



OPEN

DATA DESCRIPTOR

A telomere-to-telomere chromosome-scale genome assembly of glass catfish (*Kryptopterus vitreolus*)

Chao Bian¹✉, Dahong Li¹, Yiluan Wang¹, Zhe He¹, Wei-Ting Chen², Cheong-Meng Chong³, Hongling Zhou⁴ & Qiong Shi^{1,5}✉

Glass catfish (*Kryptopterus vitreolus*) is commonly distributed in several Asian countries, such as Thailand, Malaysia, and Indonesia. It is renowned for its near-transparent appearance, which has drawn considerable attention for biomedical research and the tropical ornamental fish industry. Here, we successfully constructed the first telomere-to-telomere (T2T) chromosome-scale genome assembly for glass catfish, by integration of PacBio HiFi, Nanopore ONT ultra-long, and Hi-C sequencing technologies. The haplotypic assembly covers approximately 687.7 Mb in length, featuring a high contig N50 of 21.3 Mb. This assembly was then anchored into 32 chromosomes, presenting a complete set of 64 telomeres and 32 centromeres. It was predicted with 252.4 Mb of repetitive sequences and annotated with a total of 24,696 protein-coding genes. Subsequent BUSCO analysis revealed high genome completeness (up to 96.4%). This high-quality T2T genome assembly not only provides a valuable genetic resource for investigating the molecular mechanisms underlying transparency, but also supports in-depth studies on functional genomics, genetic diversity, and selective breeding for this economically important fish species.

Background & Summary

Kryptopterus vitreolus (NCBI Taxonomy ID: 2012678), is a special species within the Siluridae family and the Siluriformes order. It is also named as “glass catfish” or “ghost catfish” due to its almost transparent appearance. It exhibits minimal pigmentation, displaying infrequent chromatic markings. Colors of black, metallic silver, and pale yellow are arranged linearly along its axis from cephalic to mid-body regions. Its intermuscular bones and internal organs can be directly seen through its muscle. The transparency gives it a strong adaptive edge, making this fish almost undetectable to predators. The glass catfish therefore lives smoothly in the complex river systems of some peninsular and southeastern areas of Thailand, Malaysia, and Indonesia¹.

Owing to its glass-like appearance, it holds particular appeal for aquatic enthusiasts and exhibition curators, and it is also a good model for probing pigmentation^{2,3}, cell transplantation and isolation⁴, DNA immunization⁵ and vascular system studies^{6,7}. Nevertheless, this fish has often been misidentified as either *Kryptopterus minor* or, more frequently, *K. bicirrhys*. Previous Ng & Kottelat's study¹ resolved this long-standing issue by classifying the glass catfish as a distinct species, named as *K. vitreolus*. Their report also updated *Kryptopterus* classification, noting the glass catfish's clear and small build versus the more non-transparent and gigantic *K. bicirrhys*.

In our previous glass catfish genome study⁸, we reported a draft chromosome-level genome assembly with identification of some key genes, such as *tyrp1* and *edn3*, which play potential roles in pigmentation formation.

¹Laboratory of Aquatic Genomics, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, 518057, China. ²School of Life Sciences, Jiaying University, Meizhou, Guangdong, 514015, China. ³State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macau, 999078, China. ⁴Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518116, China. ⁵Shenzhen Key Lab of Marine Genomics, BGI Academy of Marine Sciences, BGI Marine, Shenzhen, Guangdong, 518081, China. ✉e-mail: bianchao@szu.edu.cn; shiqiong@genomics.cn; shiqiong@szu.edu.cn

Parameter	T2T assembly	Previous assembly ⁸
HiFi reads (Gb)	75.0	15.1
Hi-C reads (Gb)	101.7	146.6
ONT reads (Gb)	18.5	/
Transcriptome reads (Gb)	11.7	7.5
Genome size (Mb)	687.7	652.4
Contig N50 (Mb)	21.3	1.2
BUSCO	96.4%	90.3%
Protein BUSCO	92.5%	84.0%
Repeat ratio	36.7%	35.2%
Gene number	24,696	23,344
Average CDS length (bp)	1,827.3	2,266.1
Average intron length (bp)	1,483.2	1,077.2
Average exon length (bp)	188.4	220.7
Average gene length (bp)	15,135.6	15,856.5
Functional gene number	24,307	22,537

Table 1. Statistics of the T2T genome assembly and annotation (in this study) with comparison to the previous report.

