



# Chromosome-Level Assembly of the Chinese Hooksnout Carp (*Opsariichthys bidens*) Genome Using PacBio Sequencing and Hi-C Technology

Xiaojun Xu<sup>1†</sup>, Wenzhi Guan<sup>1†</sup>, Baolong Niu<sup>1†</sup>, Dandan Guo<sup>1</sup>, Qing-Ping Xie<sup>1</sup>, Wei Zhan<sup>1</sup>, Shaokui Yi<sup>2\*</sup> and Bao Lou<sup>1\*</sup>

<sup>1</sup>Institute of Hydrobiology, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, <sup>2</sup>School of Life Sciences, Huzhou University, Huzhou, China

## OPEN ACCESS

### Edited by:

Roger Huerlimann,  
Okinawa Institute of Science and  
Technology Graduate University,  
Japan

### Reviewed by:

Dong-Neng Jiang,  
Guangdong Ocean University, China  
Tao Zhou,  
Xiamen University, China

### \*Correspondence:

Shaokui Yi  
yishaokui@foxmail.com  
Bao Lou  
loubao6577@163.com

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

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

Received: 02 October 2021

Accepted: 23 December 2021

Published: 19 January 2022

### Citation:

Xu X, Guan W, Niu B, Guo D, Xie Q-P,  
Zhan W, Yi S and Lou B (2022)  
Chromosome-Level Assembly of the  
Chinese Hooksnout Carp  
(*Opsariichthys bidens*) Genome Using  
PacBio Sequencing and Hi-  
C Technology.  
Front. Genet. 12:788547.  
doi: 10.3389/fgene.2021.788547

**Keywords:** *Opsariichthys bidens*, PacBio sequencing, Hi-C technology, chromosome-level assembly, *Opsariichthyinae*

## INTRODUCTION

Chinese hooksnout carp (*Opsariichthys bidens*) is an endemic Cypriniformes minnow in East Asia, and mainly distributed in China. Notably, this common minnow has undergone a long and complex taxonomic history. In 1960s, it was classified in Cyprinidae, Leuciscinae, *Opsariichthys* (Wu, 1964). With the advances on the application of molecular characters for the fish systematics in 1990s, *O. bidens* was assigned into Cyprinidae, Danioninae, *Opsariichthys* (Chen, 1998). Subsequently, its taxonomic status was revised several times (Mayden et al., 2009; Fang et al., 2009; Tang et al., 2010, 2013; Liao et al., 2011; Stout et al., 2016; Huang et al., 2017). According to the latest phylogenetic classification of bony fishes, *O. bidens* has been assigned into Xenocyprididae, *Opsariichthyinae*, *Opsariichthys* (Betancur-R et al., 2017), which has been adopted by the NCBI database ([www.ncbi.nlm.nih.gov/Taxonomy](http://www.ncbi.nlm.nih.gov/Taxonomy)) and FishBase ([www.fishbase.org](http://www.fishbase.org)).

