsha River (China). *Animals* **12**, 3412 (2022).
13. National Forestry and Grassland Administration, China Biodiversity Conservation and Green Development Foundation. *Redlist of China's Biodiversity* 3rd edn. (China Forestry Publishing House; 2021).
14. Zhou, C. *et al*. Diversification of Sisorid catfishes (Teleostei: Siluriformes) in relation to the orogeny of the Himalayan Plateau. *Sci. Bull.* **61**, 991–1002 (2016).
15. Peng, Z., He, S. & Zhang, Y. Phylogenetic relationships of glyptosternoid fishes (Siluriformes: Sisoridae) inferred from mitochondrial cytochrome *b* gene sequences. *Mol Phylogenet Evol* **31**, 979–987 (2004).
16. Liu, H. *et al*. Draft genome of *Glyptosternon maculatum*, an endemic fish from Tibet Plateau. *GigaScience* **7**, giy104 (2018).
17. Xiao, S. J. *et al*. Genome and population evolution and environmental adaptation of *Glyptosternon maculatum* on the Qinghai-Tibet Plateau. *Zool Res* **42**, 502–513 (2021).
18. Chen, W. *et al*. Two telomere-to-telomere gapless genomes reveal insights into *Capsicum* evolution and capsaicinoid biosynthesis. *Nat Commun* **15**, 4295 (2024).
19. Zhang, Y. *et al*. Telomere-to-telomere *Citrullus* super-pangenome provides direction for watermelon breeding. *Nat Genet* 1–12, https://doi.org/10.1038/s41588-024-01823-6 (2024).
20. Bai, M. *et al*. The telomere-to-telomere (T2T) genome of *Peucedanum praeruptorum* Dunn provides insights into the genome evolution and coumarin biosynthesis. *GigaScience* **13**, giae025 (2024).
21. Yang, H. *et al*. High-quality assembly of the T2T genome for *Isodon rubescens* f. *lushanensis* reveals genomic structure variations between 2 typical forms of *Isodon rubescens*. *GigaScience* **13**, giae075 (2024).
22. Russo, A. *et al*. Genome of the early spider-orchid *Ophrys sphegodes* provides insights into sexual deception and pollinator adaptation. *Nat Commun* **15**, 6308 (2024).
23. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
24. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. (Cambridge, United Kingdom, 2010).
25. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).
26. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).
27. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**, 1432 (2020).
28. 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).
29. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460 (2018).
30. Wingett, S. *et al*. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res* **4**, 1310 (2015).
31. Durand, N. C. *et al*. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* **3**, 95–98 (2016).

32. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* **3**, 99–101 (2016).
33. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
34. 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 (2020).
35. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **9**, e112963 (2014).
36. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**, 573–580 (1999).
37. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
38. Ou, S. & Jiang, N. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mobile DNA* **10**, 48 (2019).
39. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
40. Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research* **34**, W435–W439 (2006).
41. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
42. Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).
43. Gong, G. *et al.* Chromosomal-level assembly of yellow catfish genome using third-generation DNA sequencing and Hi-C analysis. *Gigascience* **7**, giy120 (2018).
44. Shao, F. *et al.* Chromosome-Level Genome Assembly of the Asian Red-Tail Catfish (*Hemibagrus wyckioides*). *Front. Genet.* **12**, (2021).
45. Jiang, W. *et al.* Whole-Genome Sequencing of the Giant Devil Catfish, *Bagarius yarrelli*. *Genome Biol. Evol.* **11**, 2071–2077 (2019).
46. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
47. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295 (2015).
48. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
49. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research* **25**, 955–964 (1997).
50. Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Research* **33**, D121–D124 (2005).
51. *NCBI Sequence Read Archive* http://identifiers.org/ncbi/insdc.sra:SRP552180 (2025).
52. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_048932625.1 (2025).
53. Wang, H. *Euchiloglanis kishinouyei* Genome sequencing and assembly. *Figshare* https://doi.org/10.6084/m9.figshare.28241969 (2025).
54. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
55. 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).
56. 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 (2015).
57. Huang, N. & Li, H. compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics* **39**, btad595 (2023).

Acknowledgements

This research was supported by the Breeding Program for Endemic Fish Species in the Jinsha River (T-2022-04).

Author contributions

R.Y. and H.L. conceived the study. J.L. and H.W. coordinated the study and collected the samples. H.W., X.S. and Y.F. processed the samples and performed the experiments. Y.Y., H.W. and X.Y. performed bioinformatics analysis. H.W. and X.Y. wrote the manuscript with significant contributions. H.L. and R.Y. revised the manuscript. All authors read, edited, and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

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