



OPEN ACCESS

EDITED BY
Jun Hong Xia,
Sun Yat-sen University, China

REVIEWED BY
Jie Mei,
Huazhong Agricultural University, China
Zong Xian Zhu,
Sun Yat-sen University, China

*CORRESPONDENCE
Zuogang Peng,
pzg@swu.edu.cn

[†]These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

RECEIVED 06 June 2022
ACCEPTED 15 July 2022
PUBLISHED 10 August 2022

CITATION
Xu Y, Shao F, Chen W, Ni L and Peng Z (2022), A chromosome-level genome of the helmet catfish (*Cranoglanis boudierius*). *Front. Genet.* 13:962406. doi: 10.3389/fgene.2022.962406

COPYRIGHT
© 2022 Xu, Shao, Chen, Ni and Peng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A chromosome-level genome of the helmet catfish (*Cranoglanis boudierius*)

Yuan Xu^{1†}, Feng Shao^{1†}, Weitao Chen², Luyun Ni¹ and Zuogang Peng^{1,3*}

¹Key Laboratory of Freshwater Fish Reproduction and Development (Ministry of Education), Southwest University School of Life Sciences, Chongqing, China, ²Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou, China, ³Academy of Plateau Science and Sustainability, Qinghai Normal University, Xining, China

KEYWORDS

helmet catfish, chromosome-level assembly, comparative genomics, Hi-C, HIFI

Introduction

It is said that the history of life has been written in the genome (Blaxter et al., 2022), and indeed, many significant questions pertaining to the evolution and ecology of life on earth will only be addressable when whole-genome data representing divergences at all branches in the tree of life are available. Genomes will serve as core databanks for inferring the phylogeny of all life and contribute to gaining a comprehensive understanding of species biology (Blaxter et al., 2022). Consequently, accruing genomic resources of important species in tree of life has significant implications for evolutionary studies and other fields of biological research.

The fish family Cranoglanididae (Teleostei: Siluriformes) is notable in that it comprises a single genus containing just three species (Fricke et al., 2022), one of which is the helmet catfish (*Cranoglanis boudierius*), an economically important aquaculture fish endemic to China. They are only distributed in the Pearl River Basin in South China and mainly live on shrimps and small fishes. With respect to commercial value, this species has a number of favorable traits, including a high proportion of edible flesh, high nutritional value, strong disease resistance, and fewer intermuscular spines (Wu et al., 2010). However, the wild resources of this fish have declined markedly in recent years, as a consequence of overfishing, and in 2011 it was placed on the IUCN Red List of Threatened Species (www.iucnredlist.org). Previous studies on *C. boudierius* have focused on morphology (Diogo et al., 2002; Luckenbill and Lundberg, 2009), physiology (Zhou et al., 2003; Chen et al., 2016), genetic diversity (Cheng et al., 2009), and molecular characterization of the complete mitochondrial genome (Peng et al., 2006). However, despite this extensive research, the lack of a high-quality chromosome-level reference genome for *C. boudierius* has significantly hindered in-depth research on the evolution, breeding, and conservation of this fish.

To rectify this deficiency, in this study, using Illumina, Nanopore, PacBio, and Hi-C technologies, we constructed the first chromosome-level genome for *C. boudierius*, thereby providing an available genomic resource for future studies on the conservation, artificial breeding, and evolution of this fish.

Data description

All Illumina, Nanopore, and PacBio sequencing data generated in this study are presented in [Supplementary Table S1](#). Estimates of genome size and heterozygosity were performed using a total of 36.82 Gb of filtered Illumina sequencing data based on 17-mer analysis. The main peak was located at a depth of 34 with a predicted genome size of 931.11 Mb and an estimated percentage heterozygosity of 0.5% ([Supplementary Table S2](#); [Supplementary Figure S1](#)).