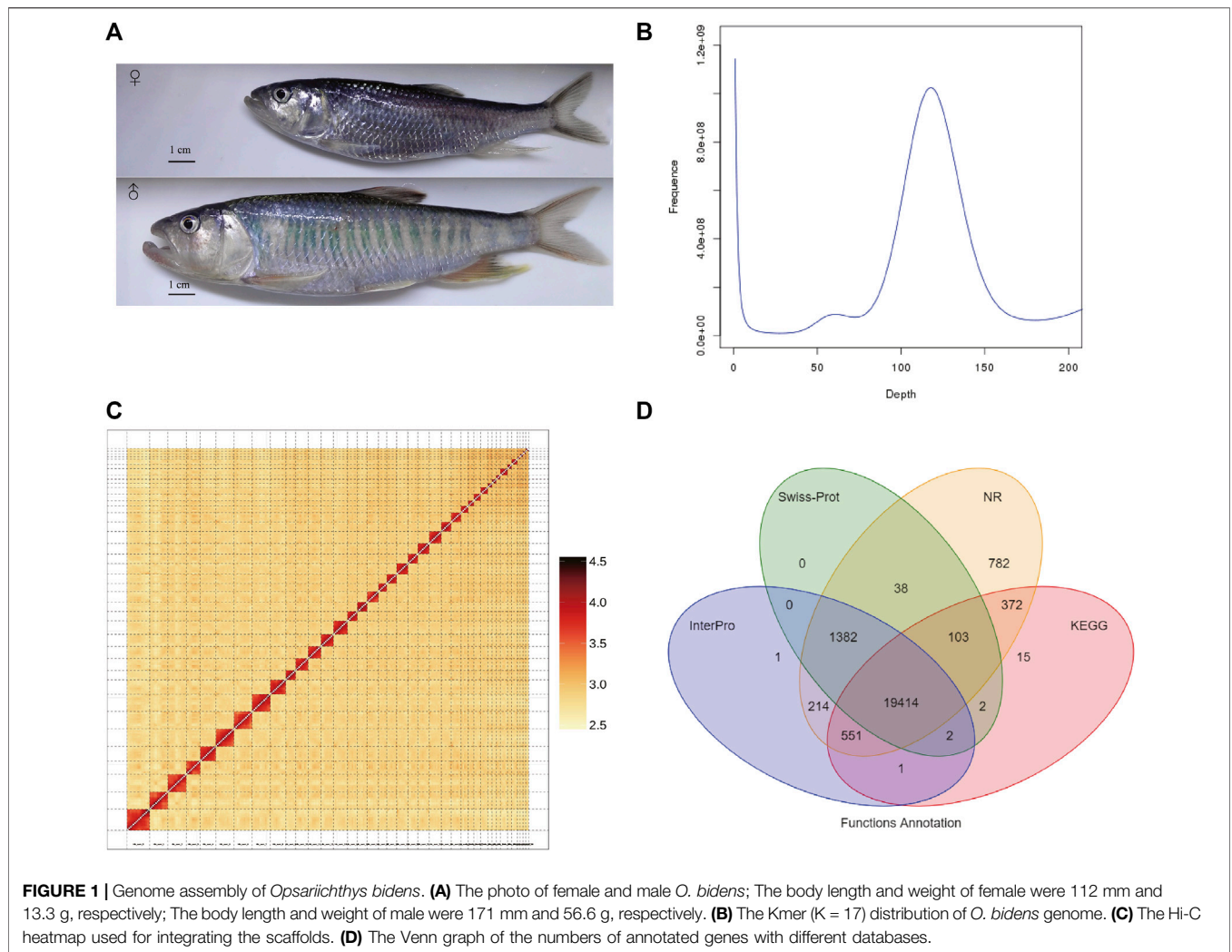
For the desirable texture and flavor of the flesh, *O. bidens* has relatively high economic values. Artificial breeding of *O. bidens* began in 2008 (Jing, 2009), and the previous studies focused on the embryonic development (Jin et al., 2017), flesh nutrition content (Zhang Q. K. et al., 2019) and spermatogenesis (Tang et al., 2020) were reported in recent years. Due to the high price, disease resistance, and wide-range temperature adaptation, *O. bidens* (Figure 1A) has become an emerging commercial fish species.

Remarkably, *O. bidens* has obvious sex dimorphism (Lian et al., 2017). In aquaculture practice, the adult males are usually twice as large as the female siblings, and have gorgeous nuptial coloration, which brings to high ornamental property as a popular ornamental fish species. Hence, a high-quality genome sequence would facilitate the development of sex-specific markers and sex control breeding.

In this study, the chromosome-level assembly of *O. bidens* was constructed using PacBio sequencing and Hi-C technology. To the best of our knowledge, this is the sequenced genome with the largest chromosome number ( $2n = 78$ ) in diploid Xenocyprididae (Arai, 2011). The genome resource will facilitate the studies of taxonomy, evolution, and genetic breeding of *O. bidens*.