However, this assembly contains too many gaps with low BUSCO values that led to a lot of missing fragments for shortcoming in genome completeness and details. Here, with large advancements in high-throughput sequencing technology, such as extra PacBio HiFi and Nanopore ONT (Oxford Nanopore Technologies) ultra-long sequencing techniques, we constructed a highly improved telomere-to-telomere (T2T) chromosome-scale genome assembly (with much higher scaffold N50 and better BUSCO) for glass catfish, which is complete in genome sequence with more details about telomeres and centromeres. This new genomic resource not only offers a high-quality assembly to work as a foundational genome reference for in-depth characterization and breeding of glass catfish, but also supports comparative and molecular studies on the regulation of pigmentation in various vertebrates.

Methods

Sample collection and DNA extraction. Three glass catfish individuals (1.0 g) were obtained from a local aquarium fish market in Guangzhou, Guangdong, China. We extracted their genomic DNA (gDNA) from muscle tissue by using a TIANamp Genomic DNA Extraction Kit (Tiangen, Beijing, China) for subsequent whole genome sequencing⁹. A Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and a Nanodrop spectrophotometer (Thermo Fisher Scientific)¹⁰ were used respectively to measure gDNA concentration and purity, while gDNA integrity was verified on a 1.0% agarose gel.

Library construction for genome and transcriptome sequencing. For PacBio HiFi long-read sequencing, DNA libraries were prepared by utilization of SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA), adhering meticulously to PacBio's established protocols. Subsequently, these libraries were sequenced on a PacBio Sequel II platform¹¹. High-accuracy consensus reads were generated using the Circular Consensus Sequencing (CCS, SMRT Link v11.0) software¹², yielding 75.0 Gb of HiFi data (Table 1) with an N50 length of 19.9 kb.

For ONT ultra-long sequencing, a library was prepared using an SQK-ULK001 kit according to the guidance from the manufacturer (Oxford Nanopore Technologies, UK). This library was then sequenced on a PromethION flow cell (Oxford Nanopore Technologies), and about 18.5 Gb of ultra-long reads (Table 1) with an N50 length of 56.2 kb were yielded. Subsequently, these data were processed by using NECAT v200221 with default parameters¹³.

For high-throughput chromosome conformation capture (Hi-C) sequencing, we extracted gDNA, digested chromatin using the restriction enzyme MboI, and then conducted proximity ligation according to protocols outlined in previous studies¹⁴. In brief, gDNA was cross-linked, digested, biotin-labeled, ligated, and fragmented to 500 bp, followed by purification with streptavidin magnetic beads. Library quality and insert size were assessed by using a Qubit 3.0 Fluorometer and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), respectively. Libraries were sequenced on an Illumina NovaSeq. 6000 platform (Illumina, San Diego, CA, USA), generating ~101.7 Gb of 150 bp paired-end reads.

For transcriptome sequencing, total RNA was isolated from muscle tissue by utilizing TRIzol reagent (Invitrogen, Carlsbad, CA, USA). RNA integrity was assessed by using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA), and only those samples with an RNA Integrity Number (RIN) above 7.0 were proceeded for subsequent library preparation. Libraries were constructed employing the DNA nanoball (DNB) technology on a DNBSEQ T7 platform (MGI, BGI Shenzhen, China), and sequencing was performed on a MGISEQ-2000 platform (MGI) to generate 11.7 Gb of 150-bp paired-end reads.