For *de novo* assembly of the *C. boudierius* genome, we used 92.50 Gb of ONT long reads generated using Oxford Nanopore Technologies to construct a preliminary assembly, and subsequently used 13.62 Gb of PacBio HiFi reads and 36.82 Gb of Illumina short reads to polish the preliminary assembly, yielding 881.33 Mb of genomic DNA sequences, comprising 231 contigs with a contig N50 length of 20.01 Mb ([Supplementary Table S3](#)). RNA-seq short reads from different tissues were aligned to the assembly genome using HISAT2, with the percentage of aligned reads ranging from 92.02 to 93.98% ([Supplementary Table S4](#)). Completeness of the *C. boudierius* genome was assessed based on BUSCO analysis using the Actinopterygii_odb10 database, with 95.71% (3,484) of BUSCO genes being identified in the *C. boudierius* genome ([Supplementary Table S5](#)), thereby indicating the high level of completeness of the genome assembly.

A total of 138.25 Gb of clean reads obtained from the Hi-C library were used for chromosome-level genome assembly. We constructed a high-quality chromosome-level *C. boudierius* genome of 876.63 Mb, with 214 contigs being successfully anchored and orientated on 38 chromosomes, ranging in size from 13.73 to 39.43 Mb ([Supplementary Table S6](#)). Among them, 10 chromosomes were gapless, such as LG07, LG15, and LG17 ([Supplementary Table S6](#)). A genome-wide Hi-C interaction heatmap indicated that the interaction signal strength around the diagonal was significantly stronger than that at other position within each chromosome ([Figure 1A](#)). To the best of our knowledge, *C. boudierius* has the highest number of chromosomes among all Siluriformes for which chromosome-level genomes have been published. Comparative gene collinearity analysis indicated that despite interspecific differences in chromosome numbers, most chromosomes showed good collinearity ([Figure 1B](#)), thereby further verifying the reliability of the genomic data generated in this study.

The *C. boudierius* genome was found to comprise 49.55% repetitive sequences, with transposons accounting for 43.87% of the assembled genome ([Supplementary Table S7](#)), among which, the largest proportion (29.2%) consists of DNA transposons, followed by LTRs (7.29%), LINES (6.32%), and SINEs (1.07%). On the basis of a combination of homologous, *de novo*, and RNA-seq prediction approaches, we annotated 21,715 protein-coding genes, with average gene, coding sequence, and exon and intron lengths of 20,496.52,