## Data

A total of 135.07 Gb raw data were obtained from the Illumina X Ten platform for genome size estimation. The estimated genome size of *O. bidens* is about 899.69 Mb, and the heterozygous rate of



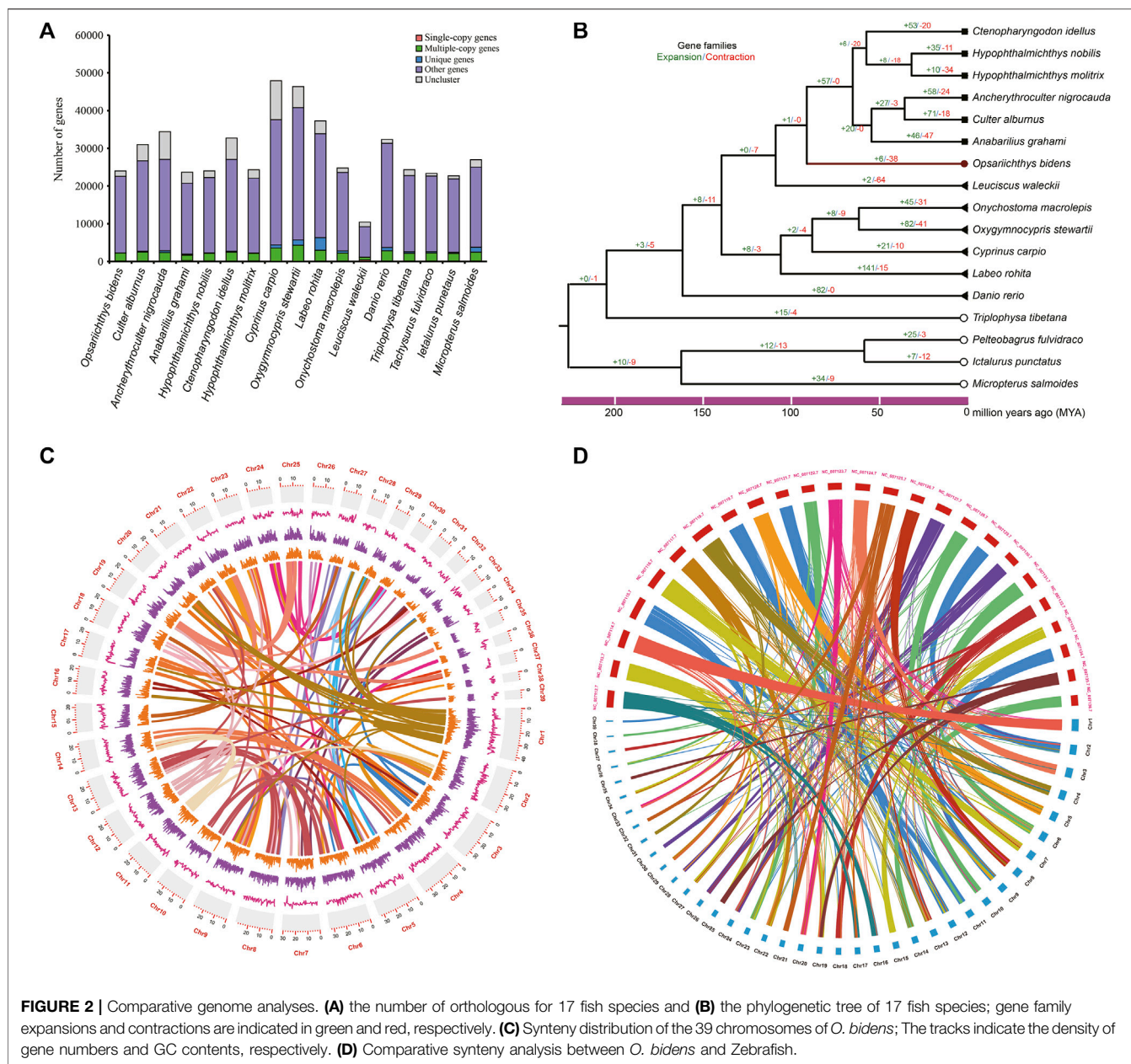
genome was 0.36%. (**Figure 1B**). Meanwhile, 167.17 Gb long reads were generated by PacBio Sequel platform. The average length of long reads was 21,861 bp, and the N50 of long reads was 34,896 bp. The long reads were *de novo* assembled into 403 contigs with total length of 818.75 Mb. The N50 of the assembled contigs was 4.71 Mb and the largest contigs was 22.26 Mb in length.