Genome assembly and quality evaluation. A genome assembly of glass catfish was initially obtained by using NextDenovo¹⁵ (v2.2-beta.0, <https://github.com/Nextomics/NextDenovo>) with the following parameters:

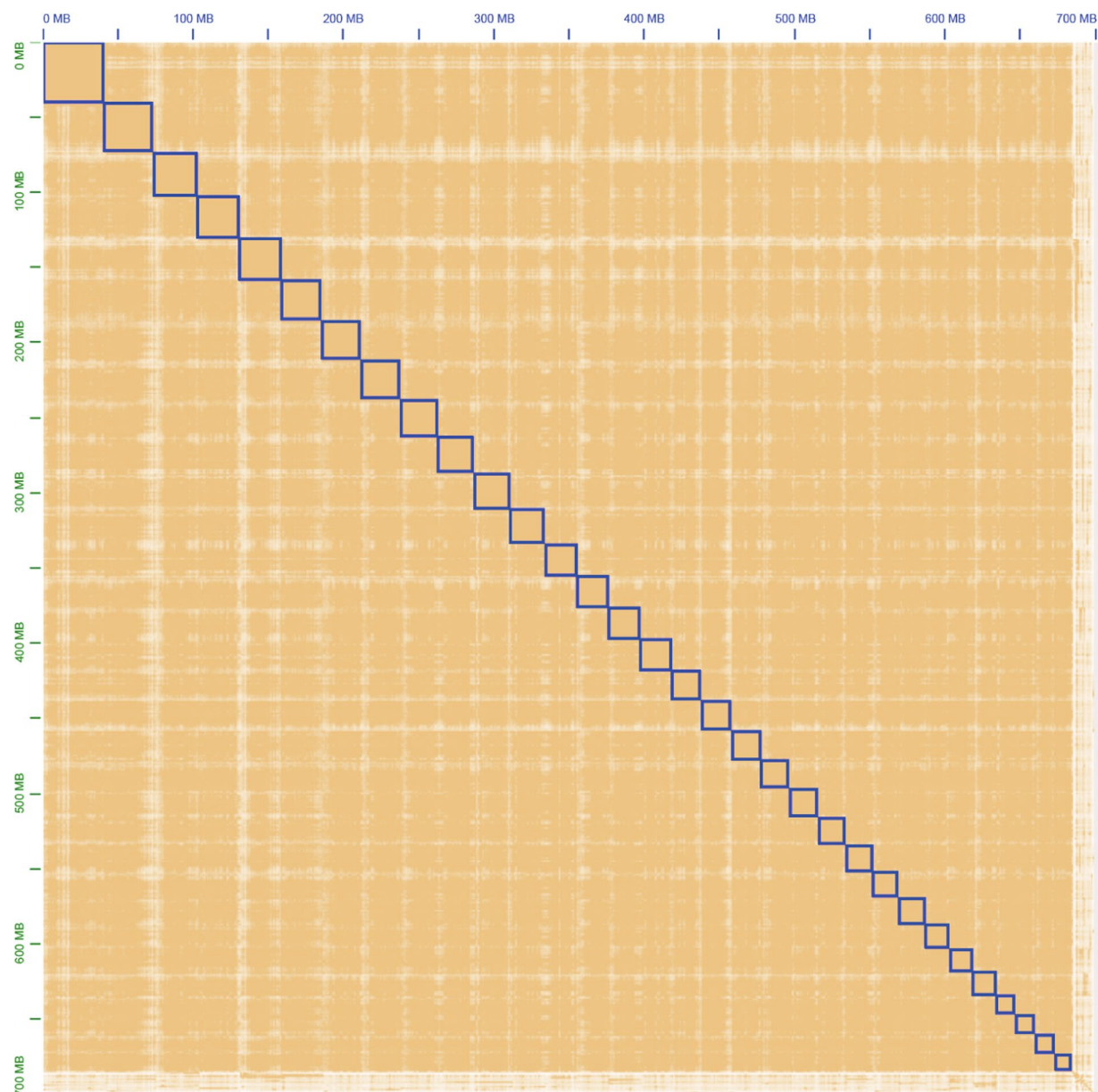


Fig. 1 A chromosomal heatmap derived from Hi-C data. A total of 32 chromosomes (blue boxes) were successfully anchored.

-rerun 3 -read_cutoff 1k -pa_correction 2, resulting in primary contigs with a total size of ~698.3 Mb and a contig N50 of 5.1 Mb. Subsequently, Hi-C sequencing data were aligned to these contigs using Bowtie2 v2.2.5 (configured with --very-sensitive -L 30 --score-min L, --0.6, --0.2-end-to-end -reorder)¹⁶, and then effective linkage information was identified by using HiC-Pro v2.8.1¹⁷ under default settings, retaining only valid contact pairs to support anchoring of contigs into chromosomes.