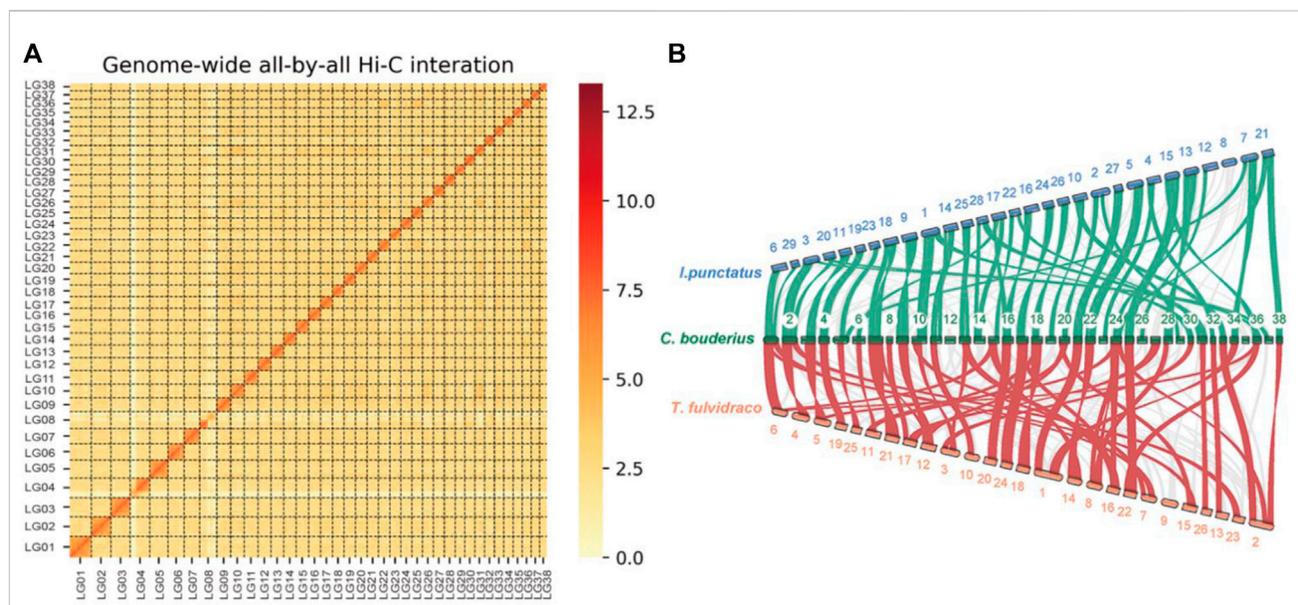


FIGURE 1

Characteristics of the *Cranoglanis boudierius* genome. **(A)** A Hi-C contact map of the *C. boudierius* genome. LGs 1–38: *Lachesis* groups 1 to 38, representing the 38 chromosomes of *C. boudierius*. **(B)** Gene collinearity between *C. boudierius* and the two related fish species *Tachysurus fulvidraco* and *Ictalurus punctatus*. The green and red links denote block scores exceeding 200 between *I. punctatus* and *C. boudierius*, and between *C. boudierius* and *T. fulvidraco*, respectively. *C. boudierius* chromosomes are numbered in ascending order.