We used BUSCO analysis to determine the completeness of genome assembly, and the result showed that this assembled genome contained 96.6% complete BUSCOs, including 91.1% complete and single-copy BUSCOs and 5.7% complete duplicated BUSCOs. Meanwhile, the evaluation using CEGMA showed that the completeness of assembly was 97.18%. After polishing with the Illumina short reads using NextPolish (Hu et al., 2020), the total length of assembled contigs was 818.75 Mb. The N50 of these contigs was 4.66 Mb.

Subsequently, 95.64 Gb Hi-C data was generated by Illumina NovaSeq 6000 platform and used for chromosome-level assembly. After quality control of Hi-C reads with HiCUP software (Wingett et al., 2015), a total of 2,210,719 valid pairs were detected and 95.66% unique Di-tags were obtained. With the Hi-C data, 82 contigs were anchored into 39 chromosomes with a total length of 814.71 Mb

(**Figure 1C**), and the length of anchored chromosomes ranged from 6.77 to 42.84 Mb. Finally, the *O. bidens* genome was 818.78 Mb in length with N50 value of 25.29 Mb. To further validate the assembly completeness, we mapped the short reads to the final assembly, and the mapping rate was 98.76%.

A total of 42.39% of the genome (347.06 Mb) were identified as repetitive elements. The most abundant transposable elements (TEs) were long terminal repeats (LTRs, 35.12% of the genome), followed by DNA transposons (4.35%) and long interspersed elements (2.19%). Meanwhile, 23,992 protein-coding genes were annotated. The mean gene length was 16,469.11 bp. The average of CDS length was 1,670.54 bp, and the average number of exons per gene was 9.79. The comparative analysis of gene prediction with other fish species was performed (Figure S1). The function annotation of these protein-coding genes showed that 95.4% were annotated by at least one of the public databases (**Figure 1D**). Meanwhile, 1.07 Mb of the genome were annotated as ncRNAs, among which miRNA, tRNA and rRNA accounted for 0.084% of the genome. We performed the BUSCO analysis with the predicted protein-coding genes, and the result showed that a total of 93.5% complete BUSCOs were present with the gene annotation.



The single-copy orthologous genes of 17 fish species were identified (Figure 2A), and the phylogenetic tree was constructed with 170 single-copy orthologous genes (Figure 2B), and the result showed that *O. bidens* was grouped with the species in families of Leuciscinae and Culterinae, indicating a closer relationship with these species. A total of 6 and 38 gene families significantly expanded and contracted in *O. bidens*, respectively. The expanded and contracted gene families contained 43 and 45 genes, respectively.

To further evaluate the quality of genome assembly, we compared *O. bidens* genome with zebrafish genome. The conservation synteny among the 39 chromosomes was shown in Figure 2C, and a total of 1,306 blocks were detected among the chromosomes. The gene synteny between *O. bidens* and zebrafish

genomes is shown in Figure 2D. The chromosomes of *O. bidens* exhibited high homology with the zebrafish chromosomes, and several chromosomes of zebrafish were corresponding to two mini chromosomes of *O. bidens*, indicating that the large number of chromosome of *O. bidens* may originated from the chromosomal break of ancestral chromosomes.

## MATERIALS AND METHODS

### Sampling, Library Construction, and Sequencing

A healthy female individual was collected from our fish base in Zhejiang Province, China. The muscle, blood, kidney, heart,



brain, liver and ovary tissues were sampled and immediately frozen and stored in liquid nitrogen until extracting the genomic DNA and total RNA. High-quality DNA samples were extracted using the DNA Isolation Reagent Kit (TaKaRa, China) from muscle tissue. DNA quality and integrity was evaluated with 1% agarose gels. Firstly, a DNA sequencing library with insert size 350 bp was constructed following the instructions of Illumina DNA Prep kit. The library was sequenced on the Illumina HiSeq X Ten System using 150 bp paired-end mode in Novogene, Co. Ltd., Beijing. Meanwhile, PacBio SMRT libraries were prepared according to the manufacturer's instructions, and the libraries were sequenced using a PacBio Sequel System. Additionally, total RNAs were extracted from the muscle, kidney, heart, brain, liver and ovary tissues using RNAiso kit (TaKaRa, China). The RNA sequencing library was constructed with the PacBio Iso-Seq Express Template Prep Kit 2.0 (Pacific Biosciences, United States) and sequenced using PacBio Sequel system. The Hi-C library was prepared from muscle tissue of the same individual following the standard protocol described previously (Belton et al., 2012). The constructed Hi-C library was sequenced with Illumina NovaSeq 6000 system.