To orient, order, and cluster contigs into chromosome-level scaffolds, Juicer v1.5 (parameter: chr num 32)¹⁸ and 3D-DNA v170123 (parameters: -m haploid -r 2)¹⁹ were employed. Visualization and manual adjustments were facilitated to adjust mis-assemblies and eliminate redundant contigs by using Juicebox v1.11.08²⁰. This primary chromosome-level genome assembly contained 421 gaps. To achieve a gap-free T2T chromosome-level assembly, LR_GapCloser v1.0 (parameters: -t 35 -m 1000000 -v 10000)²¹ and TGS-GapCloser v1.0.1 (parameter: -min_match 2000)²² were sequentially applied to fill these gaps. The final haplotypic genome assembly anchored 32 chromosomes (Figs. 1 and 2) with a total length of 687.7 Mb (Fig. 1, Table 2), which represents ~98.5% of the assembled contigs. The contig N50 value of these anchored chromosomes was increased to 21.3 Mb, which is much higher than our previous report⁸ (1.2 Mb; Table 1).

To evaluate genome completeness and accuracy, we utilized Benchmarking Universal Single-Copy Orthologs v5.2.2 (BUSCO)²³ with the Actinopterygii odb10 dataset as the reference. Moreover, genome quality was assessed by the K-mer-based tool Merqury v1.3 (<https://github.com/marbl/merqury>)²⁴ with a K-mer size of 17.

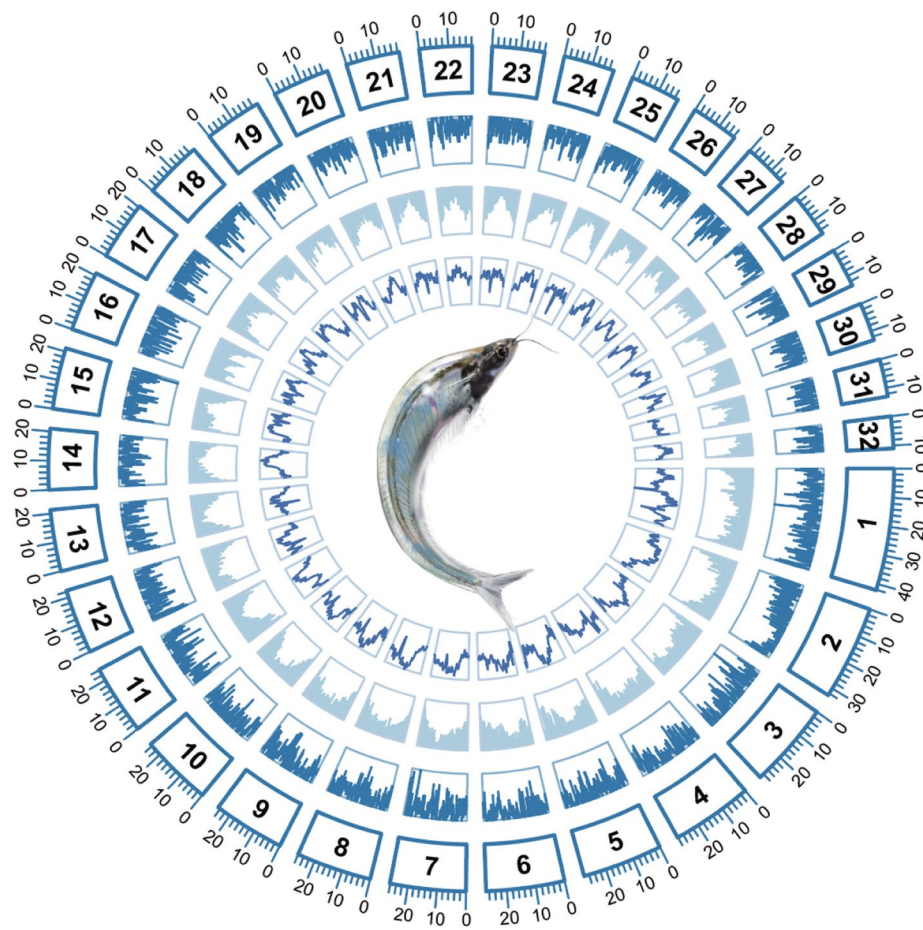


Fig. 2 A Circos image of the 32 anchored chromosomes within the assembled haplotypic genome. Data from the outside to the inside include the length and number of each chromosome, gene density distribution (bin = 0.1 Mb), GC content (bin = 0.1 Mb), repetitive sequence density (bin = 0.1 Mb), and a fish photo (in the center).