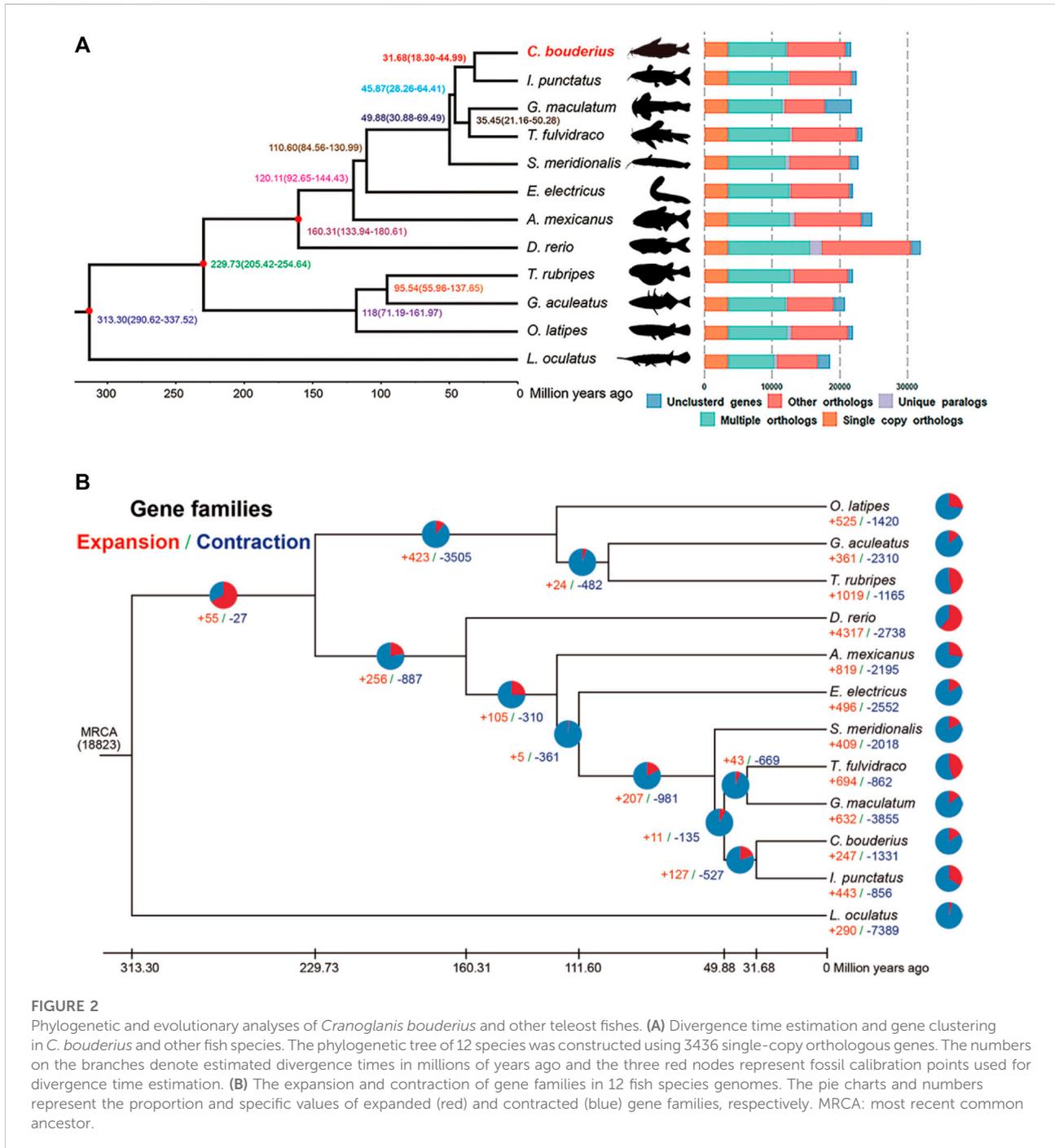


FIGURE 2

Phylogenetic and evolutionary analyses of *Cranoglanis boudierius* and other teleost fishes. (A) Divergence time estimation and gene clustering in *C. boudierius* and other fish species. The phylogenetic tree of 12 species was constructed using 3436 single-copy orthologous genes. The numbers on the branches denote estimated divergence times in millions of years ago and the three red nodes represent fossil calibration points used for divergence time estimation. (B) The expansion and contraction of gene families in 12 fish species genomes. The pie charts and numbers represent the proportion and specific values of expanded (red) and contracted (blue) gene families, respectively. MRCA: most recent common ancestor.

1,761.49, 167.97, and 1,974.86 bp, respectively, which are consistent with the distribution trends of these parameters in the released genomes used for annotation (Supplementary Figure S2). To assess the completeness of genome annotation we also applied BUSCO in conjunction with Actinopterygii_odb10 database, with the results indicating that 96.07% of the conserved single-copy genes in the *C. boudierius* genome were completely predicted

(Supplementary Table S8). In addition, non-coding RNA annotation revealed that the genome contains a total of 1,950 miRNAs, 5,792 tRNAs, and 795 rRNAs, whereas with respect to functional annotations, we succeeded in annotating 21,085 genes using at least one of the referenced databases (KEGG, KOG, NR, SwissProt, and GO), accounting for 97.10% of the predicted protein-coding genes (Supplementary Table S9).

The phylogenetic relationships between *C. boudierius* and other vertebrates, and the respective divergence times were deduced based on an analysis of 3,436 single-copy orthologous genes from the genomes of 12 teleost fishes (Figure 2A). Our findings indicated that *C. boudierius* is most closely related to species within the family Ictaluridae, which is consistent with the assessments of previously published phylogenetic studies (Sullivan et al., 2006). The five species of Siluriformes were found to form a monophyletic group, and Siluriformes and Gymnotiformes were identified as sister groups. In addition, we estimated that *C. boudierius* diverged from *Ictalurus punctatus* approximately 31.68 million years ago (Mya), whereas the time of the divergence between Siluriformes and Gymnotiformes was around 110.60 Mya.

It is speculated that the expansion and contraction of gene families may play important roles in environmental adaptation, and on the basis of gene family analysis, we identified 247 and 1,331 gene families that have undergone expansion and contraction, respectively, since the aforementioned divergence from *I. punctatus* (Figure 2B). Among these gene families, 45 and 38 families were identified as being significantly expanded and contracted, respectively (Supplementary Table S10). Further enrichment analysis revealed that the expanded genes were enriched in 21 KEGG pathways ($p < 0.05$) (Supplementary Table S11), most of which are associated with olfaction, immunity, and nutrient metabolism. These findings accordingly provide a new perspective for further studies on growth, metabolism, and adaptation to a benthic habitat. Similarly, contracted genes were identified as being enriched in 18 KEGG pathways ($p < 0.05$) (Supplementary Table S12), the most highly represented of which were “Complement and coagulation cascades,” “Tight junction,” and “Pathogenic *Escherichia coli* infection.”