## Genome Size Estimation, Genome Assembly and Polishing

The raw data generated by Illumina platform was filtered with fastp v0.20.0 program (Chen et al., 2018). Frequencies of  $K$ -mers ( $K = 17$ ) were counted using Jellyfish (Marcais and Kingsford 2012). GenomeScope v1.0 (Vurture et al., 2017) was used to estimate size, repeat content and heterozygosity of the genome with maximum  $K$ -mer coverage of 10,000. The genome size was calculated as: size =  $K$ -mer number/peak depth. The genome assembly was performed using the FALCON assembler v2.1.0 (Chin et al., 2016), and the assembled contigs were polished with Illumina reads using NextPolish v1.4.0 software (Hu et al., 2020). The assembly completeness was evaluated by Core Eukaryotic Genes Mapping Approach (CEGMA) (Parra et al., 2007) and Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.2.2 software (Simão et al., 2015) using the Actinopterygii geneset (v10.0). Subsequently, the Hi-C reads were aligned to the assembly using the Juicer v1.6.2 (Durand et al., 2016a). The contigs were ordered and anchored with Hi-C data using the allhic program (Zhang X. et al., 2019), and manually adjusted using the Juicebox Assembly Tools v1.11.08 (Durand et al., 2016b).

## Genome Annotation

Repetitive elements in the genome were identified using RepeatMasker (Chen, 2004) and RepeatModeler with default settings. The modeled repeats were classified into their subclasses using the Repbase v20.08 database (<http://www.girinst.org/repbase/>). Tandem Repeat was extracted using TRF (<http://tandem.bu.edu/trf/trf.html>) *ab initio* prediction. A custom library generated by a combination of Repbase and

the *de novo* TE library which was processed by uclust to yield a non-redundant library was supplied to RepeatMasker for DNA-level repeat identification. Gene prediction was conducted through a combination of homology-based, *ab initio*, and transcript-based prediction methods. The full-length transcripts generated using PacBio Iso-Seq pipeline were used for transcript-based prediction. The transcripts were aligned to the genome using PASA program. Protein sequences of fish species including *Ctenopharyngodon idellus*, *Cyprinus carpio*, *Carassius auratus*, *Danio rerio*, and *Onychostoma macrolepis* were used as queries to search against the genome using tBLASTN. A *de novo* gene prediction was performed with Augustus v3.2.3 (Stanke et al., 2006), GlimmerHMM v3.04 (Majoros et al., 2004) and SNAP (Korf, 2004). The gene model was predicted by combination of three methods with EvidenceModeler v1.1.1 (Haas et al., 2008). Gene functional annotation was performed by aligning predicted protein-coding genes to the public databases using BLASTP and InterProScan70 v5.31 (Mulder and Apweiler, 2007), including NCBI NR, Swiss-prot, Pfam, Gene Ontology (GO), InterPro, and Kyoto Encyclopedia of Genes and Genomes (KEGG).