This analysis compared K-mer frequency distributions between the genome assembly and whole-genome sequencing data based on HIFI data, producing good metrics for genome completeness and quality values (QV).

Repeat element annotation. In this T2T genome assembly of glass catfish, repetitive elements (REs) were identified through integration of *de novo* and homology-based annotation methods. For the *de novo* annotation, RepeatModeler v2.0.1²⁵ and LTR-FINDER v1.0.6²⁶ were employed to construct a customized repeat library. Subsequently, RepeatMasker version 4.1.0²⁷ was utilized with default settings to annotate REs within the assembled genome, employing Repbase transposable element (TE) library version 21.0²⁸ as the reference. In the homology-based methodology, Tandem Repeats Finder v4.07 was employed with fine-tuned parameters (2 7 7 80 10 50 2000 -d -h)²⁹ to discern tandem repeat sequences. Concurrently, RepeatMasker v4.1.0²⁷ and RepeatProteinMask v4.1.0²⁷ were utilized to annotate transposable elements (TEs) within the assembled genome, with default configurations. Centromeric and telomeric sequences were identified through application of the quarTeT software³⁰.

The final analysis revealed that approximately 252.4 Mb of the glass catfish genome consisted of repetitive sequences (Fig. 3), constituting 36.7% of the total genome length (Table 1). The quarTeT results indicate that this genome assembly possesses intact telomere and centromere sequences on corresponding chromosomes, including a complete set of 64 telomeres and 32 centromeres (see more details in Fig. 3).

Gene prediction and functional annotation. To develop a comprehensive gene set for the glass catfish genome, we employed an integrated approach that combines homology-based annotation and transcriptome-derived annotation to accurately predict protein-coding genes. For the homology-based annotation, protein sequences of five representative teleost species were downloaded from the NCBI database. These species included *Danio rerio* (zebrafish; GCF_000002035.6), *Oryzias latipes* (medaka; GCF_002234675.1), *Takifugu rubripes* (Japanese puffer; GCF_901000725.2), *Tetraodon nigroviridis* (green spotted puffer; GCA_000180735.1) and *Gasterosteus aculeatus* (stickleback; GCA_016920845.1). These protein sequences were aligned to the assembled glass catfish genome by utilizing tBLASTn³¹ with an e-value threshold of 10^{-5} . After obtaining alignments,

Chr ID	Chr Length (bp)	Gap number	Telomere number	GC (%)
chr1	42,155,310	0	2	40.3
chr2	32,706,494	0	2	39.93
chr3	29,618,802	0	2	39.94
chr4	28,117,082	0	2	39.96
chr5	27,755,407	0	2	40.13
chr6	27,125,815	0	2	40.29
chr7	26,202,976	0	2	40.19
chr8	26,073,133	0	2	40.02
chr9	24,844,797	0	2	40.05
chr10	24,761,284	0	2	40.36
chr11	24,164,310	0	2	40.83
chr12	22,382,318	0	2	39.83
chr13	21,286,781	0	2	40.3
chr14	20,999,178	0	2	39.73
chr15	21,128,034	0	2	40.24
chr16	20,828,305	0	2	40.97
chr17	20,468,587	0	2	40.48
chr18	19,943,660	0	2	39.78
chr19	19,840,860	0	2	39.4
chr20	19,602,435	0	2	40.14
chr21	18,547,154	0	2	40.3
chr22	18,184,955	0	2	40.69
chr23	18,054,451	0	2	40.6
chr24	17,579,007	0	2	40.12
chr25	17,189,146	0	2	40.38
chr26	16,438,643	0	2	40.17
chr27	15,494,567	0	2	40.02
chr28	13,577,254	0	2	39.94
chr29	13,248,963	0	2	40.34
chr30	13,248,963	0	2	41.06
chr31	12,954,188	0	2	41.17
chr32	11,046,560	0	2	40.12

Table 2. Summary of the complete T2T chromosomes (chr) in the assembled genome.

gene structure refinement was conducted by using GeneWise v2.2.0³², employing the following parameters: –blast_eval 1e-5, –align_rate 0.5, and –extend_len 500.