Materials and methods

Ethics statement

All animal experiments conducted in this study were reviewed and approved by the Ethics Committee of Southwest University (NO.20200715-01).

Sampling, library construction, and sequencing

For the purposes of genome and transcriptome sequencing, we used samples obtained from a female *C. boudierius* (Supplementary Figure S3) captured from the main channel of the Pearl River in Guangdong Province, China (23°8'27"N, 111°45'35"E). Blood samples were collected for DNA extraction performed using a QIAGEN® Genomic kit (Qiagen,

Hilden, Germany). A subsequently constructed DNA library (200–400 bp) was sequenced using the MGISEQ 2000 platform (Korostin et al., 2020). Low-quality reads in raw data were filtered out using fastp (Chen et al., 2018) (default parameters). An ONT library was prepared from the long DNA fragments selected using the BluePippin system (Sage Science, United States), with subsequent sequencing using a PromethION sequencer (Oxford Nanopore Technologies, United Kingdom). The FAST5 files generated using a Nanopore sequencer were converted to FASTQ format using Guppy (Wick et al., 2019), and the raw data with mean_qsocre_template values < 7 were filtered. Using the PacBio protocol (Pacific Biosciences, CA, United States), we constructed a SMRTbell library by shearing genomic DNA with g-TUBE (Covaris, United States). Target size libraries (10–20 kb) were sequenced using a PacBio Sequel II sequencer to generate HiFi reads.

RNA from 11 tissue types of the aforementioned fish (adipose fin, brain, fins, gills, gonad, heart, head kidney, kidney, liver, muscle, and spleen) was extracted using an RNeasy Plus Mini Kit (Qiagen), and validated RNA samples were used for complementary DNA library construction using a TruSeq Sample Preparation Kit (Illumina) and Illumina HiSeq4000 sequencer. Equal amounts of qualified RNA from each of the 11 tissue types were pooled and reverse-transcribed using an SQK-PCS109 sequencing kit (Oxford Nanopore Technologies) for ONT libraries preparation and sequencing (Nanopore PromethION). To prevent the potential influence of adapter sequences and low-quality reads in subsequent analyses, the raw reads were filtered as described for DNA.

Genome features based on K-Mer analysis and genome assembly

Clean reads of sequences obtained using the MGISEQ 2000 platform were used for estimation of genome size and heterozygosity based on 17-mer frequency distribution analysis using jellyfish software (Marcais and Kingsford, 2011).

Long reads generated using the Nanopore platform were used for NextDenovo-based *de novo* assembly (reads_cutoff:1k, seed_cutoff:20k, <https://github.com/Nextomics/NextDenovo>). To enhance the accuracy of assembly, polishing of the short and long reads was performed using Racon (<https://github.com/isovic/racon>) and Nextpolish (<https://github.com/Nextomics/NextPolish>), respectively, with default parameters.

Chromosome-level genome assembly using Hi-C and assessment

For chromosome-level assembly of the *C. boudierius* genome, genomic DNA was extracted from kidney samples for Hi-C

library construction, and sequencing was performed using the MGISEQ 2000 platform. The Hi-C data thus obtained was initially filtered using fastp (default parameters), and the clean paired-end reads were then mapped to the primary genome using Bowtie2 (Langmead and Salzberg, 2012) to generate unique mapped paired-end reads. Valid interacting paired reads were obtained using HiC-Pro from uniquely mapped paired-end reads. A chromosome-level assembly was generated using LACHESIS (Burton et al., 2013) with the following parameters: CLUSTER MIN RE SITES = 100; CLUSTER MAX LINK DENSITY = 2.5; CLUSTER NONINFORMATIVE RATIO = 1.4; ORDER MIN N RES TRUNK = 60; ORDER MIN N RES SHREDS = 60. A genome-wide Hi-C heat map was constructed based on the interaction between different scaffolds.