## Phylogenetic Analysis and Species Divergence Time Estimation

To investigate the phylogenetic status of *O. bidens*, we retrieved genome data of 16 fish species, including *Cyprinus carpio* (GenBank: GCA\_000951,615.2), *Ictalurus punctatus* (GenBank: GCA\_001660625.1), *Danio rerio* (GenBank: GCA\_000002035.4), *Ancherythroculter nigrocauda* (NGDC: GWHAAZV00000000), *Micropterus salmoides* (GenBank: GCA\_014851395.1), *Pelteobagrus fulvidraco* (GenBank: GCA\_003724035.1), *Hypophthalmichthys molitrix* (CNGB : CNP0000974), *Hypophthalmichthys nobilis* (CNGB : CNP0000974), *Culter alburnus* (GenBank: GCA\_009869775.1), *Oxygymnocypris stewartii* (GenBank: GCA\_003573665.1), *Anabarilius grahami* (GenBank: GCA\_003731715.1), *Labeo rohita* (GenBank: GCA\_017311145.1), *Onychostoma macrolepis* (GenBank: GCA\_012432095.1), *Leuciscus waleckii* (GenBank: GCA\_900092035.1), *Triplophysa tibetana* (GenBank: GCA\_008369825.1), and *Ctenopharyngodon idellus* (<http://www.ncgr.ac.cn/grasscarp/>) from public databases. All-to-all BLASTP was employed to identify the similarities among filtered protein sequences in these species with an E-value cutoff of  $1e^{-5}$ . We identified orthologous gene clusters using the OrthoMCL pipeline (Li et al., 2003). Protein sequences from the single-copy gene families were used for phylogenetic tree reconstruction. MUSCLE (Edgar, 2004) was used to generate multiple sequence alignments for protein sequences with default parameters, and the ambiguously aligned positions were trimmed using Gblocks (<http://molevol.cmima.csic.es/castresana/Gblocks.html>). The alignments of each family were concatenated to a super alignment matrix. The alignment matrix was used for phylogenetic tree reconstruction through maximum likelihood methods. The phylogenetic tree was constructed using RAxML v7.2.9 (Stamatakis, 2014) with 1,000 bootstrap replicates.

Divergence time between species was estimated using MCMCtree with model of JC69 in PAML (Yang, 2007). The divergence time calibration of *Oxygymnocypris stewartii* and *Cyprinus carpio* were obtained from the TimeTree website (<http://www.timetree.org/>). The likelihood analysis for gene gain and gene loss was identified using CAFE v4.2 (De Bie et al., 2006) with  $p < 0.05$ .

## Synteny Analysis

Synteny analysis of intra-genome was carried out using the MCScanX pipeline (Wang et al., 2012), output were converted to blocks by in-house Perl scripts. Circos (Krzywinski et al., 2009) was used to display the syntenic blocks. We identified syntenic blocks of genes between *O. bidens* and *D. rerio*. For the comparison, we carried out an all-to-all BLAST search of annotated protein sequences and ran MCScanX with the parameters “-s 10 -b 2”.

## DATA AVAILABILITY STATEMENT

The sequences of genome assembly are available in the National Genomics Data Center (NGDC) with accession number GWHBEIO00000000. The newick file of phylogenetic tree generated by RAxML is available in figShare with doi: [https://figshare.com/articles/dataset/phylogenetic\\_tree\\_generated\\_by\\_RAxML/17085437/1](https://figshare.com/articles/dataset/phylogenetic_tree_generated_by_RAxML/17085437/1). The karyotype image is available in figShare with doi: [https://figshare.com/articles/figure/karyotype\\_image\\_of\\_O\\_bidens/17161865/1](https://figshare.com/articles/figure/karyotype_image_of_O_bidens/17161865/1).

## REFERENCES

- Arai, R. (2011). *Fish Karyotypes: A Check List*. Tokyo: Springer.
- Belton, J.-M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: A Comprehensive Technique to Capture the Conformation of Genomes. *Methods* 58, 268–276. doi:10.1016/j.ymeth.2012.05.001
- Betancur-R, R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., et al. (2017). Phylogenetic Classification of Bony Fishes. *BMC Evol. Biol.* 17, 162–201. doi:10.1186/s12862-017-0958-3
- Chen, N. (2004). Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Curr. Protoc. Bioinformatics* 5 (4.10), 1–4. doi:10.1002/0471250953.bi0410s05
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560
- Chen, Y. Y. (1998). *Fauna Sinica Osteichthys Cypriniformes II*. Beijing: Science Press.
- Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. (2016). Phased Diploid Genome Assembly With Single-Molecule Real-Time Sequencing. *Nat. Methods* 13 (12), 1050–1054. doi:10.1038/nmeth.4035
- 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
- Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., et al. (2016a). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cel Syst.* 3, 99–101. doi:10.1016/j.cels.2015.07.012
- Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., et al. (2016b). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cel Syst.* 3, 95–98. doi:10.1016/j.cels.2016.07.002

## ETHICS STATEMENT

The animal study was reviewed and approved by the Animal ethics committee of Zhejiang Academy Of Agricultural Sciences.