Approximately 11.7 Gb of transcriptomic data were assembled into unigenes using Trinity v2.5.1³³, and then gene structures were predicted using PASA v2.3.3³⁴. The outcomes derived from above two methodologies were subsequently combined via the MAKER software³⁵ (version specifics: max_dna_length = 300000, minimum_contig_size = 500, prediction_flanking_region = 500, annotation_error_threshold = 1.0, split_hit_length = 30000, allow_single_exon_genes = True, minimum_single_exon_gene_length = 250). For subsequent functional annotation, predicted protein sequences were subjected to a comparative analysis against three public databases, SwissProt³⁶, TrEMBL³⁷ and KEGG³⁸, by employing BLASTP with a stringent e-value threshold of < 1e-5. Gene Ontology (GO)³⁹ annotation was additionally ascertained through application of InterProScan⁴⁰.

Finally, a total of 24,696 protein-coding genes were annotated, with an average mRNA length of 15,135.6 bp. Each gene had an average of 9.7 exons with an average exon length of 188.4 bp, an average intron length of 1,483.2 bp, and an average coding sequence (CDS) length of 1,827.3 bp. Approximately 98.4% (24,307 genes) of the predicted genes were annotated with at least one database, highlighting the high completeness and reliability of the glass catfish gene annotation (Table 1).

Data Records

The final genome assembly of glass catfish was publicly accessible in the NCBI database under accession number PRJNA1169629⁴¹. The genomic data are stored in the NCBI GenBank with the entry number GCA_044706155.1⁴². Annotated coding sequences and protein sequences have been submitted to Figshare (<https://doi.org/10.6084/m9.figshare.28333385>)⁴³. Additionally, raw reads obtained by PacBio, ONT and Hi-C reads are available in the NCBI database under accession numbers SRR32247860⁴⁴, SRR32247861⁴⁵ and SRR32247859⁴⁶.