We assessed the completeness of the assembly using two strategies, namely, BUSCO (Simao et al., 2015) and RNA-seq read mapping. Core genes of the Actinopterygii database were searched against the *C. boudierius* genome using BUSCO, and RNA-seq reads were mapped to the genome using HISAT2 (Kim et al., 2015) with default parameters. Furthermore, to assess the accuracy of the scaffolding results, we performed gene collinearity analysis between *C. boudierius* and two related species, *I. punctatus* and *Tachysurus fulvidraco*, using MCScanX (Wang et al., 2012).

Repeat sequence annotation

Tandem repeats were annotated using GMATA (Wang and Wang, 2016) and TRF (Benson, 1999) with default parameters. With respect to the annotation of transposable elements, we applied two strategies, namely, *de novo* and homology-based prediction, using MITE-hunter (Han and Wessler, 2010), LTR_finder (Xu and Wang, 2007), and RepeatModeler (Bedell et al., 2000) with default parameters for *de novo* prediction. For homology-based prediction, we initially merged the ab initio library and Repbase (Bao et al., 2015) TE library as a final repeat sequence library, and the aligned this to the genome using RepeatMasker (Tarailo-Graovac and Chen, 2009). The outcomes of the two predictions were integrated as the final non-redundant repeat sequence annotation.

Gene prediction and functional annotation

Three independent methods, namely, homologous, *de novo*, and RNA-seq prediction, were used for gene prediction. For homology-based prediction, protein sequences from the genomes of the following nine vertebrates were aligned to the *C. boudierius* genome using GeMoMa (Keilwagen et al., 2016): channel catfish (*I. punctatus*), Chinese yellow catfish (*T. fulvidraco*), giant devil catfish (*Bagarius yarrelli*), humans (*Homo sapiens*), Japanese

puffer (*Takifugu rubripes*), medaka (*Oryzias latipes*), red-tail catfish (*Hemibagrus wyckiooides*), Tibetan catfish (*Glyptosternon maculatum*), and zebrafish (*Danio rerio*). For *de novo* prediction, we used AUGUSTUS (Stanke et al., 2008) with default parameters and a training set generated from RNA-seq reads assembled using StringTie (Pertea et al., 2015) and PASA (Haas et al., 2003). For transcriptome-based prediction, the transcripts assembled based on full-length and short reads were merged with open reading frames and predicted using PASA. The final integrated gene set was produced using EvidenceModeler (Haas et al., 2008), with sequences containing transposons being removed using TransposonPSI (<http://transposonpsi.sourceforge.net/>). The accuracy of gene prediction was assessed based on BUSCO analysis using the Actinopterygii_odb9 database, and the NR (Marchler-Bauer et al., 2011), KOG (Tatusov et al., 2003), KEGG (Kanehisa et al., 2014), and SwissProt (Bateman et al., 2021) databases were used to functionally annotate these predicted genes via BLASTP searches with an E-value cutoff of 1e-5. In addition, Gene Ontology (GO) annotation was performed using InterProScan (Zdobnov and Apweiler, 2001) with default parameters.

For the annotation of non-coding RNAs, we used tRNAscan-SE (Lowe and Eddy, 1997) and RNAmmer (Lagesen et al., 2007) to annotate tRNAs and rRNAs, respectively, whereas microRNAs and small nuclear RNAs were respectively predicted using INFERNAL (Nawrocki and Eddy, 2013) and Rfam database (Daub et al., 2015).

Comparative genomic analyses

For the purpose of identifying orthologous gene families using OrthoMCL (Li et al., 2003), in addition to the protein sequences of *C. boudierius*, we downloaded those for the following 11 species from the NCBI and GIGADB databases: channel catfish, Chinese yellow catfish, electric eel (*Electrophorus electricus*), Japanese puffer, medaka, Mexican tetra (*Astyanax mexicanus*), southern catfish (*Silurus meridionalis*), spotted gar (*Lepisosteus oculatus*), three-spined stickleback (*Gasterosteus aculeatus*), Tibetan catfish, and zebrafish.