## AUTHOR CONTRIBUTIONS

XX and BL conceived the study. WG, DG, and WZ collected samples. SY and BN performed the bioinformatics analyses. XX and SY wrote the manuscript. Q-PX revised the manuscript. XX, WG, and BN contributed equally to this work. All authors read and approved the final manuscript.

## FUNDING

This study was financially supported by grants of Zhejiang provincial Department of Science and Technology (No.2020C02014).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.788547/full#supplementary-material>

**Supplementary Figure S1** | The comparative analyses of CDS length, exon length, exon number, gene length and intron length with 6 species. *Cau*, *Cca*, *Cid*, *Dre*, *makouyu* and *Oma* indicate *Carassius auratus*, *Cyprinus carpio*, *Ctenopharyngodon idellus*, *Danio rerio*, *Opsariichthys bidens*, and *Onychostoma macrolepis*, respectively.

- Edgar, R. C. (2004). MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* 32, 1792–1797. doi:10.2460/ajvr.69.1.8210.1093/nar/gkh340
- Fang, F., Norén, M., Liao, T. Y., Källersjö, M., and Kullander, S. O. (2009). Molecular Phylogenetic Interrelationships of the South Asian Cyprinid genera *Danio*, *Devario* and *Microrasbora* (Teleostei, Cyprinidae, Danioninae). *Zool. Scr.* 38, 237–256. doi:10.1111/j.1463-6409.2008.00373
- Haas, B. J., Salzberg, S. L., Zhu, W., Perter, 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
- Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: A Fast and Efficient Genome Polishing Tool for Long-Read Assembly. *Bioinformatics* 36, 2253–2255. doi:10.1093/bioinformatics/btz891
- Huang, S. P., Wang, F. Y., and Wang, T. Y. (2017). Molecular Phylogeny of the opsariichthys Group (Teleostei: Cypriniformes) Based on Complete Mitochondrial Genomes. *Zool. Stud.* 56, e40–52. doi:10.6620/ZS.2017.56-40
- Jin, D. L., Zhang, Q. K., Wang, Y. F., Zhu, Y. M., Zhang, Y. M., Wang, J. P., et al. (2017). Observation of Embryonic, Larva and Juvenile Development of *Opsariichthys bidens*. *Oceanol. Limnol. Sin.* 48 (04), 838–847. doi:10.11693/hyh20170200034
- Jing, J. T. (2009). Artificial Breeding and Aquaculture experiment of *Opsariichthys bidens* of the Yalu River. *China Fish.* 6, 32–34. doi:10.3969/j.issn.1002-6681.2009.06.017
- Korf, I. (2004). Gene Finding in Novel Genomes. *BMC bioinformatics* 5, 59–9. doi:10.1186/1471-2105-5-59
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: An Information Aesthetic for Comparative Genomics. *Genome Res.* 19, 1639–1645. doi:10.1101/gr.092759.109
- 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