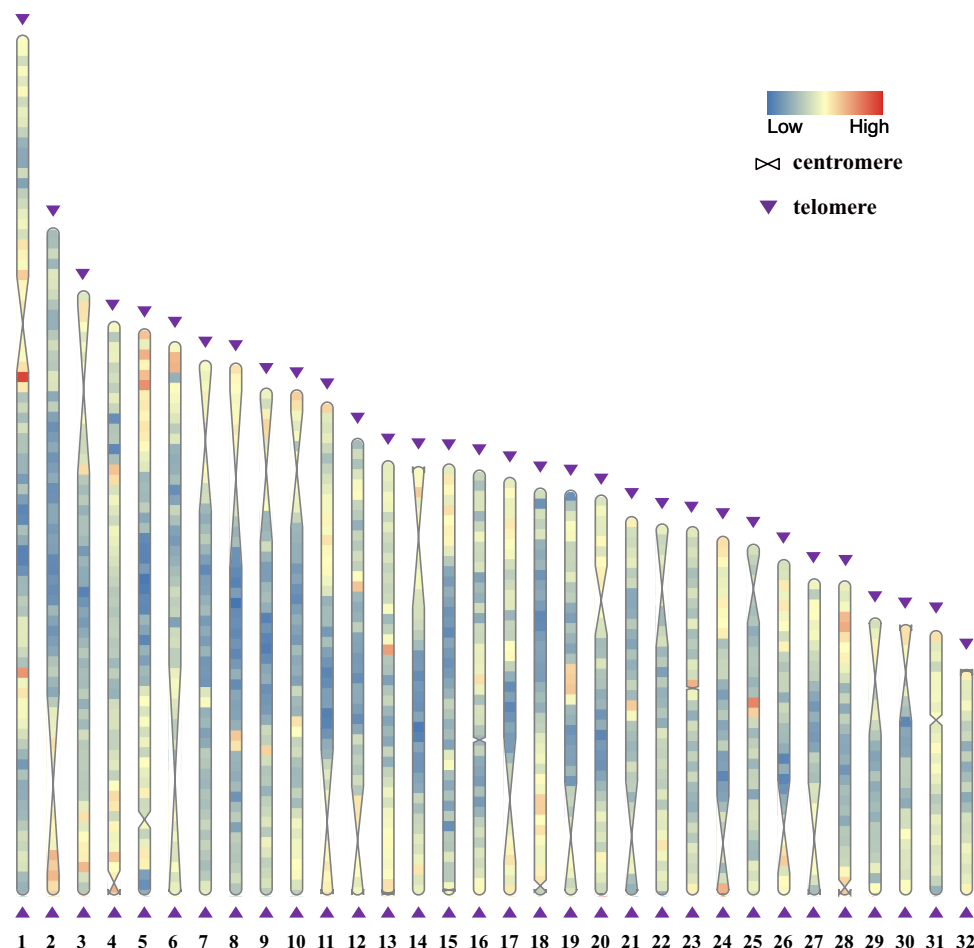


Fig. 3 Localization of telomeres and centromeres and the distribution pattern of repetitive sequences in the assembled genome of glass catfish. Arrows represent the complete set of 64 telomeres at both ends of the 32 chromosomes. One centromere is localized within each chromosome. Repetitive sequences are marked with different colors (a ruler in the top right is provided as the density reference).

Technical Validation

The PacBio HiFi reads, aligned to the genome assembly utilizing Minimap2, achieved an impeccable mapping rate of 100.0%, and the HIC reads also gained a 100% mapping rate. Completeness was further evaluated with BUSCO v.5.2.2, using the Actinopterygii database (3,640 single-copy orthologs; OrthoDB v.10) as the reference. In summary, 96.4% (3,510) of BUSCO genes were complete, with 95.1% (3,462) being identified as single-copy, 1.3% (48) as duplicated, and a mere 0.6% (21) as fragmented. Protein-level BUSCO assessment with the Actinopterygii dataset revealed 92.5% (3,367) completeness, with 90.8% (3,306) single-copy, 1.7% (61) duplicated, and 2.4% (88) fragmented orthologs, confirming a high-quality gene annotation. A Merquy QV score of 39.7 further validated the high quality and reliability of our constructed genome assembly.

Code availability

All scripts and pipelines used for the genome assembly and gene annotation followed the standard manuals and protocols of the applied bioinformatics software. No specific code was developed for this study.

Received: 10 February 2025; Accepted: 14 March 2025;

Published online: 23 March 2025

References

1. Ng, H. H. & Kottelat, M. After eighty years of misidentification, a name for the glass catfish (Teleostei: Siluridae). *Zootaxa* **3630**, 308–316 (2013).
2. Hayashi, H. & Fujii, R. Pharmacological profiles of the subtypes of muscarinic cholinergic receptors that mediate aggregation of pigment in the melanophores of two species of catfish. *Pigment Cell Res* **7**, 175–183 (1994).
3. Hayashi, H. & Fujii, R. Possible involvement of nitric oxide in signaling pigment dispersion in teleostean melanophores. *Zoolog Sci* **18**, 1207–1215 (2001).
4. Han, J. *et al.* Establishment of glass catfish (*Kryptopterus bicirrhus*) fin-derived cells. *Cell Biol Int* **18**, e00008 (2011).
5. Dijkstra, J. M. *et al.* Luciferase expression 2 years after DNA injection in glass catfish (*Kryptopterus bicirrhus*). *Fish Shellfish Immun* **11**, 199–202 (2001).