The coding sequences obtained for single-copy gene families found in all 12 of these teleost species were aligned using MAFFT (Yamada et al., 2016). Poorly aligned sequences were filtered using Gblocks (Talavera and Castresana, 2007) and a maximum likelihood tree was then constructed using RAxML (Stamatakis, 2014) with a GTRGAMMA substitution model and 1000 bootstrap replicates. Divergence time estimation was implemented using MCMCTree in PAML (Yang, 2007) software, and three established fossil divergence times derived from the TimeTree database (<http://www.timetree.org/>) were used for calibration (Supplementary Table S13).

Gene families identified using OrthoMCL and divergence times estimated for *C. boudierius* and related species were used to assess the possible expansion and contraction of orthologous gene families using CAFÉ (De Bie et al., 2006).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

The animal study was reviewed and approved by Ethics Committee of Southwest University.

Author contributions

ZP contributed to design the study and edit/review the manuscript. YX performed the major part of bioinformatic analyses and drafted the manuscript. FS contributed to sample collection, data analysis and drafted the manuscript. WC and LN contributed to sample collection and data curation. All authors have read and approved the final manuscript.

Funding

This work was supported by the grant from the National Natural Science Foundation of China (31872204).

References

- Bao, W. D., Kojima, K. K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6, 11. doi:10.1186/s13100-015-0041-9
- Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., et al. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi:10.1093/nar/gkaa1100
- Bedell, J. A., Korf, I., and Gish, W. (2000). MaskerAid: A performance enhancement to RepeatMasker. *Bioinformatics* 16, 1040–1041. doi:10.1093/bioinformatics/16.11.1040
- Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi:10.1093/nar/27.2.573
- Blaxter, M., Archibald, J. M., Childers, A. K., Coddington, J. A., Crandall, K. A., Di Palma, F., et al. (2022). Why sequence all eukaryotes? *Proc. Natl. Acad. Sci. U. S. A.* 119, e2115636118. doi:10.1073/pnas.2115636118
- Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R. L., Kitman, J. O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31, 1119–1125. doi:10.1038/nbt.2727
- Chen, J. T., Zhou, A. G., Xie, S. L., Wang, C., Lv, Z. J., and Zou, J. X. (2016). Comparative proteomic identification of mature and immature sperm in the catfish *Cranoglanis boudierius*. *Plos One* 11, e0151254. doi:10.1371/journal.pone.0151254
- Chen, S. F., Zhou, Y. Q., Chen, Y. R., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, 884–890. doi:10.1093/bioinformatics/bty560
- Cheng, F., Xie, S. G., Ye, W., and Ye, F. L. (2009). Genetic analysis of *cranoglanis boudierius* by molecular marker AFLP. *Acta Hydr. Sin.* 33, 539–545. doi:10.3724/sp.j.1035.2009.00539
- Daub, J., Eberhardt, R. Y., Tate, J. G., and Burge, S. W. (2015). Rfam: Annotating families of non-coding RNA sequences. *Methods Mol. Biol.* 1269, 349–363. doi:10.1007/978-1-4939-2291-8_22
- De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). Cafe: a computational tool for the study of gene family evolution. *Bioinformatics* 22, 1269–1271. doi:10.1093/bioinformatics/btl097
- Diogo, R., Chardon, M., and Vandewalle, P. (2002). Osteology and myology of the cephalic region and pectoral girdle of the Chinese catfish *Cranoglanis boudierius*, with a discussion on the autapomorphies and phylogenetic relationships of the Cranoglanididae (Teleostei: Siluriformes). *J. Morphol.* 253, 229–242. doi:10.1002/jmor.10000
- Fricke, R., Eschmeyer, W. N., and Van der Laan, R. (2022). Eschmeyer's catalog of fishes: Genera, species, references. <http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>. Electronic version (accessed July 12, 2022).
- Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Hannick, L. I., et al. (2003). Improving the arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31, 5654–5666. doi:10.1093/nar/gkg770
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using evidencemodeler and the program to assemble spliced alignments. *Genome Biol.* 9, R7. doi:10.1186/gb-2008-9-1-r7
- Han, Y. J., and Wessler, S. R. (2010). MITE-hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* 38, e199. doi:10.1093/nar/gkq862
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205. doi:10.1093/nar/gkt1076