- Lian, Q. P., Mi, G. Q., and Liu, S. L. (2017). Sexual Dimorphism in Morphological Traits of *Opsariichthys bidens*. *XianDai NongYe Keji* 22, 226–229. doi:10.3969/j.issn.1007-5739.2017.22.126
- Liao, T.-Y., Kullander, S. O., and Fang, F. (2011). Phylogenetic Position of Rasborin Cyprinids and Monophyly of Major Lineages Among the Danioninae, Based on Morphological Characters (Cypriniformes: Cyprinidae). *J. Zool. Syst. Evol. Res.* 49, 224–232. doi:10.1111/j.1439-0469.2011.00621.x
- Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: Two Open Source Ab Initio Eukaryotic Gene-Finders. *Bioinformatics* 20, 2878–2879. doi:10.1101/gr.122450310.1093/bioinformatics/bth315
- Marçais, G., and Kingsford, C. (2012). Jellyfish: A Fast K-Mer Counter. *Tutorialis e Manus* 1, 1–8.
- Mayden, R. L., Chen, W.-J., Bart, H. L., Doosey, M. H., Simons, A. M., Tang, K. L., et al. (2009). Reconstructing the Phylogenetic Relationships of the Earth's Most Diverse Clade of Freshwater Fishes—Order Cypriniformes (Actinopterygii: Ostariophysi): A Case Study Using Multiple Nuclear Loci and the Mitochondrial Genome. *Mol. Phylogenet. Evol.* 51, 500–514. doi:10.1016/j.ympev.2008.12.015
- Mulder, N., and Apweiler, R. (2007). InterPro and InterProScan. *Methods Mol. Biol.* 396, 59–70. doi:10.1007/978-1-59745-515-2\_5
- Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a Pipeline to Accurately Annotate Core Genes in Eukaryotic Genomes. *Bioinformatics* 23, 1061–1067. doi:10.1093/bioinformatics/btm071
- Simão, 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., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: Ab Initio Prediction of Alternative Transcripts. *Nucleic Acids Res.* 34, W435–W439. doi:10.1093/nar/gkl200
- Stout, C. C., Tan, M., Lemmon, A. R., Lemmon, E. M., and Armbruster, J. W. (2016). Resolving Cypriniformes Relationships Using an Anchored Enrichment Approach. *BMC Evol. Biol.* 16, 244–256. doi:10.1186/s12862-016-0819-5
- Tang, D., Gao, X., Lin, C., Feng, B., Hou, C., Zhu, J., et al. (2020). Cytological Features of Spermatogenesis in *Opsariichthys Bidens* (Teleostei, Cyprinidae). *Anim. Reprod. Sci.* 222, 106608. doi:10.1016/j.anireprosci.2020.106608
- Tang, K. L., Agnew, M. K., Hirt, M. V., Lumbantobing, D. N., Raley, M. E., Sado, T., et al. (2013). Limits and Phylogenetic Relationships of East Asian Fishes in the Subfamily Oxygastrinae (Teleostei: Cypriniformes: Cyprinidae). *Zootaxa* 3681, 101–135. doi:10.11646/zootaxa.3681.2.1
- Tang, K. L., Agnew, M. K., Hirt, M. V., Sado, T., Schneider, L. M., Freyhof, J., et al. (2010). Systematics of the Subfamily Danioninae (Teleostei: Cypriniformes: Cyprinidae). *Mol. Phylogenet. Evol.* 57, 189–214. doi:10.1016/j.ympev.2010.05.021
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: Fast Reference-free Genome Profiling from Short Reads. *Bioinformatics* 33, 2202–2204. doi:10.1093/bioinformatics/btx153
- Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCSScanX: a Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Res.* 40, e49. doi:10.1093/nar/gkr1293
- Wingett, S. W., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., et al. (2015). HiCUP: Pipeline for Mapping and Processing Hi-C Data. *F1000Res* 4, 1310. doi:10.12688/f1000research.7334.1
- Wu, X. W. (1964). *Cyprimid Fishes in China*. Shanghai: Shanghai Science and Technology Press
- Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi:10.1093/molbev/msm088
- Zhang, Q. K., Zheng, X. B., Tang, D. J., Zhu, Y. M., Wang, J. P., and Zhu, J. Q. (2019). Analysis and Evaluation of Nutritional Components in Muscle of Cultured *Opsariichthys bidens*. *J. Ningbo Univ.* 32 (04), 15–19. doi:10.3969/j.issn.1001-5132.2019.04.003
- Zhang, X., Zhang, S., Zhao, Q., Ming, R., and Tang, H. (2019). Assembly of Allele-Aware, Chromosomal-Scale Autopolyploid Genomes Based on Hi-C Data. *Nat. Plants* 5, 833–845. doi:10.1038/s41477-019-0487-8

**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.

Copyright © 2022 Xu, Guan, Niu, Guo, Xie, Zhan, Yi and Lou. 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.