6. Jensen, L. *et al.* Nitric oxide permits hypoxia-induced lymphatic perfusion by controlling arterial-lymphatic conduits in zebrafish and glass catfish. *PNAS* **106**, 18408–18413 (2009).
7. Rummer, J. *et al.* Function and control of the fish secondary vascular system, a contrast to mammalian lymphatic systems. *J Exp Biol* **217**, 751–757 (2014).
8. Bian, C. *et al.* Chromosome-level genome assembly of the glass catfish (*Kryptopterus vitreolus*) reveals molecular clues to its transparent phenotype. *Zool Res* **45**, 1027–1036 (2024).
9. Yang, Y. *et al.* Gap-free chromosome-level genomes of male and female spotted longbarbel catfish *Hemibagrus guttatus*. *Sci Data* **11**, 572 (2024).
10. Wen, Z. *et al.* Chromosome-level genome assemblies of vulnerable male and female elongate loach (*Leptobotia elongata*). *Sci Data* **11**, 924 (2024).
11. Rhoads, A. & Au, K. PacBio Sequencing and Its Applications. *Genom, Proteom Bioinf* **13**, 278–89 (2015).
12. Chin, C.-S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat methods* **10**, 563–569 (2013).
13. Chen, Y. *et al.* Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun* **12**, 60 (2021).
14. Zhu, W. *et al.* Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific Arabidopsis hybrid. *Genome Biol* **18**, 157 (2017).
15. Hu, J. *et al.* NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol* **25**, 107 (2024).
16. Langmead, B. *et al.* Fast gapped-read alignment with Bowtie 2. *Nat methods* **9**, 357–359 (2012).
17. Servant, N. *et al.* HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259 (2015).
18. Durand, N. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95–98 (2016).
19. Hai, D. *et al.* A High-Quality Genome Assembly of Striped Catfish (*Pangasianodon hypophthalmus*) Based on Highly Accurate Long-Read HiFi Sequencing Data. *Genes* **13**, 923 (2022).
20. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99–101 (2016).
21. Xu, G. C. *et al.* LR_GapCloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience* **8**, giy157 (2019).
22. Xu, M. Y. *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).
23. Simao, F. A. *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
24. Rhie, A. *et al.* Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**, 245 (2020).
25. Abrusán, G. *et al.* TEclass-a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* **25**, 1329–1330 (2009).
26. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**, W265–268 (2007).
27. Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **Chapter 4**, Unit 4 10 (2004).
28. Jurka, J. *et al.* Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**, 462–467 (2005).
29. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573–580 (1999).
30. Lin, Y. *et al.* quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Hortic Res-England* **10**, uhad127 (2023).
31. Mount, D. W. Using the basic local alignment search tool (BLAST). *Cold Spring Harbor Protocols* **2007**, pdb.top17 (2007).
32. Birney, E. *et al.* GeneWise and genomewise. *Genome Res* **14**, 988–995 (2004).
33. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512 (2013).
34. Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654–5666 (2003).
35. Cantarel, B. L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**, 188–196 (2008).
36. Boeckmann, B. *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **31**, 365–370 (2003).
37. Kulikova, T. *et al.* The EMBL nucleotide sequence database. *Nucleic Acids Res* **32**, D27–D30 (2004).
38. Ogata, H. *et al.* KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **27**, 29–34 (1999).
39. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29 (2000).
40. Hunter, S. *et al.* InterPro: the integrative protein signature database. *Nucleic Acids Res* **37**, D211–215 (2009).
41. NCBI Bioproject <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1169629> (2024).
42. Bian, C. NCBI GenBank https://identifiers.org/ncbi/insdc.gca:GCA_044706155.1 (2024).
43. Bian, C. A telomere-to-telomere chromosome-level genome of glass catfish, *Kryptopterus vitreolus*. Figshare <https://doi.org/10.6084/m9.figshare.28333385> (2025).
44. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR32247860> (2024).
45. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR32247861> (2024).
46. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR32247859> (2024).

Acknowledgements

This study was supported by National Key Research and Development Program of China (no. 2023YFE0205100) and Shenzhen Natural Science Foundation (no. JCYJ20241202124511016).

Author contributions

Q.S. and C.B. conceived this project; C.C., W.C. and C.B. collected samples; C.B., D.L. and Y.W. assembled the genome and performed annotation; C.B., Z.H. and H.Z. analyzed the data. C.B., D.L. and Y.W. wrote the manuscript; Q.S., C.C. and H.Z. 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. or Q.S.

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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025