Acknowledgments

We thank the Fish10K Genome Project (Fish10K) for the support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.962406/full#supplementary-material>

- Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* 44, e89. doi:10.1093/nar/gkw092
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). Hisat: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi:10.1038/nmeth.3317
- Korostin, D., Kulemin, N., Naumov, V., Belova, V., Kwon, D., and Gorbachev, A. (2020). Comparative analysis of novel MGISEQ-2000 sequencing platform vs Illumina HiSeq 2500 for whole-genome sequencing. *PLoS One* 15, e0230301. doi:10.1371/journal.pone.0230301
- Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi:10.1093/nar/gkm160
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/Nmeth.1923
- Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi:10.1101/gr.1224503
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi:10.1093/nar/25.5.955
- Luckenbill, K. R., and Lundberg, J. G. (2009). CAT scan-based images of the skeleton of the asian catfish *cranoglanis* (Siluriformes: Cranoglanididae). *Proc. Acad. Nat. Sci. Phila.* 158, 297–299. doi:10.1635/053.158.0115
- Marcais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770. doi:10.1093/bioinformatics/btr011
- Marchler-Bauer, A., Lu, S. N., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., Dewese-Scott, C., et al. (2011). Cdd: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 39, D225–D229. doi:10.1093/nar/gkq1189
- Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi:10.1093/bioinformatics/btt509
- Peng, Z. G., Wang, J., and He, S. P. (2006). The complete mitochondrial genome of the helmet catfish *Cranoglanis boudierius* (Siluriformes: Cranoglanididae) and the phylogeny of otophysan fishes. *Gene* 376, 290–297. doi:10.1016/j.gene.2006.04.014
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi:10.1038/nbt.3122
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi:10.1093/bioinformatics/btv351
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033
- Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntetically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24, 637–644. doi:10.1093/bioinformatics/btn013
- Sullivan, J. P., Lundberg, J. G., and Hardman, M. (2006). A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using *rag1* and *rag2* nuclear gene sequences. *Mol. Phylogenet. Evol.* 41, 636–662. doi:10.1016/j.ympev.2006.05.044
- Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. doi:10.1080/10635150701472164
- Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.* 4, 4, doi:10.1002/0471250953.bi0410s25
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: An updated version includes eukaryotes. *Bmc Bioinforma.* 4, 41. doi:10.1186/1471-2105-4-41
- Wang, X. W., and Wang, L. (2016). Gmata: an integrated software package for genome-scale SSR mining, marker development and Vlewing. *Front. Plant Sci.* 7, 1350. doi:10.3389/fpls.2016.01350
- Wang, Y. P., Tang, H. B., Debarry, J. D., Tan, X., Li, J. P., Wang, X. Y., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi:10.1093/nar/gkr1293
- Wick, R. R., Judd, L. M., and Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* 20, 129. doi:10.1186/s13059-019-1727-y
- Wu, X. M., Zhou, L. B., and Zeng, X. L. (2010). Analysis on the flesh rate and the muscle nutrition component in *Cranoglanis boudierius*. *J. Hydroecology* 3, 131.
- Xu, Z., and Wang, H. (2007). LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. doi:10.1093/nar/gkm286
- Yamada, K. D., Tomii, K., and Katoh, K. (2016). Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees. *Bioinformatics* 32, 3246–3251. doi:10.1093/bioinformatics/btw412
- Yang, Z. H. (2007). Paml 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi:10.1093/molbev/msm088
- Zdobnov, E. M., and Apweiler, R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847–848. doi:10.1093/bioinformatics/17.9.847
- Zhou, L. B., Liu, X. C., Lin, H. R., and Ye, W. (2003). Changes in pituitary and serum gonadotropin levels in helmet catfish (*Cranoglanis boudierius*) at different stages of ovarian development. *Acta Zool. Sin.* 49, 399